Traitement visuel rapide de scènes naturelles chez le singe, l'homme et la machine : une vision qui va de l'avant...

Arnaud Delorme

THESE

présentée en vue de
l'obtention du titre de

DOCTEUR
de
L'UNIVERSITE PAUL SABATIER

Spécialité : sciences cognitives

Arnaud DELORME

Traitement visuel rapide de scènes naturelles chez le singe, l'homme et la machine :

une vision qui va de l'avant...

Soutenue le 26 octobre 2000 devant la commission d'examen

Y.	**BURNOD**	Rapporteur
M.	**FABRE-THORPE**	Directeur de thèse
M.	**IMBERT**	
J.	**LORENCEAU**	Rapporteur
K.	**O'REGAN**	
S.	**THORPE**	

Centre de Recherche Cerveau et Cognition UMR 5549 CNRS-UPS
Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 TOULOUSE

À ma femme, Séti

Remerciements

Merci avant tout à Michèle Fabre-Thorpe, toi qui m'a accueilli en thèse et m'a supporté ces dernières années dans les deux sens du terme. Merci aussi pour le pointillisme quasi-inquisiteur dont tu as fait preuve dans tes relectures. Michèle, à bientôt!

Merci aussi à toi Simon Thorpe pour ta gentillesse et pour m'avoir dévoilé l'élément indispensable pour faire de la recherche : avant tout avoir de bonnes idées. Salut, boss!

Je tiens également à remercier messieurs les rapporteurs, Yves Burnod et Jean Lorenceau pour avoir accepté de relire ce travail et à messieurs les examinateurs Kevin O'Regan et Michel Imbert d'avoir consenti à être membre du jury de ma thèse. Merci aussi à Jean Bulier (et doublement à Michel Imbert) pour leur accueil chaleureux au laboratoire.

Eudora et Synapse, mes fidèles compagnons de travail, merci pour avoir supporté, de longues heures durant, mes sautes d'humeur et la perversion dont j'ai parfois fait preuve pour arriver à mes fins (Synapse, je sais que nous avons eu parfois quelques mots à propos de certains logiciels, mais c'est oublié).

Merci à toi, Guillaume, notre héraut, pour ton enthousiasme exubérant. Bonne chance pour la suite !

Catherine, merci pour avoir bien voulu conjuguer tes talents afin de compléter mon abécédaire orthographique déficient!

Merci aussi aux étudiants en thèse, en DEA, vous les vaillants soldats de la science. Merci également au reste du cénacle du laboratoire, qui forme un groupe d'où transsude l'ambiance propice à l'émulation scientifique.

Merci à mes parents, qui ont su motiver mon inclinaison tendancieuse à toujours faire ce dont j'avais envie. Merci enfin à mes grands-parents, à mes fréros, Fabrice et Laurent, et à mes amis, Bruno, Patrice, Pierro, Emilie, Léti, Rufin, Vanessa, Laurent pour .....le reste.

TABLE DES MATIERES

Introduction

L'énigme du cerveau est l'un des défis les plus ambitieux lancés à la science qui, aujourd'hui, possède enfin des outils d'investigation performants pour en résoudre certains mystères. Dès le milieu du 20ème siècle, l'approche cybernétique a posé les bases d'une étude scientifique des processus cognitifs. Depuis, les mentalités ont beaucoup évolué et l'étude des processus cognitifs réunit un éventail de disciplines impressionnant allant de l'étude du comportement des neurones à la philosophie. Les sciences cognitives englobent la psychologie, la psychanalyse, la sociologie, la phénoménologie, l'anthropologie, la philosophie et la psychologie expérimentale, l'intelligente artificielle, la linguistique et les neurosciences qui, elles-mêmes, aborde ce problème grâce à plusieurs approches : l'électrophysiologie, l'imagerie cérébrale, la neuropharmacologie, la neuronatomie et les neurosciences computationnelles.

Au sein de l'organisation cérébrale, le système visuel est particulièrement intéressant car, du fait de son organisation hiérarchique, il établit un lien fort entre des représentations à différents niveaux. On peut aborder l'étude du système visuel au niveau de neurones individuels ou de populations de neurones - allant de quelques milliers de neurones en champs de potentiels à quelques millions en imagerie cérébrale - et tenter d'établir un lien entre le comportement des sujets et l'activité de ces neurones. Il est également possible en psychologie expérimentale d'analyser les réponses de sujets à des stimuli visuels et d'inférer les processus neuronaux sous-jacents. Des modèles permettent ensuite de vérifier ces hypothèses et d'établir un pont entre le comportement des neurones et celui des individus. Ces approches combinées, à différents niveaux, seront sans aucun doute nécessaires pour mettre à jour les processus à l'origine de la formidable efficacité du système visuel et c'est dans ce cadre que se place mon travail de thèse. M'intéressant plus aux populations de neurones qu'aux comportements des neurones individuels, j'ai utilisé une approche de psychologie expérimentale et basé mon travail de modélisation à la fois sur les résultats que j'ai obtenus en psychophysique et sur les contraintes de structure et de connectivité au sein du système visuel.

J'ai tenté modestement dans ma thèse d'appréhender les processus mis en jeu dans le système visuel dans le cadre de la détection et de la reconnaissance de scènes brièvement flashées. Dans une première partie, j'introduirai un cadre conceptuel aux expériences et aux modèles que j'ai réalisés. Je traiterai en particulier des origines des sciences cognitives et de ce qu'elles nous ont déjà apporté du point de vue de l'étude scientifique. Je résumerai succinctement

ensuite les données, principalement électrophysiologiques, qui décrivent le comportement des neurones du système visuel. Ces travaux seront utilisés à la fois pour interpréter les résultats des expériences en psychologie expérimentale que j'ai menées et pour construire mes modèles.

Dans une seconde partie, je présenterai les expériences que j'ai réalisées. Le paradigme expérimental de la tâche go-nogo permet d'étudier le traitement d'une catégorie abstraite - dans nos expériences, la catégorie animale - dans des conditions bien contrôlées. La tâche, bien qu'extrêmement simple, fait appel à des traitements très complexes dans le système visuel et l'on connaît encore très mal la façon dont ils s'effectuent. Mes études ont porté sur les processus de catégorisation chez les sujets humains mais aussi chez les singes macaques. Les différences et les similitudes entre les deux espèces peuvent en effet nous renseigner sur les différentes stratégies utilisées par les systèmes visuels des primates[1]. Mon approche consiste à contraindre les conditions dans lesquelles est effectuée la catégorisation pour déterminer le rôle à la fois des propriétés intrinsèques des images - couleur, luminance, nombre d'animaux présents, parties visibles de leurs corps, espèce de l'animal... - mais aussi de leurs propriétés extrinsèques - condition de présentation, effet de séquence, familiarité du stimulus, consigne... L'analyse de l'influence de ces caractéristiques sur la catégorisation me permet d'inférer, dans une certaine mesure, la dynamique de traitement des images par le système visuel.

Par la suite, dans une troisième partie, je tenterai d'intégrer ces contraintes, ainsi que celles imposées par la biologie, au sein de modèles simplifiés du système visuel. Mon but est à la fois de conserver une certaine plausibilité biologique et d'obtenir des performances similaires à celle du système visuel en ce qui concerne le traitement d'image. Je me pencherai en particulier sur les mécanismes qui pourraient permettre la mise en place d'une sélectivité très rapide des neurones à la fois dans les aires visuelles de bas niveau, où les décharges des neurones sont sélectives à des formes très simples, et dans les aires visuelles de haut niveau, où les décharges des neurones sont sélectives aux objets présentés. Je tenterai également d'introduire des lois d'apprentissage conformes à la fois à la biologie et à la dynamique du réseau, et j'étudierai les propriétés émergentes des réseaux de neurones ainsi définis.

Dans une dernière partie, j'analyserai les résultats obtenus par rapport aux théories existantes sur le traitement visuel. Le système visuel, au-delà de son aspect de calcul purement automatique, est également la porte qui peut permettre à la simulation d'atteindre des états

[1] Chez le singe, il est éventuellement possible d'enregistrer directement le comportement des neurones du système visuel alors que, chez l'homme, ces techniques ne sont pas envisageables. Savoir si le comportement des deux espèces, et par conséquent le traitement visuel sous-jacent, est similaire est donc critique pour l'étude du système visuel humain.

cognitifs comme la sémantique des images, l'attention et pourquoi pas les émotions et la conscience visuelle. Je discuterai donc dans quelle mesure il est envisageable de simuler ces types de phénomène. La compréhension de ces mécanismes et de leur interaction sera, à mon avis, l'un des enjeux majeurs des prochaines années pour les sciences cognitives.

I

Comment aborder l'étude du système visuel

Avant toute étude scientifique, il convient de fixer à la fois le cadre conceptuel dans lequel on se situe et de décrire les bases nécessaires à la compréhension des études que l'on va mener. Cette première partie est donc double : le premier chapitre décrit en particulier pourquoi les systèmes autonomes et autopoïétiques me semblent être la meilleure définition que l'on puisse donner des êtres vivants. Cette approche est relativement en décalage par rapport au reste de mes travaux, et par rapport aux neurosciences en général qui postulent l'intervention de processus mécanistes dans le traitement visuel. Cependant, je ne pense pas que ces deux approches - systèmes autonomes et neurosciences computationnelles - soient incompatibles et je tenterai en conclusion de resituer mes travaux dans ce cadre. Ce premier chapitre aborde également l'histoire des sciences cognitives, leur naissance et le rejet de leurs origines : la cybernétique. Je tenterai d'en tirer quelques leçons pour l'avenir des sciences cognitives en général et des neurosciences en particulier. Le second chapitre est plus traditionnel et pose les connaissances nécessaires à nos études en définissant brièvement ce que l'on connaît aujourd'hui du traitement dans le système visuel, à la fois grâce à l'apport de l'électrophysiologie et à celui de l'imagerie fonctionnelle.

1

Cadre épistémologique

Tout organisme est une mélodie qui se chante elle-même.

Uexküll (1984)

Avant toute étude scientifique, il convient de définir le cadre conceptuel et épistémologique auquel on adhère. Chaque scientifique possède sa propre idée de ce que doit être la science et de l'objectivité scientifique en général : pour moi, et je pense que ce sentiment est largement partagé, l'objectivité scientifique est par essence basée sur les faits expérimentaux. La science serait un savant mélange de rigueur et de bonnes idées, les idées servant à imaginer des expériences et la rigueur à les réaliser. Que dire de plus ? Est-il vraiment nécessaire d'aller plus loin ?

En fait, nous verrons que l'objectivité totale est un leurre et qu'implicitement, même si nous nous réclamons d'une impartialité totale, notre démarche est empreinte d'*a priori* dont nous ne pouvons nous défaire. Il s'agit alors de nous en accommoder et de ne pas ignorer les limites des techniques, notamment conceptuelles, que nous utilisons.

Nous verrons également comment dépasser une analyse purement mécaniste des processus neuronaux à l'aide de l'étude des systèmes et quelles sont les implications de ce type d'approche systémique sur la notion de représentation cognitive et neuronale.

Mes principales inspirations pour écrire ce chapitre ont été les travaux de Varela (1989) et de Dupuy (1999) pour replacer la vie, les systèmes cognitifs et même les interactions sociales

entre individus dans le cadre conceptuel des systèmes, qui est à mon avis à ce jour le plus approprié en sciences cognitives. Ma formation en science cognitive et les nombreux articles que j'ai pu lire dans ce domaine sont autant de sources d'informations complémentaires qui m'ont aidé à me forger un sentiment sur ce sujet.

1.1 - L'objectivité scientifique est une chimère

En partant de la simple assertion "je suis un menteur", on peut saisir le problème de la circularité. Si je dis "vrai" alors je dis "faux" car "je suis un menteur". Si je dis "faux", alors je dis "vrai" car "je ne suis pas un menteur". Cette phrase auto référentielle est contradictoire au sens de la logique du premier ordre. Cependant on peut également voir cette assertion auto référentielle comme étant un processus circulaire : du faux, l'on passe au vrai puis à nouveau au faux...

De la même façon, le processus de la vie est autoréférentiel et circulaire. Prenons l'exemple d'une cellule : les molécules de la membrane cellulaire définissent la frontière de la cellule, ce qui permet la production de molécules par la cellule et en particulier des molécules de la membrane qui définissent la frontière (figure 1.1). On a ici un processus cyclique analogue à celui de l'assertion du menteur. Ce processus s'apparente au problème de la poule et de l'œuf : qui était le premier de la poule et de l'œuf : en fait, là n'est pas la question, le système n'est qu'un et il implémente un processus cyclique.

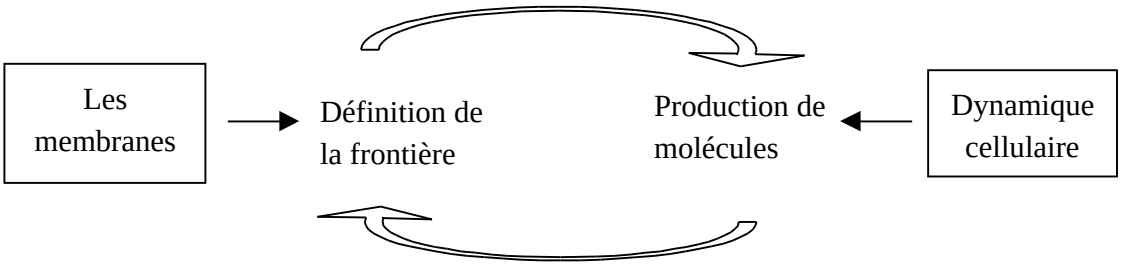


Figure 1.1 : processus circulaire permettant à une cellule de conserver son identité.

On peut resituer la cognition, l'interaction de l'homme avec son environnement, dans un cadre similaire. L'homme spécifie son propre environnement - composé du monde extérieur, des autres êtres vivants - autant que l'environnement spécifie la nature de l'homme : on existe par le regard des autres, par la cohérence du monde extérieur qui influencent notre propre organisation. La cognition humaine est un processus circulaire dans le sens où nos comportements organisent autant le monde extérieur que le monde extérieur organise nos comportements. L'objectivité scientifique, dans le sens où les résultats sont indépendants de

l'observateur, est donc une chimère car le monde que nous tentons de spécifier est une partie de nous-mêmes. Cette approche découle de la phénoménologie, développé par Husserl (1859-1938) au début du siècle, pour laquelle il faut absolument replacer la première personne au centre de l'étude scientifique. Varela (1996) a également développé une neurophénoménologie, où il tente d'intégrer "la première personne" à l'étude du cerveau.

Une conséquence directe de cet état de fait est que la compréhension d'un processus cognitif signifie également la modification de notre comportement vis à vis du monde et donc celle du monde lui-même. Dans ce cadre conceptuel, la recherche scientifique n'a de limites que celles de l'imagination humaine : au fur et à mesure que nous avons l'impression de comprendre le monde il change avec nous et de nouveaux problèmes émergent. La maîtrise totale des processus intervenant dans la nature serait donc un objectif inatteignable.

1.2 - Un modèle récurrent : les systèmes

La capacité principale des êtres vivants est de pouvoir maintenir leur identité, c'est-à-dire en quelque sorte de maintenir les relations entre les composants d'un système. Cette définition est valable aussi bien pour la cellule que pour les êtres multicellulaires et l'homme. On parle alors de systèmes autopoïétiques qui engendrent et spécifient continuellement leur propre organisation. Les molécules se renouvellent continuellement dans une cellule dont la définition ne peut se limiter à la liste de ses composants matériels. Sa spécification fonctionnelle - son identité - dépend plutôt des relations que les composants entretiennent entre eux. Prenons l'exemple d'un patient subissant une opération chirurgicale de greffe d'un cœur artificiel. Le système (l'individu) survivra si le cœur artificiel entretient des relations avec les autres composants du système (circulation sanguine, système immunitaire et neuronal) qui sont compatibles avec celles que remplissait l'organe initial. Ce qui est critique pour que le système conserve son identité (c'est-à-dire que le patient survive) est moins la structure de ses éléments (qui peuvent être faits de plastique pour le cœur artificiel) que les relations qu'entretiennent les éléments du système entre eux.

La théorie des systèmes autopoïétiques (Varela, 1989), que je viens de décrire en quelques lignes, est à mon avis l'une des plus riches pour expliquer l'émergence de nouveaux domaines, de nouveaux mondes si l'on veut. Partant d'un domaine vierge - par exemple la terre avant que la vie n'apparaisse[1] - des systèmes autopoïétiques peuvent émerger, par exemple les premières

[1] On peut cependant considérer la terre comme étant elle-même un système autopoïétique définissant un nouveau domaine.

bactéries. L'interaction entre ces nouveaux systèmes va définir un nouveau domaine dans lequel les règles de la physique classique moléculaire n'auront plus cours. Par exemple dans le cas simple de deux organismes unicellulaires qui entrent en contact, les interactions qu'ils mettent en jeu, les systèmes de maintien respectifs de leur identité font que ces interactions ne sont pas comparables avec celles qu'ils peuvent avoir avec la matière inorganique. De nouveaux types d'interactions apparaissent et définissent un nouveau domaine avec ses propres lois.

Les systèmes autopoïétiques se retrouvent dans des disciplines très variées. Les domaines dans lesquels ils évoluent ont chacun leurs composants (molécules, cellules, neurones, individus, corporations, états) et chacun des règles d'interaction entre leurs composants. On a vu émerger en psychiatrie par exemple la théorie systémique qui tente d'expliciter les règles d'interaction à l'intérieur d'une famille (ici le système est la famille). Les interactions entre les individus composant le système définiraient un système complexe avec sa propre "monnaie d'échange" (dont les cadeaux, les services rendus, la souffrance font partie). Grâce à ces interactions, le système - la famille - conserve son identité. On parle ici plus d'homéostasie que d'autopoïèse mais l'idée de conservation de l'identité est la même (i.e. Monroy, 1996).

Les systèmes autopoïétiques peuvent être organisés de façon hiérarchique : par exemple, la cellule, l'organe, l'individu sont différents systèmes imbriqués l'un dans l'autre. On remarque que le système de niveau supérieur ne semble pas se préoccuper de la survie de ses unités mais plutôt de la sienne propre. Au sein du corps humain, certaines cellules sont sacrifiées pour le développement de l'organisme[2]. Dans le cerveau, des milliers de neurones meurent chaque jour sans que la cognition semble dramatiquement affectée. D'une façon similaire, au sein des systèmes familiaux, le suicide de l'un des individus intervient parfois pour tenter de préserver l'intégrité du système.

Pour en finir avec les systèmes autopoïétiques, il convient de citer le plus énigmatique d'entre eux, qui est sans doute celui de la conscience de soi. Pour Damasio (1999), la conscience réflexive est le niveau homéostatique le plus intégré de la cognition - il emploie le terme homéodynamique pour la désigner. Nous reviendrons sur ce sujet dans la discussion, notamment en ce qui concerne la conscience visuelle.

Nous allons maintenant, au sein de ce cadre conceptuel, voir de quelle façon les notions de représentation et de communication, notamment dans le système nerveux, peuvent trouver leur place.

---

[2] C'est le mécanisme de mort cellulaire programmée (apoptose).

1.3 - Représentation et communication

La notion de communication et d'information est centrale en science cognitive, en particulier dans le système visuel. Une grande partie des travaux en électrophysiologie tente d'estimer la quantité d'information qu'un neurone encode à partir d'un stimulus. Si elle est importante, on dira que l'activité du neurone intervient dans la représentation du stimulus.

Cependant la communication, au sens de transfert d'information, entre deux systèmes vivants est un leurre. Prenons l'exemple de deux unités A et B. A déforme B qui déforme A qui lui-même a une action en retour sur B. On est ici en présence d'une "communication" entre A et B mais à aucun moment ni A, ni B n'a voulu transmettre de l'information. Dans le cas de la communication hormonale, la cellule capte une hormone et répond à ce signal par un comportement prototypique. Cependant, du point de vue de la cellule, il s'agit uniquement d'une réaction intrinsèque au domaine dans lequel elle évolue. La cellule ne reçoit pas un message hormonal; elle réagit à une perturbation conformément aux spécifications fonctionnelles qui lui permettent de garder son identité au sein de son environnement.

De la même manière, la notion de représentation neuronale apparaît caduque. La notion de représentation au sens strict implique que le système (neuronal par exemple) construise une image relativement fidèle des objets dans le monde extérieur. Une vue extrême de cette approche est relativement bien illustrée par le neurone "grand-mère" qui, par sa sélectivité extrême, ne représenterait qu'une seule personne. Cette vue anthropomorphique apparaît pour le moins simpliste : même si des catégories peuvent émerger *a priori* du système, elles résultent autant des contraintes de l'environnement que de la structure et des interactions entre les composants du système. Ces catégories ont donc peu de chances de découler uniquement des propriétés de notre environnement. Pour le système, ces catégories peuvent être absente de toute pertinence.

En fait, le système subit uniquement des modifications de ses états possibles. Cependant du sens peut émerger de ces processus homéostatiques. La métaphore de l'aviateur qui pilote en aveugle, se fiant uniquement aux instruments de bord, illustre bien cette idée. Au compliment "oh, que tu as bien piloté !", il répond "non, j'ai juste activé des manettes en suivant certaines séquences logiques". Le monde interne du système - son domaine - peut donc créer du sens sans que la présence de l'expérimentateur ne soit, directement ou indirectement, nécessaire. Il est très difficile d'imaginer de tels processus justement parce qu'ils ne font pas partie de notre logique de penser. Je décrirai par exemple, dans la partie consacrée à la modélisation, un modèle du système visuel où les interactions et les

renforcements des connexions entre les neurones sont les seules lois d'interaction entre les composants. Ce système (simplifié car on a supprimé la majorité des interactions entre neurones) s'organise automatiquement pour détecter des orientations dans les images. Dans ce système, je suis parti des règles d'interaction entre les neurones et des contraintes de l'environnement et j'ai vu émerger un comportement complexe de cet amalgame compatible avec celui observé pour les neurones réels.

Par analogie, certains phénomènes, comme la mémoire et l'apprentissage, deviennent de faux problèmes. Dans le cas de la mémoire, on peut imaginer que des perturbations font évoluer le système dans une direction plutôt qu'une autre. Dirions-nous d'une voiture accidentée sur un arbre qu'elle a appris cet arbre ? Évidemment non ! Et pourtant, cette métaphore illustre parfaitement ce qui peut se passer pour le système : face à un événement imprévu, le système modifie sa trajectoire pour conserver son identité[3]. Dans l'exemple précédent, la voiture a en quelque sorte "mémorisé l'arbre". Dans le cas du système nerveux, il est possible que la façon dont il réagit lui permette d'anticiper une nouvelle occurrence de cet événement et donc de mieux préserver son identité. En aucun cas cependant, le système n'apprend au sens anthropomorphique du terme.

Pour un observateur extérieur, le niveau d'intégration du corps humain, des organes, de leur approvisionnement en énergie, de leur mode de communication..., toute cette extraordinaire complexité, ne peut être que le résultat d'un ingénieur surdoué. Chaque organe semble remplir une fonction pour l'organisme. Mais de la même façon au sein d'un organe, le rein par exemple, toutes les cellules semblent œuvrer uniquement au fonctionnement de cet organe. La théorie autopoïétique implique que chaque organe, chaque cellule, a en fait sa vie propre et tente de maintenir son équilibre interne dans l'environnement dans lequel il évolue plutôt que de participer à la vie de l'organisme. Ainsi est assurée l'extrême stabilité du système et les éléments sont à ce point interchangeables qu'il est possible, par exemple chez l'homme, de greffer pratiquement n'importe quel organe. De la même façon dans le cerveau, les neurones ne tenteraient pas d'interpréter le monde extérieur mais plutôt de développer leur dynamique interne. Les neurones ne s'activeraient pas pour remplir une fonction précise, mais plutôt pour survivre; leur activité est le garant de cette survie puisqu'elle leur permet de recevoir des facteurs de croissance indispensables en provenance des synapses. Sur la base de

[3] Varela désigne par le terme ontogenèse, l'histoire - la trajectoire dans le temps - du système.

ces interactions locales, des comportements complexes et du sens émergent au sein des réseaux de neurones[4].

1.4 - Approche opérationnelle/approche symbolique

La théorie des systèmes a également des implications sur les méthodes de recherche. Très schématiquement, deux types d'analyses s'opposent, d'une part les analyses de type opérationnel et d'autre part les analyses de type téléologique.

Le premier type d'analyse est une analyse du bas vers le haut (*bottom-up*), partant des lois d'interaction au niveau local entre les éléments d'un système, elle tente d'inférer les propriétés macroscopiques et globales du système. Dans l'approche opérationnelle, on tente de déterminer comment les propriétés macroscopiques du système émergent à partir des interactions locales. Dans l'approche téléologique, on tente plutôt de déterminer comment implémenter ces propriétés macroscopiques.

Le second type d'analyse s'effectue du haut vers le bas (*top-down*). On part alors de la finalité du système pour tenter de construire et d'organiser des composants permettant d'expliquer son comportement. Cette seconde approche est typiquement symbolique et procédurale. La biologie est censée implémenter l'équivalent d'un algorithme et l'on parle sans vergogne de communication et de transfert d'information entre les éléments du système.

Ces deux approches sont complémentaires, mais il ne faut pas oublier que la vraie - et la seule approche pertinente de mon point de vue - est l'analyse opérationnelle dont font partie les systèmes autopoïétiques que j'ai présentés précédemment. L'exemple qui illustre bien la complémentarité des deux approches est celui de la génétique. D'un point de vue didactique, l'explication des processus de transcription par exemple est très fortement ancrée au niveau symbolique. Une partie d'une molécule d'ADN est transcrite en ARN puis traduite en protéine. Les termes de transcription et de traduction font explicitement appel à des procédures déterministes de type téléologique. Les processus sous jacents sont cependant tout autres, faisant intervenir des équilibres enzymatiques complexes. Ne pas prendre en compte ces phénomènes probabilistes, signifierait mettre de côté une partie de l'explication. Il en est de même pour l'image de l'hormone se fixant à son récepteur (la clef dans la serrure) ou de l'interaction antigène/anticorps.

---

[4] Une vue poussée à l'extrême de ces idées est "la théorie du gène égoïste" de Dawkins (1990). Pour l'auteur, la vie n'a aucun sens, les êtres vivants n'étant là que comme "moyens efficaces" pour assurer la transmission des gènes qui les constituent.

Approches opérationnelle et symbolique ne sont pas présentes seulement en biologie. En physique par exemple, je pense que la physique quantique est une approche de type opérationnelle alors que la théorie de la relativité est avant tout une approche symbolique. L'une est partie des propriétés des atomes et l'autre des observations macroscopiques du mouvement des astres. L'effort extraordinaire que consacrent aujourd'hui les physiciens à faire le lien entre ces deux théories, qui sont contradictoires sur de nombreux points, est à mon avis similaire à ce qui se passe aujourd'hui en sciences cognitives pour rapprocher le comportement des organismes et les propriétés électrophysiologiques des neurones. Les sciences humaines et la psychologie expérimentale ont une approche *top-down* symbolique alors que l'électrophysiologie à une approche inverse. La modélisation, quant à elle, tente de relier ces deux extrêmes.

Il ne faut donc pas rejeter en masse les approches symboliques car elles ont parfois une capacité prédictive supérieure aux approches opérationnelles. Varela (1989) fournit une définition d'un symbole acceptable. Un symbole acceptable est un symbole qui reste à l'intérieur de la clôture opérationnelle d'un système. Cela signifie que le symbole ne fait pas référence à un élément extérieur au système. Par exemple, les tribases (ou codons) au sein de l'ADN, chacune codant pour un acide aminé, sont, jusqu'à preuve du contraire, un excellent symbole pour mieux comprendre et expliquer la dynamique du système cellulaire. Cette approche symbolique présente un aspect pédagogique indéniable, mais ne constitue pas l'explication du système en elle-même.

Dans le cas d'une approche purement symbolique, il faut être conscient que l'explication est incomplète et qu'il faudra ensuite donner une implémentation aux symboles, une explication au système. Cet aspect est la plupart du temps négligé, les auteurs se contentant parfois de fournir une explication biologiquement plausible. L'implémentation exacte de leur modèle ne semble pas les concerner plus avant. Partant d'une approche plutôt symbolique, c'est pourtant sur ce dernier point que j'ai concentré mes travaux de modélisation.

Dans la partie consacrée à la modélisation de ma thèse, je suis parti d'une approche symbolique, le codage asynchrone de l'information imaginé par Simon Thorpe (1990) et le codage par rang dont les propriétés computationnelles sont exceptionnelles. J'ai tenté de relier cette théorie avec le comportement des neurones réels, allant d'une description de bas niveau sur le comportement d'un seul de ces neurones, à une description de très haut niveau décrivant l'interaction de millions de neurones pour traiter le contenu d'images naturelles. Tout au long du manuscrit, on pourra voir que je me rattache aux deux types de représentation, symbolique et opérationnelle. Je tiens cependant à préciser une fois encore, que de mon point de vue,

seule l'approche opérationnelle permet d'expliquer le système. Dans le dernier paragraphe, je retrace l'histoire de la cybernétique, qui est en fait, comme nous allons le voir, à l'origine de toutes ces questions.

1.5 - Aux origines de la modélisation : la cybernétique

> *Recently, traditionnal AI - Artificial Intelligence - has been challenged by an old charge under a new name, "connectionism".*
>
> Anya Hurlbert et Tomaso Poggio (1988)

La cybernétique est un mouvement né dans les années 1950 qui a presque disparu au cours des années 1970, écrasé par l'intelligente artificielle, et renaît en partie aujourd'hui sous une nouvelle forme que l'on appelle généralement "connexionisme" ou "néo-connexionisme", le préfixe néo se référant en quelque sorte à la cybernétique. Les systèmes autopoïétiques dont nous avons traités dans les paragraphes précédents font également partie de cette nouvelle vague. Je me considère moi-même comme un néo-connexionniste à part entière, ce qui signifie que dans le cadre de mes travaux, les réseaux interconnectés de neurones sont capables de rendre compte, par leur dynamique d'interactions complexes, des traitements effectués au sein du cerveau.

Bien que l'approche connexionniste ne semble rien avoir d'original aujourd'hui quand on connaît les découvertes concernant la puissance de calcul des réseaux connexionnistes (Rosenblatt, 1961; Hopfiled, 1982; Kohonen, 1982), une telle approche était il y a encore quelques années très mal ressentie tant du point de vue de l'intelligence artificielle que des neurosciences. Les raisons sont avant tout historiques.

Dans les années 1950, la cybernétique voyait le jour, science qui avait pour vocation d'expliquer les processus mentaux en termes de dynamique au sein de réseaux de neurones formels[5]. Cette approche avait pour but de réconcilier biologie et mathématiques tout en englobant psychologie expérimentale et sciences humaines. Bien que les objectifs soient, *a*

[5] Aujourd'hui le sens initial du terme cybernétique a été détourné. Dans le petit Robert, on peut lire : la cybernétique est la "La science constituée par l'ensemble des théories relatives au contrôle, à la régulation et à la communication dans l'être vivant et la machine". Je m'intéresserai donc en particulier à la cybernétique des années 50, dans laquelle, peu de gens le savent, les réseaux de neurones étaient déjà présents et les questions déjà particulièrement axées sur les processus mentaux.

priori, les mêmes que les sciences cognitives aujourd'hui, les nombreuses erreurs commises à cette époque font que les sciences cognitives renient totalement cet héritage. La principale différence entre la cybernétique et les sciences cognitives tient à la détermination des premiers cybernéticiens à tout ramener aux processus physiques et mathématiques.

Schématiquement, deux vues se sont opposées dans les années 50 concernant la cognition. D'un côté les idées de Von Neuman (1903-1957) pour lequel la cognition était orientée directement vers la résolution de problèmes. Von Neuman est par exemple l'instigateur de l'ordinateur séquentiel. Cette approche est encore dominante aujourd'hui et a mené aux recherches en intelligence artificielle que l'on connaît. À l'opposé, les idées de McCulloch (1898-1972) et Wiener[6] (1894-1964) qui incarnent le mouvement cybernétique : pour eux la cognition résultait du changement d'état de systèmes autonomes, changement d'état qui permettaient d'effectuer des traitements complexes. McCulloch était neuropsychiatre de formation et incarne parfaitement la philosophie des modèles que je présenterai. S'inspirant de la biologie, il a tenté de faire émerger des comportements au sein de réseaux de neurones formels capables de rendre compte de processus cognitifs. À l'époque, on souhaitait cependant à tout prix expliquer ces processus en termes mathématiques. Aujourd'hui, le chercheur est plus modeste face à son modèle, étudiant ses comportements sans vouloir absolument en formaliser la dynamique.

L'échec de la cybernétique tient à deux raisons principales. La première est que, du fait de son ancrage très fort dans les mathématiques, l'information et les processus dynamiques, la cybernétique a été absorbée et écrasée par l'intelligence artificielle symbolique. Les pères de l'intelligence artificielle moderne sont pour certains des descendants de la cybernétique qui constitue à leurs yeux l'âge de pierre, une période à oublier. Les résultats incroyables de l'intelligence artificielle à ses débuts, tant au niveau théorique qu'en ce qui concerne les applications, ont poussé des chercheurs[7] à affirmer dans les années 1960 que d'ici 10 ans le problème de la machine pensante serait résolu. Une anecdote qui fait toujours sourire les milieux scientifiques est celle de cet étudiant au MIT, engagé pour un stage d'été par Marvin Minsky[8] à la fin des années 60, et à qui l'on avait demandé de reproduire les traitements effectués par le système visuel à l'aide d'un système expert. Inutile de dire qu'aujourd'hui ils y

[6] Le logicien et philosophe Norbert Wiener est considéré, à tort semble-t-il, comme étant le père de la cybernétique pour son introduction à la cybernétique et à la théorie de l'information (1961). Dans son ouvrage "Aux origines des sciences cognitives", Jean-Pierre Dupuy (1999) attribue à Warren McCulloch la paternité de cette entreprise.

[7] Newell et Simon par exemple.

[8] L'un des pères de l'intelligence artificielle. Minsky était par exemple l'élève de McCulloch.

travaillent encore ! En fait, la perception, un processus apparemment automatique, semblait bien plus simple à simuler que la résolution de problèmes symboliques. Tout cela a sans doute contribué au fait que la cybernétique soit tombée en désuétude durant plusieurs années.

La seconde raison de l'échec de la cybernétique est qu'elle n'a pas pu se rattacher aux sciences humaines, comme la psychologie, la sociologie ou la psychanalyse pour subsister pendant ce raz-de-marée symbolique. Au cours des 11 conférences fondatrices, les cybernéticiens, dont faisaient partie quelques représentants des sciences humaines, se sont acharnés à tenter de réduire les sciences humaines à des processus dynamiques et physiques. Tout au plus, la psychologie expérimentale pouvait-elle trouver sa place à titre consultatif sur des problèmes techniques. À l'avènement de l'intelligence artificielle, aucun représentant des sciences humaines ne s'est donc inquiété de sauver la cybernétique.

La nouvelle cybernétique ou néo-connexionisme, doit donc prendre en compte ces leçons du passé, ne pas se restreindre à une approche mathématique et savoir intégrer les sciences humaines et sociales sans vouloir absolument les englober. Nous reviendrons sur ces problèmes à travers l'étude des processus cognitifs comme les sentiments dans la conclusion.

Dans le chapitre qui suit, je vais introduire l'architecture du système visuel, ainsi que le comportement des neurones qui le composent. L'organisation et la dynamique des réponses des neurones dans le système visuel constitueront la base qui nous permettra à la fois d'interpréter nos résultats expérimentaux et de construire des modèles.

2

Neurophysiologie de la vision chez le primate

We are so familiar with seeing, that it takes a leap of imagination to realise that there are problems to be solved. ...We are given tiny, distorted, upside-down images in the eyes, and we see separate, solid objetcts in the surrouding space. From the patterns of stimulation on the retina we perceive the world of object, and this is nothing short of a miracle.

Richard Gregory (1972)

Ce chapitre introduit les bases de ce qui est aujourd'hui connu du fonctionnement, *quasi* miraculeux comme le dit si bien Gregory, du système visuel. Je m'attacherai, en particulier, à décrire la sélectivité et la dynamique de réponse des neurones dans le cortex inféro-temporal dont les réponses aux objets sont invariantes à de nombreuses transformations. Cette introduction, je l'espère, permettra d'éclairer la suite de mes travaux qui, bien qu'indirectement reliés à l'électrophysiologie, s'inspirent directement des résultats obtenus dans cette discipline[1].

[1] Les bases de l'enregistrement des neurones et de leur fonctionnement seront supposés connus. Pour une revue plus générale cf. Kandel et al. (1991).

2.1 - De la rétine vers le cortex

L'étude des propriétés des neurones du système visuel remonte à près de 40 ans et est marquée par la découverte de David Hubel et Torsten Wiesel[2]. Dans une série d'articles publiés dans les années 1960 (Hubel et Wiesel, 1962; Hubel et Wiesel, 1968), ils rapportent l'existence de neurones dans le cortex occipital qui répondent de façon sélective à la présentation d'une barre de certaine orientation dans le champ visuel.

Avant de décrire ces travaux et ceux qui ont suivi, revenons un peu en arrière. Le point de départ incontournable pour étudier le système visuel se situe à l'origine de notre expérience visuelle : l'œil. Les cellules photoréceptrices de la rétine sont sensibles à des variations de luminance locales, c'est-à-dire au nombre de photons qui peuvent les stimuler. Elles transforment le signal lumineux à certaines longueurs d'onde (i.e. couleur) en une variation de potentiel électrique. Ces cellules sont extrêmement sensibles et l'occurrence d'un seul photon est parfois suffisante pour les stimuler. Au sein de la rétine, les signaux générés par ces cellules photoréceptrices sont relayés par des cellules intermédiaires[3] jusqu'aux cellules ganglionnaires qui les transmettent, sous forme de potentiels d'action, vers le cortex visuel.

Chaque cellule ganglionnaire est sélective à un contraste de luminance centre/pourtour en intégrant les réponses de plusieurs photorécepteurs, certains excitant la cellule et d'autres l'inhibant. Cela signifie que la cellule ganglionnaire répond maximalement quand un spot de lumière est entouré d'une zone sombre - cellules centre-ON - ou l'inverse - cellules centre-OFF. Certaines cellules ganglionnaires, notamment au niveau de la fovéa (située au centre de la rétine), sont également sélectives à des contrastes chromatiques, par exemple un spot vert entouré de rouge (Kufler, 1953). Outre la distinction cellules centre-ON et centre-OFF, il existe également deux principaux types de cellules ganglionnaires. Tout d'abord, les grosses cellules magnocellulaires, caractérisées par des latences de réponse très rapides, une forte sensibilité aux variations de luminance achromatique et des champs récepteurs - zones d'intégration - relativement étendus. À l'opposé, les petites cellules parvocellulaires présentent des latences de réponse assez longues, sont sélectives à des variations de luminance chromatique et possèdent des champs récepteurs relativement petits (Hubel et Wiesel, 1972). Ces deux types de cellules ganglionnaires sont à l'origine des deux principales voies dans le système visuel, la voie magnocellulaire plus particulièrement impliquée dans le traitement du

[2] Le prix Nobel de médecine a été décerné en 1981 à David Hubel, Torsten Wiesel et Roger Sperry pour l'ensemble de leurs travaux sur le système visuel.

[3] Cellules bipolaires, amacrines et horizontales.

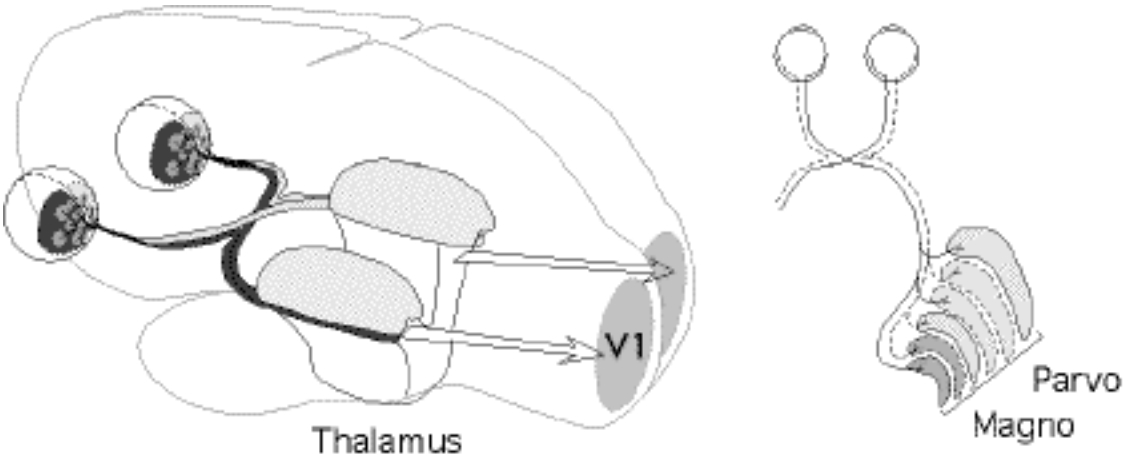


Figure 2.1 : flots d'information visuelle de la rétine vers le cortex visuel primaire (V1) en passant par le LGN dans le thalamus. À droite, la structure laminaire du LGN est représentée, dans laquelle les projections en provenance de chaque œil, des neurones magnocellulaires et parvocellulaires sont ségrégées. Adapté de Kelly (1985).

mouvement et de la position des objets, et la voie parvocellulaire intervenant dans l'identification des objets. Nous verrons que cette distinction n'est pas si tranchée : nos résultats, en accord avec des données électrophysiologiques (Ferrera et al, 1994), laissent penser que les neurones magnocellulaires interviendraient également dans la phase précoce de reconnaissance des objets.

Chez le primate, les neurones ganglionnaires projettent ensuite vers le cortex visuel en passant par le corps genouillé latéral dans le thalamus (LGN[4]). Avant d'atteindre le LGN, au niveau du chiasma optique (figure 2.1), les signaux visuels en provenance d'un hémichamp visuel donné (droit ou gauche) sont ségrégés pour être redirigés vers l'hémisphère cérébral opposé. Les neurones du LGN sont généralement considérés comme des neurones relais qui n'effectueraient pas de traitement particulier chez le primate[5]. Bien que ces neurones reçoivent de nombreuses entrées corticales, leur rôle semble être, en fonction de l'état de veille, de transmettre ou de ne pas transmettre les signaux rétiniens (i.e. Wiesel et Hubel, 1966). De plus, la structure laminaire du LGN ségrégue à la fois les projections issues des cellules ganglionnaires magnocellulaires et parvocellulaires, et les projections en provenance de l'œil ipsilatéral et de l'œil contralatéral.

[4] Lateral Geniculate Nucleus. J'adopte ici la nomenclature internationale.

[5] A la différence de chez le chat par exemple.

2.2 - Le cortex visuel primaire

Du LGN, les neurones projettent ensuite vers le point d'entrée des informations visuelles dans le cortex : le cortex visuel primaire, encore appelé aire visuelle V1. Cette aire corticale a été étudiée de façon extensive : dès 1940, il était clair que l'organisation des cellules reprenait celle de la rétine (Talbot et Marshall, 1941) et je parlerai, par la suite, d'organisation rétinotopique pour décrire ce type d'arrangement. Cependant, comme je l'ai déjà mentionné, nous devons la découverte la plus importante à Hubel et Wiesel dans les années 60 pour avoir montré que les cellules sont sélectives à des contours orientés. Quand une barre visuelle est présentée dans le champ visuel d'un animal anesthésié ou vigile, les neurones dans V1 ne répondent que pour une certaine orientation de la barre. Hubel et Wiesel (1968) rapportent également une organisation des neurones en colonnes corticales tangentes à la surface du cortex : les neurones au sein d'une colonne corticale de cortex sont sélectifs aux mêmes stimuli[6]. De plus, ils décrivent des groupes de neurones sélectivement stimulés par les signaux provenant de l'un des deux yeux : dans l'exemple précédent de la barre, le neurone est plus ou moins sélectif suivant que la barre est présentée à un œil ou à l'autre. Ils observent enfin des déplacements de sélectivité à l'orientation relativement continus d'à peu près 2,5° d'une colonne à l'autre.

Des travaux plus récents ont montré que les neurones magnocellulaires et parvocellulaires projetaient vers différentes couches du cortex : les neurones parvocellulaires projetant dans la couche $4C_\beta$ et les neurones magnocellulaires dans la couche $4C_\alpha$ (figure 2.2). Les neurones de la couche $4C_\beta$ - voie parvocellulaire - projettent en majorité vers les couches 2 et 3, dans lesquelles les neurones seraient regroupés pour répondre soit au contour, soit à la couleur. Les régions répondant aux couleurs sont appelées blobs et autour de ces blobs - régions interblobs - les neurones seraient sélectifs à des contours orientés[7] (Gouras & Kruger, 1979). Ces neurones, sélectifs à l'orientation à une position donnée dans le champ visuel, sont communément appelés "cellules simples". La modélisation des propriétés de ce type de cellule a rencontré un vif succès et je propose également dans ma thèse un modèle biologiquement plausible d'émergence de la sélectivité à l'orientation dans la seconde partie.

---

[6] Le cortex est en fait une surface laminaire repliée sur elle-même. On distingue classiquement 6 couches de neurones auxquelles on attribue des fonctions différentes, la couche numéro 1 étant celle située à la surface extérieure du cortex. Une colonne corticale, perpendiculaire à la surface, traverse toutes les couches.

[7] La présence de couleur n'est pas critique dans la réponse de ce type de neurone.

Les neurones de la couche $4C_\alpha$ - voie magnocellulaire - projettent vers ceux de la couche 4B, sélectifs à un contour orienté en mouvement dans une certaine direction. Par exemple, ces neurones peuvent décharger pour une barre orientée se déplaçant dans une direction perpendiculaire à son orientation et ne pas répondre quand la barre se déplace dans la

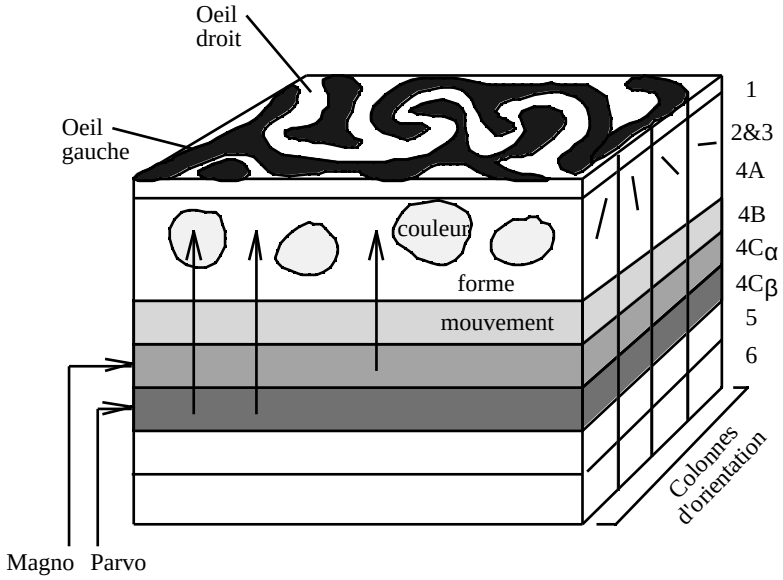


Figure 2.2 : représentation de l'organisation très spécialisée des neurones dans le cortex visuel primaire V1. La couche $4C_\beta$ reçoit majoritairement des afférences parvocellulaires du LGN et la couche $4C_\alpha$ des afférences magnocellulaires. Les neurones sont sélectifs à différentes orientations et intègrent des informations en provenance d'un œil ou de l'autre. Adapté de Livingstone et Hubel (1988).

direction opposée. On appelle généralement ces neurones "cellules complexes" par opposition aux "cellules simples". Savoir si la sélectivité de ces neurones dépend de celles des "cellules simples" ou de la seule intégration des signaux en provenance du LGN est une question encore ouverte. Les cellules dans V1, simples et complexes, sont également sélectives à la distance du stimulus présenté. Ces cellules intégrant des informations en provenance des deux yeux ne déchargent que si la barre orientée est présentée à une distance donnée et il semble même que la position des yeux module les décharges de ces neurones (Trotter et Celebrini, 1999).

Les interactions intervenant au sein des couches neuronales sont encore relativement mal comprises. On sait cependant que les neurones de deux colonnes corticales voisines interagissent fortement. On considère généralement que les interactions locales sont excitatrices, que les interactions à plus longue distance sont inhibitrices[8]. Le rôle de ces connexions est encore très controversé et est le sujet d'intenses recherches. Il est probable que ce type d'interaction organise la répartition de la sélectivité des neurones dans V1 en fonction des contraintes environnementales, c'est-à-dire des stimuli du monde extérieur (i.e. Kohonen

[8] Les connexions à très longue distance seraient excitatrices mais très sélectives et pourraient intervenir dans le traitement des textures régulières (Stemmler et al, 1995).

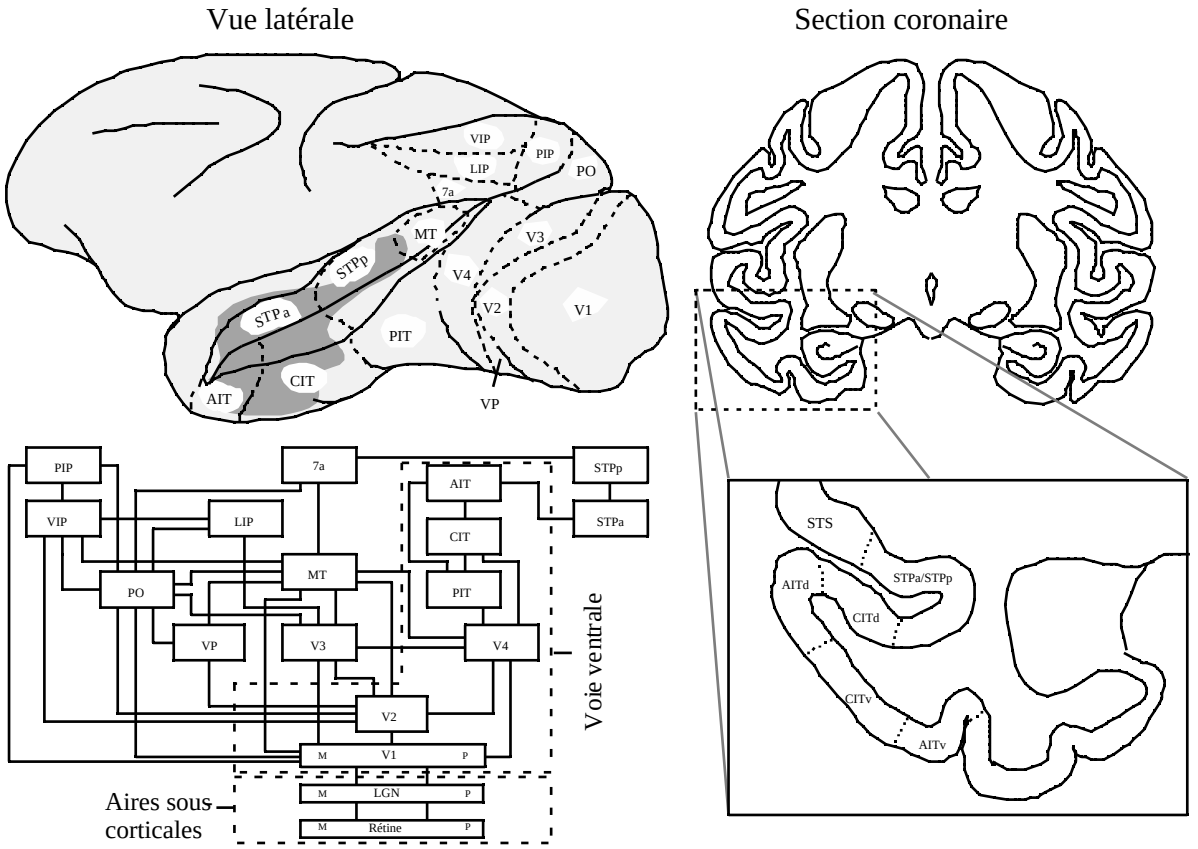
(1982) le précurseur de ce type de travaux). On verra également que ces interactions latérales ont peut-être un rôle de filtrage pour le passage des informations visuelles à l'aire neuronale de niveau supérieur (Grossberg, 1999).

Pour résumer le comportement des neurones dans V1, on peut dire qu'il existe des groupes de neurones sélectivement activés par certains attributs de l'image comme la couleur et le contour et par différents neurones afférents magnocellulaires et parvocellulaires en provenance de l'un des deux yeux. L'activité des neurones sélectifs aux contours est ensuite modulée par différents facteurs comme l'orientation, le mouvement, la profondeur, la direction du regard et probablement bien d'autres phénomènes. La compréhension des mécanismes intervenant dans V1 est vitale car toutes les aires corticales, bien que possédant des sélectivités différentes, présentent une organisation similaire, c'est-à-dire une structure laminaire à 6 couches et des interactions latérales excitatrices et inhibitrices.

2.3 - Organisation des aires visuelles

Les neurones de V1 projettent majoritairement vers l'aire visuelle V2, bien qu'il existe de nombreuses projections de l'aire V1 vers l'aire temporale médiale MT. Dans l'aire V2, les neurones se regroupent en fonction de leur sélectivité aux contours, à la couleur et à la profondeur (Livingstone et Hubel, 1982). V2 se caractérise par la présence de cellules capables de répondre à des contours illusoires : le contour n'est pas réellement présent dans l'image, mais est suggéré par la présence de deux barres colinéaires séparées par un espace. Il est possible que cette propriété se base sur la présence de cellules *end-stopped* dans V1, sélectives à l'extrémité d'un contour. Dans l'étude de von der Heidt et Peterhans (1989), 32 % des cellules répondaient à des contours illusoires alors que dans V1, leur nombre est bien plus faible (2 %). On peut imaginer que cette propriété est particulièrement utile pour la complétion de contours d'un objet qui seraient occultés par un autre objet.

V2 projette ensuite vers de nombreuses aires corticales. Dans leur article de 1988, Zeki et Shipp introduisent l'idée que chaque aire corticale serait spécialisée dans le traitement d'un attribut de l'image. Ils suggèrent que V3 serait impliquée dans le traitement des formes, V4 dans le traitement de l'orientation et de la couleur et MT et V5 (MST) dans le traitement du mouvement. Cette ségrégation est sans doute exagérée et l'on a par exemple montré depuis que les neurones de V4 étaient sélectifs à certains stimulus complexes (Desimone et Schein, 1987). Cependant, elle a tout de même le mérite d'introduire une classification des aires visuelles dont je présente une version figure 2.3. Pour revenir à V2, les neurones de V2



V1-V4: Aires visuelles 1 à 4
VP: Postérieur Ventral
MT: Temporal Médian
7a: Aire de Broadman 7a

CIT: Inféro-temporal Central
AIT: Inféro-temporal Anterieur
PIT: Inféro-temporal Postérieur
STP: Temporal Superieur Multimodal

VIP: Intra-pariétal Ventral
LIP: Intra-pariétal Latéral
PIP: Intra-pariétal Postérieur
PO: Pariéto-Occipital

Figure 2.3 : organisations des différentes aires corticales chez le singe. La zone de couleur sombre sur la vue latérale représente la zone dans laquelle des neurones sélectifs aux visages ont été enregistrés. Le diagramme indique la voie ventrale V1-V2-V4-(AIT-PIT-CIT) impliquée dans la reconnaissance des objets. À droite, la coupe coronaire présente les subdivisions du cortex inféro-temporal. Adapté de Tovee et al (1994) pour l'organisation des aires corticales et Felleman et Van Essen (1991) pour le diagramme.

projettent en masse vers V4 qui eux-mêmes projettent vers le cortex inféro-temporal (IT). L'analyse de la connectivité de ces zones montre que l'on a probablement un traitement en cascade dans IT (V4-PIT-CIT-AIT) qui s'étend peut-être même jusqu'au sulcus temporal supérieur (STPa-STPb) (Seltzer et Pandya, 1978; Baiser et al, 1991). Des lésions du cortex inféro-temporal ainsi que des enregistrements unitaires de neurones chez l'animal éveillé ont conduit certains auteurs à attribuer à cette voie un rôle dans la reconnaissance d'objets. Par opposition, la voie dorsale (V1-V2-V3-MT) serait impliquée dans la détermination de la position et du mouvement des objets. Dans une expérience princeps, Underleider et Mishkin (1982) ont montré qu'après une lésion du cortex inféro-temporal, un singe ne pouvait plus

choisir un objet en fonction de sa forme. Par opposition une lésion du cortex pariétal[9] ne permettait plus au singe, en présence de plusieurs petites trappes et d'un objet, de choisir la trappe la plus proche de l'objet. Une lésion du cortex inféro-temporal n'affectait pas les performances du singe dans la seconde tâche et une lésion du cortex inféro-pariétal n'affectait pas ses performances dans la première tâche. Les auteurs en ont conclu que le cortex inféro-temporal serait impliqué dans l'identification des objets alors que le cortex inféro-pariétal serait responsable de la localisation des objets. Plus tard, des expériences ont montré que la voie dorsale serait plus sélectivement impliquée dans la réalisation d'action sensori-motrice. L'exemple d'une patiente est très révélateur : suite à une lésion de la voie ventrale, la patiente est sévèrement handicapée pour la reconnaissance des objets. Son déficit dans la voie ventrale l'empêche, par exemple, d'insérer une feuille de papier dans une fente. Cependant elle est tout à fait capable d'effectuer l'action de poster une lettre car c'est une action sensori-motrice qui pourrait être prise en charge par la voie dorsale (Goodale et al, 1991). Ainsi est née la distinction entre voie ventrale et dorsale et, en bien des aspects, elle reprend - au niveau fonctionnel - la subdivision voie parvocellulaire et magnocellulaire (Kessels et al, 1999). Il faut cependant modérer cette distinction : il est en effet très probable que les voies dorsale et ventrale interagissent pour la reconnaissance de formes définies par de points en mouvement (Lorenceau, 1996).

Mon objectif est maintenant de détailler plus avant les données concernant les neurones sélectifs aux objets dans le cortex inféro-temporal (IT), données qui sont en relation directe avec mes travaux.

2.4 - Les neurones du cortex inféro-temporal

La taille des champs récepteurs des neurones varie dramatiquement tout au long de la voie visuelle ventrale. De 4°x4° dans V4, elle passe à 16°x16° dans PIT pour atteindre 150°x150° dans AIT (Rolls, 1992; Tanaka, 1996)[10]. La taille des champs récepteurs est directement corrélée à l'invariance de la sélectivité des neurones par rapport à la position des objets présentés. La sélectivité des neurones au sein de cette voie semble également suivre la même logique, les neurones dans V4 étant sélectifs à des objets de complexité intermédiaire - formes géométriques simples, yeux... - et les neurones de IT à des stimulus plus complexes comme nous allons le voir.

---

[9] IP dans la nomenclature que j'ai utilisée mais on trouve également le terme PG.

[10] A 1 mètre de distance, cela correspond respectivement à un écran de 7 cm^2, 28 cm^2 et 2 m^2.

On a tout d'abord montré que les neurones dans IT étaient sélectifs aux parties du corps[11]. Ils peuvent ainsi répondre sélectivement aux mains (Gross, 1972), aux visages (Gross, 1972; Perrett et Rolls, 1982) et à diverses parties du corps (Wachsmuth et al. 1994). D'autres neurones dans IT sont également sélectifs à différents objets naturels et artificiels (Tanaka, 1996; Booth et Rolls, 1998; Vogels, 1999). Il est généralement admis que les neurones dans IT sont sélectifs à tous les types d'objets que l'on peut rencontrer. Vogels (1999) a par exemple observé des neurones sélectifs aux arbres dans ces aires neuronales. Récemment Booth et Rolls (1998) ont montré qu'après un contact prolongé avec des jouets en plastique de diverses formes et de différentes couleurs, une fraction des neurones de IT répondaient de manière sélective à ces objets. Les neurones de IT répondent également très fortement sur des formes abstraites (Tanaka, 1996) : Tanaka et ses collaborateurs ont en effet enregistré les réponses des cellules suite à différentes transformations d'un stimulus élaboré[12]. Le stimulus était progressivement dégradé jusqu'à obtenir le stimulus minimum sur lequel le neurone déchargeait. 2 % des neurones dans V4 répondent sélectivement sur des stimuli élaborés, 9 % dans PIT et 45 % dans AIT. L'équipe de Tanaka a également montré que la sélectivité des neurones à travers les couches neuronales (de 2 à 6) était plus ou moins invariante au stimulus présenté (Fujita et al, 1992), de la même façon que dans V1, les neurones d'une colonne corticale sont sélectifs à la même orientation.

La sélectivité des neurones d'IT est extrêmement résistante aux variations des caractéristiques des objets. Les réponses de ces neurones ne semblent pas dépendre du fait que l'objet soit partiellement caché ou que ses contours soient uniquement définis par un mouvement relatif ou par une texture complexe (Vogels et Orban, 1996). Les neurones de IT répondent également de façon similaire sur un objet, que cet objet soit présenté en photo, sous forme de dessins au trait ou suggéré par un jeu d'ombre (Chadaide et al, 1999). Enfin leurs réponses sur un objet ne semblent pas dépendre de la taille de l'objet ni de sa position dans le champ visuel (Rolls, 1992; Ito et al, 1995; Vogels, 1999). Il faut cependant modérer ces affirmations car même si l'on admet généralement que la sélectivité des neurones de IT ne dépend pas de la taille du stimulus, il semble que cela soit vrai uniquement pour 21 % des neurones qui répondent à des variations de tailles jusqu'à 4 octaves. En fait 43 % des neurones de IT ne répondent pas à des variations de taille de 2 octaves (Ito et al, 1995). L'invariance à la position semble cependant mieux préservée (Ito et al, 1995).

La sélectivité des neurones dans IT à un objet peut être plus ou moins élevée. Concernant

[11] Cf. Rolls (1992), Tanaka (1996) et Logothetis et Sheinberg (1996) pour une revue sur le sujet.

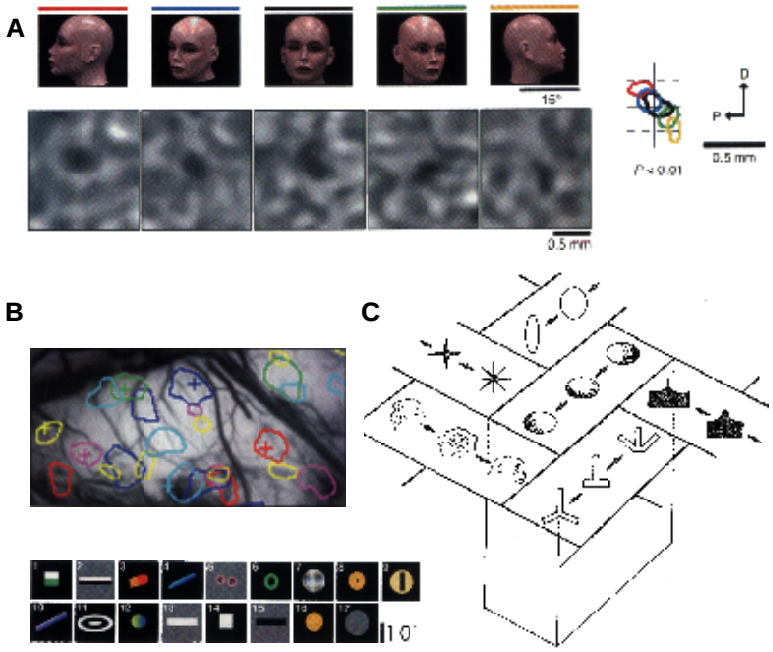


Figure 2.4 : sélectivité des neurones dans IT estimée par *optical imaging* et enregistrements unitaires. A, la sélectivité de population de neurones varie en fonction de la vue du visage (chaque visage correspond à une couleur, et l'activation neuronale est reportée dans le diagramme à droite). B, activation distribuée à la surface du cortex pour divers stimulus (chaque couleur correspond à l'un des stimuli). Les stimuli sont représentés sous la carte d'activation. C, organisation schématique de la sélectivité des neurones dans IT. Adapté de Wang et al (1996) et Tanaka (1996).

la sélectivité aux visages, certaines cellules peuvent ne répondre qu'à certaines des vues du visage d'une seule personne, parmi l'éventail des individus présentés. En ce sens, elles pourraient coder l'identité de l'individu. D'autres sont beaucoup moins sélectives et répondent à plusieurs des individus testés (Perrett et Rolls, 1982; Baylis et al, 1985; Perret et al, 1991). Certaines cellules sélectives aux visages ou aux parties du corps le sont aussi à un individu de profil ou de face. En règle générale, on peut dire que les neurones sélectifs aux visages ou à une partie du corps sont sélectifs à la vue, de face ou de côté (Wachsmuth et al, 1994; Farah, 1996) alors que les réponses des neurones sélectifs à des objets artificiels semblent plus ou moins invariantes à la vue présentée (Booth et Rolls, 1998). Cet effet semble dépendre de la familiarité du stimulus. Suite à la présentation d'un stimulus, les neurones apparaissent tout d'abord sélectifs à la vue présentée puis se dégagent de cette contrainte au fur et à mesure que les objets deviennent familiers (Logothetis et al, 1995). Récemment Wang et al (1996) ont mis au point une technique consistant à filmer avec une caméra très sensible la surface du cortex pour détecter les zones de plus forte activité. Bien que la résolution temporelle de ce type de technique soit très mauvaise[13], ils montrent que les zones d'activité correspondent à des décharges neuronales. On constate alors qu'un stimulus active plusieurs groupes de neurones et que ces zones se superposent souvent pour différents stimulus (figure 2.4).

En présence d'images naturelles, c'est-à-dire de photographies, qui sont *a priori* plus

[12] Le stimulus est dit élaboré si au moins deux formes sont nécessaires pour activer le neurone.

[13] La caméra enregistre les variations de flux sanguin dont la dynamique est bien plus lente que l'activation neuronale.

proches de l'environnement dans lequel évolue l'animal, les réponses des neurones sont faibles à la fois dans V1, V2, V3 (Galant et al, 1996) et IT (Baddeley et al, 1997), mais conserveraient la même sélectivité (DiCarlo et Maunsell, 2000). On peut interpréter cela en termes de contrastes locaux des objets et de compétition entre ces objets. Sans entrer dans des détails qui nous entraîneraient trop loin, la sélectivité des neurones dépend de la forme et de la surface du fond sur lequel l'objet est présenté (Missal et al, 1997). Dans une scène complexe, les réponses des neurones dans IT à un objet sont modulées par les réponses des neurones sélectifs aux autres objets dans la scène visuelle. Cette compétition pourrait être biaisée à la fois par le contraste des objets - les objets de fort contraste ayant tendance à prendre le pas sur les objets de faible contraste - et par l'attention spatiale (Chelazzi et al, 1998). Nous reviendrons sur ces propriétés dans la discussion générale.

Pour résumer les travaux présentés dans cette partie, le système visuel est avant tout organisé de façon hiérarchique. La taille des champs récepteurs des neurones augmente de V1 vers IT : de très locaux, ils finissent par englober tout le champ visuel. La sélectivité des neurones devient également de plus en plus fine, les neurones de V1 étant principalement sélectifs à des barres orientées alors que les neurones d'AIT sont sélectifs à des visages et à des objets complexes. Cependant la sélectivité des neurones dans IT a une limite et il ne semble pas exister de neurones sélectifs à un objet précis. Ce type de comportement semble être un compromis entre un codage hyper spécifique, qui ne serait pas suffisamment souple pour s'adapter à de nouveaux objets et un codage très distribué et grossier, où les interférences entre les neurones seraient trop fortes.

Cette section avait pour but d'appréhender le comportement des neurones dans le système visuel. Comme nous le verrons par la suite, de nombreux travaux s'accordent à dire que ces types de neurones sont responsables à la fois de la décision perceptive et de la perception consciente du sujet (Sheinberg et Logothetis, 1997; Logothetis, 1998). Dans le paragraphe suivant, nous allons nous intéresser à la dynamique de réponses des neurones dans le système visuel.

2.5 - Dynamique des réponses neuronales

Malheureusement pour l'électrophysiologiste - et pour le modélisateur - les neurones du système visuel n'ont pas une activité bimodale, émettant de nombreux potentiels d'action en présence du stimulus pour lesquels ils sont sélectifs et restant silencieux pour d'autres stimuli, tel un signal lumineux qui s'allumerait ou s'éteindrait en fonction du stimulus présenté. Les

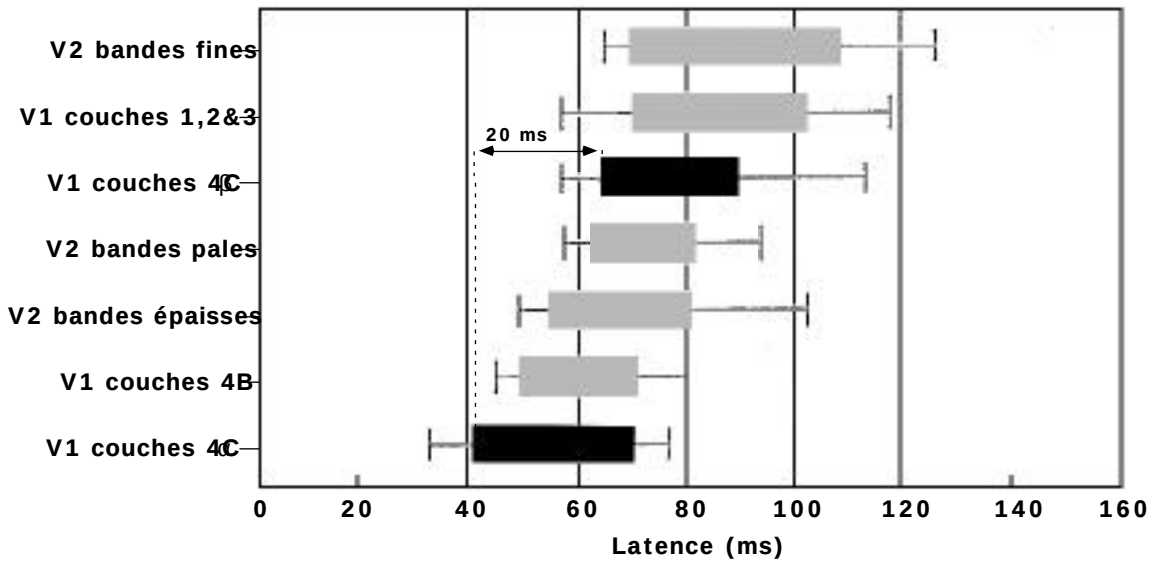


Figure 2.5 : latences de décharge dans les aires visuelles V1 et V2. Le rectangle central correspond aux centiles 25 % à 75 % (la barre verticale au centre représentant la médiane). Les barres verticales des deux cotés du rectangle indiquent les centiles 10 % à gauche et 90 % à droite. On constate que les neurones de la voie magnocellulaire (4Cα) sont environ 20 ms plus rapides que ceux de la voie parvocellulaire dans V1 (4Cβ). Les différentes bandes dans V2 correspondent aux bandes de cytochrome oxydase. Les neurones des bandes fines dans V2 sont sélectifs à la couleur et sont donc probablement sous l'influence du système parvocellulaire, ce qui explique leur lenteur. Les neurones des bandes épaisses dans V2 ont les latences les plus rapides et sont probablement activés par la voie magnocellulaire puisqu'ils projettent ensuite vers MT. Enfin, les neurones des bandes pâles seraient sélectifs aux contours et présentent des latences intermédiaires : ils pourraient être activées à la fois par des entrées magnocellulaires et parvocellulaires. Adapté de Nowak et Bullier (1997).

travaux, que j'ai résumés dans le paragraphe précédent, considèrent tous qu'un neurone est sélectif si, suite à la présentation d'un stimulus, sa fréquence de décharge augmente par rapport à son activité spontanée. Cependant, il semble que la latence de décharge des neurones, par rapport à la présentation du stimulus, soit également un indicateur de la sélectivité du neurone (Richmond et Optican, 1990; Celebrini et al, 1993). Il existe un intense débat à ce jour concernant le type de code utilisé par les neurones, latence ou fréquence de décharge par exemple, auquel je ne souhaite pas prendre parti : comme je l'ai mentionné dans le chapitre précédent, il est peu probable que les neurones "utilisent" un code précis ou "représentent" des caractéristiques de l'objet. Il y a en effet tout lieu de penser que les neurones sont de simples unités autonomes tentant de survivre avec les contraintes qui leur sont imposées par la nature en termes d'interactions avec leurs voisins. Les propriétés globales du réseau émergent des interactions locales. Au niveau global, il pourraient donc exister plusieurs codes neuronaux qui dépendraient des multiples contraintes au niveau cellulaire.

Sans donc entrer dans le débat du code "utilisé" par les neurones, il est cependant vital de caractériser la dynamique de réponse des neurones dans le système visuel. Un des résultats compatible avec l'organisation hiérarchique du système visuel est que la latence d'activation

des différentes aires corticales serait en relation avec leur niveau hiérarchique (figure 2.5). Cette affirmation est cependant à modérer. Les neurones dans la voie dorsale - majoritairement activée par les neurones magnocellulaires - présentent des latences de réponse à un stimulus flashé plus rapides que ceux de la voie ventrale. Dans le LGN, les latences de réponses des neurones magnocellulaires sont au moins 20 ms plus précoces que celles des neurones parvocellulaires (Nowak et al, 1995; Nowak et Bullier, 1997). Cet effet est également visible pour les premières décharges des neurones, 10 ms plus rapides pour les neurones magnocellulaires dans le LGN (Maunsel et al, 1999). On retrouve ce biais dans l'aire STP - à l'un des niveaux hiérarchiques les plus élevés du système visuel - où certains neurones sont sélectifs à la forme et d'autres au mouvement : les neurones sélectifs au mouvement déchargent environ 30 ms plus tôt que ceux qui sont sélectifs à la forme des objets (Oram et Perrett, 1996).

Comme je l'ai déjà mentionné, il est couramment accepté de considérer la voie magnocellulaire comme intervenant dans la sélectivité à la position et au mouvement des objets dans la voie dorsale, alors que les neurones parvocellulaires interviendraient plus spécifiquement dans la voie ventrale pour la reconnaissance des objets. Cependant les interactions entre ces deux types de neurones dans le système visuel semblent complexes[14] (Ferrera et al, 1994), et il est possible, du fait de leur différence de latence d'activation, que l'activité des neurones de la voie parvocellulaire soit modulée par celle des neurones de la voie magnocellulaire (Bullier et al, 1996; Hupe et al, 1998). Ce type de modulation interviendrait au sein même de la voie ventrale, de V2 vers V1, mais pourrait aussi intervenir entre la voie dorsale et la voie ventrale : les neurones de la voie magnocellulaire activeraient la voie dorsale qui elle-même modulerait l'activité des neurones dans la voie ventrale. Du point de vue computationnel, cela apporte une solution au problème du liage de la position et de l'identité des objets : les neurones de la voie ventrale, majoritairement sélectifs à l'identité des objets, seraient modulés par les signaux de la voie dorsale qui sont eux sélectifs à la position de ces objets (Vidyasagar, 1999).

La dynamique des réponses des neurones dans IT semble également varier au cours du temps. Les neurones dans IT peuvent être activés seulement 80-100 ms après la présentation d'un stimulus (Vogels, 1999; Perrett et al, 1982). La première partie de cette réponse semble

[14] De façon apparemment contradictoire, l'incidence d'une inactivation des neurones de la voie magnocellulaire sur l'activité des neurones de V4 - dans la voie ventrale - semble très importante : l'inactivation des neurones parvocellulaires dans le LGN réduit l'activité de 36 % alors que cette réduction atteint 47 % dans le cas d'une inactivation des neurones magnocellulaires.

être grossièrement sélective alors que la seconde serait plus spécifique. Par exemple, Sugase et al (1999) ont montré que les décharges rapides de certains neurones étaient sélectives aux visages dans leur globalité pour se concentrer ensuite - environ 50 ms plus tard – sur les visage qui présentent une certaine expression. Nous reviendrons en détail sur la dynamique de réponse des neurones du système visuel dans le cadre des modèles que je vais présenter.

2.6 - Du singe à l'homme

La plupart des résultats que j'ai présentés ont été réalisés chez le singe macaque et jusqu'ici, on a implicitement considéré que le traitement visuel chez le singe et chez l'homme était similaire. Chez l'homme, il n'est pas possible d'enregistrer directement le comportement des neurones dans le système visuel, à l'exception de très rares cas où des électrodes doivent être implantées, pour localiser un foyer épileptique par exemple (i.e. Weber et Ojemann, 1995). Pour déterminer l'activité dans les aires visuelles on utilise donc des approches indirectes comme l'imagerie par résonance magnétique (IRMf) ou les potentiels évoqués. La première technique consiste à enregistrer les variations de flux sanguin au sein des aires corticales. La résolution spatiale est de quelques millimètres, mais la résolution temporelle est très faible, de l'ordre de la seconde. Les potentiels évoqués se basent sur l'enregistrement de l'activité électrique à la surface du scalp et il est donc possible d'obtenir une résolution temporelle beaucoup plus élevée. Cependant il est souvent difficile de localiser les zones activées. On utilise en général l'IRMf pour localiser les aires cérébrales coactivées et les potentiels évoqués pour déterminer la dynamique des processus sous-jacents. Je me contenterai ici de citer quelques travaux en IRMf pour comparer les zones activées chez l'homme et chez le singe. Concernant la dynamique des processus visuels que l'on peut mettre à jour avec les potentiels évoqués et que j'ai moi-même utilisés, j'y reviendrai en détail dans la partie expérimentale.

Dans la comparaison du système visuel du singe et de l'homme, il convient de distinguer les aires homologues et les aires analogues : les aires homologues ont une parenté embryologique commune aux deux espèces alors que les aires analogues renvoient à des traits résultant d'une évolution convergente (pour une revue cf. Imbert, 1999). Les expériences récentes indiquent que le système visuel du singe et de l'homme semble être "homologue", notamment pour les aires visuelles de bas niveau de V1 à V4. On connaît depuis longtemps l'organisation rétinotopique de V1 chez l'homme (Holmes, 1818). Comme chez le singe, on observe des colonnes de dominance oculaire, c'est-à-dire des neurones qui répondent

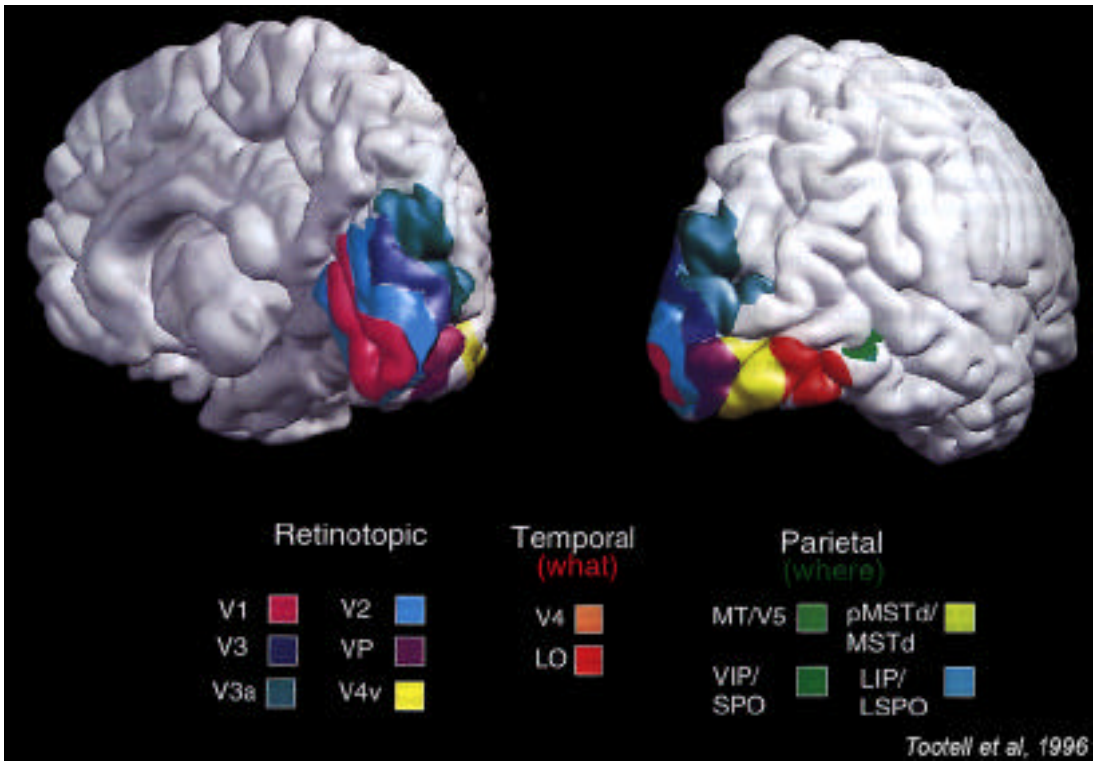


Figure 2.6 : Localisation et topographie présumées des aires visuelles chez l'homme. Les aires visuelles sont représentées en couleur à la surface du cortex à partir de données d'IRMf obtenues chez plusieurs sujets. Seul le cortex droit est représenté. La dénomination des aires corticales est identique à celle utilisée chez le singe lorsqu'une homologie topographique et fonctionnelle existe; spécifique à l'humain dans les autres cas, par exemple L[ateral] O[ccipital], L[ateral] S[uperior] P[arietal] O[ccipital] (Tootell et al, 1996).

sélectivement à un œil ou à l'autre. Elles ont d'abord été décrites anatomiquement par un marquage à la cytochrome oxydase[15], chez un patient décédé qui avait perdu tardivement l'usage d'un œil (Horton et Stryker, 1993) et ont été plus tard visualisées en IRMf chez un sujet sain stimulé de façon monoculaire (Menon et al, 1997). Autour de V1, plusieurs aires sont présentes, activées par des stimulations visuelles simples, dont on a montré en IRMf qu'elles seraient fonctionnellement homologues aux aires V2 et V3 du singe (figure 2.6; pour une revue cf. Tootell et al, 1996). De même, l'analogue de l'aire V4, sélective à la couleur et à la forme se situerait dans "l'aire de la couleur" du gyrus fusiforme (McKeefry et Zeki, 1997).

Comme on l'a vu, les principales aires du système visuel sont conservées entre l'homme et le singe. Avant d'aborder l'aire analogue d'IT chez l'homme, il convient de déterminer si, comme chez le singe, une voie ventrale, sélective à l'identité des objets, et une voie dorsale, sélective à leurs positions, est présente chez l'homme. Une expérience va en effet dans ce

[15] La cytochrome oxydase marque les neurones en activité (juste après le décès du patient). Si le patient avait perdu l'usage d'un œil, seuls les neurones correspondant à l'autre œil présentent une activité et un métabolisme normaux.

sens. Dans un protocole d'appariement, les sujets devaient alterner deux tâches (Haxby et al, 1994). Dans la première tâche, ils devaient décider lequel de deux visages était similaire au visage cible. Dans la seconde, ils devaient uniquement prendre en compte la position des photographies de visages, la cible étant indiquée par une double barre ou un point disposé dans une relation spatiale conforme à celle du visage cible. Pour résumer les résultats obtenus, quand on compare les deux tâches, on observe une double dissociation : la tâche d'appariement de forme activait sélectivement le cortex occipito-temporal alors que la tâche de localisation activait de façon sélective le cortex occipito-pariétal[16]. Cela laisse donc penser qu'il existe également chez l'homme une voie dorsale, sélective à la position des objets, et une voie ventrale, sélective à leur identité (Tootell et al, 1996; Ungerleider et al, 1998).

D'autres expériences renforcent ce résultat en montrant que l'activation de zones dans la voie ventrale, analogue du cortex inféro-temporal chez le singe, semble sélective à la présence d'objets dans le champ visuel. Comme chez le singe, on trouve chez l'homme des zones corticales sélectives à des parties du corps comme les visages, les yeux, les mains ainsi qu'aux scènes naturelles et aux objets manufacturés[17] (Kanwisher et al, 1997; Chao et al, 1999; Puce et al, 1999; Gautier et al, 2000). La présentation de visages en particulier, comparée à des présentations de maison ou de mains, active sélectivement en IRMf une partie du gyrus fusiforme[18] (Kanwisher et al, 1997; O'Craven et al, 1999). Ces mêmes zones semblent également activées lors de la présentation d'animaux ou d'objets manufacturés comme des voitures (Chao et al, 1999; Gautier et al, 2000). Une spécialisation hémisphérique pourrait également exister : dans l'hémisphère droit, les aires de la voie ventrale ne semblent pas spécialisées dans la reconnaissance spécifique de certains objets alors que cet effet semble présent dans l'hémisphère gauche, l'activation dans le lobe temporal gauche étant plus antérieure pour les objets manufacturés (Moore et Price, 1999). L'étude du comportement de patients cérébro-lésés, confirment ces résultats. Des lésions dans certaines parties du gyrus fusiforme font que le sujet n'est plus capable de reconnaître les visages (Farah, 1996). De même, il existe des patients sélectivement atteints dans la reconnaissance des objets vivants animés, des objets vivants inanimés ou des objets manufacturés. Ces déficits sélectifs laissent penser qu'il existe une certaine spécialisation des aires corticales pour la reconnaissance des différentes catégories d'objets. Nous reviendrons sur ce point en détail quand nous traiterons des modèles étendus du système visuel.

---

[16] Etude réalisée en TEP (Tomographie par Emission de Positon), analogue à l'IRMf.

[17] Pour une revue, voir Cabeza et Nyberg (2000).

[18] La présentation de photographies de maison activait par contre sélectivement l'aire parahypocampique.

Comme on l'a vu, les systèmes visuels du singe et de l'homme semblent très similaires à la fois en ce qui concerne l'organisation corticale et la sélectivité. Il convient toutefois d'apporter quelques restrictions à cette vue idéalisée. En ce qui concerne l'organisation du système visuel, chez l'homme, à la différence du singe, la voie ventrale est localisée presque entièrement sur la partie ventrale du lobe temporal. De plus, la voie ventrale n'est pas aussi étendue chez l'homme que chez le singe : chez l'homme, elle n'atteint pas les parties postérieures du lobe temporal. Ces différences suggèrent que ces régions du cortex se seraient déplacées au cours de l'évolution du fait de l'émergence de nouvelles aires spécialisées comme celle du langage (Nobre et al, 1994; Damasio et al, 1996). Concernant la sélectivité des neurones et plus particulièrement la sélectivité des neurones (ou des aires corticales) aux visages, les singes et les hommes semblent également se distinguer. Les cellules dans IT chez le singe répondent aux visages indépendamment de leur orientation, les neurones répondant très bien pour des visages présentés à l'envers (Perrett et al, 1982). A l'opposé, les résultats en psychologie expérimentale indiquent que les hommes présentent des difficultés pour reconnaître des visages présentés à l'envers (Jeffreys, 1989) et donc que les cellules dans le gyrus fusiforme, capables de telles généralisations, sont probablement absentes chez l'homme. Il est cependant possible que cette différence soit dûe à l'environnement que fréquente l'animal : on peut imaginer que la présence de neurones sélectifs aux visages indépendamment de leur orientation chez le singe est due au fait que ces animaux vivent dans les arbres et doivent donc souvent reconnaître leur congénère dans des positions inhabituelles[19]. L'environnement de l'animal biaiserait donc la sélectivité neuronale : chez le singe, certains neurones d'IT deviennent en effet sélectifs à des objets que l'animal est habitué à manipuler (Booth et Rolls, 1998).

J'ai tenté d'indiquer, aussi succinctement que possible, ce qui était connu de la sélectivité des neurones chez le singe - respectivement des aires corticales chez l'homme - par rapport aux objets du monde extérieur. Très clairement, mon but n'est pas ici d'analyser plus avant les propriétés du système visuel ni de tenter de comparer les différents modèles de traitement décrits dans la littérature. Ces sujets seront abordés dans la discussion des expériences et des

[19] Cette hypothèse semble confirmée par le fait que chez le mouton, bien que certains neurones soient sélectifs aux visages, de telles invariances ne sont pas observées (Kendrick et Baldwin, 1987).

modèles que je vais présenter. Pour le moment, je préfère ne pas biaiser le jugement du lecteur vers l'un ou l'autre des modèles. Comme je l'ai indiqué dans la partie introductive, je considère qu'il faut absolument prendre en compte les contraintes imposées au système et tenter d'analyser dans quelle mesure, des processus émergents peuvent intervenir pour traiter l'information visuelle. Pour cette raison, l'analyse des décharges des neurones m'a semblé l'approche la plus pertinente car elle prend à la fois en compte la perturbation du monde extérieur et la dynamique d'interaction entre les neurones. Bien que partant d'expériences en psychophysique, je tenterai toujours de revenir à ce niveau de description et d'interpréter les résultats obtenus en terme de dynamique de décharge neuronale.

II

Catégorisation visuelle rapide chez le primate

À la différence des expériences en psychologie expérimentale classique, les expériences que je présente ne partent pas d'une hypothèse qu'il faudrait valider ou invalider. Ma démarche s'inspire de l'électrophysiologie, dans le sens où les expériences réalisées sont simples mais contraignantes pour le système visuel, et où les données obtenues permettent d'extrapoler les traitements neuronaux sous-jacents.

En réponse à la détection d'un objet ponctuellement intéressant, notre comportement s'adapte de manière extrêmement rapide. Cette capacité est vitale : la vitesse de détection et de réaction à la survenue d'un prédateur est critique pour la survie d'un organisme. Cette réaction implique que l'animal ait su déterminer si oui ou non un stimulus visuel appartient à une catégorie donnée, c'est à dire catégoriser visuellement l'objet soudainement apparu. Cette capacité est présente chez la plupart des animaux (Hernstein 1990).

L'efficacité du système visuel des primates est remarquable (Fabre-Thorpe et al., 1998 ; Thorpe et al., 1996). Dans une tâche où les sujets doivent effectuer une catégorisation visuelle de type "go/nogo" - répondant à la présentation d'une image uniquement lorsqu'elle contient un animal, le taux de réussite est de 94 % de réponses correctes, avec un temps de réaction médian de 440 ms malgré (1) la présentation brève des stimuli (20 ms), qui empêche toute possibilité d'exploration oculaire, (2) le fait que chaque image n'est vue qu'une seule fois, ce qui empêche toute possibilité d'apprentissage, et (3) le fait que les sujets n'ont aucune information ni sur le type d'animal à chercher (mammifères, reptiles, oiseaux, poissons...), ni sur leurs positions, leurs tailles, leurs orientations, ou le nombre d'animaux présents. Chez l'homme, l'analyse des potentiels évoqués a permis de montrer que les réponses cérébrales évoquées lors des essais cibles et des essais distracteurs diffèreraient de façon significative dès 150 ms. Le traitement visuel nécessaire pour cette tâche peut donc être réalisé en moins de 150 ms. Chez le singe, la contrainte temporelle pourrait être encore plus sévère (Fabre-Thorpe, et al., 1998; Fabre-Thorpe et al, 1999) puisque pour une performance légèrement inférieure à celle de l'homme (87-90 %) leurs temps de réaction médians sont beaucoup plus courts (250 ms).

Dans ce type d'étude, le choix de la tâche est loin d'être anodin et explique probablement en partie la rapidité spectaculaire que je viens de décrire. Comme je l'ai déjà mentionné dans l'introduction, la tâche choisie est à la fois très simple et très ardue. Très simple puisque même des singes rhésus, que certains considèrent encore comme incapables de former des concepts, ont des performances que l'on peut qualifier de très bonnes. Très complexe dans le sens où, à ce jour, aucun système artificiel ou ordinateur ne peut atteindre les performances ni des singes, ni des

humains dans ce domaine. Il ne s'agit pas seulement de limite de puissance de calcul de tels modèles, mais plutôt d'un déficit conceptuel car les traitements effectués par le système visuel pour atteindre de telles performances sont encore peu connus.

Le choix de la catégorie est un point délicat La catégorie, telle qu'elle est définie par l'expérimentateur, ne reflète pas forcément une classe d'objets pertinente pour le système visuel. La catégorie "animal" sur laquelle je base mes études expérimentales semble adaptée et pour l'homme et pour le singe, puisqu'elle se confond avec celle des objets vivants animés. Des études cliniques ont en effet montré que certains patients cérébro-lésés sont incapables de reconnaître des organismes vivants, et que d'autres sont perturbés dans la reconnaissance d'objets inanimés (pour une revue cf. Boucart, 1996). Le fait que des zones cérébrales soient dédiées chez l'homme au traitement de ce type de stimulus constitue une indication supplémentaire de leur pertinence en tant que catégorie. De plus, nous avons en général une réaction de recul si nous entrons en contact avec un organisme vivant animé - même statique - que nous ne connaissons pas. La réponse motrice demandée dans notre test semble donc particulièrement adaptée au stimulus présenté, puisque le sujet doit retirer sa main du bouton dès qu'il perçoit un animal.

De façon pour le moins surprenante, la dynamique de la catégorisation est un phénomène largement ignoré et aucune étude à ce jour n'a été réellement entreprise. Les travaux menés dans ce domaine se bornent à déterminer la performance des sujets en termes de pourcentage de réponses correctes et de temps moyen de réaction. Toutefois, comme nous le verrons, il est possible d'aller beaucoup plus loin dans l'étude de la performance des sujets.

Les expériences que je vais présenter ont pour but de déterminer les facteurs critiques qui permettent d'expliquer les performances de l'homme et du singe dans la tâche de catégorisation précédemment décrite[1].

L'étude de la progression des performances d’Eudora - la guenon que j'ai personnellement entraînée - au cours de son apprentissage nous permettra de déterminer les différentes stratégies qu’elle utilise, d'analyser l’évolution de ses performances, et leur stabilisation. Nous tenterons ensuite de déterminer les caractéristiques des images qui sont critiques dans la rapidité de

[1] Les détails du protocole expérimental sont présentés en annexe et tous les tests de significativité ont été introduits en tant que notes de bas de page.

catégorisation des primates humains et non-humains. L'influence de la couleur fera l'objet d'une étude approfondie mais nous aborderons aussi celle de la luminance, du contraste, du contenu de l'image. Si l'effet individuel de ces changements est minime, il est toutefois possible de classer chacune de ces caractéristiques en fonction de leur influence sur la vitesse de catégorisation.

Toujours dans le but de déterminer les facteurs clefs de la rapidité de la catégorisation, nous tenterons d'estimer l'influence de ce que le sujet connaît de la tâche. Tout d'abord, l'influence de l'attention temporelle sera évaluée en faisant apparaître les images soit à intervalle fixe, soit à intervalle variable. Nous montrerons que cela n'a pratiquement aucune influence sur la vitesse de catégorisation. Une autre question intéressante a trait à la familiarité des sujets avec les images. Des images très familières seront mélangées à des images totalement nouvelles et nous montrerons, à la fois au niveau des TRs et des PEs, qu'une image nouvelle peut être traitée aussi rapidement qu'une image familière. Pour finir, nous tenterons d'évaluer l'influence de la tâche en elle-même. Les sujets devront effectuer deux tâches, l'une de catégorisation animal/non-animal et l'autre de détection d'une image-cible unique contenant un animal. Nous montrerons que le gain en temps de réaction est corrélé avec l'activité neuronale enregistrée en potentiels évoqués. Nous montrerons également que les deux tâches semblent recruter les mêmes zones corticales mais avec des dynamiques différentes.

1

Apprentissage d'une catégorie chez le singe

In the study of primate evolution we are studying our own kin, climbing toward the summit of our own family tree.

A.S. Romer (1959)

Les capacités de catégorisation des animaux est un sujet d'intense recherche. Il semble admis que les animaux sont capables d'effectuer diverses tâches de catégorisation, le niveau et les performances dépendant de l'espèce de l'animal (Hernstein, 1990). Jusqu'à récemment, chez le primate, on réservait les catégorisations de type conceptuel - c'est-à-dire les catégories d'éléments ne possédant pas de similarité de forme - aux hominoïdes dont le macaque ne fait pas partie.

Cependant, comme je l'ai déjà mentionné, Fabre-Thorpe et al (1998) ont montré que les macaques rhésus étaient capables d'effectuer des tâches aussi abstraites que la catégorisation d'animaux ou de nourriture dans des images naturelles, qui sont des catégories conceptuelles à part entière. Les singes commettaient de plus des erreurs similaires aux sujets humains, ce qui laisse penser que les processus de catégorisation pouvaient être en partie similaires dans les deux espèces. Les auteurs ne se sont pourtant pas intéressés à l'apprentissage et je me suis donc appliqué à étudier ce phénomène.

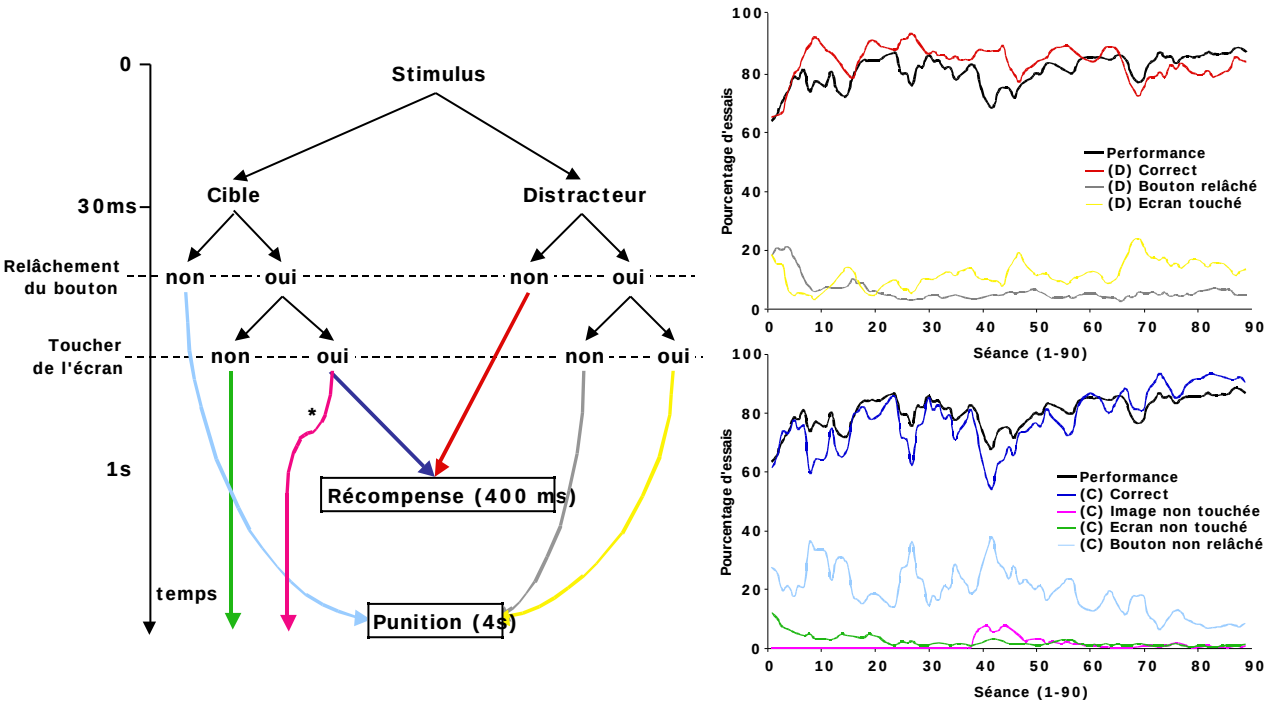


Figure 1.1 : à gauche, protocole expérimental de présentation de cibles et de distracteurs. Suivant son comportement sur les deux catégories d'images, le singe est récompensé par la distribution de jus de fraise ou puni par la réapparition de l'image pendant 4s (le signe * indique que le singe a touché l'écran en dehors de la zone où l'image était présentée). A droite, évolution de la précision d'Eudora au cours des 90 séances sur l'ensemble des images, pour les distracteurs (en haut) et les cibles (en bas). Le pourcentage de distracteurs catégorisés correctement ne semble pas évoluer alors que le taux de réussite sur les cibles augmente au cours des séance.

Dans un premier temps, j'ai tenté d'évaluer la dynamique d'apprentissage d'un tel type de catégorisation en entraînant une guenon sur la tâche animal/non animal et en utilisant un protocole strictement contrôlé. Puis j'ai reproduit les résultats de Fabre-Thorpe et al (1998).

L'apprentissage commence naturellement avec un nombre limité d'images et l'on peut se poser la question de savoir si avant de généraliser à une catégorie, le singe n'effectue pas une association stimulus/réponse.

1.1 - Matériel et méthodes - apprentissage

Le macaque que j'ai choisi d'entraîner est une femelle rhésus du nom d'Eudora âgée d'environ 3 ans à la date du début de l'apprentissage et totalement naïve à tout type de dressage. L'entraînement se décompose en trois phases :

1. Dans une première phase - 2 semaines - le singe doit simplement garder sa main droite ou gauche sur un bouton électro-sensible pendant une seconde. Si le singe réussit à effectuer cette tâche, il est récompensé par une giclée de jus de fraise envoyée par l'ordinateur

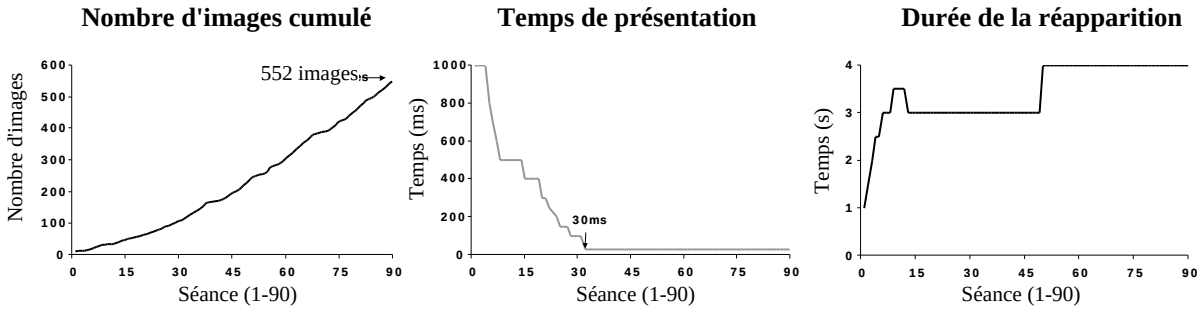


Figure 1.2: augmentation du nombre d'images au cours des séances et variation du temps de présentation des images et de la durée de leur réapparition correspondant à la punition.

enregistrant les réponses du singe. Dans le cas contraire, il n'est pas récompensé[1] (cf. annexe 1 pour le détail du dispositif expérimental).

2. Au cours d'une seconde phase - 1 mois - le singe doit appuyer sur le bouton pour faire apparaître une image contenant un animal qu'il doit ensuite toucher pour être récompensé[2]. Par la suite, des séries d'images distracteurs, sur lesquelles le singe ne doit pas répondre, sont présentées dans les mêmes conditions. Quand le singe maîtrise ces deux phases, on mélange les images et l'on passe rapidement à l'étape suivante d'introduction de nouvelles images afin que le singe ne s'habitue pas à catégoriser à l'aide d'une stratégie d'association stimulus/réponse (figure 1.1).
3. Au cours de la dernière étape de l'entraînement - 4 mois pour Eudora - on augmente progressivement le nombre d'images nouvelles introduites quotidiennement[3] et l'on raccourcit le temps de présentation des images (figure 1.2). À chaque séance, de nouvelles images sont mélangées aux images des jours précédents[4]. Pour être récompensé, le singe

[1] La chaise à primate est produite par la société Crist aux Etats-unis. Les premières séances permettent facilement de déterminer le membre supérieur que le singe préfère utiliser et d'adapter la configuration de la chaise en rapport.

[2] La durée de présentation des images est initialement fixée à une seconde et correspond au temps durant lequel le singe peut répondre.

[3] Les images sont introduites à raison de deux par jour au début (1 animal et 1 distracteur) pour progressivement atteindre 10 par jour au cours des dernières séances (5 animaux et 5 distracteurs). Je présente ici les résultats de 90 séances consécutives - 1000 premiers essais de chaque séance – l'expérience de l'animal portant initialement sur 12 images pour atteindre 564 images à l'issue des 90 séances. Dans chaque série, la proportion d'images contenant des animaux et des distracteurs est toujours de 50 %.

[4] Le nombre d'images nouvelles est complété avec des images familières (sélectionnées aléatoirement) pour former un bloc de 100 images. Au début de chaque série, 90 images familières sont proposées pour laisser au singe le temps de se stabiliser comportementalement (elles sont supprimées lors de l'analyse). Un bloc de 100 images contenant les images nouvelles est ensuite présenté. Ce bloc est remélangé et répété tant que l'animal est disposé à travailler, c'est-à-dire entre 800 et 2000 essais. De plus 16 des 90 séances étaient précédées de 100-150 essais pendant lesquels on ne présentait à l'animal que les cibles - une dizaine d'images environ - sur lesquelles il avait tendance à répondre de façon erronée (no go). Suivait une série de 100-150 essais pendant lesquels on ne présentait à l'animal que les distracteurs sur lesquels il se trompait.

doit toucher l'image quand une cible - un animal - est présenté et maintenir le bouton appuyé quand il s'agit d'un distracteur (cf. annexe 1 pour le détail du protocole).

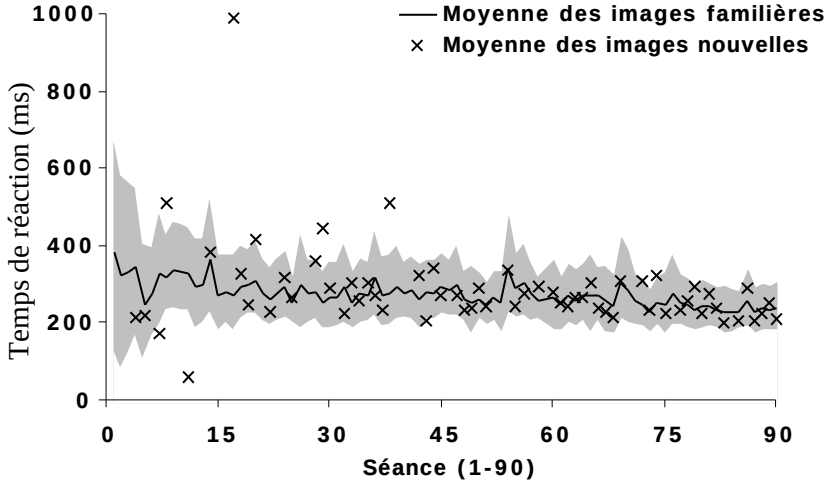


Figure 1.3 : évolution des temps de réaction en fonction des séances. La courbe en noir représente la moyenne des temps de réaction des cibles catégorisées correctement et la zone grisée représente la déviation standard. Les croix représentent la moyenne des temps de réaction des images nouvelles (étant donné le nombre restreint d'images par séance, l'écart type n'est pas représenté).

1.2 - Performance globale

En ce qui concerne l'évolution de la précision d'Eudora, on observe deux tendances. Tout d'abord, la précision augmente au cours des séances pour atteindre 87 %[5] de réponses correctes à l'issue des 90 séances. Notons toutefois que les gestes interrompus après le relâchement du bouton sont plus nombreux pour les distracteurs que pour les cibles et donc que le singe, bien que cela ne soit pas visible dans ses performances, catégorise correctement certaines de ces images (figure 1.1)[6]. La précision sur les cibles par rapport aux distracteurs varie au cours de l'apprentissage. Lors des premières séances, Eudora catégorise plus facilement les distracteurs que les cibles. Vers la 70ème séance, cette tendance s'inverse et elle devient plus précise sur les cibles que sur les distracteurs (figure 1.1).

À la 38ème séance, une nouvelle contrainte est ajoutée à la tâche : le singe ne doit plus toucher l'écran mais la zone de l'écran dans laquelle l'image a été flashée. La figure 1.1 montre qu'Eudora s'habitue rapidement - c'est-à-dire en une dizaine de séances - à cette nouvelle contrainte.

[5] Moyenne des 5 dernières séances.

[6] De plus les essais de toutes les séances sont inclus dans l'analyse et il arrivait parfois, sur des périodes d'environ 50 images, qu'Eudora soit perturbée par un bruit dans le laboratoire et qu'elle maintienne son appui sur le bouton, donc le défilement des images, tout en cessant de répondre sur les cibles.

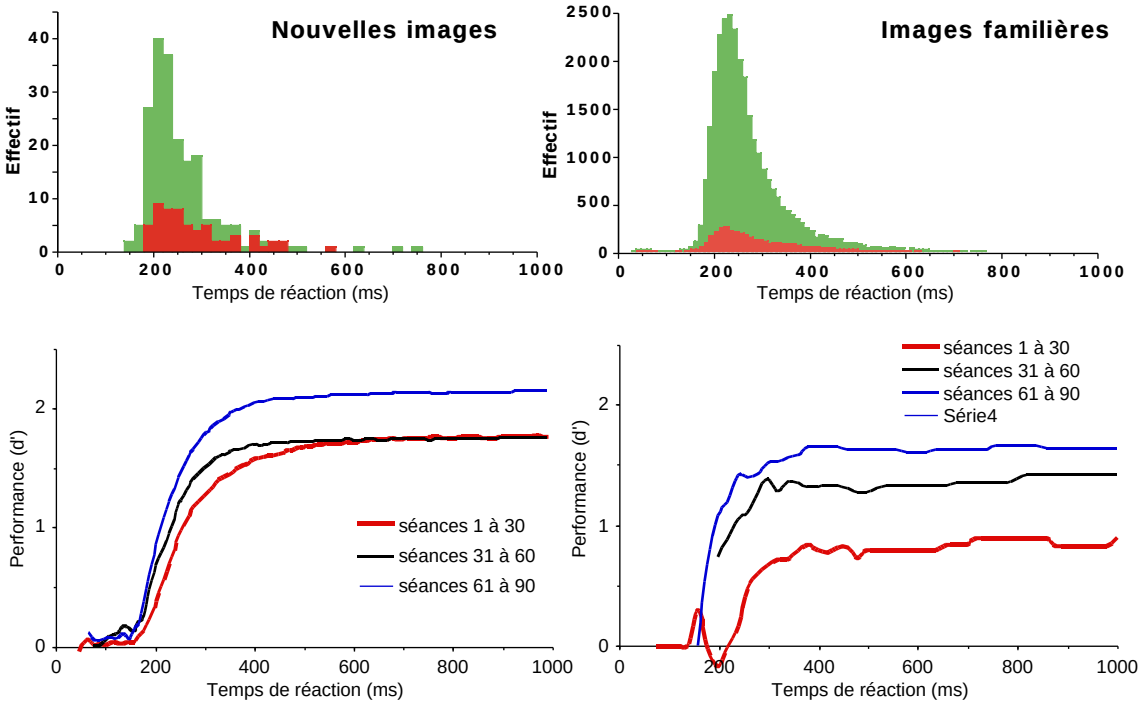


Figure 1.4 : A, répartition des temps de réaction pour les images nouvelles et les images familières sur l'ensemble des 90 séances. Les deux distributions sont très proches et s'il existe un effet, il semble être en faveur d'une plus grande rapidité pour les images nouvelles (cf. texte). B, d' en fonction du temps des 30 premières séance, des 30 suivantes et des 30 dernières séances pour les images nouvelles (à gauche) et pour images familières (à droite). Dans les deux cas, le d' devient non nul aux alentours de 150 ms. Au fur et à mesure des séances, le plateau de d' terminal est plus élevé, ce qui correspond à une augmentation de la performance de l'animal. Pour les images nouvelles, la courbe correspondant aux 30 premières séances est donnée à titre indicatif car le nombre d'images est très faible (la déviation initiale correspond par exemple à une différence de 2 images).

Concernant les réponses sur les cibles, les TRs moyens sont de plus en plus rapides[7] (figure 1.3). Cependant l'analyse des d' montre (figure 1.4) que les TRs les plus rapides évoluent peu (concernant le d', cf. Annexe 1) et apparaissent de façon constante aux alentours de 170 ms après la 30ème séance. La figure 1.5 indique l'évolution du comportement d'Eudora entre les premières et les dernières séances. Au cours des premières séances, la performance mesurée par rapport au relâchement du bouton est différente de celle calculée par rapport au toucher de l'écran. Eudora a tendance à relâcher le bouton puis à prendre la décision de toucher ou non l'écran. Elle adapte son comportement pour atteindre un stade, dans les dernières séances, puisque la précision estimée sur le relâchement du bouton et celle estimée

[7] Au cours des 15 premières séances, le TR moyen sur les cibles est de 319 ms pour descendre à 243 ms au cours des 15 dernières séances. Cette évolution est significative : par groupe de 15 séances, ANOVA, DDL=5, somme des carrés 16749822 et $p<0.0001$.

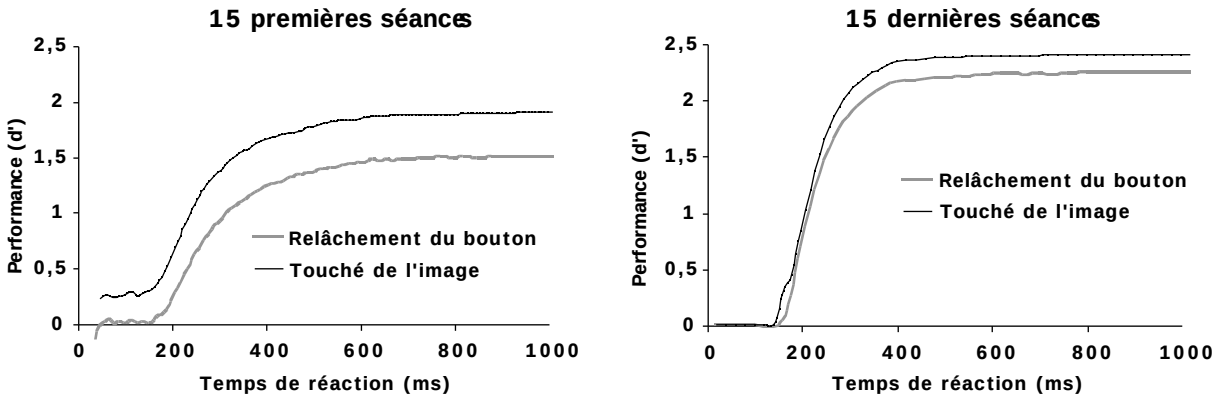


Figure 1.5 : performance pour les 15 premières séances (à gauche) et pour les 15 dernières séances (à droite). Le calcul de la performance par rapport au toucher de l'image montre que, au cours des premières séances, le singe relâche le bouton très rapidement (avant 200 ms) pour ensuite aller toucher l'écran : la décision est prise après qu'il a relâché le bouton. Cette stratégie disparaît au cours des dernières séances, le fait que le singe touche l'image étant alors fortement corrélé avec le fait qu'il ait relâché le bouton.

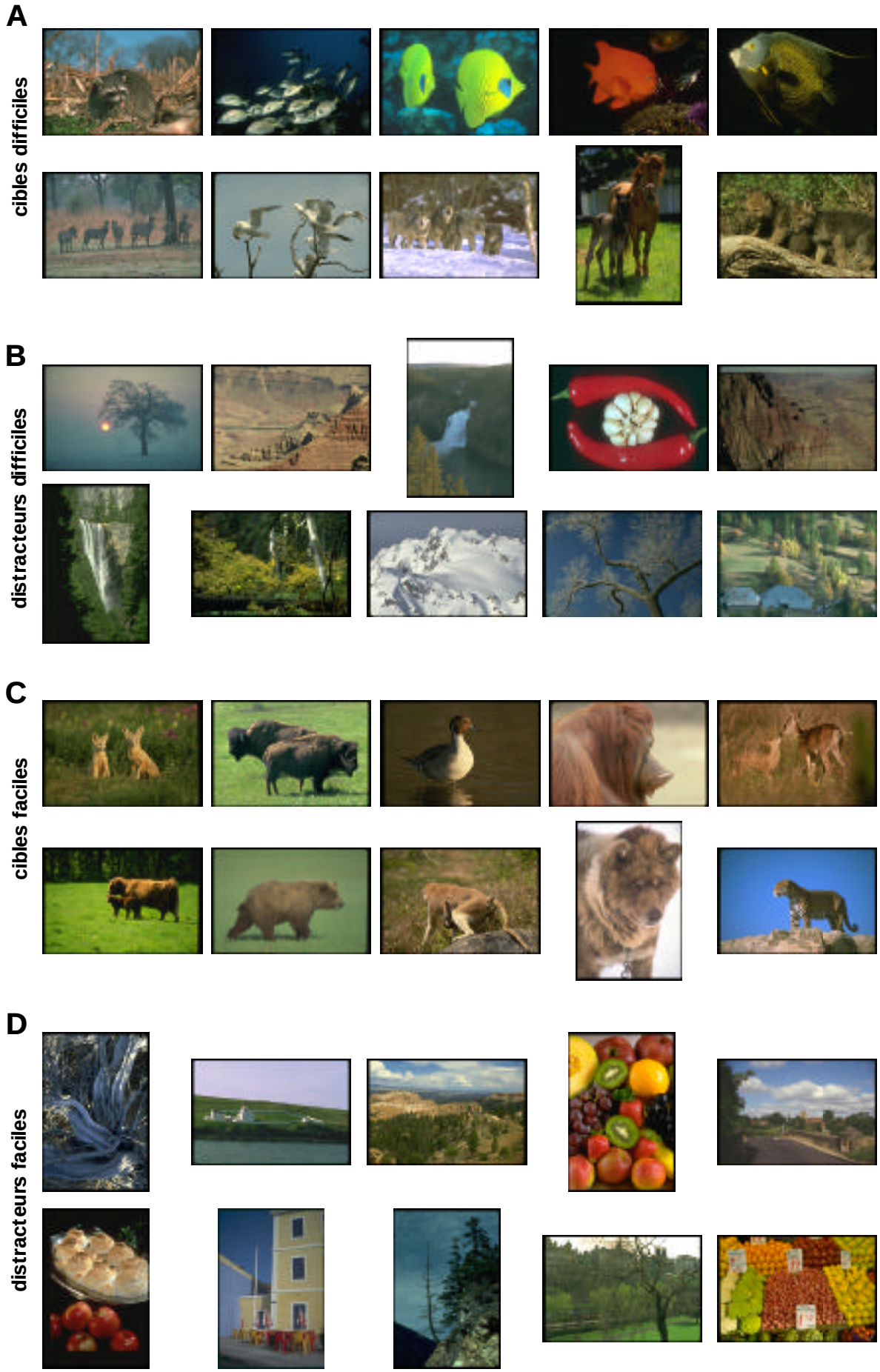
sur le touché de l'écran (au niveau de l'image) sont pratiquement identiques; comportement optimal du point de vue de l'obtention de la récompense[8].

L'analyse des erreurs au cours de l'apprentissage nous renseigne également sur le type de catégorisation effectuée par Eudora (figure 1.6). Elle continue à commettre des erreurs sur des images qui lui ont déjà été présentées plus de 100 fois. Eudora a par exemple beaucoup de mal à catégoriser les poissons en tant qu'animaux - 4 images de poissons sont parmi les 10 cibles les plus difficiles - et à des difficultés à catégoriser les chutes d'eau en tant que distracteurs[9]. À l'opposé, les images distracteurs représentant de la nourriture semblent particulièrement faciles. Dans l'ensemble toutefois, les caractéristiques des images très difficiles et très faciles ne semblent pas être fondamentalement différentes.

Figure 1.6 : cf. page suivante. A, cibles difficiles sur lesquelles Eudora a commis le plus d'erreurs (par rapport au nombre de présentations). Les cibles sont classées par ordre de difficulté de gauche à droite, la cible la plus difficile se trouvant dans le coin supérieur gauche et la 10ème cible la plus difficile se trouvant dans le coin inférieur droit. B, distracteurs difficiles suivant le même mode de classification. C, cibles faciles. Toutes ces cibles sont catégorisées avec un temps médian de moins de 218ms, la cible du coin supérieur droit étant catégorisée la plus rapidement. Le taux de réussite de ces cibles a été sélectionné pour être supérieur à 95%. D, distracteurs faciles dont le taux de réponses correctes est supérieur ou égal 98%. Dans tous les cas, les images vues moins de 20 fois par le singe ne sont pas prises en compte.

[8] On peut noter la persistance d'une légère différence pour les temps de réaction les plus rapides, car ayant relâché le bouton très tôt, Eudora semble parfois "se rendre compte" de son erreur.

[9] Les expériences que j'ai réalisées par la suite montreraient que les hommes semblent commettre le même type d'erreur sur les chutes d'eau. Serait-ce dû au mouvement présent dans l'image ?


A
cibles difficiles
B
distracteurs difficiles
C
cibles faciles
D
distracteurs faciles

1.3 - Performance sur les images nouvelles

Je vais maintenant analyser la précision d'Eudora sur les images nouvelles, c'est-à-dire les images qu'elle voit pour la première fois, et comparer ses résultats à ceux obtenus sur l'ensemble des images. Ces essais correspondent à une véritable catégorisation d'objets nouveaux au même titre que la catégorisation que peut réaliser l'homme, comme on le verra par la suite.

Du point de vue de la précision, le taux de réussite global est plus élevé que sur les images nouvelles. La performance finale est de 87 % de réponses correctes[10] à l'issue des 90 séances pour l'ensemble des images et de 84 % de réponses correctes pour les images nouvelles. Le gain en précision sur les images nouvelles au cours des séances semble dépendre pratiquement de la précision sur les cibles : alors que la précision sur les distracteurs est d'environ 75 % et varie peu au cours des séances, la précision sur les cibles augmente régulièrement partant de 20 % pendant les 15 premières séances pour atteindre environ 90 % à la fin de l'entraînement (figure 1.7). Concernant les cibles, au cours des premières séances la précision, à la 2nde et la 3ème présentation de l'image, est bien supérieure à celle de la première présentation. Cet effet s'estompe pour s'inverser puis disparaître au cours des dernières séances d'entraînement. Comme nous le verrons plus tard, Eudora semble passer d'une stratégie d'association stimulus/réponse à une stratégie faisant appel à une catégorisation complexe.

Du point de vue des temps de réaction, les TRs moyens des images nouvelles et de l'ensemble des images sont comparables tant pour les premières que pour les dernières séances[11] (figure 1.3). S'il existe une tendance, elle semble même être en faveur d'une catégorisation plus rapide des images vues pour la première fois par rapport à leurs occurrences suivantes[12]. Le fait qu'Eudora soit plus rapide au début de la séance - c'est-à-dire lorsque les images nouvelles sont vues pour la première fois explique en partie cette différence[13].

---

[10] Moyenne des 5 dernières séances.

[11] La variation des TRs pour les images nouvelles, plus élevée au cours des premières séances, s'explique en partie par le faible nombre d'images introduites au cours de ces séances.

[12] Comparaison du nombre d'images par tranche de 20 ms de la première présentation et de l'ensemble composé des 2, 3 et 4 présentations de chaque image : t-test apparié, DDL = 8, p=0,0048.

[13] J'ai tenté de vérifier cette hypothèse sur l'ensemble des images : une légère tendance est effectivement visible - augmentation de 2ms par tranche de 200 essais - mais n'explique pas complètement la rapidité d'Eudora sur les cibles nouvelles. Le fait qu'Eudora soit plus rapide sur les images nouvelles pourrait également être corrélé avec sa plus faible précision sur ces images.

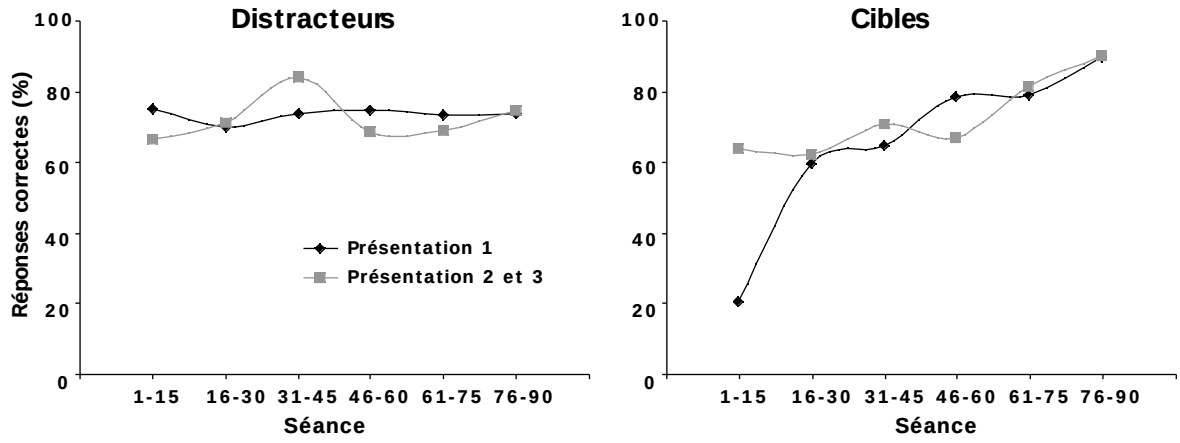


Figure 1.7 : précision sur les images nouvelles pour la première présentation de l'image et les 2nde et 3ème présentations. La précision pour les présentations 2, 3, 4 et 5 diffèrent peu (non représenté). La précision sur les distracteurs semble constante alors que celle sur les cibles augmente au fur et à mesure des séances.

Comme on l'a vu précédemment sur l'ensemble des images, l'évolution de la performance d'Eudora peut être évaluée en fonction du temps séance après séance (figure 1.4). La figure 1.6 montre que la performance globale présente la même évolution pour les images qu'elles soient nouvelles ou familières, c'est-à-dire une tendance à une précision meilleure, un TR moyen plus rapide sans que les TRs les plus rapides ne soient affectés.

1.4 - Discussion

Les résultats de l'entraînement d'Eudora dans la tâche de catégorisation d'animaux montrent qu'elle est capable d'atteindre 84 % de réponses correctes sur des images nouvelles[14], niveau de performance calculé par l'expérimentateur sur la base de l'absence - ou de la présence - d'un animal dans l'image. Ses performances globales augmentent au cours des séances ce qui semble principalement dû à une meilleure catégorisation des cibles. Concernant les images nouvelles, le comportement d'Eudora évolue également d'une stratégie d'association stimulus/réponse vers une stratégie de catégorisation plus complexe.

À l'issue de l'entraînement, la précision est meilleure sur les images familières que sur les images nouvelles. Les temps de réaction – notamment les plus rapides - ne semblent pas évoluer de façon dramatique au cours de l'apprentissage et sont similaires pour les images nouvelles et familières. Dans les expériences qui suivent, nous retrouverons ce type de résultat à la fois chez l'homme et chez le singe.

[14] Sur les 5 dernières séances (50 images nouvelles). Elle a catégorisé environ 85000 images - dont la majorité lui étaient familières - avant d'atteindre ce niveau de performance.

1.4.1 - La catégorisation chez l'animal

Hernstein (1990) distingue cinq niveaux de catégorisation chez les animaux. Tout d'abord, la discrimination ou détection d'un objet puis la discrimination d'une liste apprise d'objets. Au niveau supérieur, les catégories de bases sont des classes infinies d'objets similaires par leur forme. Les concepts - ou catégories supraordonnées - sont des classes de stimuli qui conduisent à un même comportement (par exemple la catégorie nourriture : manger). La forme des objets au sein d'un concept peut être très variée et la tâche animal/non-animal appartient à ce niveau de catégorisation, les animaux pouvant être très divers à la fois par leur forme, leur position, leur nombre... Enfin le dernier niveau de catégorisation serait celui des relations abstraites définies *a posteriori* par un individu ou par l'expérimentateur : c'est un concept créé *de novo* pour la tâche.

Concernant la catégorisation animale, la majeure partie des expériences ont été réalisées chez les primates et les oiseaux. Les catégories infinies d'éléments sont parfaitement appréhendées par les animaux. Les geais bleus par exemple sont capables de catégoriser différents types de feuilles en fonction du type de chenilles qui s'en nourrissent (feuilles déchiquetées ou régulièrement découpées), comportement en relation avec leur préférence alimentaire.

Les pigeons sont capables de catégoriser des catégories infinies comme des arbres, des photographies contenant de l'eau, ou des photographies contenant une personne particulière. Ils répondent de façon correcte même sur des images qu'ils n'ont jamais vues auparavant (Hernstein et Loveland, 1964; Hernstein et al, 1976) et les images difficiles pour les sujets humains semblent l'être aussi pour les pigeons. De plus les pigeons font la différence entre les objets du monde réel et les images qu'on leur présente (Watanabe, 1997; Bovet et Vauclair, 2000). Dans certains domaines, leur expertise dépasse même celle des hommes : si l'on présente des photographies de vues aériennes à des pigeons, ils sont capables de catégoriser les images appartenant à la même zone géographique alors que les sujets humains semble avoir des difficultés (Gray, 1987; Hernstein, 1990). Cependant, la limite de la capacité du système visuel des pigeons apparaît rapidement. À la différence des humains, des singes ou même des souris, les pigeons ne sont pas capables de compléter mentalement des parties d'objet (Sekuler et al, 1996). Si on leur présente des parties de photographies, ils ne sont pas capables de compléter un objet partiellement visible pour le catégoriser.

Les singes macaques sont également capables de catégoriser des arbres (Vogels, 1999) et même des animaux comme cela a été montré dans notre équipe (Fabre-Thorpe et al, 1998;

Delorme et al, 2000). La catégorisation d'animaux semble cependant relativement difficile comparée à des catégorisations plus simples comme celle de photographies de martins pêcheurs parmi d'autres espèces d'oiseaux (Roberts et Mazmanian, 1988). Toutefois même dans ce type de catégorisation, les singes n'utilisent pas uniquement des indices de bas niveau comme la couleur de l'oiseau. Il est très probable que le concept que les singes se font de cette catégorisation ne coïncide pas avec le concept d'animal tel que nous pouvons nous le représenter. Cette représentation est cependant sans doute proche puisque Fabre-Thorpe et al (1998) ont montré qu'à la fois les TRs et les erreurs commises par les singes et par les sujets humains sur les mêmes images étaient comparables. Les résultats que j'ai obtenus avec Eudora semblent aller dans le même sens, bien qu'une analyse plus poussée soit nécessaire.

L'effet de la catégorisation de stimuli nouveaux pose également la question de la généralisation. Chez l'homme, et tout spécialement chez l'enfant, suite à la présentation de moins de 12 figures géométriques composées de points, l'analyse des erreurs montre que les stimuli auraient tendance à être catégorisés individuellement alors qu'avec un grand nombre de stimuli, la catégorisation serait effectuée sur la base d'un prototype (Posner et Keelle, 1970). Comme on l'a vu, on observe également cette évolution chez Eudora. Au cours des premières séances, les performances à la 2^{nde} et la $3^{ème}$ présentation d'une image cible sont bien supérieures à celles de la première présentation (figure 1.7). Dans un premier temps, Eudora associe donc l'image à la réponse qu'elle donne. Par la suite cet effet disparaît et Eudora semble passer à une catégorisation plus complexe, peut-être conceptuelle.

1.4.2 - Classement spontané chez le singe

Nous allons voir que les singes ont tendance à utiliser des catégories de façon spontanée. Pour commencer, la familiarité d'un stimulus est un élément très important chez les primates. Si on apprend à des chimpanzés à catégoriser des objets en fonction de leur similarité de forme, certains objets étant familiers et d'autres moins, alors on observe que les singes ont tendance à commettre des erreurs en groupant les objets familiers entre eux (Tanaka, 1995). Cette importance de la familiarité des stimuli peut expliquer l'augmentation significative de précision entre les images familières et les images nouvelles que nous observons chez Eudora.

Une autre expérience très intéressante (Sands et al, 1982) indique que les singes semblent catégoriser implicitement les stimulus de manière sémantique. Dans cette expérience, les singes doivent apparier des images présentées séquentiellement. Une image était présentée brièvement suivie d'une autre image à une position légèrement décalée. Le singe devait

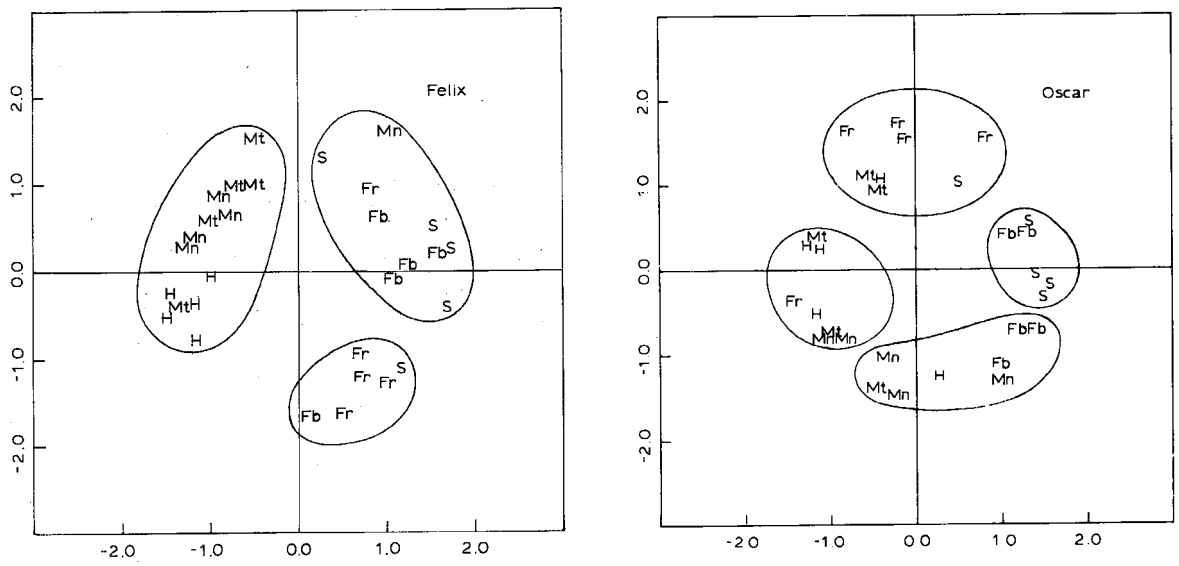


Figure 1.8 : décomposition en composantes principales des erreurs de deux singes (Felix et Oscar) sur un espace multidimensionnel dans l'expérience de Sands et al. (1982). L'axe horizontal représente la première composante principale et l'axe vertical la seconde composante principale. H=humains, M=singes, T=arbres, Fr=fruit, Fl=fleur. Les images ont tendance à se regrouper selon leur contenu sémantique.

pousser un levier dans un sens si les deux images étaient identiques et dans un autre sens si les deux images étaient différentes. Plusieurs catégories de photographies, de fleurs, de fruits, de visages de singes et d'hommes étaient présentées. Chaque image était appariée avec elle-même et avec toutes les autres. Les erreurs du singe étaient ensuite analysées[15]. Les erreurs concernant les fruits se décomposent selon deux axes : la couleur et le type de fruit - pomme ou raisin - ignorant des caractéristiques telles que la taille ou le nombre d'objets. Les singes ont également tendance à effectuer des erreurs entre les visages d'hommes et de singes (figure 1.8). Cette expérience ne pose aucun *a priori* sur le type de catégorisation qu'effectuent les sujets (sinon celui du choix des images présentées). Cela montre que les singes utilisent spontanément des catégories visuelles pour classer les objets. Il est donc tout à fait probable qu'Eudora et par extension la famille des macaques catégorisent les images en fonction de leur contenu et non de leurs caractéristiques.

Un autre type de groupement spontané effectué par les singes serait celui de l'ordre de présentation des stimuli (Orlov et al, 1999). Dans une tâche simple où le singe doit répondre sur une séquence de stimuli, il peut employer deux stratégies. La première consisterait à apparier les stimuli par paire : quand un stimulus apparaît le singe connaîtrait celui qui doit suivre. La seconde stratégie consisterait pour l'animal à se souvenir de l'ordre de présentation

[15] Seules les erreurs sont prises en compte lors de l'analyse des résultats. La matrice de confusion est analysée en effectuant une décomposition en composantes principales.

du stimulus. L'analyse des erreurs des singes montre qu'ils sont plus enclins à utiliser la seconde stratégie. Cette expérience montre également que les singes utilisent d'autres stratégies secondaires. Cela signifierait donc que les singes sont capables d'utiliser simultanément plusieurs stratégies complémentaires (cf. également Roberts et Mazmanian, 1988), et de s'adapter à la tâche qui leur est demandée.

Afin de pousser plus avant cette étude, et de déterminer dans quelle mesure les singes adaptent leur stratégie aux stimuli qui leur sont présentés, nous allons contraindre la tâche de catégorisation. Notre but sera de déterminer si oui ou non les singes utilisent des caractéristiques de bas niveau - comme la couleur des objets - pour catégoriser les images qu'on leur présente. Nous allons dans l'expérience suivante supprimer les informations chromatiques des images. Si cette caractéristique n'est pas critique pour la tâche alors cela nous fournira un argument supplémentaire pour affirmer que les singes catégorisent les images en fonction de leur contenu et qu'ils utilisent la même stratégie sur les images en couleur et en noir et blanc.

2

Le rôle de la couleur

Dans les chapitres qui suivent, nous allons tenter de déterminer les caractéristiques des images pouvant être critiques pour une catégorisation rapide, en particulier celles qui peuvent influencer le traitement précoce et ont donc une incidence sur les TR les plus rapides.

Dans un premier temps, nous nous sommes intéressés aux indices chromatiques. Le rôle de la couleur est très controversé dans la littérature concernant la catégorisation. Suivant le type de protocole - tâche go-nogo ou tâche de dénomination d'objet - le type de stimuli - photographie ou dessins - la suppression des informations chromatiques a des effets très divers. Bien qu'il existe des exceptions, il semble que la tendance qui se dégage de ces études fait état d'un rôle plus important de la couleur dans des tâches où il faut nommer les objets que dans une tâche de simple catégorisation visuelle (Ostergaard et Davidoff, 1985). Nous verrons, à la lumière de nos résultats, que la couleur n'intervient dans la catégorisation que dans la réalisation des réponses tardives. La prise en compte des indices chromatiques pourrait être lente. Dans ce cas, le fait que la couleur intervienne dans les tâches de dénomination d'objet serait lié à la lenteur relative des réponses inhérentes à ce type de tâche faisant appel à la sémantique.

Cette partie résume les travaux présentés dans l'article intégré en annexe (annexe 3). J'ai tenté dans ce résumé une approche complémentaire des résultats obtenus dans l'article, notamment en ce qui concerne l'analyse des performances en fonction du temps.

2.1 - Matériel et méthodes

Le protocole expérimental est le même que celui utilisé pour l'entraînement d'Eudora (cf. annexe 1). Les 3 singes testés effectuent la tâche go-nogo décrite précédemment : Eudora et Rouky sur la catégorie animal/non animal et Rox sur la catégorie aliment/non aliment[1]. L'expérience se déroule en deux temps :

1. Dans une première phase, les singes apprennent à catégoriser en NB les images qui leur sont familières en couleur. Les images présentées en NB sont toujours mélangées avec des images en couleur pour éviter que les singes n'utilisent des stratégies différentes pour les deux types d'images[2].
2. Dans un deuxième temps, 400 images nouvelles (200 en couleur et 200 en NB) dont la moitié de cibles et l'autre moitié de distracteurs sont présentées aux singes, à raison de 20 par jour, mélangées à des images familières. Les mêmes images sont présentées à 10 sujets humains pour chaque tâche animal/non animal et aliment/non aliment[3].

2.2 - Résultats

2.2.1 - Précision des réponses

Comme le montre la figure 2.1, les performances des 3 singes sur les 400 images nouvelles sont étonnamment bonnes tant sur les images en couleur (87,2% de réponses correctes dans l'ensemble) que sur les images en NB (87,3% de réponses correctes dans l'ensemble). Quelle que soit la tâche on n'observe aucune différence significative entre les deux conditions.

La précision des sujets humains est supérieure à celle des singes d'environ 5% dans les deux tâches. Comme l'indique la figure 2.1, la précision dans les deux tâches est meilleure d'environ 2% sur les images en couleur par rapport aux images en NB et diffère significativement dans la tâche animal. En fait la perturbation est variable d'un sujet à l'autre, certains - peu nombreux - présentant une diminution de précision en l'absence d'indices

[1] Rouky et Rox ont suivi un entraînement similaire à celui d'Eudora.

[2] A l'issue de cet entraînement on teste les singes sur 200 images familières mais qu'ils n'ont jamais vues en NB. Ces résultats indiquent que les singes sont capables de catégoriser les images familières présentées pour la première fois en NB avec des baisses de performance très faibles par rapport aux images en couleur.

[3] Pour l'homme, les 400 images sont présentées au cours de 4 sessions successives de 100 images chacune (contenant 50 cibles et 50 distracteurs).

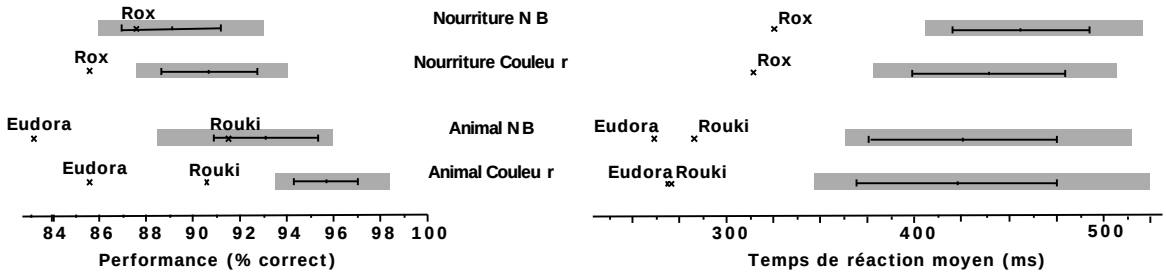


Figure 2.1 : à gauche, précision des singes et des 10 sujets humains dans les deux tâches pour les images en couleur et en NB. La moyenne et l'écart type de la précision des sujets humains sont représentées dans les barres qui elles-même indiquent la fourchette des réponses correctes sur les images en couleur (barre gris clair) et en NB (barre gris foncé). Les croix indiquent les performances de chaque animal. A droite, les TRs moyen des singes et des 10 sujets humains sur les cibles sont représentés. Le type de présentation est le même que celui utilisé pour la précision.

chromatiques alors que d'autres (les plus nombreux) apparaissent indifférents à cette manipulation[4]. L'analyse des performances sur les cibles et les distracteurs montre que dans la tâche animal, la baisse de performance sur les cibles et les distracteurs est d'environ 3% pour les images en NB. Dans la tâche de détection d'aliment, la performance sur les cibles en NB diminue de 6% alors que celle sur les distracteurs augmente d'environ 3%. La baisse de performance dans cette tâche est donc en partie due au fait que les sujets humains répondent moins sur les images présentées en NB.

2.2.2 - Temps de réaction

Chez le singe, quelle que soit la tâche, on ne constate aucune différence significative sur les TR des réponses produites sur les images en couleur et en NB. On observe cependant une tendance dans la catégorie aliment en faveur des images chromatiques puisque les images en NB sont catégorisées 10 ms plus lentement que les images en couleur[5] (figure 2.1). Comme le montre la figure 2.2, les distributions des TRs sont très proches dans les deux conditions.

L'analyse de la précision des singes en fonction du temps (figure 2.3) montre que la catégorisation ne semble pas plus rapide lorsque les indices chromatiques sont présents, tout du moins pour les réponses précoces[6]. En effet, si certaines images étaient catégorisées plus

[4] Sur l'ensemble des sujets : Chi2=13,6, d.f.=1, p=0.0002.

[5] Dans la tâche "aliment", temps moyen de réaction du singe sur les cibles: en couleur 312ms (médiane 297ms), en NB: 324ms (médiane 306ms).

[6] Dans les deux cas, cible et distracteur, les différences en termes de précision sont non significatives mais le faible nombre de réponses ne permet pas de conclure qu'il n'y a en effet aucune différence entre les images en couleur et en NB.

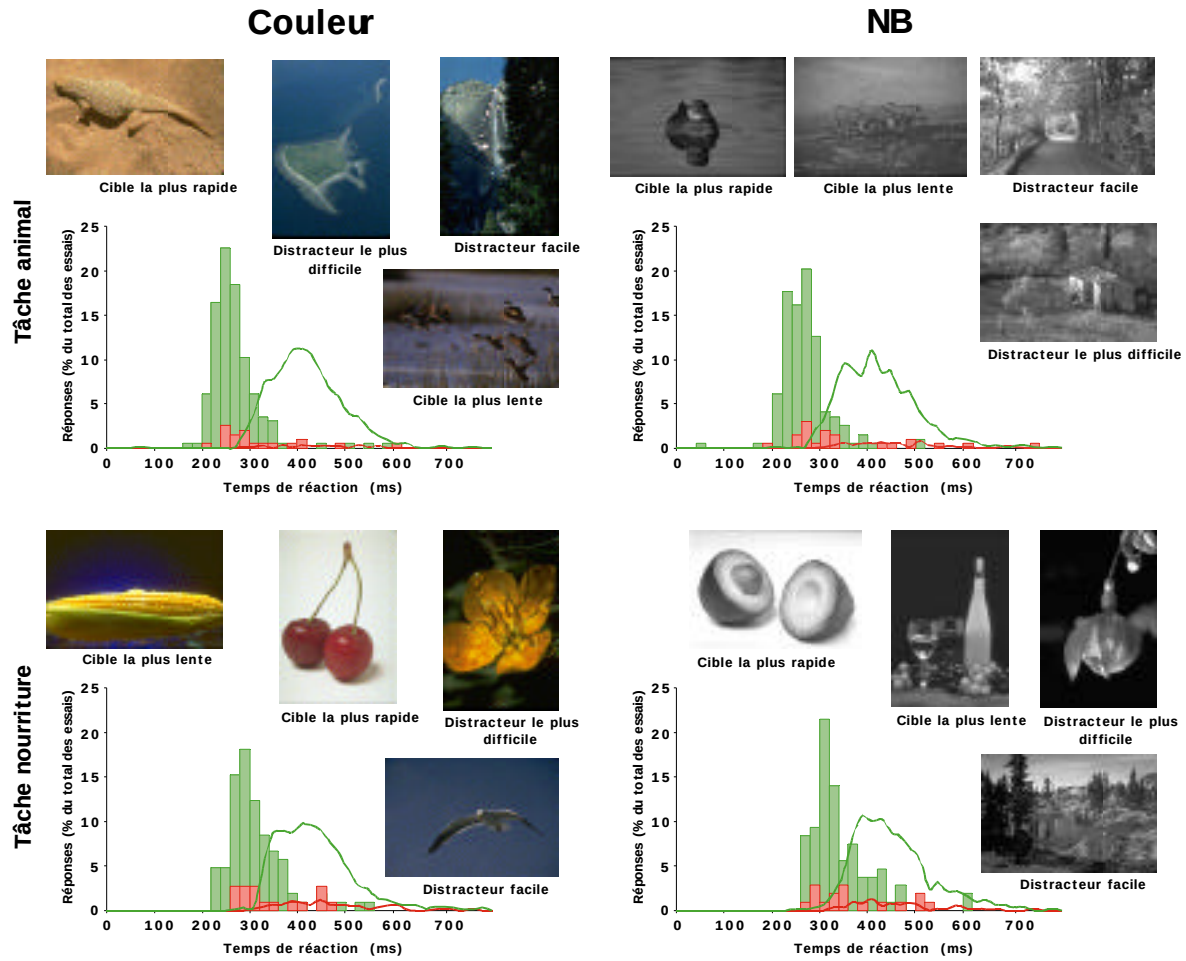


Figure 2.2 : TRs des singes et des sujets humains dans la tâche de catégorisation d'animaux (lignes supérieures) et de nourriture (lignes inférieures) pour les images en couleur (colonne de gauche) et en NB (colonne de droite). Les histogrammes en vert représentent la distribution des TRs des singes vers les cibles et les histogrammes en rouge représentent les réponses des singes sur les distracteurs (les images familières ne sont pas représentées). De la même façon les courbes en vert et en rouge indiquent la distribution des TR des sujets humains sur les cibles et les distracteurs. Au sein de chaque graphique, la cible pour laquelle la réponse est la plus rapide - à la fois pour les hommes et le/les singes - et la cible pour laquelle la réponse est la plus lente sont présentées. Pour déterminer ces images, les cibles sur lesquelles aucun sujet (humain ou simien) n'a commis d'erreur sont triées par ordre croissant en fonction des latences de réponse pour les hommes d'un côté et pour les singes de l'autre. L'image la plus rapide est celle dont la somme des rangs (pour l'homme et pour le singe) est la plus faible. L'image la plus lente est celle pour laquelle la somme des rangs est la plus élevée. Pour les distracteurs, les images sont classées par ordre décroissant suivant le nombre d'erreurs qu'elles ont induites. L'image la plus difficile est celle dont la somme des rangs, du classement pour le singe et pour l'homme, est la plus faible. Une image facile, sur laquelle aucune erreur n'a été commise est également présentée.

rapidement en couleur, les courbes de performance illustrées sur la figure 2.3 divergeraient rapidement, ce qui n'est pas le cas.

En ce qui concerne les sujets humains, et de façon similaire à ce qui était observé chez le singe, la différence entre les distributions des TRs entre les images en couleur et en NB n'est

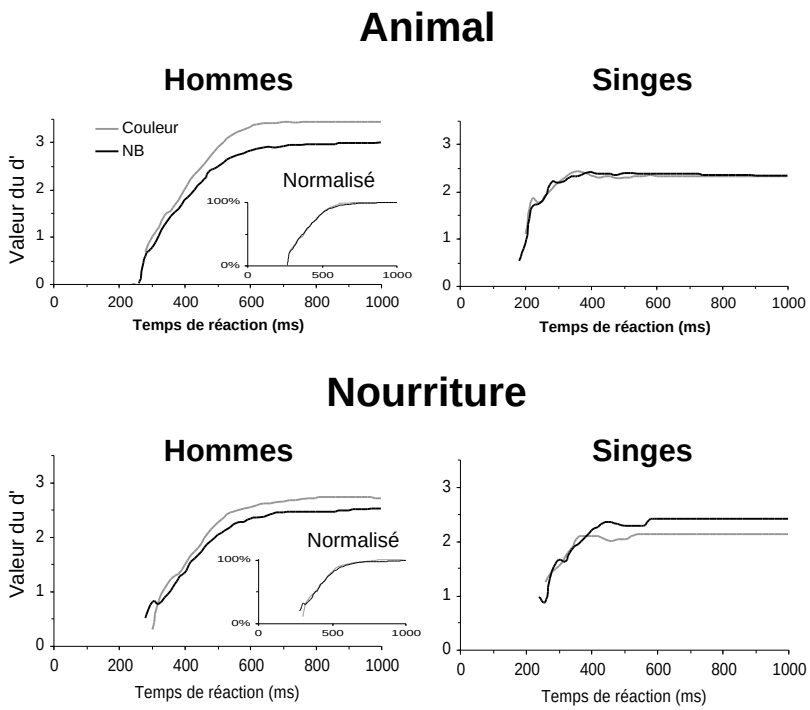


Figure 2.3 : Précision en fonction du temps dans les tâches de catégorisation d'animaux et d'aliments pour les images en couleur (courbes grises) et en NB (courbes noires). Pour les singes, il semble y avoir très peu de différences entre les images présentées en couleur et en NB. Pour les sujets humains, bien que la performance sur les images en couleur soit meilleure que celle sur les images en NB, les courbes normalisées, en particulier dans la tâche de catégorisation d'animaux, montrent que cet effet sur la précision ne dépend pas du temps.

significative que pour la catégorisation d'aliments[7] et indique un avantage relativement faible de 15 ms en moyenne pour les images en couleur (figure 2.2). La catégorisation des aliments apparaît plus difficile que celle des animaux en terme de durée de traitement, les TRs enregistrés dans la détection d'aliments étant de 30 à 40 ms plus tardifs que ceux observés dans la détection d'animaux. Notons cependant que dans les deux tâches, l'effet de la suppression de la couleur ne semble pas intervenir sur les réponses les plus rapides, inférieures à 300ms (figure 2.2). L'analyse de la précision en fonction du temps (figure 2.3) montre que la performance des sujets, pour les temps de réaction les plus rapides est similaire pour les images en couleur et en NB[8]. Bien que les courbes de performance divergent relativement tôt, le fait que les courbes normalisées soient pratiquement confondues (cf. annexe 1) indique que cette perte de performance ne dépend pas du temps : certaines images en NB sont simplement plus difficiles et ne sont pas reconnues comme étant des cibles. Cette hypothèse est compatible avec la diminution des réponses sur les cibles en NB que l'on a observée précédemment.

[7] Dans la tâche aliment, temps moyen de réaction des sujets humains sur les cibles: en couleur 437ms (médiane 427ms), en NB: 453ms (médiane 439ms). Mann Whitney U=359702, p<0.0001.

[8] S'il fallait dégager une tendance, ce serait même des temps de réaction plus rapides pour les images en NB.

2.3 - Discussion

La comparaison des performances et des temps de réaction des hommes et des singes montre que (1) comme les sujets humains, les 2 singes entraînés sur la tâche de catégorisation d'animaux sont plus rapides que celui entraîné sur la tâche de détection d'aliments, (2) comme les sujets humains, les TRs moyens des singes ne semblent affectés par la suppression de la couleur que pour la catégorie aliment. Cette augmentation des TRs atteint des valeurs compatibles chez les deux espèces :10 ms chez le singe et 15 ms chez l'homme. La couleur des images interviendrait donc de façon plus marquée dans la tâche de détection des aliments que dans celle de détection des animaux.

2.3.1 - Le rôle de la couleur et de la catégorie

Pour les images en couleur, il existe deux type de contrastes : d'une part les contrastes de luminance qui dépendent de l'intensité de la lumière et d'autre part les contrastes chromatiques qui dépendent des couleurs elles-mêmes. Des études ont montré cependant que pour des luminances éloignées des seuils perceptifs, ce qui est le cas dans notre expérience, le rôle des contrastes chromatiques était négligeable (Jordan et al, 1990; Troscianko, 1994; Cavanagh et al, 1995; Nagy et Kamholz, 1995; Syrkin et Gur, 1997). Ces expériences indiquent que si la couleur intervient, ce n’est pas en renforçant le contraste des contours de certains objets mais plutôt en conférant certaines qualités de surface aux objets.

Le macaque rhésus a un système visuel très proche de celui de l'homme en ce qui concerne le traitement de la couleur[9] et des études en électrophysiologie ont montré que les catégories de couleur présentes chez l'homme, relativement indépendantes du langage, se retrouvent dans la réponse des neurones du système visuel des macaques (Yoshioka et al, 1996).

De plus Sands et al (1982) ont montré que pour les fruits, la couleur était un élément caractéristique de la réponse des singes. Dans une tâche d'appariement de catégories d'images, les singes avaient tendance à commettre des erreurs majoritairement sur les fruits partageant la même couleur et ignoraient par exemple les caractéristiques telles que la taille ou le nombre

[9] Les macaques rhésus possèdent 3 types de cônes comme l'homme, à la différence de certains singes du nouveau monde, notamment les mâles, qui n'en possèdent que 2.

de fruits. Cette observation n'était pas vraie pour d'autres catégories de stimuli comme les visages. Cette expérience montre donc que l'utilisation de la couleur dépend de la catégorie. On retrouve ce résultat dans notre expérience puisque le fait de supprimer les informations de couleur a un effet beaucoup plus important pour la catégorie aliment que pour la catégorie animale.

La spécificité de forme pour les fruits est relativement faible et la couleur pourrait être nécessaire pour identifier certains d'entre eux. D'une manière générale, les études chez l'homme et le singe s'accordent à dire que la couleur est utilisée quand les informations relatives à la forme de l'objet présenté ne sont pas suffisantes (Price et Humphreys, 1989). Pour certains auteurs, le trichromatisme a évolué chez les primates afin qu'ils puissent distinguer par exemple leur nourriture - i.e. fruit - dans leur environnement. La couleur peut également servir à déterminer si tel fruit est pourri, vert, ou mûr et comestible (Mollon, 1989; Gegenfurtner et Riege, 2000).

Comme nous l'avons vu ici et comme nous le verrons dans le chapitre suivant, les TRs les plus rapides ne sont pas affectés par l'absence de couleur. Une hypothèse développée dans la discussion de l'article (annexe 3) est que la couleur ne serait pas nécessaire à la catégorisation ultra-rapide et que la catégorisation pourrait s'effectuer à partir des informations en provenance de la voie magnocellulaire dans le système visuel. Nous reviendrons plus en détails sur ce point dans la discussion générale des résultats expérimentaux.

2.3.2 - Des voies neuronales distinctes, le problème du liage

Une hypothèse relativement répandue dans la communauté scientifique est que le traitement de la couleur suit une voie distincte de celle de la forme (Kolers et von Grunau, 1975; Livingstone et Hubel, 1987; Legge et al, 1990; Humphreys et Boucard, 1997).

Une expérience relativement convaincante concernant cette hypothèse est que la détection d'une forme de couleur différente parmi d'autres formes similaires ne dépend pas du nombre de formes (Hanna et Remington, 1996). Cependant, si aucun attribut de couleur n'est présent et que la décision doit se baser sur la configuration de la forme, le temps de détection dépend du nombre de formes présentées. Si les attributs de forme et de couleur de chaque objet étaient intimement intriqués, les résultats seraient comparables dans les deux cas. Les expériences montrant que la couleur intervient dans les processus pré-attentifs (Verghese et Nakayama, 1994) sont également compatibles avec l'existence de deux voies distinctes, l'une pour la couleur et l'autre pour la forme.

Cette hypothèse est parfaitement illustrée dans les modèles de Grossberg (Bradski et Grossberg, 1995; Grossberg, 1994) où le traitement des formes et celui des surfaces sont indépendants et n'interagissent que ponctuellement. Ce modèle postule que des décisions peuvent être prises uniquement sur la base des caractéristiques de forme des objets présentés sans que le liage des informations de surface (donc de couleur) et de contour ne soit nécessaire. Le liage de la couleur et de la forme serait un processus relativement lent et qui n'interviendrait que tardivement. Nos résultats renforcent cette hypothèse, dans la mesure où la couleur n'intervient pas pour les temps de réaction les plus rapides.

Le rôle de la couleur dans les tâches de catégorisation rapide est relativement controversé (Ostergaard et Davidoff, 1985) alors qu'il est très net quand la catégorisation fait appel au contenu sémantique de l'image. La couleur joue, par exemple, un rôle majeur dans l'association de mots et d'images (Glaser et Glaser, 1989; Joseph et Proffitt, 1996). La couleur joue également un rôle bien plus important dans les tâches où il faut nommer l'objet par rapport aux tâches où la dénomination n'est pas requise (Biederman et Ju, 1988; Ostergaard et Davidoff, 1985; Wurm et al, 1993; Humphreys et Boucard, 1997). Plutôt qu'invoquer un accès différent au système sémantique pour les stimuli en couleur et en NB, je préfère interpréter ces résultats dans le cadre de la rapidité du traitement énoncé précédemment. Les réponses pour lesquelles la dénomination de l'objet est requise sont beaucoup plus tardives que celles obtenues dans la tâche go-nogo que nous avons utilisée. Le système visuel a donc eu le temps de traiter les informations chromatiques tardives et de lier informations de couleur et de contour.

Cette hypothèse sur la rapidité postule uniquement que la couleur est un attribut visuel qui prend plus de temps à traiter que la forme et que l'accès au lexique et à l'identification d'un objet est postérieur à sa détection.

3

Catégorisation et caractéristiques des cibles

Dans le chapitre précédent, nous avons montré que les attributs de couleur intervenaient peu dans la catégorisation ultra-rapide et nous avons vu les implications qui en découlaient pour le traitement dans le système visuel. La couleur n'est cependant que l'une des nombreuses caractéristiques des images. La luminance de l'image ou de l'objet contenu dans l'image influence aussi probablement la catégorisation. De même, le contenu de l'image, en termes de configuration de la cible, du nombre de parties caractéristiques visibles de l'animal et éventuellement du type d'animal présenté, pourrait biaiser la catégorisation.

La recherche du rôle de ces caractéristiques n'est pas innocente. Il est possible que la catégorisation passe par la détection d'éléments diagnostiques des animaux dans les images (Schyns 1999), par exemple la présence d'un œil suffirait à détecter un animal. Les bases neurophysiologiques d'une telle catégorisation sont présentes dans le système visuel : dans IT, des neurones répondent à des parties du corps, aux yeux et aux visages (Gross, 1972; Perrett et Rolls, 1982; Wachsmuth et al, 1994). À l'opposé, au lieu de placer l'objet au centre de la perception, on peut considérer que le stimulus forme un tout, de sorte que la catégorisation de l'objet dépende fortement du contexte dans lequel l'image est présentée (De Graef et al, 1992).

Déterminer les caractéristiques des images les plus critiques pour la catégorisation nous permettra d'apporter des arguments en faveur d'une de ces deux hypothèses ou de montrer qu'elles ne sont pas mutuellement exclusives.

Il semble difficile de poser la question de l'étude simultanée d'un grand nombre de caractéristiques. Habituellement, pour étudier l'effet une caractéristique donnée, on construit deux groupes d'images, l'un contenant la caractéristique étudiée et l'autre non, et l'on compare les réponses des sujets sur ces deux groupes. Nous utiliserons ici une approche différente : à partir de 200 cibles catégorisées par un grand nombre de sujets, nous avons déterminé *a posteriori* le contenu des images, et pris soin de construire - pour les comparer - des classes de caractéristiques d'effectif suffisant. De cette façon, il nous a été possible d'analyser l'effet d'un grand nombre de caractéristiques. Dans un premier temps, nous allons confirmer pour un grand nombre de sujet le rôle négligeable des informations chromatiques, puis nous analyserons l'influence d'autres caractéristiques des images dans la catégorisation rapide chez l'homme.

3.1 - Matériel et méthodes

Le protocole est identique à celui de l'expérience précédente pour les sujets humains, mais le nombre de sujets est plus important afin de rendre possible l'analyse des réponses en fonction des caractéristiques des cibles. Quarante sujets - 20 hommes et 20 femmes[1] d'âge moyen d'environ 23 ans - ont catégorisé chacun les 400 images utilisées précédemment pour étudier le rôle de la couleur. Parmi ces 400 images présentées aléatoirement, 200 images sont présentées en couleur - 100 animaux et 100 distracteurs - et 200 images sont présentées en noir et blanc - 100 animaux et 100 distracteurs. Pour les sujets novices, 200 images supplémentaires sont présentées avant le début de l'expérience proprement dite afin qu'ils se familiarisent avec la tâche.

Dans toutes les analyses qui suivent, certaines précautions ont été prises pour garantir la validité des résultats présentés. En particulier, pour tous les tests se basant sur les caractéristiques des images, toutes les classes de caractéristiques représentent au moins 10 % du total des cibles. Les tests de significativité présentés dans les figures sont indiqués au

[1] Les résultats concernant 10 de ces 40 sujets ont déjà été considérés dans la comparaison homme/singe pour la catégorisation d'images en NB et en couleur au chapitre précédent.

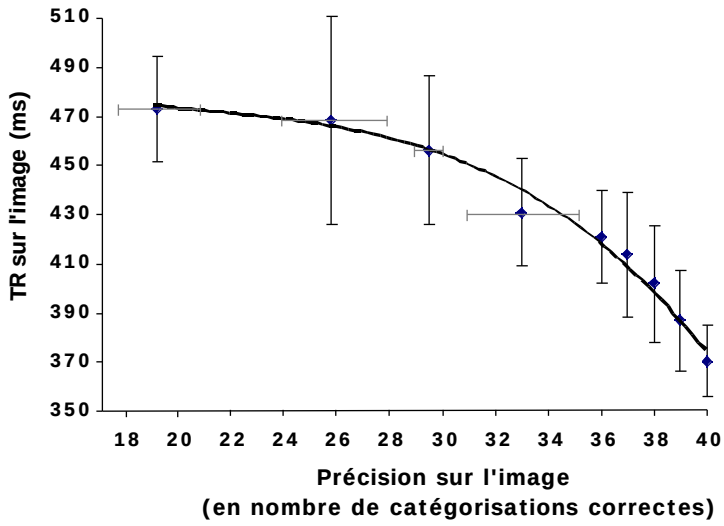


Figure 3.1 : relation entre la précision sur les images, exprimée en nombre de sujets ayant catégorisé cette image correctement et le TR moyen des sujets sur ces images. Les barres verticales représentent la déviation standard des réponses des sujets. Les images pour lesquelles plus de 4 sujets commettent des erreurs sont regroupées pour une meilleure lecture du graphique (les classes sont de même effectif et leur étendue est représentée par les barres grises horizontales). La corrélation entre la précision et le TR pour les images présentant moins de 5 erreurs est très forte (R^2=0,978).

moment où leur seuil franchit 5 %, mais ils deviennent tous fortement significatifs pour l'ensemble des réponses[2].

3.2 - Résultats

Tout d'abord, du fait du nombre important de sujets, il est possible de tenter de déterminer la corrélation existant entre la précision des sujets sur une image donnée et leur temps de réaction sur cette image. Si une cible est difficile à catégoriser, c'est-à-dire que de nombreux sujets commettent une erreur sur cette image, alors on peut s'attendre à ce que les TRs des sujets ayant correctement catégorisé cette image soient plus longs et c'est en effet ce que l'on observe (figure 3.1). Les images en couleur et en NB, prises indépendamment, suivent la même tendance[3]. Cela indique que, pour les images posant une difficulté, le TR est un bon indice pour situer l'image sur une "échelle de difficulté" valable pour l'ensemble des sujets[4].

Les premières analyses effectuées tentent d'estimer le rôle de la couleur afin de répliquer les résultats obtenus précédemment chez le singe et chez les 10 sujets humains. On observe que la précision des sujets est plus élevée pour les images en couleur tant pour les cibles que pour les distracteurs[5]. Cependant cette différence de précision n'est significative qu'à partir de

[2] Pour chaque test, j'ai pris soin de vérifier que sur l'ensemble des données, le test en question devenait significatif à moins de 0,1 %. Par chance, je n'ai pas eu à traiter de cas ambigus où le test n'atteignait pas ce seuil.

[3] La relation entre la performance et les temps de réaction semble cependant suivre une loi plus linéaire pour les images en couleur.

[4] J'ai également vérifié que les performances ainsi que la distribution des temps de réaction étaient similaires (c'est-à-dire non significativement différentes) entre les sujets hommes et femmes qui ont participés à l'expérience.

[5] Cibles en couleur 96 %; cibles en NB 93,2 %; distracteurs en couleur 93,9 %; distracteurs en NB 90,8 %.

310 ms, ce qui signifie tout de même que pour les 13 % des cibles les plus rapidement catégorisées on n'observe aucune différence entre les images en NB et en couleur (figure 3.2). Les temps de réaction sont également plus rapides pour les cibles en couleur que pour les cibles en NB[6], mais cette différence n'est significative qu'à partir de 480 ms[7]. L'absence de couleur n'a donc qu'un effet tardif sur la catégorisation compatible avec les résultats obtenus précédemment chez le singe pour la catégorie "animal"[8].

Pour les images en couleur et en noir et blanc, il me semble également intéressant d'estimer le rôle de la luminance de l'image sur la catégorisation. On peut en effet s'attendre à ce que les images de plus forte luminance soient plus rapidement catégorisées, la rétine répondant plus rapidement à ce type d'image. Pour estimer la luminance des images, la moyenne de la luminance de chaque point de l'image est calculée[9]. La figure 3.2 présente la performance des sujets en fonction du temps pour les images en couleur et en noir et blanc pour des fortes luminances et des faibles luminances[10]. Très clairement, la baisse de luminance n'a aucun effet pour les images en couleur et n'a qu'un effet tardif pour les images en noir et blanc : en NB pour de faibles luminances, la précision de l'ensemble des sujets diminue significativement après 510 ms (figure 3.2). En analysant uniquement la luminance de l'animal (isolé par découpage, cf. images dans la figure 3.3), il est également possible d'estimer le rôle de la luminance moyenne de l'objet-cible lui-même dans la performance. Le seul effet que je mets en évidence (mais est-il reproductible?) est une catégorisation plus rapide quand la luminance de l'animal est plus importante pour les images en couleur[11]. La différence en termes de temps de réaction est significative à partir de 400 ms. On observe la même tendance pour la précision des sujets, supérieure significativement dès 280 ms pour une forte luminance. En NB, seule la précision des sujets est affectée, elle diffère tardivement à 440 ms entre les images de faible et de forte luminances. Il est possible en effet qu'il soit plus

[6] 384 ms (médiane 374 ms) pour les cibles en couleur et de 387 ms (médiane 377 ms) pour les cibles en NB.

[7] Sur l'ensemble des réponses Man Whitney U-test $U=6878209$, $p<0.003$.

[8] Au chapitre précédent, il semblait que la différence entre images en couleur et NB était encore plus faible. J'attribue la différence avec cette expérience au fait que les 10 sujets de l'expérience précédente étaient très entraînés à la tâche.

[9] En noir et blanc, cela consiste à effectuer la moyenne des niveaux de gris de chaque point. En couleur, le calcul de la luminance ne peut s'effectuer par simple moyennage et j'ai donc reconverti ces images en noir et blanc avant de calculer la luminance moyenne (en utilisant le même logiciel que celui utilisé pour convertir les images en NB).

[10] Deux catégories de même effectif.

[11] Temps de réaction moyen de 374 ms (TR médian 364 ms) pour la moitié des animaux en couleur présentant une luminance plus importante (seules les images ne contenant qu'un animal sont prises en compte). Temps de réaction moyen de 388 ms (TR médian 380 ms) pour l'autre moitié des animaux de luminance plus faible.

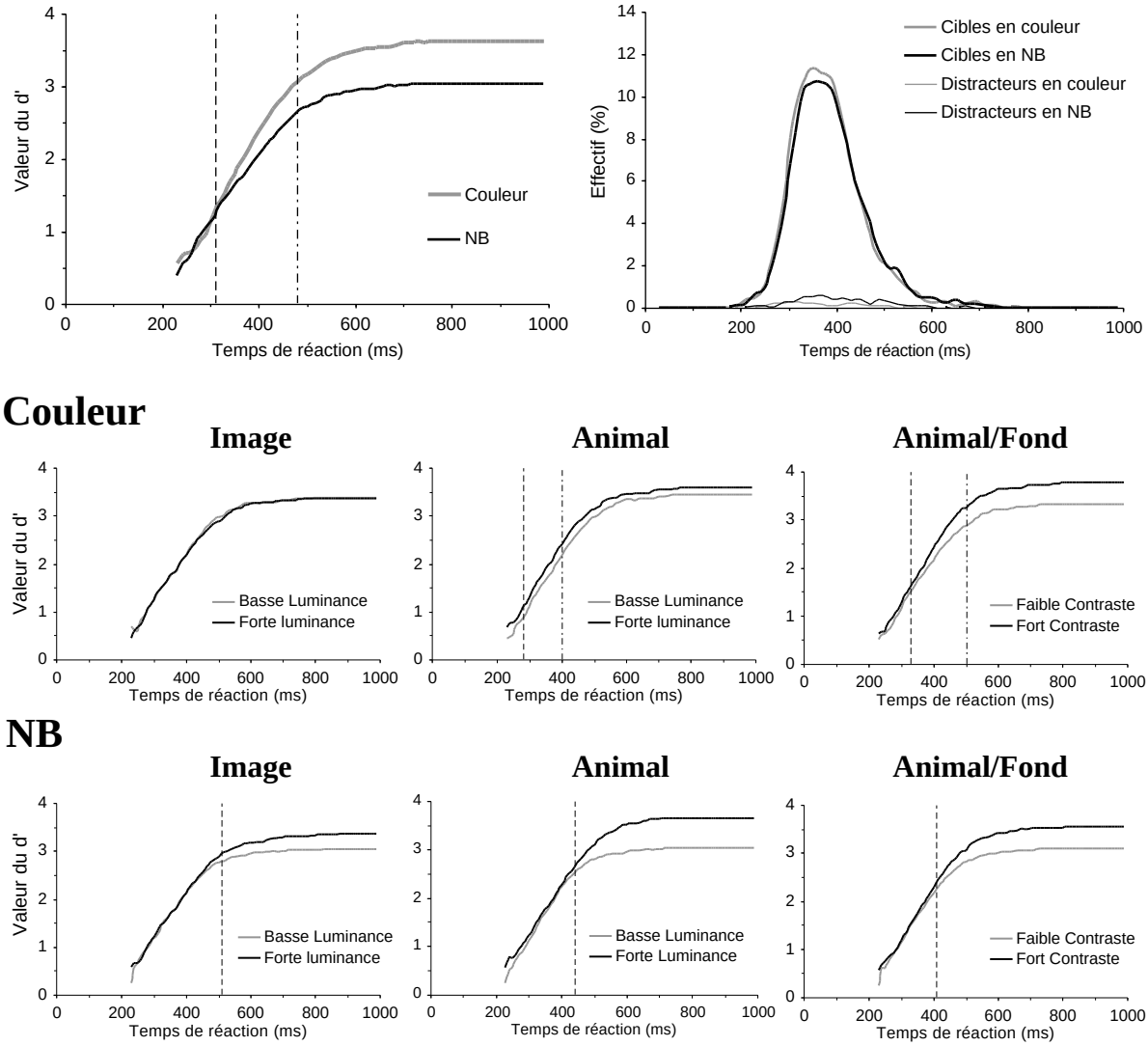


Figure 3.2 : illustration de la différence entre les images en couleur et en NB. En haut à gauche, d' pour l'ensemble des images en couleur et en noir et blanc. En haut à droite, distribution des temps de réaction pour les cibles et les distracteurs en couleur ou en noir et blanc. En bas, analyse des d' en fonction de la luminance de l'image, de la luminance de l'animal et du contraste (différence de luminance) entre l'animal et le fond. Les traits verticaux indiquent en pointillés noirs réguliers : le moment auquel un test de $\chi 2$ sur la précision des sujets devient significatif et en pointillés irréguliers le moment auquel un test Man Whitney sur les TR devient significatif. Dans le cas où les tests franchissent 5 %, la valeur pour l'ensemble des images (sans limitation de temps) est toujours inférieure à 0,001. Les différences significatives semblent parfois précoces (i.e. sous 400 ms) mais elles n'affectent pourtant qu'une partie tardive des catégorisations puisque la médiane des temps de réaction est de 375 ms.

difficile de détecter un animal dans les images sombres en NB alors que, pour les images en couleur, les informations chromatiques compensent le déficit de luminance.

Pour les images en couleur, les résultats indiquent que c'est la luminance de l'objet-cible qui importe et non la luminance globale de l'image. Les résultats sont plus difficiles à interpréter en noir et blanc puisque cet effet ne semble pas présent. Il est possible d'envisager pour les images achromatiques, que les détections rapides de cibles soient dépendantes du

contraste de la cible. J'ai donc tenté d'évaluer le contraste de l'objet par rapport au fond en calculant la différence entre la luminance du stimulus et celle du fond. La précision des réponses les plus rapides, à la fois sur les images en couleur et en NB, semble indépendante du contraste stimulus/fond qui n'intervient que pour des latences supérieures à 320 ms en couleur et à 410 ms en NB. Dans les deux cas cependant, la précision des réponses les plus lentes des sujets dépend fortement du contraste de la cible par rapport au fond, la précision étant logiquement plus faible pour les stimuli peu contrastés.

Pour tenter de synthétiser ces résultats, une luminance faible de l'image ou de l'animal-cible contenu dans l'image affecte significativement la vitesse de catégorisation des images présentées en couleur mais pas celle des images achromatiques. Il est possible qu'en couleur les propriétés de surface des objets contenus dans l'image, dont la luminance fait partie, soit plus importante pour la catégorisation.

Dans toutes les analyses qui suivent, les images en couleur et en NB ont été regroupées afin d'obtenir un nombre suffisant d'essais pour l'analyse de chaque caractéristique.

Nous allons tout d'abord nous intéresser au rôle joué par l'importance de la surface occupée par l'animal-cible dans les images. On peut imaginer qu'il est plus facile de catégoriser les images dans lesquelles les animaux occupent une grande surface plutôt que celles dans lesquelles ils n'occupent qu'une petite surface. Pour estimer le rôle de la surface des animaux dans les images, les animaux sont découpés (figure 3.3) puis la surface de l'animal est calculée en comptant le nombre de points manquants dans l'image originale[12]. Les animaux sont répartis en 4 groupes de même effectif en fonction de leur taille : "très petits", "petits", "grands" et "très grands". Il apparaît très clairement que la performance des sujets est plus faible pour les animaux "très petits" que pour les animaux "grands" (figure 3.3). Cependant, on n'observe pratiquement aucune différence entre les animaux "petits" et les animaux "très grands" tant au niveau de la performance que de la distribution des temps de réaction. En normalisant les courbes de performance, on montre que la dynamique de catégorisation semble être la même dans tous les cas, les petites cibles étant simplement plus difficiles à catégoriser que les grandes et cela presque indépendamment des temps de réaction[13]. La précision faible obtenue sur les animaux "très petits" est probablement due à la difficulté de détection de l'animal.

---

[12] Cette opération n'a été effectuée que sur les images ne contenant qu'un seul animal (144/200 cibles).

[13] Cette courbe normalisée est compatible avec la très tardive différence (significative à $p<0,05$ à 540 ms) des temps de réaction observés sur les "très petits" et les "grands" animaux.

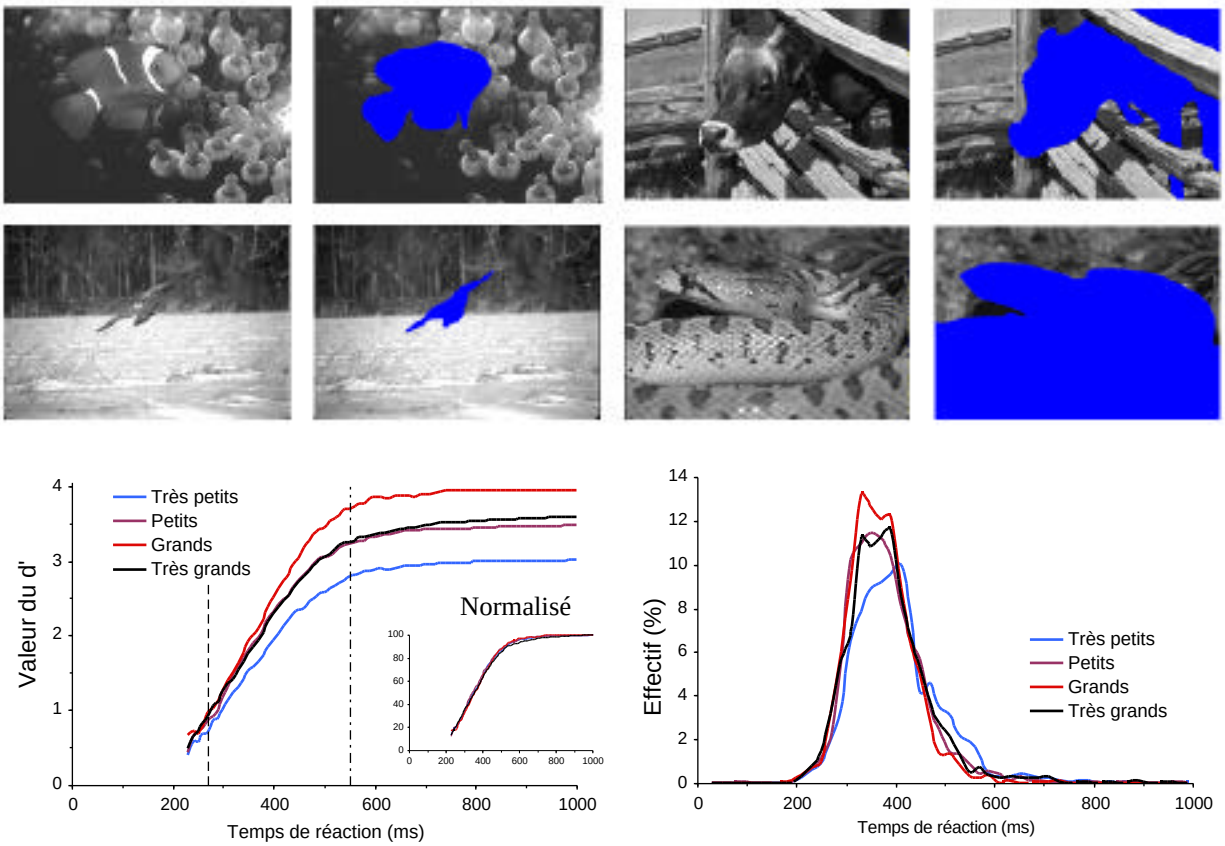


Figure 3.3 : en haut, exemple d'images découpées pour calculer la surface occupée par l'animal dans les images en NB. Les images en couleur sont converties en NB pour calculer la surface de l'animal. Les courbes, en bas à gauche, représentent les variations de la précision des sujets en fonction du temps pour 4 classes de même effectif constituées d'après la surface des animaux dans l'image. Les traits verticaux en pointillés indiquent l'instant où le résultat d'un test statistique entre le groupe "très petit" et "grand" devient significatif ($\chi 2$ sur la précision en pointillés réguliers et Man Whitney sur les TRs en pointillés irréguliers). Les courbes normalisées, identiques pour les 4 classes, montrent que c'est principalement la précision globale des sujets qui est affectée d'une classe à l'autre et donc que cet effet ne dépend que faiblement des TRs (cf. texte).

Une des raisons pour lesquelles les animaux du groupe "très grands" sont catégorisés moins aisément que les animaux du groupe "grand" pourrait découler de l'importante surface qu'ils occupent dans l'image et de leur présentation souvent partielle. Pour tester le rôle de la saillance de l'animal, les animaux sont répartis en deux classes : partiellement visibles et totalement visibles[14]. De la même façon, on peut estimer l'influence d'autres caractéristiques ayant trait à la configuration des animaux dans les images sur la difficulté de détection: le nombre d'animaux présents, la position de l'animal de profil ou de face, canonique ou inhabituelle (figure 3.4). Le fait que l'animal soit visible entièrement ou partiellement ne

[14] Au départ, j'avais considéré deux catégories d'animaux partiellement visibles, les animaux dont une partie est cachée par un objet dans l'image (par exemple la vache de la figure 2) et les animaux dont une partie sort de l'image (le serpent dans la figure 2). Une différence très tardive et très faible apparaissait entre ces deux catégories que j'ai préféré regrouper du fait de la faiblesse de leurs effectifs.

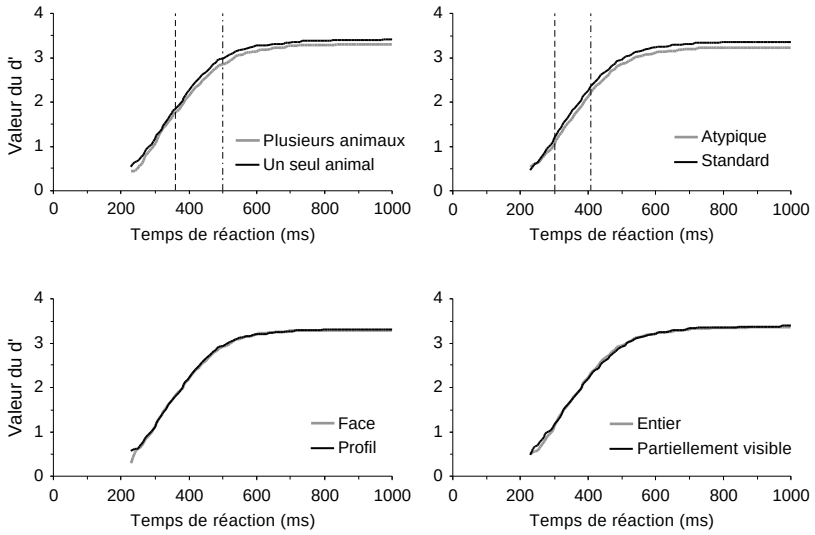


Figure 3.4 : rôle de la configuration des cibles sur la catégorisation des animaux. On analyse, en particulier, le rôle du nombre d'animaux présents dans l'image, le type de vue (standard/atypique et face/profil), et le fait que les animaux soient totalement visibles ou partiellement visibles dans l'image. Comme dans les figures précédentes, les lignes verticales en pointillés représentent les tests de significativité.

semble avoir aucune influence sur la catégorisation comme le montrent les deux courbes de performance confondues dans la figure 3.4. De même, le fait que l'animal soit de profil ou non n'influence pas le processus de catégorisation. Par contre, la performance des sujets apparaît influencée par la position canonique ou atypique de l'animal. Ils sont significativement moins précis et moins rapides lorsque les animaux sont présentés en position inhabituelle (par exemple l'oiseau couché dans la figure 3.8). De la même façon, la performance des sujets diminue avec le nombre d'animaux présents dans l'image[15]; ce résultat pouvant être attribué à la petite taille de chacun des animaux présents[16].

La taille de l'animal est un facteur important pour la reconnaissance mais peut-être est-ce lié au nombre de membres visibles, d'yeux... Lorsque l'animal occupe une grande surface dans l'image, il est par exemple probable que ses yeux seront visibles. J'ai donc tenté de déterminer l'influence de différentes caractéristiques de ce type, en particulier le nombre de membres visibles (pattes, queue ou nageoires pour les poissons), le nombre d'yeux visibles[17] et le fait que la bouche (ou le museau) soit visible ou non (figure 3.5). Dans les trois cas, la précision est affectée de façon relativement précoce, pour des réponses déclenchées avec des temps de réaction inférieurs à 300 ms. La présence des membres semble particulièrement critique, affectant la précision dès 270 ms ce qui semble indiquer une prépondérance de ce type

[15] Le nombre d'images contenant plusieurs animaux étant relativement faible (28 % des images), il ne m'a pas été possible d'estimer les variations de la performance en fonction du nombre exact d'animaux.

[16] Cependant la surface des animaux pour les images contenant plusieurs animaux n'a pas été calculée et il n'est pas possible de déterminer l'interaction entre les deux variables.

[17] Les performances sont similaires pour les images contenant 2 yeux et celles qui n'en contiennent qu'un. Les deux classes ont donc été regroupées.

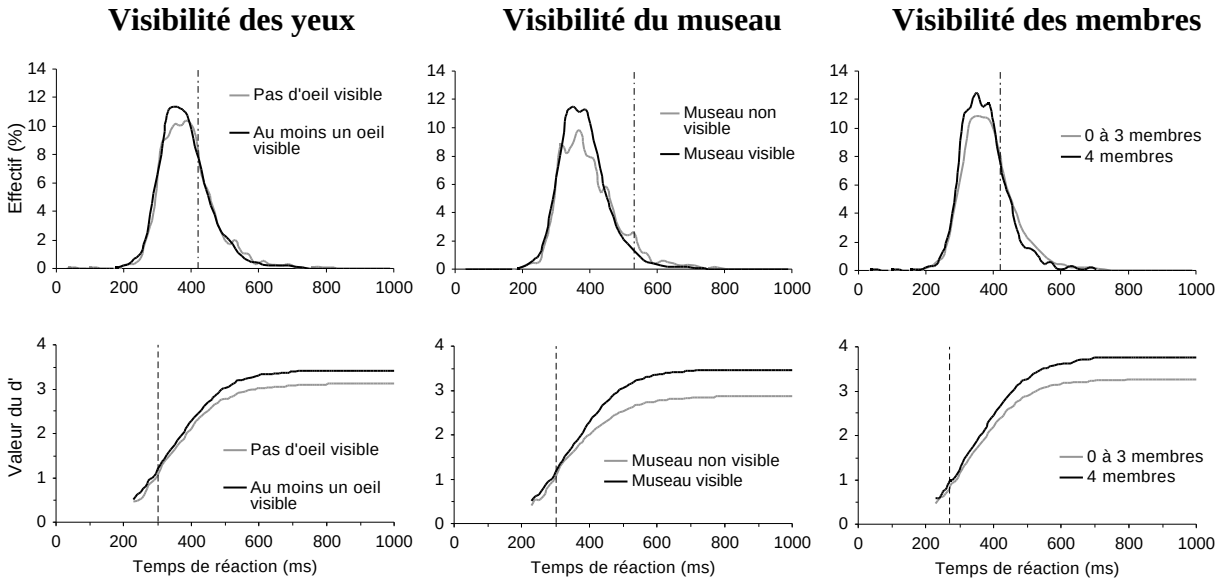


Figure 3.5 : influences de la configuration visible de l'animal sur la catégorisation : visibilité des yeux, du museau (ou de la gueule suivant le type d'animal) et enfin du nombre de membres visibles de l'animal (pattes et queue pour les mammifères; nageoires pour les poissons). Sur la ligne du haut, les TRs sont représentés (pas de temps de 20ms) et la ligne verticale en pointillés irréguliers indique le moment à partir duquel le test de Mann Whitney sur les TRs devient significatif. Sur la ligne du bas, la précision cumulée en fonction du temps est représentée pour chaque caractéristique. La ligne verticale en pointillés indique le moment à partir duquel le test de $\chi 2$ sur la précision devient significatif.

d'information pour la catégorisation ultra rapide[18]. Des différences au niveau des temps de réaction sont également présentes dans les trois cas, mais plus tardivement (figure 3.5). La présence d'un œil, de la bouche ou de membres a donc une influence sur les processus de catégorisation.

La figure 3.6 indique la répartition des images en fonction de leur contenu dans un espace multidimensionnel. Dans cette analyse, il s'agit uniquement de déterminer les cooccurrences de caractéristiques dans les images. On constate plusieurs groupes d'images, de la présence/absence de parties du corps (par exemple quand 4 pattes sont visibles, il s'agit généralement d'un seul animal) ou de la position de l'animal (un animal de profil est en général dans une position standard et visible en entier). On peut mettre en parallèle cette décomposition avec celle que l'on peut réaliser sur la base des TRs des sujets. La figure 3.7 indique la répartition des TRs pour ces caractéristiques et les groupements qui s'effectuent spontanément lors d'une analyse multidimensionnelle. Les classes d'images ainsi définies se répartissent sur 2 axes principaux, le premier représentant la contribution aux TRs et le

[18] Les courbes de performances (d') pour les classes d'animaux à 0, 1, 2 et 3 et plus de 4 membres sont pratiquement confondues et je les regroupe donc pour l'analyse. Dans le cas de plus de 4 membres visibles, le nombre d'image est très faible (4,5 % du total); dire que la performance est meilleure dans le cas de 4 membres me paraît être une interprétation abusive.

	présent	absent
une bouche	161	39
profil	135	65
partiel	67	133
position standard	134	66
plusieurs animaux	56	144
au moins 1 œil	136	64
4 membres	33	167

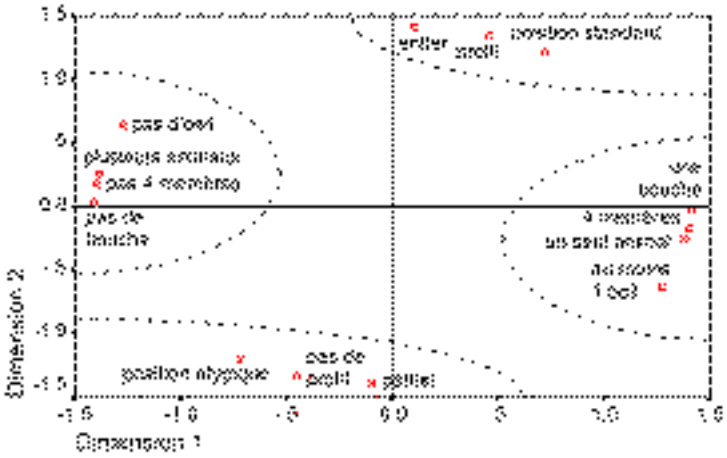


Figure 3.6 : à gauche, nombre d'images, parmi 200, dans chaque classe de caractéristique. A gauche, plongement de ces classes dans un espace multidimensionnel qui permet d'estimer leur distance (c'est-à-dire le taux de co-occurrence des caractéristiques dans les images). Les classes de caractéristiques complémentaires ("partiel"/"entier" par exemple) sont symétriques par rapport à l'origine dans cet espace. Très clairement 4 groupes, entourés en pointillés pour une meilleure lecture du graphique, se profilent : à gauche, les caractéristiques "plusieurs animaux", "œil non visible" et "bouche non visible"; à droite, "un seul animal est visible", "les 4 membres visibles", "bouche visible" et "yeux visibles"; en haut, les caractéristiques "animal entièrement visible", "de profil" et "dans une position standard"; enfin en bas, les caractéristiques "pas de profil", "dans une position atypique" et "partiellement visibles" sont regroupées. Cette description est inhérente au classement que j'ai effectué sur les images - elle ne dépend donc pas de la catégorisation effectuée par les sujets - et ne reflète que la probabilité de la présence simultanée des caractéristiques.

second un classement de l'importance des caractéristiques. La caractéristique la plus importante semble être la présence d'une bouche et la moins importante le fait que l'animal soit de profil ou de face. Cela reproduit à peu près les résultats que nous avons obtenus en analysant la performance des sujets en fonction des caractéristiques des images, la corrélation n'étant pas totale du fait que l'on considère, dans l'analyse multidimensionnelle, l'ensemble des TRs dans sa globalité sans considérer séparément les TRs les plus précoces.

Toutes les caractéristiques testées jusqu'à présent concernaient la configuration spatiale et les caractéristiques physiques des images indépendamment du type d'animal traité. Cependant, on est en droit de se demander si les mammifères par exemple, qui sont pour nous l'archétype de l'animal par excellence, ne sont pas traités plus rapidement que les poissons, les reptiles ou les oiseaux. Au cours d'une dernière analyse, je tente donc d'estimer l'influence de l'espèce de l'animal sur la détection, que ce soit un mammifère, un poisson, un reptile ou un oiseau[19]. Je différencie également plusieurs sous-catégories de mammifères : celle d'animaux assez fréquents et d'animaux peu fréquents (il s'agit en règle générale de lémuriens, de

[19] J'espère que les entomologistes ne me tiendront pas rigueur d'avoir classé les deux ou trois images de batraciens parmi les reptiles.

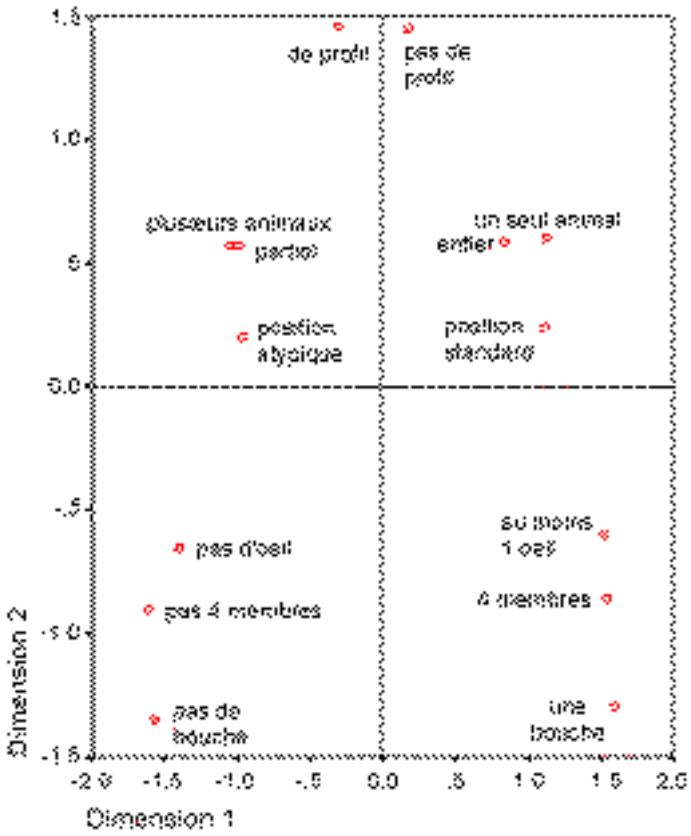
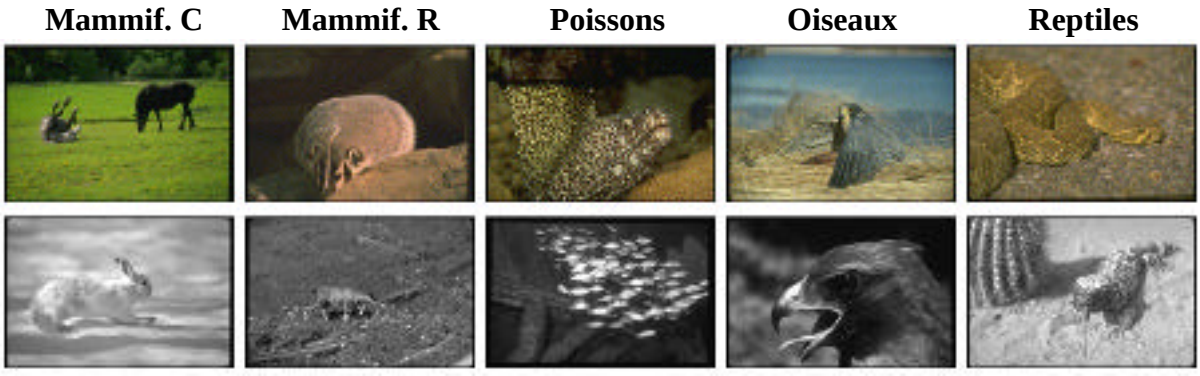
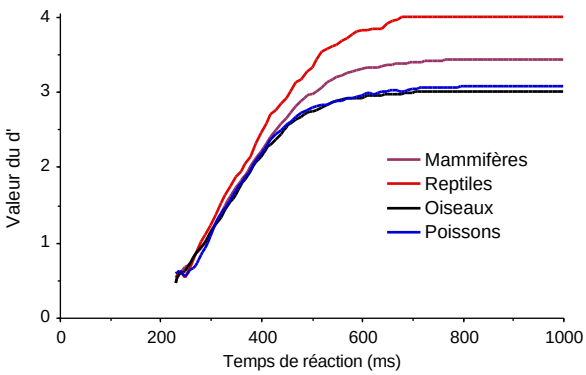


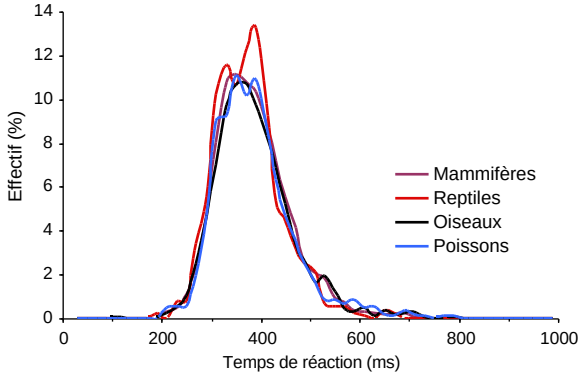
Figure 3.7 : répartition des caractéristiques dans un espace multidimensionnel calculé à partir des temps de réaction des sujets pour chaque couple de caractéristiques. Cependant, pour chaque couple de variables, les images communes ne sont pas considérées dans le calcul des moyennes de TRs et la distance en terme de TR entre deux classes ne dépend donc pas des images communes. De manière tout à fait surprenante, les dimensions principales semblent très clairement définies. La dimension 1 est probablement celle de l'influence de la caractéristique sur les TRs (lents à gauche et plus rapides à droite). La seconde dimension semble être celle des caractéristiques classées en fonction de l' importance de leur influence sur les TRs, la bouche, les yeux et les membres en bas semblant plus importants que la position de l'animal.

bovidés atypiques...). Comme critère (objectif ?) de sélection, je classe d'un côté les animaux dont je connais parfaitement le nom et d'un autre ceux dont je ne suis pas capable de déterminer le nom exact[20]. Les résultats sont surprenants puisqu'ils montrent que les reptiles sont catégorisés plus facilement que les autres types d'animaux, d'autant plus que certains reptiles ne possèdent pas de pattes dont on a montré que leur présence pouvaient permettre une catégorisation plus précoce (figure 3.8). Viennent ensuite les mammifères puis les oiseaux et les poissons. Pour ces deux derniers types d'animaux, les oiseaux et les poissons, la dynamique de catégorisation semble très similaire. En ce qui concerne les mammifères fréquents et ceux moins fréquents, aucune différence significative n'apparaissant entre les deux classes d'images, elles ne sont pas montrées. Ce dernier résultat indique que l'identité exacte de l'animal et sa fréquence importe peu dans la catégorisation. Cela renforce, comme on le verra plus loin, l'hypothèse qu'il n'est pas nécessaire de déterminer l'identité de l'animal pour le catégoriser et que le processus de catégorisation est probablement basé sur des éléments diagnostiques des animaux.

[20] Les races de chiens inhabituelles (dont je ne connais pas le nom) sont également incluses dans cette catégorie.







	χ2	Man Whitney
Reptiles/oiseaux	310	480
Reptiles/poissons	320	600
Reptiles/mammif.	310	460

	χ2	Man Whitney
Mammif./poissons	450	ns
Mammif./oiseaux	400	ns
Poissons/oiseaux	ns	ns
Mammif. connus/rares	ns	ns

Figure 3.8 : influences des types d'animaux sur la catégorisation. En haut les diverses classes d'animaux : mammifères, poissons, oiseaux et reptiles sont représentés pour les images en couleur et pour les images en noir et blanc. Les mammifères sont divisés en deux sous groupes, les mammifères communs (Mammif. C) et les mammifères rares (Mammif. R). Les courbes de d' à gauche représentent la variation de la précision des sujets en fonction du temps sur l'ensemble des images d'une catégorie et les courbes de droite les distributions des TRs. En bas, les tests de significativité sur la précision et les TRs sont présentés pour les classes d'animaux prises 2 à 2. La performance pour les reptiles est significativement plus élevée que pour les autres catégories d'animaux. De même la performance sur les mammifères est meilleure que celle observée pour les poissons et les oiseaux. Les deux sous-catégories de mammifères (communs et rares) ne présentent aucune différence, la courbe de performance en fonction du temps n'est pas présentée.

3.3 - Discussion

Les analyses effectuées ici reproduisent les résultats obtenus chez l'homme et le singe en ce qui concerne la catégorisation des images en couleur et en NB notamment pour des catégorisations très rapides (Delorme et al, 2000). Parmi les autres facteurs influençant la catégorisation rapide, la luminance de l'objet-cible pourrait être un facteur critique pour les

images en couleur[21]. La luminance de l'objet et de l'image intervient également pour les images en NB mais de façon tardive, ce qui laisse penser que ce type de caractéristique ne serait pas de première importance dans la catégorisation de ces images.

Concernant la configuration du contenu des images cibles, le nombre d'animaux et la surface qu'occupe l'animal dans l'image sont également des facteurs critiques pour la détection de l'animal, la performance étant maximale pour des tailles d'animaux suffisamment grands mais pas trop. Cet avantage ne semble pas lié avec la visibilité totale ou partielle de l'animal, sa présentation de profil ou de face. Il serait peut-être lié à une position typique (ou non) de l'animal qui influence également de façon précoce la performance des sujets.

La visibilité des parties du corps des animaux - nombre de membres, œil, bouche - semble également intervenir précocement dans les processus de catégorisation puisqu'elle affecte la précision des réponses relativement précoces (TR< 300 ms).

Enfin, l'espèce de l'animal influence aussi fortement la catégorisation. Le cerveau n'est pas entomologiste puisque les reptiles et les oiseaux très proches dans l'évolution sont catégorisés de manière très différente, les reptiles étant catégorisés très rapidement et avec une bien meilleure précision. On peut mettre en relation ces résultats avec les réactions innées de recul et de panique des singes à la vue de serpents[22] (Cook et Mineka, 1990) et il est possible que chez les sujets humains aussi, de telles réactions soient présentes[23].

3.3.1 - Catégorisation diagnostique des images

En neurosciences, l'un des grands débats est celui de savoir si la perception - et par conséquent, dans une certaine mesure, la catégorisation - est globale pour l'image ou si elle passe par ses parties, en particulier les objets contenus dans l'image. Quand on demande aux sujets ce qu'ils voient en premier dans les images flashées, tous sont unanimes pour répondre qu'ils ont perçu la scène dans son intégralité. Le type d’analyse que j'ai effectué sur le contenu des images et les répercussions sur les temps de réaction et la performance des sujets est inédit - à ma connaissance - et permet d'aborder cette question sous un jour nouveau.

---

[21] Des analyses plus poussées, qui ne peuvent pas être effectuées sur la seule base des valeurs des points mais nécessitent des appareils de mesure de luminance, seraient nécessaires pour estimer le rôle de l'énergie lumineuse dégagée par l'objet-cible et le fond de l'image.

[22] Le singe adulte qui n'a jamais vu de serpent présente également cette réaction innée.

[23] Pour revenir à la phénoménologie, il est intéressant de noter que dans notre expérience, le sujet humain retire sa main du bouton quand il voit un animal. Ce mouvement de retrait est donc peut-être plus pertinent pour les reptiles que pour les autres types d'animaux. Il serait peut-être intéressant de classer les animaux suivant le danger qu'ils peuvent représenter pour l'homme et d'analyser l'effet sur les temps de réaction.

Notre étude ne peut répondre qu'en ce qui concerne la catégorisation rapide, qui ne passe pas nécessairement par la perception consciente de l'image. Il semble que les caractéristiques de l'image et de son contenu aient une influence assez grande sur la précision et la vitesse de catégorisation. Cela penche plutôt en faveur d'une catégorisation basée sur les éléments des objets diagnostiques des animaux (cf. pour une revue Schyns 1999; Humphreys, 2000). Cela signifie que le sujet répondrait dès qu'il perçoit une partie d'un membre ou d'un œil, et pas nécessairement un animal. Cela signifie également que le traitement visuel dépend de la tâche.

D'autres travaux vont également dans le sens d'une catégorisation basée sur certaines caractéristiques des objets. Sands et al (1982), déjà mentionnés dans les chapitres précédents, analysent les erreurs des singes dans une tâche d'appariement d'images. Ils montrent que ces erreurs se décomposent sur plusieurs axes en fonction du contenu des images et de leurs caractéristiques : la taille et le nombre d'objets interviendrait peu pour les erreurs sur les images de fruits alors que le type de fruit (en grappe ou non) et leur couleur rendent compte de la majorité des erreurs. L'approche est similaire à celle que j'utilise mais les résultats ne sont pas directement comparables car les images catégorisées sont différentes. Dans notre cas le nombre d'animaux ainsi que leur taille semble intervenir dans la catégorisation. Cette étude et la nôtre indiquent cependant qu'il est possible que les caractéristiques importantes pour la détection d'objet puissent dépendre de l'objet à catégoriser.

Cependant la catégorisation ne se base pas sur les seules parties des objets : par exemple l'absence d'yeux dans les images contenant des animaux affecte les réponses mais pas de façon dramatique. La configuration des différents éléments est peut-être également très importante. Des travaux montrent que des images dont les parties sont mélangées sont catégorisées avec moins de précision et plus lentement chez l'homme (Cave et Kosslyn, 1993) et le singe (Vogels, 1999). De façon analogue, pour la reconnaissance des visages par exemple, on montre qu'elle dépend fortement de la configuration et de la distance relative des yeux et de la bouche (Cooper et Wojan, 2000). Bien que la relation ne soit pas directe avec notre étude, il semble cependant que la présence d'éléments diagnostiques et leur configuration spatiale soient deux éléments importants pour la catégorisation.

La reconnaissance diagnostique pose également la question de l'influence de la prototypicalité des objets sur la reconnaissance. Si le système visuel est biaisé vers la détection de certains modèles caractéristiques, il est probable que le fait que la position d'un objet soit typique ou non influencera la catégorisation. Des études montrent en effet que la reconnaissance des objets dépend de la configuration dans laquelle on a l'habitude de les voir (Palmer et al, 1981; Liu, 1996; Palmer, 1999). On demande aux sujets de classer des

photographies de différentes vues d'objets (cheval, voiture, maison, chaussure, réveil ...) en fonction de la typicalité de la vue. Par la suite, quand on demande à d'autres sujets de nommer le plus rapidement les objets contenus dans ces photographies, leur performance dépend de la vue présentée, les temps de réaction étant plus lents pour les vues atypiques (Palmer et al, 1981). Ces résultats sont compatibles avec les nôtres dans la mesure où des vues atypiques des objets sont catégorisées avec moins de précision et plus lentement que des vues typiques.

La question se pose donc de savoir si cet effet dépend de la familiarité de la vue : par exemple, il est possible que nous reconnaissions plus rapidement des objets s'ils sont dans une position que nous avons l'habitude de voir. Pour répondre à cette question, Edelman et Büllthof (1992) ont posé la question de la reconnaissance d'objets nouveaux, des *paperclips* constitués d'une ligne continue brisée en trois dimensions (cf. figure III.4.6 dans la partie dédiée à la modélisation pour un exemple de *paperclip*). Les objets étaient présentés initialement sous forme d'une suite de photographies qui induisaient un mouvement apparent. Par la suite, les sujets devaient déterminer s'ils avaient déjà vu ou non ces formes. Le taux de reconnaissance dépend d'une vue canonique de l'objet. Ce résultat montre qu'*a priori* la présence d'une vue canonique ne dépend pas uniquement de la familiarité de la vue, l'objet étant initialement présenté dans toutes les positions. L'existence de vues canoniques dépendrait donc également de phénomènes plus complexes, comme la relation que nous entretenons avec l'objet.

3.3.2 - L’objet et le contexte

La plupart des études en psychologie expérimentale abordent le problème du contenu de l'image par rapport au contexte dans lequel un objet est catégorisé. Dans des expériences de dénomination d'objets, il semble que la performance des sujets ne dépende pas de la présence d'objets distracteurs dans l'image, mais de la cohérence entre le fond et l’objet (Boyce et Pollatsek, 1992). Ces résultats sont compatibles avec ceux de De Graef et al (1992) qui montrent que la durée de fixation d'un objet dépend très peu de l'objet précédent. Dans cette expérience, des scènes dessinées au trait sont présentées à des sujets et le temps de fixation sur chaque objet est analysé par rapport à la cohérence avec la scène - par exemple un cochon dans une ferme est cohérent avec la scène alors qu'il ne l'est pas dans un supermarché - et par rapport à sa cohérence avec un autre objet sur lequel était initialement fixé le regard - un cochon a plus de chance d'être à côté d'une poule que d'un caddie. La durée de fixation d'un objet est censée dépendre du temps que met le système visuel pour traiter cet objet.

L'interaction entre les objets est très faible alors que l'interaction avec la scène semble relativement importante, la durée de fixation étant plus courte dans un environnement cohérent. Ces résultats indiquent que la scène serait traitée comme un tout notamment lors du début de l'exploration oculaire.

En opposition directe avec ces travaux, certaines études placent l'objet au centre de la perception. Les patients négligents[24], par exemple, semblent ne pas pouvoir percevoir le côté gauche des objets, ce qui impliquerait une représentation centrée sur l'objet (Tipper et Behrmann, 1996). De même, dans des tâches de mémorisation d'image, la cohérence des objets contenus dans l'image semble jouer un rôle prépondérant. Dans une expérience de mémoire à court terme, McKoon (1981) présente des photographies de scènes naturelles à des sujets humains. Il présente ensuite des portions de ces scènes aux sujets et leur demande si elles font partie des photographies. Il observe un amorçage positif si une partie d'une image est présentée après une autre partie de la même image initiale. Cet amorçage est plus important si les deux parties appartiennent au même objet et moins important pour deux parties dont l'une fait partie du fond et l'autre de l'objet. Cela indiquerait encore une fois que l'objet est central dans la perception et la mémorisation.

Il est cependant possible d'interpréter cette importance de l'objet en termes de propriétés de luminance des objets dans les images. Une analyse statistique des images montre que, dans les images naturelles, les caractéristiques de surface varie très peu au sein d'un même objet et de façon très importante entre deux objets (Ruderman, 1997). Si le système visuel optimise une quelconque fonction statistique, par exemple la redondance dans les images (Field, 1987), il est très probable que les objets deviendraient spontanément des entités à part entière au sein du système visuel. Cela est également compatible avec les résultats que nous avons montrés sur la luminance des objets dans les images en couleur. Alors que la luminance globale de l'image est neutre pour la catégorisation, la luminance de l'objet semble avoir un effet très précoce. Une étude plus poussée serait cependant nécessaire dans notre cas car nous n'avons pas retrouvé cet effet sur les images en NB.

Toutes ces études s'accordent pour dire que les caractéristiques des objets interviennent de façon différente dans la catégorisation. Nos résultats vont également dans ce sens, et il nous a même été possible d'établir une hiérarchie entre ces diverses caractéristiques.

[24] Le patient présente une lésion centrale dans l'hémisphère droit et n'est plus conscient de l'hémichamp visuel gauche.

Cependant, la simple présence d’une partie de l’animal ne suffit pas à induire des changements radicaux de précision et l'influence sur les TRs est souvent tardive, systématiquement supérieure à 400 ms, ce qui correspond à plus de la moitié des réponses des sujets. Ces résultats impliquent un traitement massivement parallèle dans le système visuel où l'absence d'une caractéristique pourrait être compensée par la présence d'une autre. Nous reviendrons sur ce point plus tard et sur les implications que cela peut avoir sur le traitement dans le système visuel.

Bien que nous ayons couvert un vaste ensemble de caractéristiques des images contenant des animaux, beaucoup d'autres restent à traiter. Nous avons vu ici que la tâche biaise probablement le traitement visuel vers une direction ou une autre. Nous allons donc maintenant essayer d'aller plus loin en tentant de modifier le temps de séparation entre deux images afin de déterminer le rôle d'une image sur la catégorisation de l'image subséquente.

4

L’attente du sujet

Dans le but de déterminer la durée minimale de traitement pour effectuer la tâche de catégorisation, je me suis posé le problème de l'influence du délai inter-stimulus (ISI) entre les images sur la performance des sujets et du rôle de l’image précédente dans la catégorisation.

Des études ont montré que dans le striatum et dans le cortex inféro-temporal des neurones anticipent l’apparition d’une cible ou d'un distracteur et refléteraient donc un effet d’attente (Apicella et al, 1992; Reutiman et al, 2000). Ainsi, l'activité des neurones dans IT augmente de façon linéaire pour atteindre son maximum quand le stimulus est supposé apparaître (Reutiman et al, 2000). D'autres études chez le chat indiquent que, dans une tâche visuo-motrice, l'attention focalisée et l'attente de la cible serait liée la présence d'ondes *beta* à la surface du scalp (Montaron et Fabre-Thorpe, 1996). Ces études indiquent qu'il est légitime de se poser le problème de l'attention temporelle.

A priori, on peut penser que des images présentées à intervalle fixe, à la milliseconde près, seront catégorisées plus rapidement que des images présentées à des intervalles de temps variables. En effet, dans le cas d'intervalles fixes, le sujet peut en quelque sorte optimiser son état d'attention au moment où va apparaître l'image et le fait que le sujet connaisse exactement la date d'apparition de la prochaine cible devrait induire une accélération non négligeable des

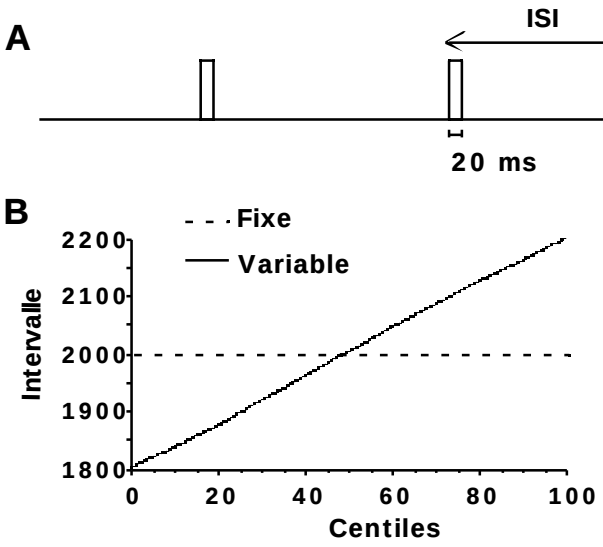


Figure 4.1 : catégorisation animal/non-animal avec délais inter-images (ISI) soit fixes, soit variables. A, chaque rectangle représente l'apparition d'un stimulus (pendant 20 ms). B, répartition des ISIs pour les deux types de séries pour l'ensemble des sujets.

temps de réaction. Nous verrons que le résultat est loin d'être aussi clair qu'on pourrait le penser.

4.1 - Matériel et méthodes

Onze volontaires - 5 hommes et 6 femmes, d'âge moyen 24 ans - effectuent une tâche de détection d'animaux dans des images naturelles. Chaque sujet voit 600 images réparties en 6 séries de 100 images contenant chacune 50 animaux et 50 distracteurs, séries qui sont mélangées aléatoirement avant chaque présentation.

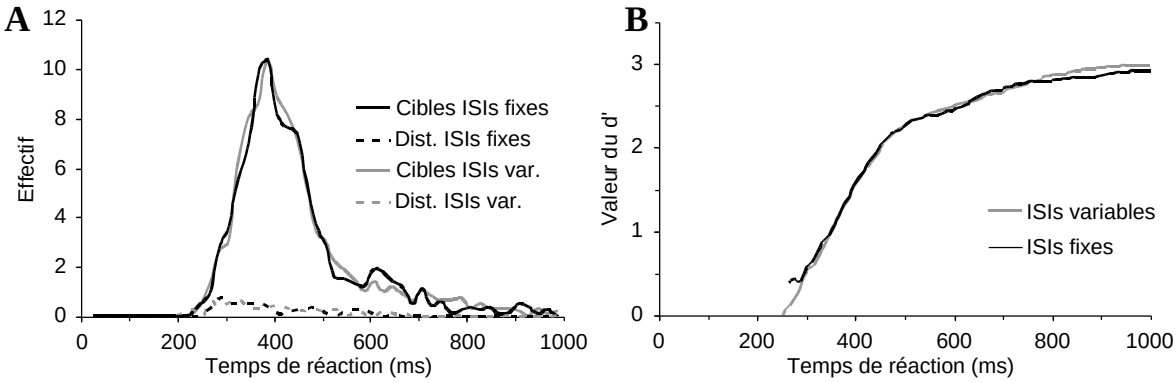
Une série sur deux est présentée avec un intervalle entre deux images d'exactement 2000 ms et une série sur deux est présentée avec un intervalle variant aléatoirement entre 1800 et 2200 ms (figure 4.1). La première série, de type ISIs fixes ou ISIs variables, alterne d'un sujet à l'autre[1].

Lors de l'analyse des résultats, les 10 premières images de chaque série, qui pourraient correspondre à une période de transition entre les deux conditions d'ISI, sont éliminées par mesure de précaution.

4.2 - Résultats

Les résultats sont surprenants puisque l'on n'observe aucune différence entre la condition où les ISIs sont fixes et celle où ils sont variables tant au niveau de la précision que des temps

[1] Six sujets ont commencé avec une série variable et 5 avec une série fixe. Afin de minimiser les effets dus à l'entraînement, les sujets qui n'étaient pas familiers avec la tâche devaient catégoriser 200 images (2 séries de 100 images, l'une à ISI fixe, l'autre à ISI variable) supplémentaires avant d'aborder le test proprement dit.



C

	ISIs fixes			ISIs Var.	
Sujets	Préc.	TR méd.	p (TR)	TR méd.	Préc.
S1	94,1	383	ns	388	96,3
S2	91,9	434	ns	431	92,2
S3	92,6	339	ns	337	93,3
S4	93,3	404	ns	412	94,1
S5	86,3	453	ns	473	83,3
S6	95,6	384	ns	389	97,0
S7	93,0	713	ns	673	93,0
S8	95,6	480	0,0009	451	95,9
S9	93,7	425	ns	441	95,6
S10	91,9	378	ns	388	90,4
S11	95,9	378	ns	373	95,6
Total	93,1	434	ns	432	93,3

Figure 4.2 : A, distribution des TRs des réponses correctes sur les cibles et incorrectes sur les distracteurs (Dist.) pour les conditions d'ISIs fixes et variables (ISIs var.). Le pas de temps est de 20 ms. B, précision en fonction du temps pour les deux conditions d'ISI. C, analyse individuelle de la précision des sujets (Préc. en %) et de leur TRs médian (TR méd. en ms) dans les conditions d'ISIs fixes et variables. La colonne du centre montre que la répartition des TRs dans les deux conditions est similaire (à une exception près, cf. texte). Le test statistique concernant la précision n'est jamais significatif.

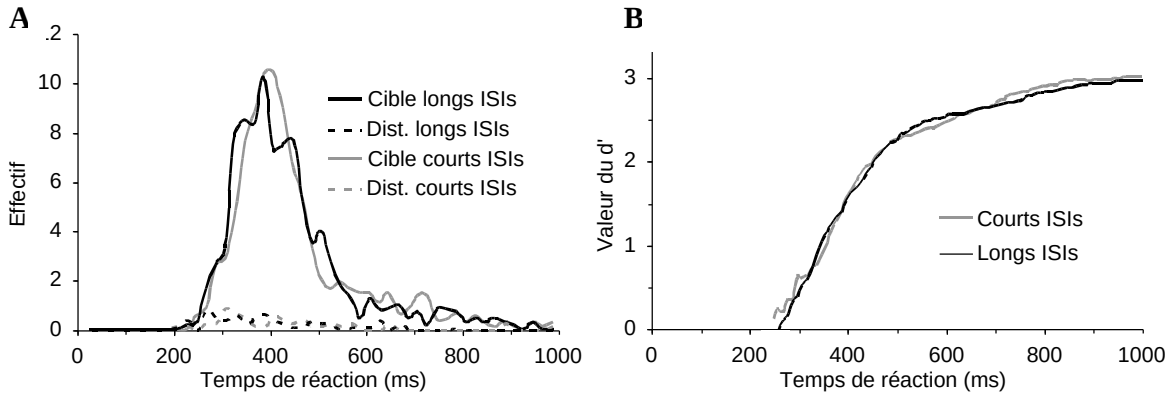
de réaction[2] (figure 4.2). La distribution des temps de réaction est également comparable dans les deux cas[3]. La vitesse de catégorisation ne semble donc pas dépendre du fait que chaque image soit présentée après un intervalle fixe ou variable. L'analyse du comportement individuel des sujets (figure 4.2) indique également que, à l'exception du sujet S8 dont le comportement est attribuable à la fatigue[4], aucune différence de précision ni de temps de réaction n'est observée. De même, quand on considère les performances sur les cibles et les distracteurs séparément, la performance des sujets ne diffère pas d'une condition à l'autre.

On est cependant en droit de se demander si, dans les séries ou l'ISI est variable, les images sont catégorisées de façon comparable ou si la catégorisation dépend de l'ISI. Pour ces séries, je tente donc d'établir une corrélation entre la durée de l'intervalle (ISI) et le temps de

[2] Temps de réaction Man Whitney U test U=940593 p=0,92; précision $\chi 2$ bilatéral, DDL=1, p=0,72.

[3] Sur les distributions de la figure 2, $\chi 2$ bilatéral DDL=52, p=0,89.

[4] Contrairement à la moyenne des sujets au fur et à mesure des séries le sujet S8 catégorise moins rapidement les cibles (dans l'ordre des séries, la médiane est égale 402 (fixe), 445 (var.), 454(fixe), 479 (var.), 494(fixe), 497 (var.)). Du fait de l'ordre des séries fixes et variables, les images des séries avec intervalles de temps variables sont donc catégorisées 10ms plus lentement pour ce sujet.



C

Sujet	courts ISIs TR médian	longs ISIs TR médian
S1	359	333
S2	393	399
S3	304	297
S4	386	355
S5	403	356
S6	357	353
S7	640	618
S8	452	440
S9	400	395
S10	354	348
S11	336	344
T-test apparié	**0,026**	

Figure 4.3 : A, répartition des temps de réaction pour les cibles et les distracteurs (Dist.) dans la condition à ISIs variables pour les images précédées d'un ISI inférieur à 2000 ms (courts ISIs) et celles précédées d'un ISI supérieur à 2000 ms (longs ISIs). Le pas de temps est de 20 ms. Bien que les longs ISIs induisent des réponses significativement plus rapides, cet effet n'est visible qu'à partir de 420 ms et disparaît après 470 ms : je ne l'ai donc pas indiqué sur la courbe. B, précision des sujets en fonction du temps (cf. texte pour la significativité). C, tableau des TRs médians des sujets pour les images traitées rapidement (TR<365ms). La distribution globale est significativement différente.

réaction. En effet, puisque la probabilité d'apparition de l'image augmente pour les intervalles longs, il est possible que les images apparaissant à ces latences soient catégorisées plus rapidement. La corrélation linéaire des ISIs et des TRs est très faible et non significative[5]. Cependant, en scindant les données en deux groupes, l'un pour les ISIs inférieurs à 2000 ms et l'autre pour ceux supérieurs à 2000 ms, une différence significative[6] apparaît entre les deux distributions de temps de réaction (figure 4.3). Lorsque les ISIs sont longs, la distribution des TRs observée est plus étalée que dans le cas d'ISIs courts. Les réponses précoces semblent plus nombreuses quand l'image est précédée d'un court ISI. L'analyse de la distribution des médianes des temps de réaction des sujets pour les réponses les plus précoces - dont l'effectif est fixé arbitrairement au quart du total des réponses - confirme cette observation[7]. L'analyse de la précision des sujets en fonction du temps (figure 4.3) nous montre qu'il n'y a que très peu de différences entre les deux conditions, de longs et de courts ISIs. S'il fallait dégager une tendance, il semble même que les courts ISIs permettent une meilleure performance, ce qui va à l'encontre des analyses sur les temps de réaction. Le fait que les courbes de performance (d')

[5] R=0,04, p=0,84.

[6] $\chi 2$ bilatéral DDL=54, p=0,008.

[7] t-test apparié, DDL=10, p=0,026.

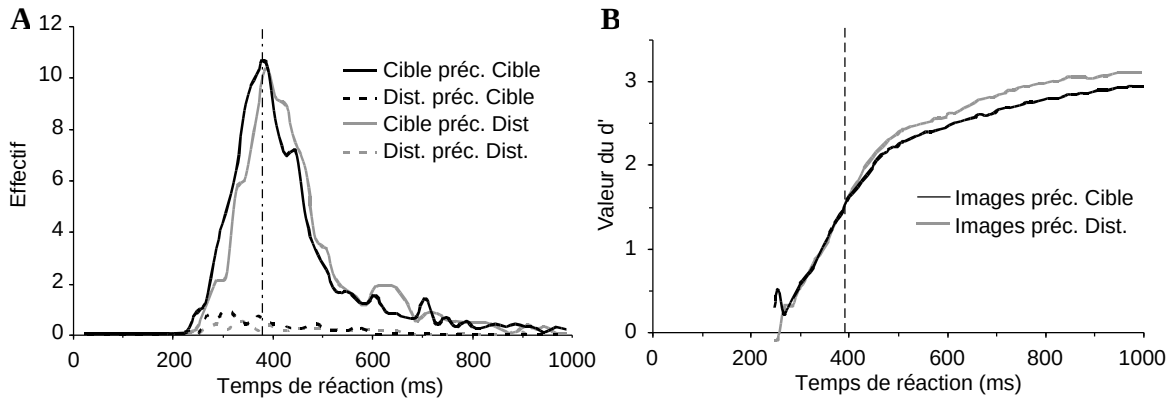


Figure 4.4 : A, répartition des TRs des sujets pour les images, cibles ou distracteurs (Dist.), précédées soit par une cible (préc.Cible), soit par un distracteur (préc. Dist.). Les deux distributions diffèrent significativement à partir de 380 ms (ligne en pointillés irréguliers). Le pas de temps est de 20 ms. B, analyse de la précision des sujets en fonction du temps pour les images précédées soit d'une cible, soit d'un distracteur. La différence ne semble intervenir que pour les TRs supérieurs à 390 ms (ligne en pointillés réguliers). Dans la partie initiale, pour les images précédées d'une cible, la performance plus élevée sur les cibles est compensée par une très faible performance sur les distracteurs (cf. texte).

soient confondues dans leur partie initiale indique que même si les sujets répondent plus rapidement, ils commettent aussi plus d'erreurs sur ces réponses, de sorte que leur performance globale n'est pas plus élevée et suit la loi d'échange précision/vitesse que nous reverrons plus loin[8].

En nous intéressant au délai qui précède la présentation d'une image donnée, nous sommes également contraints de nous intéresser au type de stimulus qui précédait cette image, ne serait-ce que pour déterminer si les deux facteurs, ISI et stimulus précédent, interagissent. Pour cette analyse, les séries présentées à ISIs fixes et à ISIs variables ont été regroupées[9]. On observe de nouveau un effet sur la performance régi par la loi d'échange précision/vitesse (figure 4.4). L'image précédée d'une cible est soit un distracteur soit une cible. Pour les distracteurs on observe une augmentation du nombre d'erreurs mais un gain en temps de réaction quand l'image est précédée d'une cible[10]. En revanche pour les cibles, on observe uniquement un gain en temps de réaction : les cibles précédées d'une autre cible sont catégorisées en moyenne 20 ms plus rapidement que les cibles précédées d'un distracteur[11]. La

[8] Il y aurait alors changement de β (complémentaire du d') qui représente en quelque sorte un seuil à atteindre pour prendre une décision.

[9] Dans les analyses qui suivent, j'ai vérifié que chaque tendance était présente dans les deux conditions d'ISIs fixes et variables.

[10] Précision : $\chi 2$ bilatéral, DDL=1, p=0,0013. TR : Mann Whitney U test U=3744, p<0,038.

[11] Précision : aucune tendance n'est observée. TR : médiane de 423 ms si la cible est précédée par un distracteur et de 402 ms si elle est précédée par une cible, Mann Whitney U test U=804654, p<0,0001. Cet effet, cependant, n'est significatif que pour la moitié des sujets et semble totalement absent chez certains sujets.

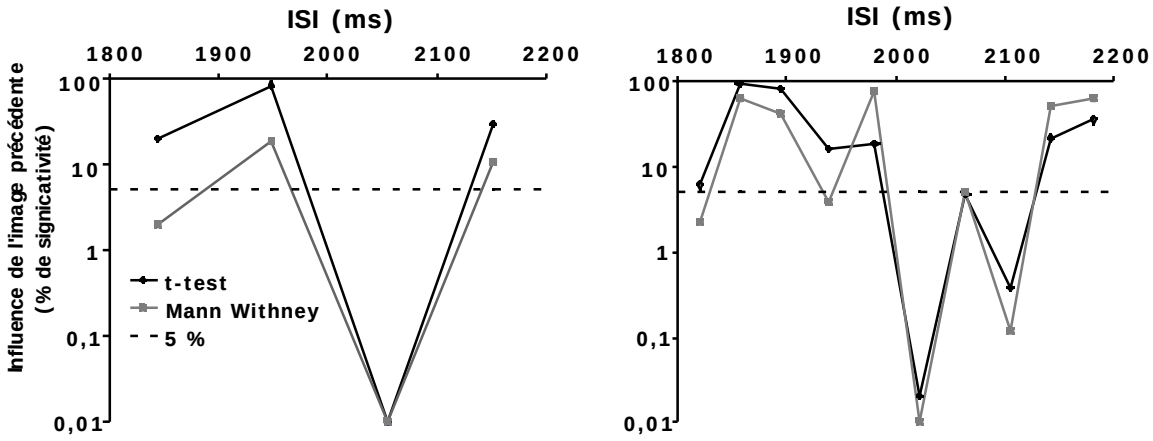


Figure 4.5 : effets de l'image précédente en fonction de l'intervalle entre deux images (division en 4 groupes à gauche et en 10 groupes à droite) estimés à l'aide de deux tests de significativité (t-test et test de Man Whitney). Les points représentent le résultat du test statistique qui détermine si les cibles précédées d'une autre cible sont catégorisées significativement plus rapidement que les cibles précédées d'un distracteur. Très clairement l'effet de l'image précédente est maximal juste après 2000 ms.

courbe de précision en fonction du temps montre qu'avant environ 400 ms, la performance des sujets est confondue : même si les sujets sont plus rapides quand l'image est précédée d'une cible, ils commettent également plus d'erreurs sur les réponses précoces et leur performance globale ne s'en trouve donc pas améliorée. De façon tardive, on constate que la précision sur les images précédées d'un distracteur est plus élevée (figure 4.4) : cela est dû au fait que pour les images précédées d'une cible, le nombre d'erreurs précoce sur les distracteurs était important.

Après avoir étudié isolément l'effet de l'image précédente, j'ai tenté de déterminer l'interaction qui peut exister entre ce facteur et l'ISI : l'amplitude de l'effet de l'image précédente peut dépendre du temps séparant la présentation de 2 images. Pour les séries présentées avec ISIs variables, les réponses des sujets ont été réparties en fonction de la durée de l'intervalle en considérant non pas deux classes (ISIs courts et longs) comme précédemment, mais 4 et même 10 classes d'ISIs. La figure 4.5 indique que l'effet de l'image précédente n'est pas réparti de façon homogène sur ces classes. L'effet de l'image précédente semble maximal pour les classes d'ISIs correspondant à un peu plus de 2000 ms. Cela signifie que pour les ISIs d'un peu plus de 2000 ms, pour les images précédées d'une cible, les réponses produites sont significativement plus rapides que lorsque les images sont précédées d'un distracteur. L'effet observé est important : pour des effectifs semblables, le test de significativité (t-test) montre que l'expérimentateur qui accepte cet effet a 94 % de chance de se tromper pour des ISIs inférieurs à 2000 ms mais seulement 0,02 % juste après 2000 ms (figure 4.5).

Cette valeur de 2000 ms correspond à l'ISI des séries où les images sont présentées à intervalle de temps fixe. Il est nécessaire de déterminer dans quelle mesure l'effet observé dépend directement de l'alternance des séries à ISIs variables et des séries à ISIs fixes. Pour ce faire, les images ont été scindées en 2 groupes, l'un correspondant à la première moitié des essais de chaque série variable et l'autre correspondant à la deuxième moitié des essais des mêmes séries. On s'attend à observer un effet plus fort sur les premières images, qui suivent une série fixe, que sur les autres images. Le résultat est décevant puisque, quel que soit le groupe, l'intensité de l'effet sur les temps de réaction est similaire[12]. S'il fallait dégager une tendance, il semble même que l'accélération due à la présence d'une cible avant l'image à catégoriser soit plus importante pour la deuxième moitié des essais de chaque série à ISIs variables. Cela ne met pas hors de cause les ISIs fixes dans le biais observé sur les séries à ISIs variables. Pour résoudre ce problème, il faudrait que les sujets catégorisent uniquement des séries à ISIs variables[13].

4.3 - Discussion

Cette expérience montre que s'il existe une attention temporelle, qui permettrait de focaliser l'attention dans le temps pour obtenir une meilleure performance, elle n'intervient pas ici car on n'observe aucune différence entre les séries à ISIs fixes et les séries à ISIs variables tant du point de vue des temps de réaction que de la précision. Cependant, pour les images présentées avec des ISIs variables, la valeur de l'ISI n'est pas neutre sur le processus de catégorisation : il semble que les images précédées d'un ISI supérieur à 2000 ms soient catégorisées plus rapidement que celles présentées après des ISIs de moins de 2000 ms.

Dans une certaine mesure ce résultat est dépendant de l'effet lié au statut - cible ou distracteur – de l'image précédente, effet plus important dans le cas de longs ISIs. Si l'image à catégoriser est précédée d'une cible, les TRs sont plus rapides, mais la précision est moindre[14]

---

[12] Man Whitney U test, U=61133, p=0,0013 pour le premier groupe; U=51667, p=0,0003 pour le second groupe.

[13] Les séries contrôles de l'expérience de détection/catégorisation, que l'on verra par la suite, sont de ce type. Un effet est visible juste vers 1950 ms pour 4, 10, 20 et 40 groupes d'ISI mais de façon moins marquée malgré un nombre de sujets et d'images plus important. Mon sentiment serait que cet effet de l'interaction de l'ISI et de l'image précédente est présent chez chaque sujet de façon différente, de sorte que ce que l'on recueille au niveau de la population ne signifie pas grand-chose (maximum d'effet à 1950 ms dans l'expérience de détection/catégorisation ou après 2000 dans celle-ci). Dans l'expérience que je viens de présenter, la présence de séries à ISIs fixes pourrait synchroniser en quelque sorte les sujets, renforçant l'effet de l'image précédente à 2000 ms d'ISI.

[14] Une ANOVA entre les deux processus par rapport aux temps de réaction montre que les deux effets ne sont pas indépendants (ANOVA, DDL=1, p=0,05).

et cet effet semble plus important quand l'intervalle entre les images est supérieur à 2000 ms. Nos résultats montrent que l'attention temporelle - c'est-à-dire l'effet de l'ISI sur la performance des sujets - dépend de l'effet de séquence, c'est-à-dire en quelque sorte de la préactivation de la catégorie animal.

4.3.1 - Amorçage du système visuel

On pourrait penser que la présence de séries à intervalles fixes de 2000 ms amorce le système visuel et optimise le traitement pour des images présentées à 2000 ms. Des études en électrophysiologie vont dans ce sens en montrant une augmentation de l'activité de certains neurones au moment où un stimulus est attendu (Apicella et al, 1992; Reutiman et al, 2000). Cependant**,** aucune différence n'a été observée dans les performances de nos sujets pendant les séries avec intervalles fixes et celles où l'intervalle de temps variait.

L'effet le plus important concerne l'influence de l'image précédente qui varie fortement au cours de l'ISI. Ainsi, le temps séparant deux images intervient peu dans les performances de catégorisation mais intervient fortement dans l'interaction avec l'image précédente. Cela est compatible avec l'activation graduelle de neurones de IT pendant l'ISI (Reutiman et al, 2000). Si ces neurones interagissent avec des neurones sélectifs aux animaux ou sont directement sélectifs eux-mêmes, leur activité serait biaisée par le stimulus précédent. Bien que cette interprétation puisse sembler pour le moins rapide, elle fournit une explication plausible de l'interaction entre l'ISI et l'image précédente que nous observons.

Il convient toutefois de poser des limites à cette interprétation. Une augmentation linéaire de l'activité des neurones, qui pourrait rendre compte tout à la fois de l'attente du sujet et de la nature de l'image catégorisée précédemment, ne permet pas d'expliquer nos résultats. En effet, l'influence de l'image précédente devrait être monotone en fonction de l'ISI, c'est-à-dire qu'elle devrait être faible pour des images séparées par de courts ISIs et plus importante pour des images séparées par de longs ISIs. Or la relation observée est clairement non monotone, et il serait nécessaire pour l'expliquer qu'un pic d'interaction entre l'ISI et l'effet de séquence soit présent à 2000 ms.

Sans entrer en contradiction avec l'interprétation en termes d'activité neuronale que je viens de présenter, il est également possible d'interpréter ces résultats en termes probabilistes. On peut imaginer que le système visuel se base sur la probabilité d'apparition de l'image pour amorcer le traitement de l'image suivante. Ce type d'interprétation s'accorde bien avec le modèle probabiliste de type de celui de Mozer (Mozer et al, 2000). Dans ce modèle, des

unités codant les objets obéissent à une loi d'échange précision/vitesse et la dynamique de ces unités peut être biaisée de différentes façons par des facteurs externes. Nous reviendrons plus en détails sur ce modèle par la suite.

Après avoir tenté d'interpréter nos résultats, nous allons voir maintenant dans quelle mesure nos résultats s'accordent avec ceux présents dans la littérature.

4.3.2 - Attention focalisée dans le temps

De façon surprenante, l'attention temporelle et l'influence de la nature des stimuli dans les séquences utilisées sont relativement peu étudiées. La raison principale est, à mon avis, que l'attention visuelle spatiale est un domaine très riche et encore relativement peu connu sur lequel se focalisent toutes les recherches. Certaines études font cependant intervenir l'attention focalisée dans le temps dans le cadre de la dynamique de l'attention spatiale. Un indice visuel précédant l'apparition d'une cible de quelques dizaines de millisecondes à la même position que celle-ci permet une catégorisation plus rapide (Eriksen et Collins, 1969; Egeth et Yantis, 1997). Ce études montrent que l'effet n'est visible que pour des intervalles de temps entre l'indice et la cible supérieurs à 50 ms. Cependant, la position de la cible est variable et, l'indice indiquant la position où va apparaître la cible, le gain en temps pourrait donc être attribué à l'attention spatiale plutôt qu'à l'attention temporelle.

Les autres études qui prennent en compte l'attention focalisée dans le temps sont en général réalisées dans le système auditif. L'attention focalisée dans le temps pourrait être au cortex auditif ce que l'attention spatiale est au cortex visuel. De façon similaire à l'expérience d'attention spatiale focalisée dans le système visuel que je viens d'indiquer (Eriksen et Collins, 1969), un son puissant avant le stimulus augmente les performances dans une tâche complexe de catégorisation (Hackley et al, 1999). Les TRs sont plus courts quand le son est présent mais les erreurs semblent légèrement plus importantes, ce qui indiquerait un échange précision/vitesse du type de celui que nous avons montré. Bien que les auteurs interprètent ce résultat en termes d'augmentation de la vigilance, étant donné que le stimulus sonore se termine 83 ms avant le début de la présentation du stimulus, il est également possible qu'un effet d'attention temporelle automatique soit responsable de l'augmentation de la performance. Ce résultat serait en contradiction avec le nôtre car nous avons montré que même si le sujet a les moyens de déterminer de façon exacte le moment où le stimulus doit apparaître, ses performances ne sont pas augmentées. Il est cependant possible que le fait que l'indice auditif soit présenté quelques ms avant le stimulus joue un rôle non négligeable. Des

expériences sont en cours dans notre équipe pour tester le rôle de tels indices, de manière visuelle cette fois.

D'autres travaux sont plus en adéquation avec nos résultats. Dans une tâche purement auditive, de façon totalement contre-intuitive, Madey et Gilovich (1993) ont montré que, quand l'attention n'est pas focalisée dans le temps, les sujets ont une meilleure performance. Malgré les différences entre notre tâche et la leur, nos résultats indiquent également que le fait qu'une cible apparaisse à intervalles très réguliers ne permet pas d'amorcer le système visuel pour obtenir une catégorisation plus rapide ou une meilleure performance.

4.3.3 - Interaction entre ISI et effet de séquence

D'une façon générale, les études s'accordent pour dire qu'il existe différents types d'attentes. Pour l'équipe de Sommer (Matt et al, 1992; Leuthold et Sommer, 1993; Sommer et al, 1999), dans une tâche auditive ou visuelle où le sujet doit distinguer deux stimuli, le rôle de l'image précédente est différent de celui que nous avons montré. La présence d'une cible augmente la rapidité des réponses des sujets, mais aussi leur précision. Cette augmentation de précision est probablement due au fait que dans ce type de tâche où le stimulus est bimodal, le sujet doit simplement détecter si le stimulus qui lui est présenté est différent ou identique à celui qu'il vient de traiter. La tâche est donc triviale lorsque deux stimuli identiques se succèdent. Ces travaux indiquent également que l'ISI ne semble pas interférer avec les effets induits par la probabilité d'apparition du stimulus. L'amplitude et la latence de l'onde P300 à la surface du scalp dépendrait de la probabilité d'apparition du stimulus et notamment des stimuli précédents. Cependant cette onde n'est que très peu affectée quand on modifie l'ISI séparant deux stimuli (Matt et al, 1992; Polich et Bondurant, 1997). Une expérience comportementale renforce ce résultat (Leuthold et Sommer, 1993). Dans cette expérience, les sujets doivent catégoriser des séries de sons de différentes fréquences avec différents intervalles de temps entre les stimuli (ISI). À la différence de nos résultats, les effets séquentiels - influence d'un stimulus sur le suivant - ne dépendent pas de l'ISI.

Cependant les ISIs utilisés dans cette dernière expérience sont relativement longs (plus de 5s) et des expériences récentes comparables à celle que je viens de présenter sont plus en adéquation avec nos résultats. Pour Sommer et al (1999), dans une tâche de catégorisation visuelle bimodale, l'effet de la cible précédente sur la vitesse de catégorisation dépend de l'ISI. À courts ISIs - 40 ms après la réponse du sujet - on observe une diminution du temps de réaction des réponses uniquement sur les cibles. À longs ISIs (500 ms après la réponse du

sujet), comme dans notre expérience, la présence de cibles successives induit une diminution du temps de réaction à la fois sur les cibles et les distracteurs. Cependant, dans ces études, les essais à courts ISIs ne sont pas mélangés avec les essais à longs ISIs et ce n'est pas l'attention focalisée en fonction du temps que les auteurs étudient mais plutôt la dynamique temporelle de l'effet du stimulus précédent.

La principale différence entre ces expériences et les nôtres reste que l'effet de séquence, quand deux cibles se suivent, induit une augmentation des TRs mais également une augmentation de la précision. Nos résultats impliquent plutôt un échange précision/vitesse, la performance restant constante puisque la diminution des temps de réaction quand une cible est précédée d'une autre cible, s'accompagne d'une augmentation des erreurs sur les distracteurs. Je reviendrai plus en en détails sur cet aspect dans la discussion générale concluant la partie expérimentale.

Dans l'expérience que je viens de présenter, les images sont totalement nouvelles pour les sujets. Étant donné le fort effet de séquence que nous avons montré, il est possible que les performances de catégorisation puissent être fortement affectées lorsque les sujets sont confrontés à des images familières, vues et donc traitées de nombreuses fois. Le prochain chapitre analyse comment les performances de catégorisation rapide peuvent être affectées en fonction de la nouveauté ou de la familiarité de l'image à traiter. Nous verrons que les effets sont bien différents de ce à quoi l'on pourrait s'attendre.

5

Familiarité du stimulus

Les expériences précédentes ont posé la question des caractéristiques des images pouvant expliquer la rapidité du traitement des images dans une tâche de catégorisation mais également celle de l'influence de la tâche, en s'attachant à rechercher les effets attribuables à la variation des ISIs et ceux liés à l'organisation séquentielle des stimuli. Un facteur qui n'a pas encore été exploré est celui de l'expérience qu'à le sujet des stimuli qu'il catégorise. On peut ainsi se poser la question de savoir si une parfaite connaissance de certaines images ne pourrait pas permettre aux sujets d'accélérer encore leur vitesse de catégorisation. Cette caractéristique influence probablement les traitements au sein du système visuel. En effet, nous ne réagissons pas de la même façon aux stimuli que nous connaissons - personnes, objets personnels... - et à ceux que nous voyons pour la première fois.

Ce type de propriété des images a été abondamment étudié dans la littérature sous la forme des processus d'amorçage. Il est couramment admis que les stimuli, qui sont vus pour la première fois, sont catégorisés plus lentement et de façon moins précise que les stimuli vus ne serait-ce qu'une seule fois auparavant (Tulving et Schacter, 1990; Ochsner et al, 1994). De même, les corrélats neurophysiologiques existent : les neurones dans le cortex périrhinal répondent sélectivement aux stimuli nouveaux (Bogatz et al, 2000).

Tous ces éléments nous conduisent à penser que la catégorisation d'images nouvelles au sein d'images familières sera grandement affectée tant au niveau du TR que de la précision.

Nous verrons très clairement que cette affirmation est à moduler. Les résultats que je présente ici ont fait l'objet d'un article (Fabre-Thorpe et al, 2000; annexe 3) et j'ai tenté ici une approche complémentaire, notamment en exploitant plus avant l'analyse des processus précoces[1].

Dans les expériences précédentes, les seules données comportementales des sujets sont parfois ambiguës et difficiles à interpréter. Nous allons voir ici que l'enregistrement conjoint des potentiels évoqués permet d'aller plus loin dans l'étude de la catégorisation rapide : les potentiels confirment et renforcent les observations comportementales.

5.1 - Matériel et méthodes

Quatorze sujets humains (7 hommes et 7 femmes) participent à l'expérience qui se déroule sur 3 semaines, c'est-à-dire sur 15 séances. Durant les 13 premières séances, qui correspondent à la phase de familiarisation avec certaines images, la tâche des sujets est d'effectuer une catégorisation animal/non animal sur 200 images (100 cibles et 100 distracteurs), identiques d'un jour sur l'autre mais mélangées de façon aléatoire. Par la suite, ces images seront désignées comme étant "les images familières"[2]. À partir de la 13ème séance, les sujets doivent catégoriser des séries de 100 images où 50 images familières sont mélangées aléatoirement avec 50 images nouvelles[3]. En tout, au cours des deux dernières séances, 1200 images nouvelles sont mélangées aux 200 images familières du sujet, chaque image familière étant présentée 6 fois[4] (figure 5.1). De cette façon, le nombre d'images familières et nouvelles est identique au cours de ces séances[5]. Les potentiels évoqués sont

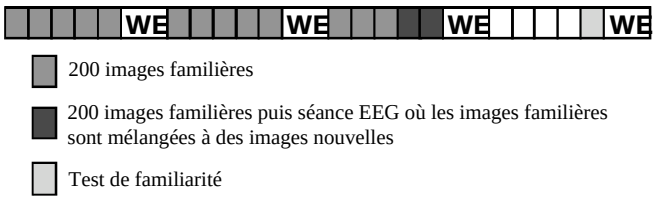


Figure 5.1 : protocole de l'expérience sur 3 semaines; 13 jours de catégorisation d'images familières suivis de 2 jours de test où les EEGs sont enregistrés. Les images familières sont catégorisées seules puis mélangées avec des images nouvelles. À la fin de la quatrième semaine, on teste si le sujet reconnaît ces images familières parmi d'autres images nouvelles.

[1] Pour les résultats de certains tests de significativité, se reporter à l'article.
[2] 7 groupes de 200 images familières correspondent à 7 couples homme-femme de sujets.
[3] 50 % de cibles et 50 % de distracteurs à la fois pour les images familières et nouvelles.
[4] Au total, 1400 images sont utilisées et tous les sujets voient ces images, soit en tant qu'images familières, soit en tant qu'images nouvelles.
[5] Le nombre total de séries au cours des deux dernière séances est donc de 24 séries.

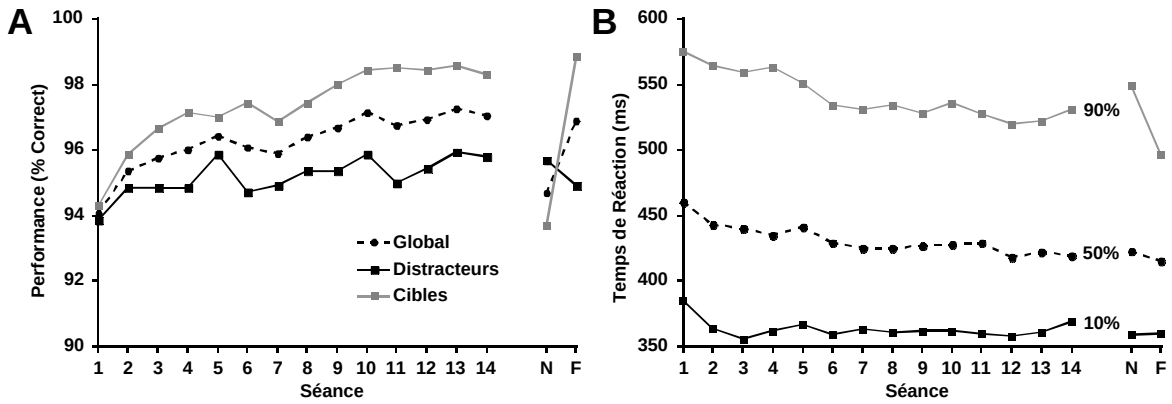


Figure 5.2 : A, évolution de la performance de l'ensemble des 14 sujets au cours des 15 séances de présentation d'images identiques. L'augmentation globale de la performance est principalement due à une meilleure catégorisation des cibles. À l'extrême droite de la courbe, la performance globale pour les images nouvelles (N) et les images familières (F) (pendant les deux séances finales de test) est représentée. B, évolution des temps de réaction de l'ensemble des 14 sujets au cours des 14 séances de présentation d'images identiques. Les temps de réaction les plus rapides n'évoluent pratiquement pas après la première séance. À l'extrême droite de la courbe, les temps de réaction pour les images nouvelles et familières (pendant les deux séances finales de test) sont représentés. L'effet est absent pour les TRs les plus rapides et maximal pour les TRs les plus tardifs.

également enregistrés durant ces séances afin de comparer potentiels évoqués par les images nouvelles et par les images familières.

Afin de contrôler si les sujets sont effectivement capables de reconnaître les images familières sur lesquelles il ont été entraînés, une semaine après ces deux séances de test, les images familières (3x200) sont à nouveau mélangées à 600 images nouvelles supplémentaires, la tâche du sujet est alors de répondre lorsqu'il détecte une image comme familière. On montre que la performance des sujets est très bonne et qu'ils reconnaissent donc les images sur lesquelles ils ont été entraînés (cf. annexe 3).

5.2 - Résultats

Au cours de la période d'apprentissage qui s'étale sur 3 semaines, la précision globale des sujets sur les images qu'il voit chaque jour augmente d'environ 3 %, cette amélioration est principalement due à une meilleure précision sur les cibles, la précision sur les distracteurs évoluant très peu (figure 5.2).

Pendant les 2 séances de test où images nouvelles et familières sont présentées aléatoirement, la précision est plus élevée sur les images familières (96.9 % de réponses correctes contre 94.7 % pour les images familières)[6]. Cette différence de performance est

[6] $\chi 2$ bilatéral, DDL=1, $p<0,0001$. L'analyse des comportements individuels des sujets montre que la différence est significative à $p<0,05$ pour 7 sujets.

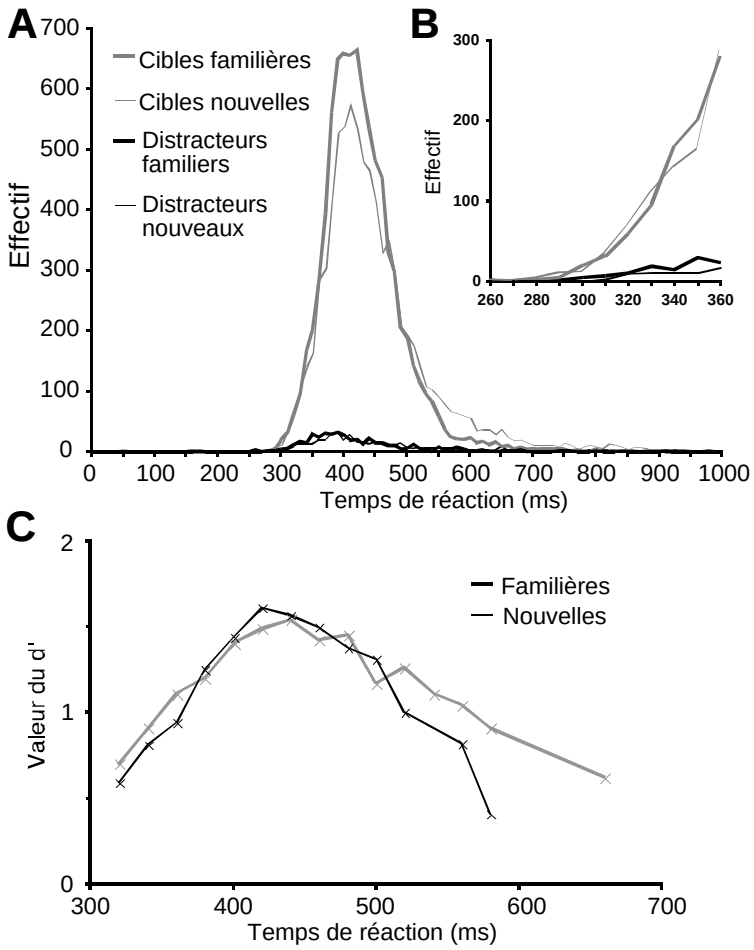


Figure 5.3 : A, distribution des temps de réaction sur les images familières et les images nouvelles pour l'ensemble des 14 sujets pendant les deux séances finales de test. Le pas de temps est de 10 ms. B, grossissement d'une partie de la distribution présentée en A qui montre que la distribution des temps de réaction les plus rapides est identique pour les images familières et les images nouvelles. C, représentation de l'évolution de la précision des sujets au cours du temps. Le calcul prend en compte à la fois la précision sur les distracteurs et la précision sur les cibles par bloc de 20ms. Il s'agit ici d'un d4 instantané, à la différence de la plupart des courbes de performance présentées, chaque point est indépendant du précédent. Pour les temps de réaction les plus rapides, la performance n'est pas plus élevée pour les images familières. S'il existe un effet, il serait plutôt en faveur des images nouvelles (cf. texte).

encore une fois uniquement due aux cibles, la précision sur les distracteurs ayant même tendance à être meilleure pour les images nouvelles (figure 5.2). En ce qui concerne la vitesse des traitements sous-jacents, les temps de réaction des sujets évoluent vers des valeurs plus brèves au cours des premières séances de familiarisation avec les images. Cependant cette évolution affecte peu (ou pas!) les temps de réaction les plus précoces (figure 5.2), l'effet majeur de la familiarisation des sujets avec les images est enregistré sur les TRs les plus longs. L'analyse du décours temporel de la précision (figure 5.3C) indique que, pour les temps de réaction les plus rapides, la performance n'est pas plus élevée pour les images familières que pour les images nouvelles. En fait, il semble même que l'effet soit inversé avant 360ms. Très clairement ce résultat va à l'encontre d'une catégorisation plus précise et plus précoce pour les images familières

De façon succincte, la familiarisation avec certaines images permettrait à un sujet d'y détecter plus facilement les cibles difficiles. La précision est augmentée (certaines cibles sont détectées alors qu'elles ne l'étaient pas à l'origine), et seuls les TRs les plus longs sont raccourcis (certaines cibles dont la détection nécessitait des traitements temporellement plus

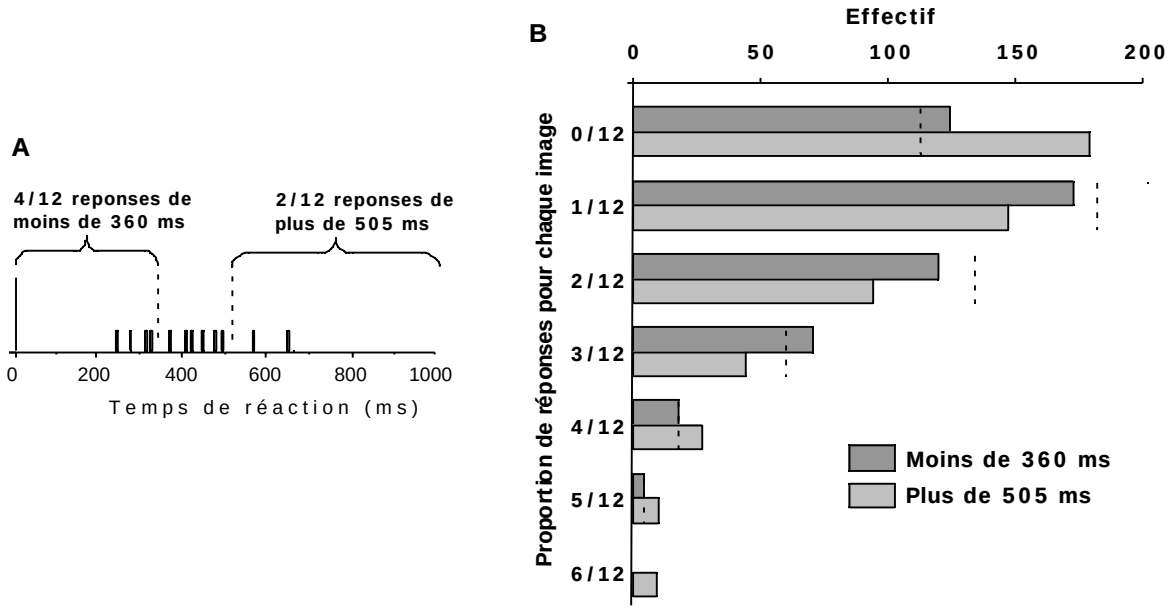


Figure 5.4 : répartition des images nouvelles dans la classe des TRs précoces (<360 ms) et dans celle des TRs tardifs (>505 ms) par rapport aux réponses des 12 sujets (sur un total de 14) qui les ont catégorisées en tant qu'images nouvelles. Seules les images sur lesquelles aucun sujet n'a commis d'erreur sont prises en compte. A, exemple de répartition arbitraire des deux classes pour une image. Il convient de justifier le choix de ces deux classes : en dessous de 360 ms il n'est pas possible de différencier les réponses sur les images nouvelles des réponses sur les images familières (cf. texte). Par analogie, la classe des réponses les plus tardives a été construite pour avoir un effectif de réponses identique à celui de la classe des TR précoces (724). B, répartition des réponses des sujets pour chacune des deux classes, TR < 360 ms (gris foncé) et TR > 505 ms (gris clair). Par exemple, les barres d'histogramme 0/12 comptabilisent le nombre d'images pour lesquelles aucune des 12 réponses n'était produite avec un TR < à 360 ms (gris foncé) ou avec un TR > 505 ms (gris clair). Ce graphique semble être bien complexe pour un résultat mineur, à savoir que la répartition des images dans les deux classes, précoce et tardive, n'est pas symétrique. En fait, il est également possible de comparer ces distributions avec une distribution aléatoire des images. Le trait en pointillés représente le nombre d'images obtenu à partir d'une distribution aléatoire (distribution binomiale, 12 tirages avec une probabilité de 0,18 correspondant aux 724 images catégorisées en moins de 360 ms ou en plus de 505 ms). La distribution des réponses inférieures à 360 ms diffère très peu d'une distribution aléatoire et donc le fait qu'une image soit catégorisée rapidement semble indépendant des caractéristiques de l'image. Pour les réponses lentes en revanche, la distribution des réponses diffère significativement ($\chi 2$ bilatéral, DDL=6, $p<0,05$) de la distribution aléatoire et donc les images ne sont pas toutes équivalentes du point de vue de la catégorisation tardive.

"gourmands" sont détectées plus rapidement). Cette hypothèse est compatible avec une analyse détaillée des distributions des TRs enregistrés sur les images familières et nouvelles. Les TRs moyens (et médians) sont significativement plus courts sur les images familières que sur les images nouvelles d'environ 20 ms[7], mais une fois encore, les TRs les plus précoces ne sont pas affectés comme le montre (i) l'analyse des 10 % de réponses les plus rapides (figure 5.2B), (ii) l'histogramme des temps de réaction (figure 5.3A et 5.3B) et (iii) l'analyse de la

[7] TR moyen de 424 ms pour les images familières et de 444 ms pour les images nouvelles. Test de Man Whitney bilatéral, U=294300695, p<0,0001. L'analyse du comportement individuel des sujets montre que la tendance est présente pour tous les sujets et atteint un niveau significatif à p<0,05 pour 11 d'entre eux.

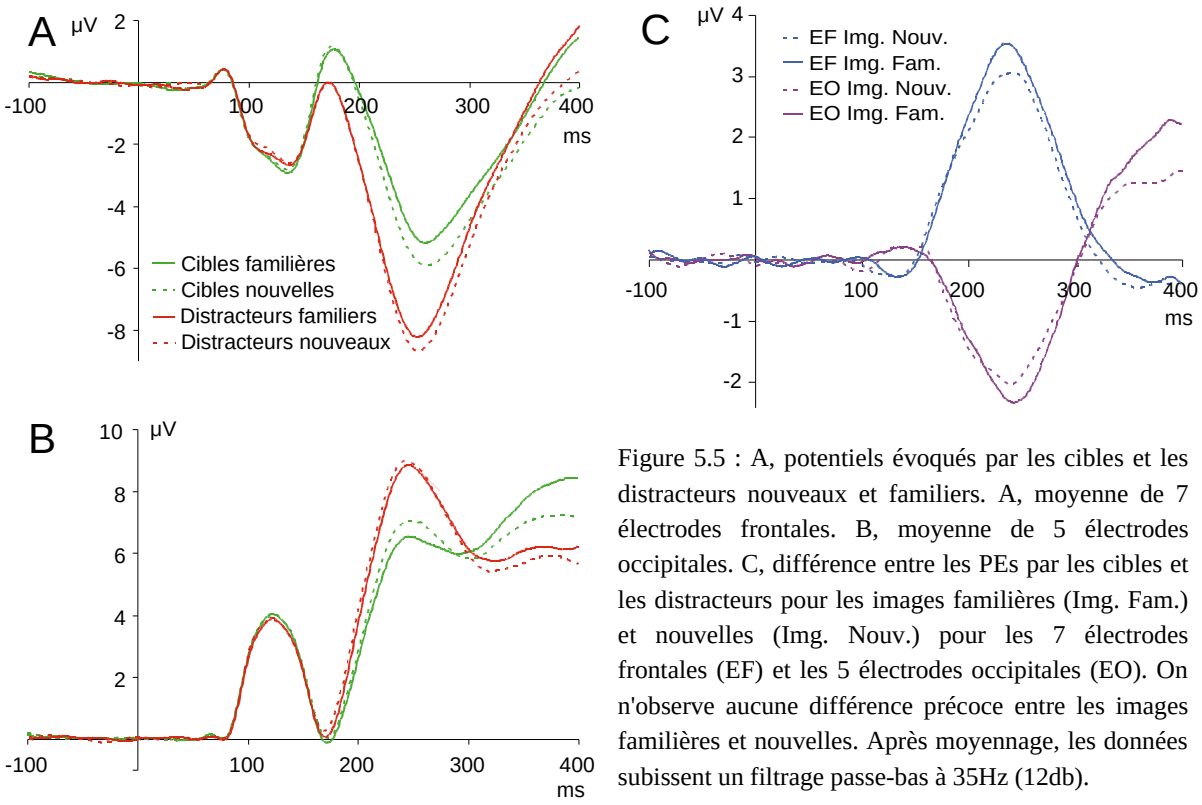


Figure 5.5 : A, potentiels évoqués par les cibles et les distracteurs nouveaux et familiers. A, moyenne de 7 électrodes frontales. B, moyenne de 5 électrodes occipitales. C, différence entre les PEs par les cibles et les distracteurs pour les images familières (Img. Fam.) et nouvelles (Img. Nouv.) pour les 7 électrodes frontales (EF) et les 5 électrodes occipitales (EO). On n'observe aucune différence précoce entre les images familières et nouvelles. Après moyennage, les données subissent un filtrage passe-bas à 35Hz (12db).

précision au cours du temps (figure 5.3C). L'effet de la familiarité est maximal pour les réponses les plus tardives, plus nombreuses quand l'image est nouvelle comme le montre l'histogramme des temps de réaction (figure 5.3). Ce biais serait causé par une minorité d'images atypiques. Cet effet apparaît clairement en analysant la répartition des réponses des sujets sur les différentes images (figure 5.4). Les réponses précoces se distribuent uniformément sur les images : cela signifie qu'il n'existe pas a priori d'images induisant de façon spécifique des TRs très courts. Par contre les réponses tardives ne semblent pas uniformément distribuées, de sorte qu'il doit exister des images spécifiquement plus difficiles (la figure 2 de l'article présenté en annexe 3 présente quelques illustrations d'images de ce type[8]). Ces deux résultats ne sont pas contradictoires dans le sens où la minorité d'images catégorisées lentement par un grand nombre de sujets influence peu la majorité d'images neutres du point de vue des réponses précoces.

Pour en terminer avec l'analyse des résultats, il convient d'analyser les potentiels évoqués (PE) enregistrés au cours de la tâche (cf. annexe 1). Les PEs ont été uniquement enregistrés au cours des deux dernières séances pour tenter de déterminer les différences qui pouvaient

[8] Bien qu'elles soient associées à de longs TRs, des sujets ont commis des erreurs sur ces images et elles n'étaient donc pas inclues dans l'analyse présentées à la figure 4.

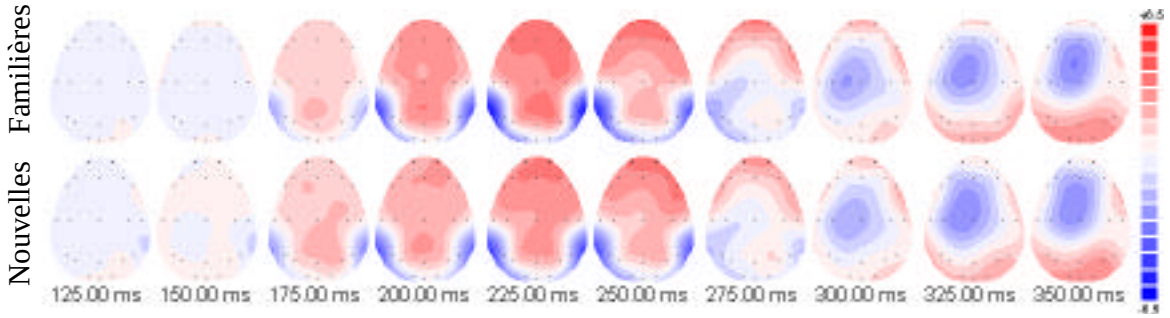


Figure 5.6 : évolution de l'activité cérébrale différentielle (différence entre les PE par les cibles et les distracteurs) pour l'ensemble des électrodes entre 125 ms et 230 ms. La ligne du haut représente cette évolution pour les images familières et la ligne du bas représente l'évolution pour les images nouvelles. L'échelle indique les voltages en µV et les points noirs à la surface du scalp représentent la position des électrodes. Les activations sont très semblables pour les deux conditions, notamment au niveau temporal entre 175ms et 250ms.

exister dans le traitement visuel des images nouvelles et des images familières. Ici, l'analyse à la fois les PEs sur les cibles et les PEs sur les distracteurs ne présentent aucune différence entre les images familières et les images nouvelles pendant les 200 premières millisecondes qui suivent la présentation du stimulus (figure 5.5). À 200 ms, les PEs deviennent plus amples pour les images nouvelles, à la fois pour les cibles et les distracteurs. En revanche, lorsque l'on considère la différence entre les PE moyennés séparément pour les cibles et les distracteurs ce sont les images familières qui - à partir de 200 ms post-stimulus - induisent la différence la plus ample (figure 5.5 C).

Cette différence à 200 ms est postérieure à celle qui correspond au contenu de l'image proprement dit - c'est-à-dire la présence d'un animal dans l'image - qui se situe aux environs de 150 ms. Comme dans l'étude originale, la différence entre les PEs moyennés sur les cibles et ceux moyennés sur les distracteurs montre que les deux traces, confondues jusqu'à 150 ms après la présentation du stimulus, divergent ensuite de façon abrupte (Thorpe et al. 1996). À 150 ms, des processus sélectifs à la présence d'une cible sont donc mis en jeu. Des processus encore plus précoces entre 100 et 130 ms semblent même différencier les essais sur les cibles de ceux sur les distracteurs[9]. Nous verrons au chapitre suivant quelle peut être leur signification. Les PEs nous permettent également d'établir une localisation grossière des processus mis en jeu. Quand on s'intéresse à l'ensemble des électrodes, on observe une forte négativité au niveau du cortex occipito-temporal qui pourrait correspondre à la détection de la cible par le système visuel (figure 5.6). Bien que l'interprétation en termes d'aires cérébrales

[9] La petite déflexion dans la direction opposée est en fait significative dès 113 ms à $p<0,02$ (t-test bilatéral de 113 à 136 ms, DDL=13, $t>2,65$).

activées à partir des PEs soit sujette à caution[10], cette zone est bien connue en IRMf (Kanwisher et al, 1997; Chao et al, 1999; Gautier et al, 2000) pour être associée à certains objets comme des visages ou des animaux. Il est donc possible que ces aires neuronales s'activent sélectivement quand un animal est présent. On note également que la différence entre les deux conditions, images familières et images nouvelles est pratiquement indiscernable pour la fenêtre temporelle que nous avons choisie (figure 5.6).

5.3 - Discussion

Ces résultats montrent que les réponses les plus précoces dans la tâche de catégorisation ne peuvent pas être accélérées même si les stimuli présentés sont bien connus des sujets. L'analyse de la précision en fonction du temps, des temps de réaction et des potentiels évoqués montrent à quel point le traitement visuel précoce est similaire pour des images très familières et pour des images totalement nouvelles. Ce résultat est surprenant si l'on tient compte de la très abondante littérature sur les effets d'amorçage. De nombreuses études montrent en effet que la catégorisation répétée d'une cible induit une meilleure performance et une diminution des temps de réaction (pour une revue cf. Tulving et Schacter, 1990; Ochsner et al, 1994). En fait, on observe tout de même un effet d'amorçage dans notre étude (une augmentation de 2.2 % en performance et une diminution de 20 ms en moyenne des temps de réaction). Cependant cet effet n'est présent que pour les temps de réaction les plus longs associés aux images difficiles.

5.3.1 - Dynamique de l'amorçage

Nous avons montré que l'effet de l'amorçage n'est visible que tardivement. Qu'en est-il du processus d'amorçage en lui-même ? Il semble que pour qu'il y ait amorçage, même tardivement, le système visuel doive traiter les stimuli pendant un certain temps. Subramaniam et al (2000) ont montré que lors de la présentation de séquences très rapides (76 ms par image), l'effet d'amorçage n'était pas observé. Les sujets devaient détecter une image au sein d’une séquence. Les distracteurs dans la séquence étaient ensuite utilisés comme cible et le fait que l'image soit présentée comme distracteur dans une séquence précédente n'avait

[10] Pour tenter de localiser les aires cérébrales activées, il est préférable d'utiliser un logiciel de simulation de sources électriques neuronales comme BESA (annexe 1). Dans cette expérience on obtient une activation du cortex occipito-temporal sélective à la présence (ou à l'absence) d'un animal. Je préfère présenter ici l'activation à la surface du scalp car elle est, à mon avis, tout aussi claire (figure 5.6).

aucune incidence sur la détection de cette image en tant que cible. Pour des délais inter-images plus longs (plus de 126 ms), les auteurs observent effectivement un effet d'amorçage, la précision des sujets sur les cibles qui avaient été amorcées étant meilleure. Les auteurs s'inspirent de résultats en électrophysiologie pour interpréter leurs résultats. La majorité de l'information sur une photographie de visage est contenue dans les premières 50 ms de décharge de neurones dans le cortex inféro-temporal (Tovee et Rolls, 1995; Sugase et al, 1999) mais ces neurones continuent à décharger pendant au moins 350 ms. Subramaniam et al (2000) pensent que cette activation tardive, qui n'est pas présente dans leur expérience puisque le traitement visuel de l'image suivante interrompt le traitement de l'image précédente, est nécessaire pour observer un effet d'amorçage[11].

Dans le même cadre, à partir de notre expérience, il apparaît naturel de supposer que l'effet de l'amorçage - et non plus le processus proprement dit - a lieu également au cours de cette activation tardive des neurones de IT et que c'est pour cette raison, qu'aucun effet d'amorçage n'est observé pour les temps de réaction les plus rapides. L'activation tardive des neurones de IT semble donc nécessaire à la fois dans la mise en place de l'amorçage et à son expression[12].

Malgré la littérature très abondante sur l'effet d'amorçage, aucune étude à ma connaissance ne traite de la dynamique de ce processus. La raison est, à mon avis, avant tout culturelle. Les psychologues qui étudient ce genre de phénomène s'intéressent encore peu à la dynamique sous-jacente des réseaux neuronaux et donc à la dynamique elle-même des processus. Bien que souvent elles ne posent pas le problème en termes d'amorçage ou de familiarité, nous allons aborder quelques études où les potentiels évoqués ont été enregistrés afin de comparer leurs résultats aux nôtres.

5.3.2 - Amorçage et EEG

Concernant la dynamique de catégorisation des images familières, Friedman et al (1990) a présenté à 28 sujets des dessins dont la moitié avait été préalablement présentée avant la séance, les sujets devant déterminer s'ils avaient déjà vu le dessin. Les auteurs enregistrent

[11] Cette interruption n'est pourtant pas immédiate. Dans des conditions de présentations sérielles rapides, Keyser et al (2000) ont récemment montré que les réponses des neurones de IT conservaient leurs sélectivités au stimulus pendant 60 ms alors même que chaque image masque l'image précédente. Cet effet est de plus encore visible pour des durées de présentation de seulement 14 ms par image.

[12] On pourrait penser que dans l'expérience de Subramaniam et al (2000), seule l'expression de l'amorçage est déficiente. Cependant, dès que le sujet détecte l'image cible, il semble traiter cette image de façon extensive (Potter, 1999), de sorte que l'effet d'amorçage s'il existe devrait être visible.

une différence entre les PEs des deux types de dessins, nouveau ou familier, qui apparaît 250 ms après la présentation de l'image. Cet effet est également observé en EEG pour les visages, 240 ms après la présentation de l'image (Hertz et al, 1994). Ces résultats, compatibles avec les nôtres, indiquent donc que l'effet de familiarité se manifeste bien après le processus de détection des animaux-cibles dont on a montré qu'il est beaucoup plus rapide, aux alentours de 150 ms (Thorpe et al, 1996; Fabre-Thorpe et al, 2000).

Une étude en EEG renforce cette hypothèse (Jemel et al, 1999). Les auteurs présentent aux sujets un premier visage de personnage connu sans les yeux pendant 200 ms puis lui présentent soit une paire d'yeux - avec ou sans le reste du visage - correspondant ou non au premier visage présenté. La tâche des sujets est simplement d'indiquer si la paire d'yeux correspond au visage présenté initialement sans les yeux. Cette expérience mélange amorçage à court terme (avec le visage initial) et dynamique d'appariement d'objet (le visage et les yeux). Les auteurs enregistrent également l'EEG et de façon très surprenante aucun effet de l'image précédente - l'amorce - n'est visible avant 300 ms sur les PEs. Même quand la tâche fait intervenir explicitement l'amorce, les sujets devant répondre en fonction de la première image présentée, l'effet ne devient visible que très tardivement. L'influence de la tâche ne permettant pas d'accélérer l'effet de l'amorce, il semble que les traitement visuels sous-jacents soient majoritairement automatiques.

Il est possible que l'amorçage intervienne à un niveau relativement élevé dans le système visuel, peut-être dans V4 (Bar et Biederman, 1999), et fasse nécessairement intervenir un accès lexical ou sémantique : un effet d'amorçage mot-image est visible en PE 250 ms après la présentation de l'image, c'est-à-dire à la même latence où nous avons observée l'effet de la familiarité de l'image (Pratarelli, 1994). Etant donné la lenteur de ces processus, il est probable qu'ils mettent en jeu des processus itératifs complexes. Nous reviendrons sur ce sujet plus en détails dans la discussion générale.

5.3.3 - Adaptation

Nous avons montré que, par rapport aux images familières, les PEs à 250 ms étaient plus amples à la fois sur les cibles nouvelles et sur les distracteurs nouveaux. On peut rapprocher ces résultats d'autres études : après 6 à 8 présentations d'un stimulus, la réponse des neurones dans le cortex inféro-temporal du singe diminue de plus de moitié par rapport à la première présentation (Li et al, 1993). C'est également un phénomène bien connu d'adaptation principalement utilisé en IRMf (Buckner et al., 1998).

Certaines études en IRMf, en cours de réalisation, utilisent même ce paradigme expérimental pour déterminer si une amorce (ayant des similarités avec l'objet-cible) a été traitée par le système visuel (professeur Nancy Kanwisher au MIT, communication personnelle). Dans une étude réalisée en PET, les auteurs ont présenté des séries d'images familières et des séries d'images nouvelles (Tulving et al, 1996). Ils montrent que les zones activées sélectivement par les images nouvelles sont le cortex hippocampique et les lobes temporaux. Cette dernière activation est compatible avec une réponse plus faible des neurones dans ces régions comme cela est observé en IRMf (Buckner et al., 1998).

Notre étude montre que cet effet ne dépend pas de l'image présentée car l'atténuation du signal sur les images familières est à la fois présente pour les cibles et les distracteurs. Cependant quand on considère la différence entre les cibles et les distracteurs, cette différence est plus importante pour les stimuli familiers que pour les stimuli nouveaux. Une explication simple de ce processus serait que, les sujets focalisant leur attention sur les cibles, la diminution d'activité pour les cibles familières serait plus importante que pour les distracteurs (figure 5.5). Cependant cette explication est incomplète car elle ne prend pas en compte les réponses comportementales des sujets. Comme nous le montrerons dans le chapitre suivant, la réponse différentielle est corrélée avec la performance des sujets et la hauteur du pic, plus important pour les images familières, pourrait être liée à une meilleure performance pour des temps de réaction intermédiaires.

L'effet de familiarité que nous avons montré n'intervient qu'à des latences relativement tardives. Afin d'aller plus loin dans cette étude, nous avons voulu savoir si cela était également vrai dans le cas où l'image-cible était extrêmement familière, c'est-à-dire qu'elle constituait l'unique cible au sein de la série. Dans ce cas précis la tâche n'est plus une tâche de catégorisation mais une tâche de détection d'une cible unique contenant un animal. L'influence des informations descendantes sur le système visuel est maximalisée puisque toutes les caractéristiques de l'image sont connues. Nous verrons que la comparaison de ces deux tâches est riche d'enseignements.

6

Détection et catégorisation

L'expérience précédente indique dans quelle mesure la familiarité avec les stimuli peut influencer les processus de catégorisation et nous avons montré que l'apprentissage intense de certaines images ne permet pas de diminuer le temps minimal de catégorisation.

Cela implique que les informations descendantes, par exemple en provenance du cortex frontal, permettant d'anticiper la catégorisation d'une cible ne jouent aucun rôle visible pour les temps de réaction les plus rapides. J'ai déjà mentionné pourquoi les nombreuses expériences sur l'amorçage et l'attention pouvaient laisser supposer le contraire (Egeth et Yantis, 1997). La question est donc maintenant de déterminer dans quelle mesure un amorçage maximal du système peut influencer le temps de traitement des scènes naturelles.

Dans une tâche plus simple de catégorisation de rond et de carré, Aubertin et al (1999) dans notre équipe ont montré qu'à la fois les temps de réaction et les PE précoces étaient comparables à ceux de la tâche animal[1]. Les stimuli étaient des rectangles gris dans lesquels 10 carrés (ou 10 ronds) apparaissaient avec des niveaux de gris variables à des positions aléatoires. A l'aide du protocole go-nogo habituel, les sujets devaient répondre lorsque le stimulus contenait des carrés. Même dans cette tâche pourtant très simple où le sujet doit

[1] Ce résultat reste à reproduire. Si les réponses les plus précoces sont enregistrées à la même latence, la divergence des TRs sur les cibles (animal ou forme géométrique) est très rapide et la qualité de l'enregistrement des potentiels évoqués était insuffisante pour véritablement conclure.

uniquement détecter une forme géométrique, il n'est pas capable d'accélérer ses temps de réaction. Pourtant, les informations permettant d'effectuer la tâche sont théoriquement disponibles dès V1 si l'on considère l'activité des cellules *end-stopped*. Dans toutes ces expériences, l'image-cible varie et l'on est donc en droit de se demander si ce n'est pas cette variation permanente des caractéristiques des cibles qui interdit l'amorçage du système visuel pour une catégorisation plus rapide.

Dans l'expérience que je présente ici, nous avons donc décidé de supprimer toute variabilité dans les cibles, ce qui signifie qu'au sein d'une même série, les distracteurs sont variés mais la cible à catégoriser est une image unique. Comme on s'y attend nous avons choisi comme cible unique des images d'animaux et nous avons tenté de comparer les performances de catégorisation dans la tâche "animal" et celles obtenues sur une image unique d'animal. De cette façon, les cibles dans les deux tâches possédant les mêmes caractéristiques, il nous est plus facile de comparer à la fois les performances des sujets et les PEs.

Les résultats que nous avons obtenus sont très intéressants : comme nous le verrons, la tâche de catégorisation animal est effectuée avec un retard d'environ 30 ms à la fois au niveau des TRs et des PEs. Cependant, la localisation des sources neuronales impliquées dans ces tâches semble très voisine et seule la dynamique d'activation des différentes structures impliquées semble varier d'une tâche à l'autre[2].

6.1 - Matériel et méthodes

Quatorze sujets, 7 hommes et 7 femmes participent à l'expérience (âge moyen 27 ans allant de 21 à 56 ans). Les sujets doivent réaliser deux tâches, l'une de catégorisation d'animaux et l'autre de détection d'image unique. La procédure et les détails expérimentaux de présentation des images et d'enregistrement des potentiels évoqués sont strictement identiques à ceux décrits dans l'expérience précédente. Des images sont flashées pendant 20 ms et les sujets doivent relâcher un bouton lorsqu'il s'agit d'une cible et garder le bouton appuyé dans le cas contraire.

Les séances se composent de séries de 100 images et les sujets doivent alternativement effectuer la tâche contrôle de catégorisation d'animaux et la tâche de détection de cible unique. La tâche contrôle de catégorisation d'animaux est identique à celle de l'expérience

[2] J'ai réalisé ce travail avec Guillaume Rousselet (1999) qui a effectué la majeure partie du recueil des données.

précédente : 100 images nouvelles dont la moitié contiennent des animaux sont présentées aux sujets. Pour la tâche de détection d'une cible, 50 distracteurs sont mélangés aléatoirement à 50 copies d'une même image-cible sur laquelle les sujets doivent répondre. Chaque série-test est immédiatement précédée d'une phase d'entraînement pendant laquelle le sujet "apprend" l'image-cible utilisée dans la série. La cible est présentée sous formes de séries de flashs identiques à ceux utilisés pendant la tâche alternant avec des présentations statiques pendant lesquelles le sujet peut explorer les détails de l'image[3].

Dans la tâche de détection, 3 types d'image-cibles uniques ont été utilisés : le premier est constitué d'animaux dits "faciles à catégoriser", le second d'animaux dits "difficiles à catégoriser" et le dernier contient des images sans animaux. Les classes d'animaux faciles et difficiles sont créées à partir des résultats de l'expérience précédente[4].

Au total, un sujet est testé sur 10 séries de la tâche de catégorisation et 15 séries de la tâche de détection réparties sur deux jours[5]. Conjointement aux réponses des sujets, les PEs sont enregistrés. Le détail de la procédure d'enregistrement des temps de réaction et des potentiels évoqués est présenté en annexe (annexe 1). Les 15 séries de la tâche de détection se divisent en 5 séries pour chaque sous-type d'images (animaux faciles, difficiles et non animaux). Cette répartition permet d'avoir autant de cibles d'images d'animaux dans la tâche de catégorisation que dans la tâche de détection. Nous étions en effet particulièrement intéressés par la comparaison des résultats des deux tâches avec des images d'animaux comme cible. Dans l'analyse des résultats, je me bornerai donc à comparer les deux tâches pour les images-cibles contenant des animaux. Le cas des images cibles unique ne contenant pas d'animaux ne sera considéré qu'à la fin, à titre de contrôle.

[3] 3 séries de 5 flashes à 1 seconde d'intervalle, entrecoupées de deux présentations statiques de 1 s.

[4] Les images d'animaux faciles à catégoriser ont été correctement catégorisées par tous les sujets et avec les TRs moyens les plus faibles. Les images d'animaux difficiles à catégoriser sont celles qui ont induit le plus d'erreurs. Parmi les images n'ayant provoqué qu'une seule erreur, ont été sélectionnées celles pour lesquelles les TRs moyens des réponses correctes étaient les plus élevés.

[5] Sur les deux jours consécutifs d'enregistrement (12 séries le premier jour et 11 séries le second), les séries se succèdent afin d'alterner les tâches et les types d'images uniques.

6.2 - Résultats

6.2.1 - Précision des sujets

La précision de l'ensemble des sujets est extrêmement bonne dans les deux tâches. Elle est pourtant supérieure d'environ 5 % dans la tâche de détection[6]. La différence entre la tâche de détection et celle de catégorisation est significative[7] et l'analyse des performances individuelles montre que cet effet est présent pour chacun des sujets.

Comme on l'a vu dans les chapitres précédents, la performance des sujets dépend à la fois de leur précision sur les cibles et de leur précision sur les distracteurs. La contribution des cibles et des distracteurs est inégale et semble inversée dans les deux tâches. Dans la tâche de détection, les erreurs sont principalement causées par les distracteurs (97,5 % de distracteurs corrects contre 99,7 % de cibles correctes). Par opposition, dans la tâche de catégorisation, les erreurs sont majoritairement causées par les cibles (93,9 % de distracteurs corrects contre 92,4 % de cibles). Ces différences sont hautement significatives dans les deux cas[8].

Concernant les deux types d'images, animaux faciles et animaux difficiles, dans la tâche de détection, la performance est significativement plus faible pour les images d'animaux difficiles[9]. Cette différence est due à de plus nombreuses erreurs sur les distracteurs. On observe également une diminution de la précision sur les images-cibles mais cela seulement quand on considère les TRs les plus rapides[10].

L'analyse des erreurs des sujets dans la tâche de détection est également d'un grand intérêt. Le sujet peut fonder sa détection d'une cible unique sur des indices visuels très bas-niveau et, au sein d'une série, on est en droit de s'interroger sur le degré de similarité entre image-cible et distracteurs ayant induit une erreur. Nous avons utilisé un algorithme permettant de déterminer ce degré de similarité (cf. figure 6.1 pour le détail de l' algorithme) et nous avons effectivement observé que les images sur lesquelles les sujets commettent des erreurs ne sont

---

[6] 93,1 % de réponses correctes dans la tâche de catégorisation; 98,4 % dans la détection d'animaux difficiles et 98,9 % pour les animaux faciles.

[7] Test chi2 bilatéral ddl=1, $p<0.0001$.

[8] Test chi2 bilatéral ddl=1, $p<0.0001$.

[9] La précision sur les distracteurs atteint 97.9% pour la détection d'animaux faciles et 97.1% pour la détection d'animaux difficiles; test chi2 bilatéral ddl=1, $p=0.022$.

[10] Quand on considère les cibles catégorisées en moins de 370 ms, la différence entre la détection d'images contenant des animaux faciles par rapport aux images contenant des animaux difficiles est hautement significative (test chi2 bilatéral ddl=1, $p<0.0001$). Cependant la performance sur les images difficiles augmente pour les temps de réaction plus longs; la performance globale est donc pratiquement identique pour les images contenant des animaux soit faciles soit difficiles.

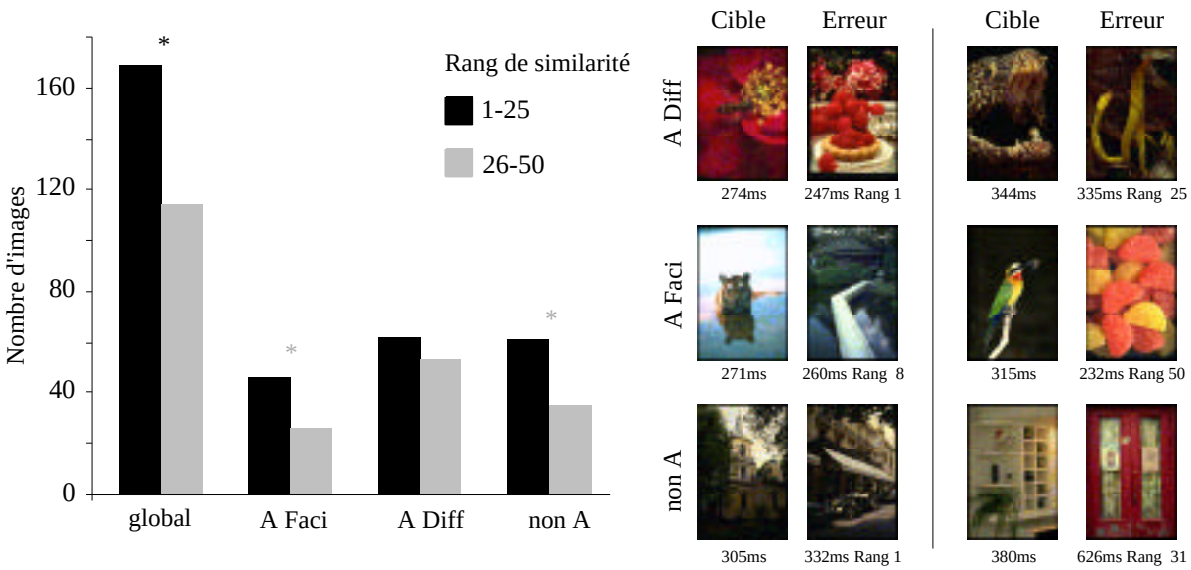


Figure 6.1 : Analyse des erreurs dans la tâche de détection. Pour chaque image unique, les erreurs sont triées en fonction de leur similarité avec l'image cible. Chaque image-cible est présentée parmi 50 distracteurs ; à chaque erreur est donc attribuée un rang de 1 à 50, 1 étant l'image la plus similaire avec l'image cible et 50 la plus dissimilaire. Le logiciel iMatch a été utilisé pour classer les distracteurs en fonction de leur ressemblance avec la cible. À gauche, on constate que pour l'ensemble des erreurs, il y a statistiquement plus d'images similaires avec la cible (rangs 1-25 en noir) que d'images dissimilaires (rangs 26-50 en gris). Cette tendance est présente pour les trois types de cibles uniques, animaux faciles (A Faci), animaux difficiles (A Diff) et non animaux (non A). L'étoile noire indique un niveau de significativité inférieur à 5 % et l'étoile grise inférieur à 10 %. À droite, sont illustrées diverses erreurs pour chaque type d'image. Sous chaque cible est indiqué le temps de réaction moyen des réponses correctes et sous chaque distracteur le temps de réaction de l'erreur et le rang de similarité du distracteur avec l'image. À partir du classement effectué par l'algorithme, la colonne de gauche illustre des erreurs effectuées vers des distracteurs considérés comme semblables à la cible par l'algorithme. La colonne de droite présente des distracteurs associés à des rangs élevés et donc potentiellement dissimilaires à la cible. Ces images ont cependant été choisies pour souligner les imperfections de l'algorithme, l'œil humain pouvant percevoir des similarités de forme (en haut et en bas) et de couleur (au milieu). La similarité des erreurs avec la cible est donc potentiellement plus élevée que celle que renvoie l'algorithme choisi.

pas indépendantes de l'image-cible associée (figure 6.1). Cela indique donc que, dans une certaine mesure, les sujets utilisent des propriétés de bas niveau de la cible pour la détecter.

6.2.2 - Temps de réaction

Les temps de réaction moyens sont plus rapides d'environ 60 ms dans la tâche de détection que dans la tâche de catégorisation d'animaux. Cette différence est hautement significative pour chacun des sujets[11].

En ce qui concerne la dynamique de la performance des sujets dans la tâche de détection, la différence entre le comportement sur les cibles et les distracteurs est significative dès 220

[11] TR moyen de 344 ms dans la tâche de détection - médian 337 ms; TR moyen de 417 ms dans la tâche de catégorisation - médian 400 ms. Test de Mann Whitney U bilatéral : U=12197972; p<0.0001. Pour chaque sujet, la différence est également significative à p<0.0001.

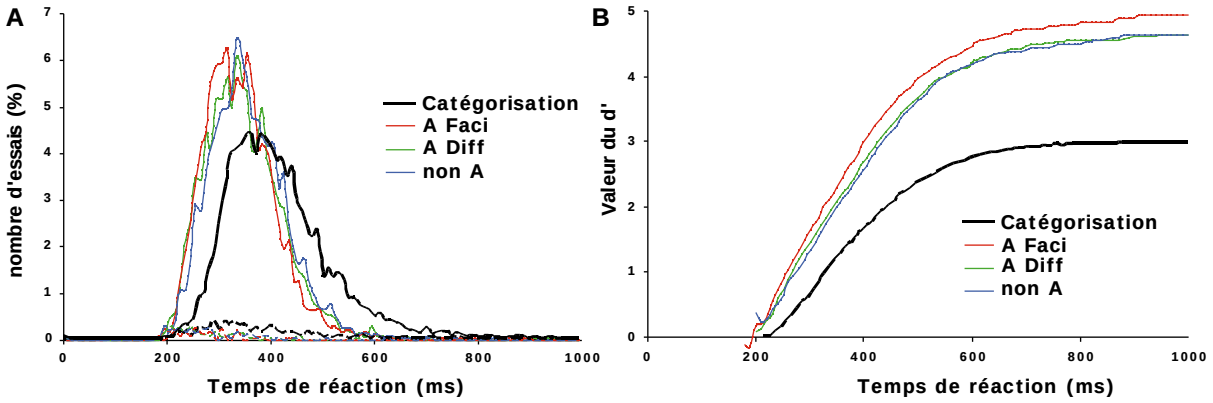


Figure 6.2 : A, comparaison des distributions des TRs de l'ensemble des sujets pour la tâche de catégorisation et la tâche de détection de cibles uniques. Les performances sur les trois types de d'image-cibles uniques, contenant des animaux faciles (A Faci), difficiles (A Diff) et ne contenant pas d'animaux (non A) sont représentées. Pour un pas de temps donné (10ms) les réponses go sont exprimées en pourcentage du total d'essais de même type (par exemple les cibles dans la tâche de catégorisation). On note que la tâche de catégorisation est retardée d'environ 30-40 ms par rapport à la tâche de détection et que la différence entre les types de cibles uniques est très faible. B, d' cumulé illustrant la performance dans les deux tâches. Ces courbes confirment le retard de la tâche de catégorisation sur la tâche de détection. Le plateau terminal de chaque courbe dépend directement de la performance des sujets; la performance est donc bien meilleure dans la tâche de détection que dans celle de catégorisation. De même la performance semble plus élevée dans le cas des image-cibles uniques contenant des animaux faciles par rapport aux autres types de cibles uniques.

ms[12] , dans la tâche de catégorisation cette différence n'apparaît qu'à 260 ms[13], soit avec un retard de 40 ms. Si les réponses sur les cibles sont en moyenne de 60 ms plus rapides dans la tâche de détection, ce gain n'est plus que de 40 ms quand on s'intéresse uniquement aux réponses les plus précoces. Pour les temps de réaction les plus rapides, une analyse de l'évolution de la performance au cours du temps, qui prend en compte les réponses sur les cibles et sur les distracteurs, indique également un gain en termes de durée de traitement d'environ 40 ms pour la tâche de détection par rapport à la tâche de catégorisation (figure 6.2). Cette différence, comme on le verra par la suite, semble directement corrélée avec la différence calculée entre les potentiels évoqués moyennés séparément sur les cibles et sur les distracteurs (figure 6.4).

La comparaison entre les deux conditions d'images uniques dans la tâche de détection, faciles et difficiles, fait également apparaître des différences significatives, les images d'animaux faciles qui était catégorisées plus précisément que celles d'animaux difficiles sont

[12] Test chi2 bilatéral réalisé pour chaque intervalle de 10 ms, ddl=1; p<0.0001

[13] Cf. note précédente, test chi2 bilatéral ddl=1; p=0.0007

également catégorisées plus rapidement[14]. Cette tendance de plus est significative pour la majorité des sujets[15].

6.2.3 - Analyse des potentiels évoqués

Les potentiels évoqués nous permettent une meilleure analyse des processus visuels qui interviennent dans les deux tâches, de détection et de catégorisation. Quelle que soit la tâche, les potentiels évoqués sur les distracteurs sont pratiquement identiques : ils ne diffèrent que tardivement, après 300 ms (figure 6.3). Il est possible d'interpréter cette différence comme une plus forte inhibition des zones motrices pour les distracteurs de la tâche "animal"[16] ou à une attention plus soutenue sur l'image pour tenter d'y détecter un animal. Sur les PEs moyennés sur les cibles, les tracés divergent d'abord précocement vers 100-120 ms (figure 6.3). Pour cette onde précoce, la négativité est plus marquée pour la tâche de catégorisation et celle de détection d'animaux faciles. Un processus vient inverser cette négativité ; il se manifeste plus tardivement dans la tâche de catégorisation que dans la tâche de détection quel que soit le type d'image-cibles. Ce pic positif atteint une amplitude qui apparaît liée à la difficulté de la tâche, elle est minimum pour la tâche de catégorisation et maximum pour la tâche de détection des animaux faciles.

L'activité différentielle mise en évidence en soustrayant, dans chacune des tâches, les PE moyennés sur les distracteurs des PE moyennés sur les cibles fait apparaître une différence précoce en fonction de la tâche : catégorisation et détection et en fonction du type d'image-cibles : animaux faciles ou difficiles. Cette différence est significative et de plus grande amplitude dans la tâche de catégorisation[17], de moyenne amplitude dans la tâche de détection d'animaux faciles[18] et de très faible amplitude dans la tâche de détection d'animaux difficiles[19]. De plus, la différence entre les deux types d'images dans la tâche de détection atteint un niveau significatif à des latences similaires[20]. Cette première partie est donc probablement sélective à la présence d'un animal dans l'image. Cette hypothèse est confirmée

---

[14] Temps de réaction moyen de 341ms - médian 334 ms - pour les animaux faciles et de 348 ms - médian 340 ms - pour les animaux difficiles; test Mann Whitney U bilatéral U=5773772, p<0.0001.

[15] Elle est significative à 5 % pour 9 d'entre eux.

[16] La localisation des sources dans BESA indique que cette différence semble être due à des processus frontaux ou pariétaux/frontaux, c'est-à-dire proches du cortex pré-moteur.

[17] Test t bilatéral apparié ddl=13, p<0.02; occipitales : 98 ms; frontales : 120 ms; c. f. annexe 1 pour la méthode de calcul.

[18] Test t bilatéral apparié ddl=13, p<0.05; occipitales : 100ms; frontales : 112 ms.

[19] Ns.

[20] Test t bilatéral apparié ddl=13, p<0.05; occipitales : 120 m l; frontales : 127 ms.

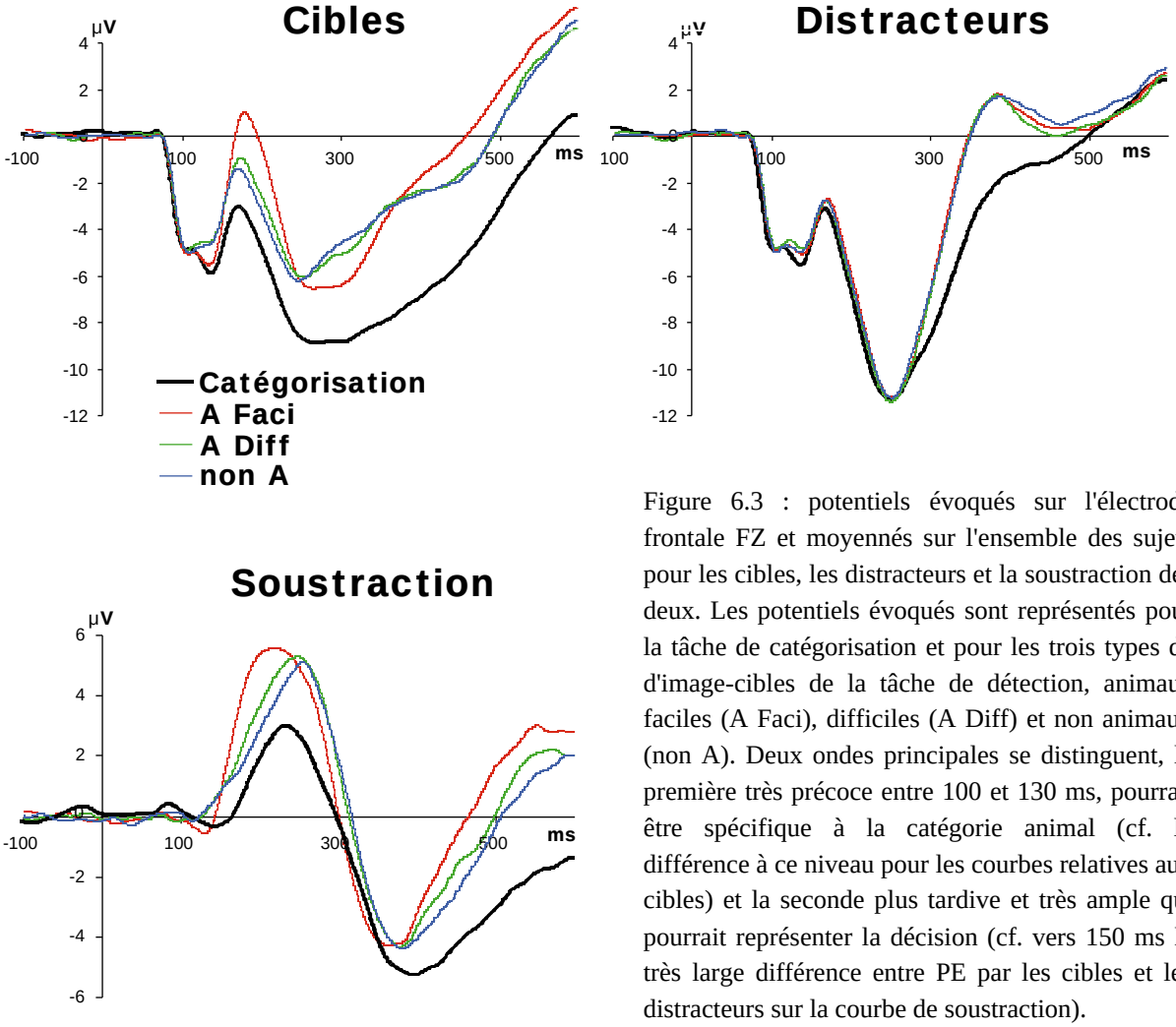


Figure 6.3 : potentiels évoqués sur l'électrode frontale FZ et moyennés sur l'ensemble des sujets pour les cibles, les distracteurs et la soustraction des deux. Les potentiels évoqués sont représentés pour la tâche de catégorisation et pour les trois types de d'image-cibles de la tâche de détection, animaux faciles (A Faci), difficiles (A Diff) et non animaux (non A). Deux ondes principales se distinguent, la première très précoce entre 100 et 130 ms, pourrait être spécifique à la catégorie animal (cf. la différence à ce niveau pour les courbes relatives aux cibles) et la seconde plus tardive et très ample qui pourrait représenter la décision (cf. vers 150 ms la très large différence entre PE par les cibles et les distracteurs sur la courbe de soustraction).

par la tâche de détection contrôle dans laquelle les cibles sont des images ne contenant pas d'animaux où ce pic n'est pas présent. D'autres travaux dans notre équipe rapportent des PEs sélectifs aux animaux à ces latences (VanRullen et Thorpe, 2000b).

Une différence plus tardive atteignant son amplitude maximale vers 250 ms, apparaît également sur les PEs différentiels (figure 6.3). Cette onde différentielle est plus ample et plus précoce dans la tâche de détection que dans la tâche de catégorisation. Sa latence est d'environ 140 ms dans la tâche de détection[21], alors qu'il faut attendre 170 ms pour la tâche de catégorisation[22]. Au niveau de cette onde différentielle la tâche de détection précède de 30 ms celle de catégorisation. Cette différence est significative légèrement plus tôt pour les images d'animaux difficiles que pour celles d'animaux faciles. Il est probable cependant que ce retard

[21] Test t bilatéral apparié ddl=13, p<0.02; occipitales : 135ms; frontales : 148 ms.
[22] Test t bilatéral apparié ddl=13, p<0.02; occipitales : 169 ms; frontales : 179 ms.

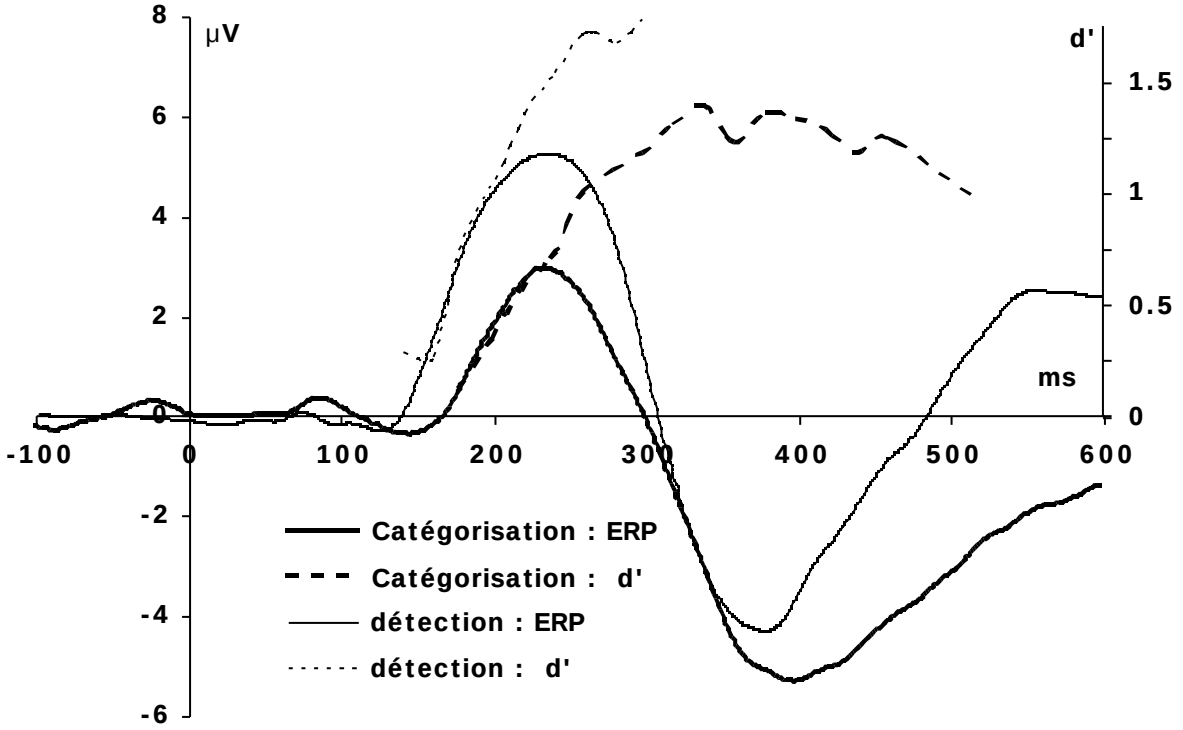


Figure 6.4 : Activités différentielles entre les PEs enregistrés sur l'électrode FZ et moyennés séparément sur les essais cibles et les essais distracteurs de l'ensemble des sujets pour la tâche de catégorisation (lignes épaisses) et la tâche de détection sur les images d'animaux (lignes fines). En pointillés sont représentées les courbes de performance instantanée des sujets (d'). Ces courbes ont été décalées vers des latences plus courtes de 60 ms pour illustrer la façon dont elles se superposent aux PEs. La performance est indiquée par des pointillés fins (détection) ou des pointillés épais (catégorisation). Dans chaque tâche, les potentiels évoqués vers 150 ms apparaissent fortement corrélés avec la performance.

soit causé par la première déflection de signe opposé qui, comme on l'a vu, est plus ample pour les images d'animaux faciles que pour celles d'animaux difficiles[23].

Il est toujours intéressant de pouvoir lier activité cérébrale et performances comportementales. La figure 6.4 illustre la superposition qui a pu être réalisée entre la courbe de d' instantanée – représentant la performance instantanée des sujets dans une tâche donnée et pour un pas de temps précis – et l'activité cérébrale différentielle obtenue dans chacune des tâches. Dans le cas de la tâche de détection, les PEs et la courbe de performance sont environ 40 ms plus précoces que dans la tâche de catégorisation[24]. Les deux courbes de d' se superposent aux PE différentiels par un simple décalage de ces courbes vers des latences plus courtes de 60 ms. Il y a donc de bonnes raisons pour corréler cette onde différentielle à un processus décisionnel.

[23] Test t bilatéral apparié ddl=13, $p<0.02$; occipitales : 152 ms; frontales : 151 ms pour les animaux dits faciles et test t bilatéral apparié ddl=13, $p<0.02$; occipitales : 134 ms; frontales : 145 ms pour les animaux dits difficiles.

[24] Il est également intéressant de noter que la pente des pics à 250 ms des PEs différentiels dans la tâche de catégorisation et de détection sont corrélés avec la pente de la courbe de performance.

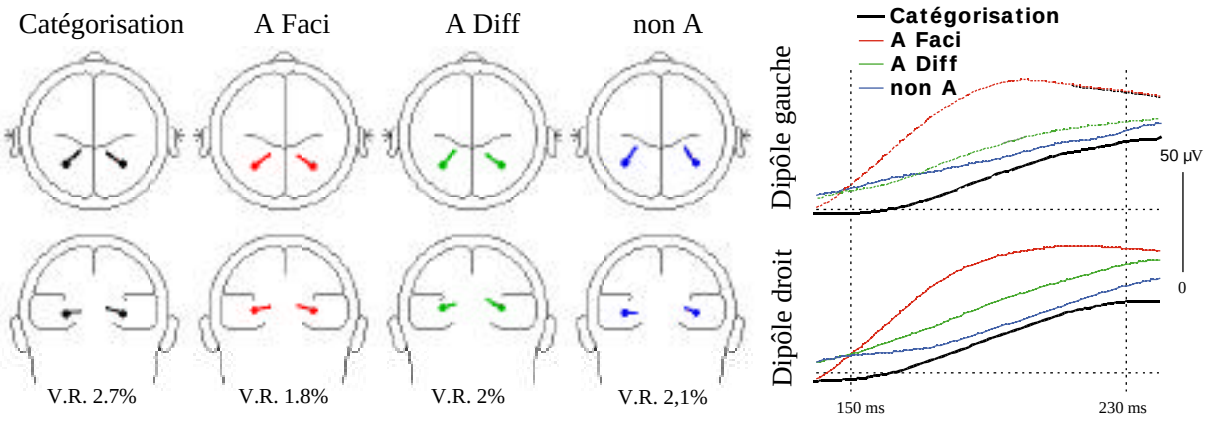


Figure 6.5 : analyse de sources effectuée sur les activités différentielles dans les deux tâches, de catégorisation et de détection. Pour la tâche de détection, les trois types de cibles sont considérées séparément. Deux dipôles sont contraints en symétrie et l'on optimise leurs positions et leurs amplitudes pour rendre compte au mieux du signal obtenu sur le scalp entre 150 et 230 ms (cf. l'annexe 1 pour le détail de la localisation des sources). Dans les trois cas, la position des dipôles est très proche et la variabilité observée au niveau du scalp s'explique par une dynamique temporelle différente des dipôles. Dans le cas d'images cibles ne contenant pas d'animaux, l'orientation du dipôle gauche semble cependant se distinguer. On notera également que malgré la contrainte en symétrie et la large fourchette temporelle, les variances résiduelles sont très faibles. La structure cérébrale à l'origine des sources dans les deux tâches est donc probablement identique.

Il est intéressant d'analyser la latence à partir de laquelle les courbes différentielles obtenues dans chacune des tâches diffèrent entre elles. Cette divergence atteint un niveau significatif dès 140 ms[25]. Les images utilisées pour les tâches de détection et de catégorisation étant similaires, cette différence serait donc liée à la tâche. Elle pourrait refléter les différences qui existent en termes de mécanismes cérébraux sous-jacents spécifiques à chacune des deux tâches dans le traitement des scènes naturelles.

Dans le but de déterminer la (les) structure(s) cérébrale(s) impliquée(s) dans la génération de cette activité différentielle, j'ai également tenté, pour chacune des deux tâches, de localiser les sources électriques dans le cerveau permettant d'expliquer les potentiels observés à la surface du crâne (cf. annexe 1 pour le détail du modèle). Je me suis restreint à la localisation des sources rendant compte des différences entre les potentiels évoqués sur les cibles et les distracteurs (figure 6.5). Dans tous les cas, les modèles obtenus indiquent que ces sources se localisent au niveau des cortex occipito-temporaux. Cela signifie que l'amplitude du signal dans ces zones diffère entre les cibles et les distracteurs. La localisation des sources est très similaire dans la tâche de catégorisation et dans la tâche de détection. Cela laisse supposer que les aires corticales mises en jeu sont identiques dans les deux cas et que les processus pourraient alors être similaires. Dans ce cas, seule la dynamique d'activation de ces zones

[25] Test t bilatéral apparié ddl=13, $p<0.02$; occipitales : 141 ms; frontales : 158 ms.

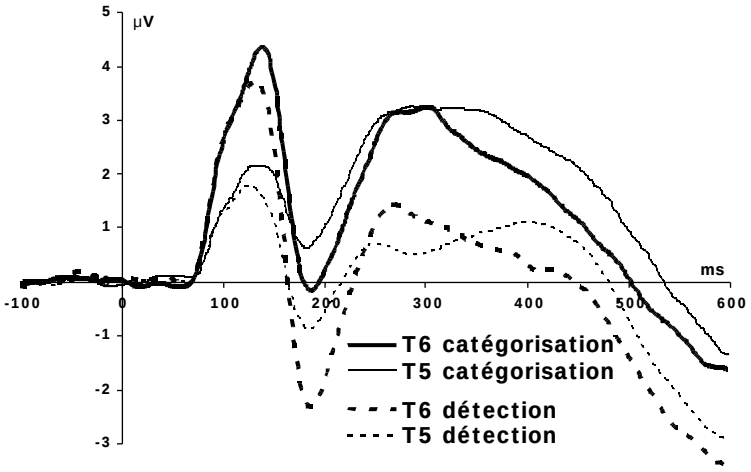


Figure 6.6 : illustration de la latéralisation dans les deux tâches de détection et de catégorisation. Les potentiels représentent l'activité sur les deux électrodes temporales (T5 à gauche et T6 à droite). Les potentiels diffèrent entre les deux hémisphères et cela de façon très similaire dans chacune des tâches. Cela indique à nouveau que les localisations de traitement sont probablement les mêmes dans les deux cas.

cérébrales diffèrerait d'une tâche à l'autre. La comparaison des signaux en provenance des deux hémisphères au niveau d'électrodes temporales renforce cette hypothèse (figure 6.6). Dans chacune des tâches, on observe une forte asymétrie entre les signaux enregistrés sur l'hémisphère gauche et sur l'hémisphère droit. Quelle que soit la tâche, l'amplitude des signaux est beaucoup plus importante au-dessus de l'hémisphère droit. Que cette asymétrie soit similaire dans chacune des tâches laisse penser que les mêmes zones neuronales sont recrutées.

6.3 - Discussion

Les résultats obtenus au cours de cette expérience sont divers : tout d'abord la durée de traitement additionnel nécessaire à une tâche de catégorisation complexe et abstraite d'images naturelles par rapport à une simple tâche de détection d'une image donnée, est seulement de 30 ms si l'on considère les PEs différentiels et au maximum de 40 ms en termes de TRs. S'il était possible d'anticiper cette réduction de temps de traitement pour la tâche de détection, l'évaluation du surcoût comportemental dû à la tâche de catégorisation apparaît faible au regard de la différence de complexité des deux tâches. Il semble de plus que l'activité cérébrale différentielle soit corrélée à la performance des sujets et par conséquent au processus décisionnel. Le second résultat concerne cette décision perceptive finale dont plusieurs arguments nous permettent de dire qu'elle implique la même structure cérébrale dans les deux tâches. Ces arguments nous permettent de supposer que la différence observée sur la latence des potentiels évoqués différentiels serait due à une dynamique d'activation différente des voies visuelles impliquées. Ces résultats ont à mon avis des répercussions importantes pour expliquer la rapidité du traitement et imposent de sévères contraintes aux modèles de traitement de l'information dans le système visuel. Cependant, j'aborderai ce sujet, qui est en

fait le sujet commun de toutes les expériences que j'ai présentées, dans la discussion générale qui clôt la partie expérimentale.

6.3.1 – Influences descendantes

Nos résultats montrent que les mêmes aires corticales seraient impliquées dans la décision perceptive qu'il s'agisse d'une tâche de détection ou d'une tâche de catégorisation. En accord avec ces données, les études en PET et en IRMf fonctionnelle indiquent que des zones proches de celles que l'on peut localiser dans notre tâche sont activées dans les tâches de catégorisation et de détection (i.e. Gauthier et al, 1999; Haxby et al, 1996; Roland & Gulyás, 1995). Toutefois, du fait de la faible résolution temporelle de ces techniques, en IRMf et PET, il n'est possible de voir que les processus de plus forte amplitude ou de plus longue durée. Bien qu'on puisse supposer que cela soit le cas pour l'onde qu'on observe à 250 ms, il n'est pas certain que la comparaison soit possible avec notre expérience[26]. La localisation de la décision perceptive dans une zone cérébrale identique dans les deux tâches est un résultat qui est loin d'être trivial. On aurait pu croire par exemple que, dans la tâche de détection, les zones cérébrales impliquées dans la décision soient de plus bas niveau que dans la tâche de catégorisation. Par exemple l'aire V4 serait impliquée dans la mémoire à court terme des couleurs, les neurones de cette aire restant actifs lors d'une tâche d'appariement (McKeefry et Zeki, 1998).

La raison tient à mon avis aux stimuli que nous avons choisi d'utiliser : des images naturelles. Dans la tâche de détection, distracteurs et images-cibles partagent donc de nombreuses caractéristiques de bas niveaux comme les fréquences spatiales, les contrastes, les couleurs et la forme. La préactivation ne peut donc se baser entièrement sur les caractéristiques des images et doit faire intervenir des niveaux de représentation suffisamment abstraits pour permettre une réponse correcte. Ce niveau est bien évidemment celui du contenu de l'image. Dans le cas où les cibles flashées sont des images ne contenant pas d'animaux, il est possible que la décision perceptive soit plus difficile du fait de la plus grande similarité des distracteurs avec la cible, tant du point de vue de la forme que du contenu de l'image. Cet argument tient également pour les images d'animaux dites difficiles. Cette hypothèse s'apparente à la détection d'éléments diagnostiques dans l'image (Schyns, 1999)

[26] Les techniques de TEP et d'IRMf permettent de localiser les zones fortement irriguées dans le cerveau et l'on suppose que cette irrigation est directement corrélée avec l'activité des neurones sous-jacents. Ces techniques cependant ne permettent pas d'atteindre une résolution temporelle supérieure à la seconde et les dynamiques temporelles d'activation suite à la présentation d'un stimulus sont difficiles à aborder.

dans laquelle la tâche permet de biaiser la catégorisation vers telle caractéristique ou telle autre des images. Dans une tâche de dénomination de scènes de ville où les stimuli sont construits à partir d'une combinaison des traits grossiers d'une image et des traits fins d'une autre, les sujets sont capables de répondre soit sur une image (i.e. contours fins), soit sur l'autre (i.e. contours grossiers) (Schyns et Oliva, 1994). De même les sujets peuvent chercher des objets de différentes tailles dans les images (Kinchla, 1974**;** Kinchla, 1992).

Les résultats concernant la similarité des images sur lesquelles les sujets commettent des erreurs dans la tâche de détection (figure 6.1) contredisent en apparence cette hypothèse, les sujets commettant plus d'erreurs sur les distracteurs similaires du point de vue de la couleur et de la forme avec la cible. Toutefois, il est possible que le maintien en mémoire de la cible se fasse dans les aires visuelles de haut niveau et biaise l'activité dans les aires de bas niveau en fonction des propriétés de la cible. Ce type de mémoire à court terme serait tourné vers l'objet et indirectement vers ses caractéristiques. Cela est compatible avec la différence précoce que l'on observe à 100-120 ms, sélective aux animaux. Ce processus se localiserait dans les aires visuelles de bas niveau comme V1 (VanRullen et Thorpe, 2000b) ce qui est compatible avec l'hypothèse d'une action descendante des aires visuelles de haut niveau.

Cette hypothèse s'accommode relativement bien des résultats que nous avons obtenus dans la tâche de détection sur les images faciles et difficiles. Les images faciles à catégoriser semble également être facile à détecter et il est donc possible que cet avantage soit dû à des propriétés de bas niveau des images comme la luminance ou les contrastes locaux. Nous avons en effet vérifié que les erreurs dans la tâche de détection semblent plus fortement corrélées aux caractéristiques des images pour les animaux faciles par rapport aux animaux difficiles (figure 6.1). Ces images seraient plus rapidement traitées car elles sont plus en adéquation à la fois avec la préactivation intervenant dans la tâche de détection et avec l'amorçage intervenant dans la tâche de catégorisation. Dans la tâche de détection et la tâche de catégorisation, il est possible de préactiver le système visuel pour l'optimisation de la détection de certaines caractéristiques des cibles. Dans la tâche de détection de cible unique, il est également possible de préactiver la détection de ces caractéristiques à certaines positions dans l'image. En un sens, la préactivation, dans le cas des cibles uniques, pourrait à la fois faire intervenir la voie ventrale, impliquée dans l'identification des caractéristiques des objets, mais aussi la voie dorsale qui interviendrait dans la position de ces caractéristiques. Cette double préactivation pourrait alors expliquer la différence de rapidité entre la tâche de détection et la tâche de catégorisation.

Toutefois, cela ne signifie pas que les influences descendantes doivent provenir d'une structure située à un haut niveau hiérarchique : le niveau dont elles sont issues dépend à mon avis de la tâche à effectuer. Par exemple dans une tâche de catégorisation simple barres horizontales vs barres verticales, la préactivation pourrait intervenir entre aires visuelles de bas niveau. Si l'on avait demandé au sujet de catégoriser des images en couleur parmi des images en NB, les aires cérébrales activées n'auraient peut-être pas été les mêmes, la préactivation aurait pu dépendre de V4 (McKeefry et Zeki, 1998; Fize, 2000).

Les implications de ce type de préactivation pour amorcer les traitement dans le système visuel sont nombreuses. Nous avons montré dans notre expérience que l'effet lié au type de tâche réalisée était visible dès 140 ms, c'est-à-dire approximativement la latence à laquelle l'onde de décision émerge dans la tâche de décision. L'absence d'un effet lié à la tâche avant 140 ms ne signifie pas forcément que des processus de préactivation précoces dans les aires de bas niveau n'interviennent pas. Pour reprendre l'hypothèse d'amorçage des aires de bas niveau (ou de reconnaissance de formes simples) par les aires de haut niveau : l'amorçage dans le cas de la tâche de détection serait plus focalisé que dans la tâche de catégorisation, les formes à reconnaître étant mieux définies. Toutefois, il est probable qu'en moyennant les PEs sur un grand nombre d'image-cibles uniques, on ne puisse plus distinguer les différences spécifiques de chacune des deux tâches du point de vue de la préactivation.

Même s'il est probable que la préactivation provienne d'aires visuelles de haut niveau, la décision dans la tâche de détection de cible unique ferait plus fortement intervenir les aires de bas niveau. Cette différence de préactivation dans les deux tâches pourrait expliquer pourquoi la décision perceptive est plus rapide dans la tâche de détection d'une cible unique. Le niveau de confiance que le sujet a de sa réponse, même s'il est probablement inconscient, serait atteint plus rapidement dans le cas de la tâche de détection. Dans le paragraphe suivant, nous allons donc tenter d'analyser les résultats que nous avons obtenus dans le cadre de la théorie d'échange entre la précision que le sujet peut obtenir et la rapidité de ses réponses.

6.3.2 - La loi d'échange précision/vitesse

Plus le sujet est rapide et moins il est précis et inversement, plus il prend du temps pour répondre et plus la probabilité qu'il réponde correctement est élevée (Fitts, 1954). La dynamique de l'activation cérébrale serait en relation avec la performance des sujets à un moment donné. La figure 6.4 est en accord avec la loi d'échange précision/vitesse. Cette idée trouve un écho dans les travaux de Mozer et al (2000) qui implémente en cascade des

processus probabilistes obéissant à une loi d'échange précision/vitesse pour expliquer des données comportementales sur l'amorçage de mots. Nous reviendrons en détail sur ces travaux dans le chapitre suivant.

À partir de cette hypothèse, il est également possible d'interpréter la différence entre la tâche de détection et la tâche de catégorisation que nous avons obtenue. Les neurones responsables de la décision perceptive pourraient être sélectivement activés par les images d'animaux dans la tâche de catégorisation. Comme j'en ai déjà longuement parlé, la tâche de détection ferait plus sélectivement intervenir les aires corticales de bas niveau pour biaiser la réponse de ces neurones. Si la couleur et la forme soit très différentes de l'image cible, le biais peut être très important. Si cependant les caractéristiques visuelles des distracteurs sont très proches de celles des cibles, la décision se basera sur le contenu de l'image.

La courbe de précision/vitesse est donc décalée pour certaines images. Pour cette raison et pour vérifier cette hypothèse, il aurait été intéressant de demander à un autre groupe de sujets de répondre sur les distracteurs plutôt que sur les cibles uniques, le sujet devant mémoriser la cible et répondre uniquement sur les distracteurs. On aurait ainsi pu déterminer lesquelles des images parmi les distracteurs étaient les plus difficiles et quels étaient leurs points communs avec l'image unique.

Dans l'expérience que j'ai présentée, l'échange précision/vitesse est visible sur la performance des sujets. De plus, nous avons montré que le développement des ondes débutant vers 150 ms dans les deux tâches semble fortement corrélé avec la performance des sujets (figure 6.4). Si l'on considère un retard de 60 ms pour l'intégration motrice, alors les courbes de performance et celles des potentiels évoqués différentiels se superposent parfaitement dans les deux tâches. Il est probable qu'à ces latences des biais en provenance du cortex préfrontal soient présents. Chez un patient présentant une lésion préfrontale unilatérale, les signaux EEG entre les deux hémisphères divergeaient dès 150 ms suite à la présentation d'un stimulus, (Barcelo et al, 2000). Le cortex préfrontal pourrait donc intervenir dans l'onde que l'on observe autour de 150 ms et dont la latence est corrélée avec les TRs des sujets, et qui est en relation avec des processus à très haut niveau d'intégration, comme la décision. Cette onde pourrait ainsi résulter de l'implication de multiples aires neuronales et pas uniquement des cortex occipito-temporaux localisés à l'aide de BESA.

L'expérience que j'ai présentée ici impose également de très fortes contraintes temporelles sur la catégorisation ultra-rapide. Dans la discussion générale de l'ensemble des résultats expérimentaux, nous tenterons d'inférer les implications de nos données sur les traitements qui s'effectuent dans le système visuel et en particulier sur le traitement *feedforward*.

7

Conséquences de la catégorisation ultra-rapide

Avant d'aborder une discussion générale sur les implications des données que j'ai présentées, je souhaite revenir brièvement sur la catégorisation du singe. Chez l'homme, la question du niveau de catégorisation ne se pose pas vraiment puisque l'on demande explicitement aux sujets de catégoriser les animaux. Chez le singe, l'apprentissage remplace la consigne verbale. Si on ne peut réellement savoir ce que le singe a appris, nous avons cependant un certain nombre d'arguments pour penser que l'animal effectue réellement une catégorisation animal/non-animal. Tout d'abord, dans les articles princeps de Fabre-Thorpe et al (Fabre-Thorpe et al 1998; Fabre-Thorpe et al, 1999), les animaux effectuent le même type d'erreur et sur les mêmes images que les sujets humains. De plus, les images sur lesquelles les TRs des singes sont les plus lents sont également celles sur lesquelles les TRs des sujets humains sont les plus lents. L'apprentissage d'Eudora nous renseigne également sur la stratégie utilisée par le singe pour effectuer sa catégorisation. Nous avons pu constater que d'une stratégie d'association stimulus/réponse, Eudora évoluait vers une stratégie plus complexe : la précision sur les images nouvelles contenant des animaux - qu'elle n'avait jamais vues auparavant - est en effet très faible lors des premières séances (20 %) et très

élevée lors des dernières séances (90 %). Le fait que la suppression des informations chromatiques des images n'ait aucun effet sur les réponses les plus rapides à la fois chez l'homme et chez le singe montre également que les deux espèces se basent sur les caractéristiques de contour des cibles dans l'image pour réaliser cette catégorisation rapide (Delorme et al, 1999; Delorme et al, 2000). Enfin, l'expérience réalisée chez l'homme, qui indique que la détection d'un animal dans l'image ne dépend pas de la familiarité de l'image, suggère que le traitement visuel impliqué est relativement automatique et pourrait donc déjà être présent chez d'autres espèces animales. Même si le "concept" d'animal ou la représentation des catégories diffère chez le singe et chez l'homme, l'hypothèse selon laquelle il existe une large superposition de ces représentations est donc plausible.

Il est cependant possible que les similarités entre hommes et singes aient été accentuées par les contraintes de rapidité inhérentes à la tâche utilisée. Si les processus sous-jacents à la catégorisation rapide sont similaires chez le singe et l'homme, la plus grande rapidité du singe pourrait être attribuée à la loi d'échange précision/vitesse : le singe étant moins précis que l'homme. Une autre explication prend en compte la taille de son cerveau : une grande partie du temps de traitement étant dévolu au transfert des informations d'une structure cérébrale à l'autre, et les vitesses de propagation intracorticales étant particulièrement lentes : 1 m/s (Nowak et Bullier, 1997), la petite taille du cerveau du singe pourrait lui permettre d'économiser un temps considérable.

7.1 - Catégorisation animal/non-animal

Après cette brève mise au point sur la catégorisation chez le singe, je voudrais axer principalement l'interprétation qui peut être faite des résultats que nous avons obtenus sur les processus de traitement sous-jacents. Nos résultats montrent une certaine robustesse des réponses précoces à toute manipulation expérimentale et fournissent des arguments en faveur d'un traitement visuel massivement parallèle et automatique, et pourquoi pas basé sur une seule vague d'information visuelle se propageant vers l'avant : c'est-à-dire *feedforward*. Dans toutes les expériences que j'ai présentées, l'effet des caractéristiques des images sur les réponses précoces est mineur, à la fois pour les caractéristiques intrinsèques des images, qui se réfèrent au contenu de l'image, et pour leurs caractéristiques extrinsèques, qui dépendent des conditions dans lesquelles elles sont présentées. Le cas le plus critique semble être celui de la tâche où la détection de cibles uniques parmi des distracteurs permet un gain d'environ 30 ms à la fois sur les TRs et les PEs. L'influence d'autres caractéristiques comme la couleur,

la luminance de la cible et le contenu des images, semble assez faible. Concernant les caractéristiques de contenu des images, l'influence de la position - typique ou atypique - de l'animal, la présence des yeux ou des 4 membres est visible de façon plus précoce que l'ensemble des autres caractéristiques. Cependant pour ce type de caractéristiques, l'effet sur les TRs précoces, s'il existe, est très faible et n'excède en aucun cas 10 ms. Pour la catégorisation rapide, quelques conditions sont totalement neutres comme la familiarité des images et l'effet de séquence. Dans l'expérience de familiarité des images au chapitre II.5, les TRs rapides ne sont pas accélérés par un entraînement intensif. Que l'on considère les TRs rapides ou le PE différentiel, le traitement visuel semble être aussi rapide pour les images nouvelles que pour les images très familières. Cet effet est pour le moins surprenant car l'abondante littérature sur l'amorçage indique que l'expérience du sujet avec l'image devrait faciliter la catégorisation. En fait, on observe bien un effet d'amorçage[1], mais cet effet n'est pas présent sur les réponses les plus précoces. De même, l'effet de séquence au chapitre II.4, c'est-à-dire l'incidence du stimulus qui précède l'image à catégoriser, perturbe les TRs en accord avec la loi d'échange précision/vitesse et n'affecte donc pas la performance globale sur ces réponses. Dans le cas où deux cibles se suivent, la réponse sur la seconde cible est statistiquement plus rapide que si elle avait été précédée d'un distracteur mais la probabilité que le sujet commette une erreur est supérieure, de sorte que les deux effets se compensent.

D'autres études effectuées au laboratoire sont complémentaires de celles que j'ai réalisées et renforcent ces résultats. Concernant la tâche en elle-même, en EEG, des travaux ont comparé la tâche de catégorisation animal à d'autres tâches de catégorisation: carré vs rond[2] (Aubertin et al, 1999), images en couleur vs NB (Fize, 2000), aliments vs non-aliments (Fabre-Thorpe et al, 1998; Delorme et al, 2000) et véhicules vs non-véhicules (VanRullen et Thorpe, 2000a; VanRullen et Thorpe, 2000b). Toutes ces études indiquent que les réponses les plus rapides, observées dans la détection des cibles de ces diverses catégories, apparaissent à la même latence que dans la tâche de catégorisation animal/non-animal. De plus, dans tous les cas, les PEs moyennés sur les cibles et les distracteurs diffèrent à environ 150 ms après la présentation du stimulus[3], ce qui indiquerait que le processus de décision est similaire (cf. chapitre II.6). Les images contenant des animaux ne semblent donc pas subir un sort qui leur serait particulier lors de leur traitement par le système visuel.

---

[1] 20 ms de gain pour les TRs et 2,2 % d'augmentation de précision.

[2] Cf. la première note du chapitre précédent concernant cette expérience.

[3] A l'exception de la détection du caractère achromatique d'image NB présentées parmi des images en couleur qui semble retardée de 50 ms à la fois sur les PEs et les TRs.

D'autres expériences menées au laboratoire renforcent un type de traitement massivement automatique pour ces catégories. Ces études tentent en particulier de déterminer le rôle de l'attention dans la tâche de catégorisation animal (Fize, 2000; Thorpe et al, 2000). Elles montrent que précision et temps de réaction ne sont affectés que de façon minime lorsque les images sont présentées à 3,5° d'excentricité. Les PEs à 150 ms pourraient être légèrement retardés (Fize, 2000). Plus récemment, il a été montré que l'homme est capable de catégoriser des images contenant des animaux avec une précision supérieure au niveau de la chance à plus de 75° d'excentricité (Thorpe et al, 2000) alors même qu'il doit partager son attention sur l'ensemble du champs visuel horizontal. A ce niveau d'excentricité, l'acuité est faible et le sujet répond de façon instinctive. Cela renforce encore la caractéristique automatique des processus sous-jacents à la catégorisation d'animaux et montre que ce type de tâche requiert peu (ou pas !) d'attention focalisée, bien que celle-ci influence légèrement la performance des sujets en accord avec les données présentes dans la littérature (i.e. Egeth et Yantis, 1997).

Toutes ces études montrent que le traitement visuel d'une image doit être massivement parallèle et que les informations visuelles provenant du stimulus suivent des chemins précablés et automatiques dans le système visuel de façon majoritairement *feedforward*. Ces processus "vers l'avant" sont cependant bien plus complexes et sophistiqués qu'on a pu le penser auparavant.

7.2 - Catégorisation *feedforward*

Les résultats que je viens de présenter sont compatibles avec les données électrophysiologiques. Tout d'abord dans IT, la sélectivité des neurones aux objets est présente très rapidement après la présentation du stimulus (80-100 ms). Etant donné le nombre d'aires neuronales à traverser pour atteindre ce niveau, ces résultats suggèrent que les processus de *feedback* n'ont pas le temps d'intervenir pour établir cette sélectivité (Celebrini et al, 1993; Oram et Perrett, 1992, Tovee et Rolls, 1995; Keysers et al, 2000). Une grande partie de l'information visuelle semble être disponible dès les premières décharges neuronales : au moins 50% de l'information sur le stimulus dans V1 et IT semble concentrée dans la première vague de décharges (Heller et al, 1995; Sugase et al, 1999) et pour Tovee et Rolls (1995), 67 % de l'information du stimulus est encodée durant les 50 premières millisecondes de la décharge d'un neurone de IT.

7.2.1 Éléments diagnostiques

Les analyses que j'ai effectuées pour tenter de déterminer les caractéristiques de l'image importantes pour la réalisation d'une catégorisation rapide montrent que l'incidence de chacune des différentes caractéristiques étudiées est très faible. Les différentes parties du corps des animaux semblent cependant avoir des effets différents, la présence des membres par exemple ayant une influence plus importante que les autres caractéristiques. Ces caractéristiques sont diagnostiques de la présence d'un animal, ce qui signifie que la présence de l'un de ces éléments indique *a priori* la présence d'un animal (cf. Schyns, 1999, pour une revue). Le traitement visuel serait en fait fortement biaisé pendant la tâche. La présence d'un œil ou d'une tête indique par exemple avec une forte probabilité qu'un animal est présent dans l'image. Il est probable que la décision soit prise dès que l'un de ces éléments est détecté dans l'image. Cela permettrait d'expliquer pourquoi il nous a été possible de classer l'influence de la présence ou de l'absence de certaines caractéristiques sur les temps de réaction. Si ce type de modèle se vérifie, il ne laisse que peu de place aux effets de feedback dynamique dans le traitement visuel, qui est pourtant une hypothèse couramment répandue (i.e. Ullman, 1995). Il n'est pas nécessaire de considérer l'ensemble des caractéristiques des animaux pour les reconnaître, la détection des différentes caractéristiques étant effectuée en parallèle et la première à être détectée, ou la plus représentative de la présence d'un animal, pouvant directement mener à la décision. L'analyse des erreurs sur les faux animaux dans l'expérience de familiarité est compatible avec ce type de traitement automatique[4] et d'autres expériences sont en cours pour tenter de vérifier cette hypothèse. Toutefois, certaines des images distracteurs, notamment dans les deux dernières expériences présentées, représentaient des individus humains qui partagent de nombreuses caractéristiques avec les animaux. Les sujets catégorisaient correctement ces images et une catégorisation basée sur la détection d'éléments diagnostiques ne peut donc, à elle seule, expliquer les résultats que nous avons obtenus.

Certains argumenteront que la détection d'éléments diagnostiques nécessite un *feedback* des aires visuelles de haut niveau vers les aires de bas niveau. Cependant, pour ce type de mécanisme que j'ai traité en détail au chapitre II.3, le terme de *feedback* est inapproprié. Le terme de *feedback* signifie "effet en retour" et il fait donc intervenir un processus itératif. Pour

[4] Concernant l'expérience du chapitre II.5 sur la familiarité des stimuli, des faux animaux (ballon de foire, montgolfière, statue...) avaient été insérés dans les séries. Les sujets humains commettaient régulièrement des erreurs sur ces images. Il est intéressant de noter qu'ils commettaient toujours ces erreurs quand ces distracteurs étaient familiers et qu'ils savaient parfaitement qu'ils ne devaient pas répondre dessus. Cela semble indiquer un traitement automatique sur lequel l'inhibition volontaire n'a que peu de prise.

la détection d'éléments diagnostiques dans les images, la présélection de certaines populations de neurones dans le système visuel avant la présentation de l'image n'implique pas nécessairement de traitement en boucle mais plutôt des influences descendantes des aires visuelles de haut-niveau vers celles de plus bas-niveau qui seraient mises en place pour toute la durée du test. Les traitements *feedforward* pourraient donc être modulés par cette préactivation des voies visuelles.

Cette hypothèse, d'une catégorisation basée sur des éléments diagnostiques présents dans les images cibles, s'accorde parfaitement avec le modèle de Humphreys et Forde (2000) de traitement dans le système visuel. Il semble que le temps nécessaire pour nommer un animal soit plus long que le temps nécessaire pour nommer un objet manufacturé[5] (Humphreys et Forde, 2000), alors que le temps de catégorisation des animaux serait plus rapide que celui d'objets manufacturés (Humphreys et Forde, 2000; VanRullen et Thorpe[6], 2000a). Pour expliquer ces résultats, Humphreys et Forde affirment que, étant donné que les animaux possèdent des similitudes de forme (patte, œil...), la compétition pour l'identification prend plus de temps que pour les objets manufacturés, qui sont en général de formes différentes. Par contre, concernant la catégorisation, l'identité de l'animal importe peu et l'activation d'éléments diagnostiques suffit à détecter la présence d'un animal : les TRs pourraient donc être relativement rapides. Les objets artificiels ne possédant pas de telles similitudes de formes et donc de propriétés diagnostiques, le système visuel devrait passer par l'identification de ces objets pour les catégoriser, les réponses étant donc plus tardives. Du point de vue de l'évolution, cela a un sens d'utiliser à chaque fois que cela est possible des processus basés sur les éléments diagnostiques contenus dans les images cibles, et un traitement *feedforward*. Cela permet d'accélérer les réactions des organismes et de leur assurer une plus grande chance de survie.

7.2.2 - Voie magnocellulaire et parvocellulaire

Un autre argument en faveur d'un traitement *feedforward* provient de l'expérience où les indices de couleur ont été supprimés. Comme on l'a vu dans la partie introductive, les

[5] Je reviendrai en détails sur ce modèle dans la conclusion.

[6] Bien que les auteurs montrent que, statistiquement, les TRs les plus rapides pour les moyens de transport (avion, voiture, hélicoptère, bateau...) ne sont pas plus lents que ceux pour les animaux, s'il existe une tendance, elle serait en faveur d'une catégorisation plus rapide dans la tâche animal/non-animal. On peut expliquer cet absence d'effet du fait que les objets de la classe des véhicules possèdent – comme pour les animaux - de nombreuses caractéristiques communes (roue, pare-brise...).

informations – achromatiques – véhiculées par la voie magnocellulaire se propagent plus rapidement d'environ 20 ms que celles de la voie parvocellulaire (Nowak et al, 1995; Nowak et Bullier, 1997). De plus, une inactivation réversible de la voie magnocellulaire dans le LGN induit une diminution de 47 % de l'activité des neurones dans V4, supérieure à la réduction de 36 % provoquée par une inactivation de la voie parvocellulaire[7]. Cela montre une claire participation de la voie magnocellulaire à la reconnaissance des objets[8] (Ferrera et al, 1994).

Même si les informations fournies par le système magnocellulaire sont moins précises en termes de résolution spatiale que celles en provenance du système parvocellulaire, il est possible que le traitement puisse s'effectuer sur la base de ces informations grossières. Des expériences ont en effet montré que des sujets humains peuvent réaliser des tâches de reconnaissance sur des photographies avec de très faibles résolutions (de l'ordre de 5 par 8 pixels pour la reconnaissance de visages connus et de 5 par 5 pour leur détection; communication personnelle du professeur Pawan Sinha au MIT). Dans notre expérience, quand la couleur, majoritairement véhiculée par la voie parvocellulaire, est supprimée, cela semble n'avoir aucun effet sur les temps de réaction rapides. On peut donc supposer que la voie magnocellulaire véhicule la majorité des informations visuelles utilisées pour ces réponses rapides. De même, le fait que la luminance joue un rôle important et de façon relativement précoce - tout du moins pour les images en couleur - semble indiquer que la dynamique d'activation des neurones à l'entrée du système visuel est critique pour la rapidité de détection, les luminances plus élevées induisant des décharges plus précoces dans la rétine (Albrecht, 1995; Gawne et al, 1996). L'effet de la taille des animaux semble également être un argument de poids. Les sujets humains ont en effet beaucoup de mal à détecter les animaux très petits et on peut supposer que la résolution des neurones magnocellulaires n'est pas suffisante pour que la détection ait lieu rapidement.

7.2.3 - Échange précision/vitesse

Les expériences présentant de façon très rapide des séries d'images - RSVP[9] - viennent également renforcer notre argument en faveur d'un traitement *feedforward*. Nous sommes, par exemple, capables de détecter une cible même si 10 images sont présentées par seconde (Potter, 1999), ce qui implique que le traitement dédié à chaque image ne dépasse pas 100 ms.

---

[7] V4 est une aire corticale impliquée dans le traitement des objets dans la voie ventrale.

[8] On ne sait pas encore dans quelle mesure la voie magnocellulaire active directement les aires corticales de la voie ventrale ou indirectement (en passant par la voie dorsale).

[9] Rapid Serial Visual Presentation.

Dans une superbe expérience, Keysers et al (2000) présentent à un singe une série de visages de leurs congénères. Les auteurs commencent par enregistrer un neurone dans IT et déterminent un stimulus sur lequel le neurone répond de façon optimale. Ils présentent ensuite aléatoirement cette image en RSVP parmi d'autres images de visages et observent que le neurone est significativement plus activé sur le visage auquel il est sélectif. L'effet est toujours présent lorsque dans la séquence de stimuli, chacun n'est présenté que pendant 14 ms, soit à une fréquence de 71 images par seconde. Le neurone répond environ 100 ms après la présentation de l'image, ce qui signifie que, au moment où le neurone répond sur l'image à laquelle il est sélectif, 6 autres images sont en train d'être traitées par le système visuel. Cela implique clairement un processus *feedforward* dans lequel le traitement dans une couche ou une aire neuronale est communiqué en cascade à la suivante.

Ce type de raffinement de la sélectivité a fait l'objet d'un modele. Mozer et al (2000) ont en effet imaginé les neurones comme des unités probabilistes. Chaque unité est sélective et sa sélectivité augmente au cours du temps. Les informations sont copiées au niveau suivant de sorte qu'au niveau supérieur, le niveau d'activation des unités ne dépend que de celles des niveaux inférieurs - par exemple le premier niveau peut représenter les caractéristiques de l'image et le second celles des objets. L'avantage d'une telle représentation est qu'il n'est pas nécessaire de prendre en compte les interactions entre les objets car elle est incluse en quelque sorte dans la règle d'augmentation de la sélectivité au cours du temps. Mozer et al (2000) utilisent ce modèle de façon statique, c'est-à-dire qu'ils ne raisonnent que sur la moyenne des TRs dans les expériences. Je pense cependant qu'il est plus approprié de mettre en parallèle la dynamique de réponse des sujets avec la dynamique d'activation des unités. On peut en effet mettre en relation les courbes présentées à la figure 7.1 avec les courbes de performance (d') que j'ai présentées précédemment. L'activation d'une unité sélective aux objets dans le modèle de la figure 7.1 peut être biaisée, soit par l'augmentation de la probabilité initiale de l'unité, soit par la dynamique de l'activation de l'unité. Ces deux processus introduisent des modifications différentes de la courbe de sélectivité initiale : le premier induit un décalage de cette courbe vers la gauche, c'est-à-dire vers des réponses plus précoces, et pourrait être attribué à une préactivation des voies visuelles. Le second phénomène provoque une amplification de la courbe de sélectivité et est donc assimilable à un effet de *feedback* dynamique. Ainsi, dans la dernière expérience, le décalage entre les courbes de performance dans la tâche de détection et celle de catégorisation peut être attribué à une augmentation de la probabilité initiale attribuées aux objets cibles et donc à une préactivation, comme nous l'avons déjà discuté au chapitre précédent. Par contre dans l'expérience de familiarité des

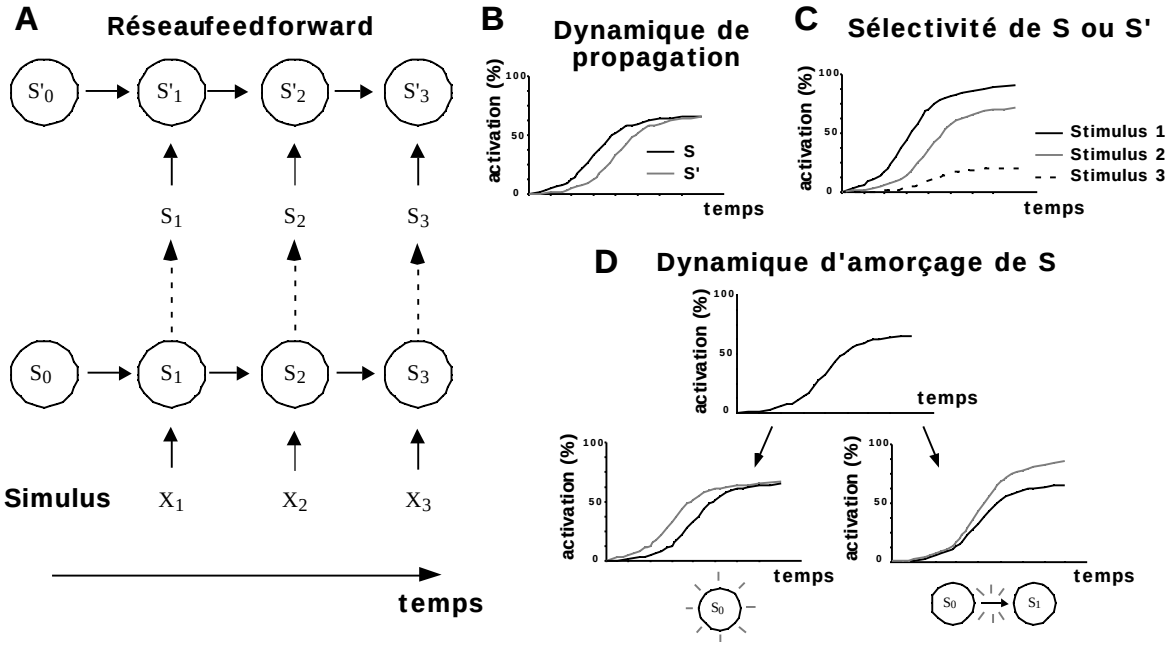


Figure 7.1 : modèle de Mozer et al. A, deux unités organisées hiérarchiquement sont représentées en fonction du temps (S_i étant le niveau d'activation de S au temps *i*). La première unité S détecte par exemple les attributs de l'image à un endroit donné et la seconde S' la présence d'un objet particulier (il peut exister par exemple une unité par objet). Ce modèle statistique est *feedforward* dans le sens où l'unité S' ne dépend que de son niveau d'activation au pas de temps précédent et d'une entrée de S, considérée par l'unité S' comme une constante. B, la dynamique de propagation pour S et S' est indiquée : S' est en retard par rapport à S car située à un niveau hiérarchique supérieur. C, sélectivité de S ou S' pour différents stimuli : le niveau de l'activation de l'unité dépend de l'objet présenté. D, au cours de présentations précédentes du stimulus, il peut y avoir modification de la dynamique : amorçage du système. Il existe deux possibilités: soit, à gauche, la probabilité initiale S_0 ou S'_0 augmente, ce qui induit un décalage de la courbe de sélectivité, soit, à droite, la dynamique d'activation de S ou S' est plus rapide, ce qui provoque une amplification de la réponse. Adapté de Mozer et al (2000).

images, les courbes de performance pour les images familières et nouvelles étant initialement confondues, il s'agirait plutôt d'un renforcement des connexions entre les différents états au cours du temps d'une unité et donc d'un processus de *feedback*, mais qui, dans notre cas[10], serait tardif. Cette description semble en parfaite adéquation avec nos résultats et il n'y aurait qu'un pas à franchir pour considérer les neurones comme des unités probabilistes, une hypothèse qui ne semble pas totalement être dénuée de sens (Koechlin et al, 1996; Koechlin et al, 1999).

7.2.4 - Le rôle du feedback

Bien que toutes mes données supportent un modèle *feedforward*, il n'y a aucune raison d'imaginer que le traitement visuel s'arrête là. Les processus itératifs de *feedback* pourraient

[10] Mozer et al raisonnent en général sur les temps de réaction moyen, mais je pense que les courbes de d' et d'EEG sont parfaitement appropriées à ce type d'analyse.

intervenir tardivement dans le raffinement de la sélectivité des neurones ou des aires corticales. Cette hypothèse est compatible avec les modèles de traitement séquentiels/parallèles (Humphreys et al, 1988). Dans ces modèles, un traitement grossier et sommaire, par exemple basé sur les informations de la voie magnocellulaire, s'affine avec le temps (Parker et al, 1992). Au sein du système visuel hiérarchique, les informations imprécises sont propagées aux niveaux supérieurs pour construire progressivement une représentation du contenu de l'image sur laquelle se baserait la décision. Ainsi les décisions les plus rapides pourraient dépendre uniquement des informations achromatiques portées par la voie magnocellulaire, les informations chromatiques véhiculées par la voie parvocellulaire n'étant pas encore disponibles. Les décisions plus tardives feraient, quant à elles intervenir des processus itératifs : l'expérience sur la familiarité des stimuli au chapitre II.5 montre en fait, que le traitement tardif peut-être accéléré dans le cas d'images familières, un résultat qui indique que certains traitements probablement itératifs peuvent être raccourcis. De même la présence de la plupart des caractéristiques des images, notamment la couleur présentent un effet tardif. Il semble en fait que les décharges plus tardives mettent en jeu les processus de mémorisation ou d'amorçage (Subramaniam et al, 2000). Les processus conscients en particulier font probablement intervenir des effets en retour pour le remplissage des surfaces (Hupé et al, 1998; Bringuier et al, 1999) où le liage dynamique de l'activité des neurones codant les différents objets est nécessaire au percept conscient (von der Malsburg, 1995). Ce type d'interaction pourrait faire intervenir des synchronisations entre des aires neuronales distinctes (Rodriguez et al, 1999; Tallon-Baudry et Bertrand, 1999).

Il n'y a cependant aucune raison de penser que la perception consciente soit nécessaire pour effectuer la catégorisation d'animaux. Suite à la présentation d'une image flashée, quand on demande aux sujets de rapporter ce qu'ils ont perçu en tout premier, ils rapportent avoir vu la scène visuelle dans sa globalité. Le percept conscient implique probablement des traitements bien plus complexes que ceux qui sont nécessaires pour effectuer la tâche. Les réponses sur les "faux" animaux dont j'ai déjà parlé semblent également aller dans ce sens car bien que le sujet se rende systématiquement compte de son erreur, il continue à catégoriser ces "faux" animaux comme des cibles. Le système visuel peut également segmenter des chiffres de façon totalement inconsciente (Dehaene et al, 1998). Les auteurs présentaient un chiffre sous la forme d'une chaîne de caractères - par exemple "neuf" - suivit très rapidement d'un masque et d'un autre chiffre - cette fois en notation arabe "6" . Le sujet devait déterminer si le premier chiffre était plus petit ou plus grand que le second. Pour des temps de présentation très courts du premier chiffre, le sujet dit ne pas être capable de le voir.

Cependant, dans une situation de choix forcé, il effectue la tâche avec une précision significativement au-dessus du niveau de la chance. Cela signifie que le premier chiffre est traité sans pour autant atteindre une représentation consciente. Il semble que la latence de la perception du sens ou la perception consciente soit d'environ 300-500 ms chez l'homme, c'est-à-dire souvent bien après que les sujets ont répondu (Sperling et Reeves, 1980; Libet et al, 1981; Curran, 1993). Ces travaux renforcent donc l'hypothèse que les processus de feedback interviennent de façon tardive.

La sélectivité des neurones est en accord avec cette hypothèse. Dans V1, le feedback au sein de réseaux récurrents locaux pourrait servir à renforcer des entrées thalamiques de très faible amplitude, à filtrer certaines entrées, à remplir des contours ou à faire jouer les effets de contexte (Hupé et al, 1998; MacEvoy et al, 1998; Bringuier et al, 1999; Grossberg, 1999). Les auteurs sont cependant d'accord pour affirmer que ces processus sont très lents et prennent plusieurs dizaines de millisecondes. Ils n'interviendraient donc pas dans la catégorisation rapide. A un niveau hiérarchique plus élevé dans IT, l'expérience de Sugaze et al (1999) que j'ai déjà mentionnée, indique que l'information encodée par le neurone sur le stimulus s'effectue en deux temps. Lors des premières décharges, le neurone répond à la présence d'un visage alors que 50 ms plus tard, il répond à certaines caractéristiques de ce visage comme l'identité de l'individu ou le type d'émotion exprimée. Dans l'expérience de Keysers et al (2000) dont j'ai longuement parlé, plus le délai séparant deux images est important et plus la réponse du neurone est sélective : cela peut signifier qu'avec le temps, on passe d'un traitement grossier et très bruité à un traitement plus précis. Ces résultats renforcent en fait les modèles de raffinement de la sélectivité dans lesquels, du fait par exemple de la compétition entre les neurones et donc du feedback, la sélectivité fine est postérieure à une sélectivité grossière.

7.3 - Contraintes à imposer à la modélisation

Il convient maintenant de déterminer explicitement les contraintes à imposer à un modèle du système visuel pour qu'il soit compatible avec la dynamique que nous avons mise à jour.

1. La première contrainte est que le système visuel doit être capable de catégoriser des images achromatiques aussi rapidement que des images en couleur. Un modèle du système visuel minimaliste doit donc être capable de catégoriser les objets en NB et pouvoir s'affranchir de la couleur.

2. La seconde contrainte est que le traitement visuel doit pouvoir s'effectuer de façon *feedforward*. Je pense que j'ai introduit suffisamment d'arguments en faveur d'un tel processus. Le modèle devra donc intégrer l'organisation hiérarchique du système visuel et la propagation de l'activité neuronale devra se faire dans un seul sens, des aires de bas-niveau, vers les aires de haut-niveau. Au sein d'une couche neuronale, les neurones ne pourront pas faire appel à des processus de feedback dynamique.
3. La vitesse de propagation dans le traitement visuel impose également des contraintes sur la dynamique de réponse des neurones**.** Thorpe et Imbert (1989), en suivant un raisonnement limpide, montrent à la fois la nécessité d'un traitement *feedforward* et tentent d'expliquer la dynamique de traitement dans le système visuel pour qu'un neurone devienne sélectif à un objet complexe dans IT, seulement 100 ms après la présentation de l'image. Étant donné le nombre d'étapes neuronales à traverser pour atteindre IT (environ une dizaine), la fréquence maximale de décharge des neurones (100 décharges par seconde), et en comptant à peu près 10 ms par étape d'intégration synaptique, chaque neurone ne peut pas décharger plus d'une fois (figure 7.2). Un modèle de traitement visuel, devrait donc pouvoir être capable de rendre compte de la sélectivité des réponses dans le système visuel avec uniquement une seule décharge par neurone.
4. Dans ces conditions, il est nécessaire pour les neurones d'utiliser un autre moyen de codage de l'information visuelle que la fréquence de décharge, qui nécessite plusieurs décharges par neurone. Il est toujours possible d'utiliser un grand nombre de neurones, et d'estimer la fréquence de décharge sur cette population. Cependant, cela nécessitera un

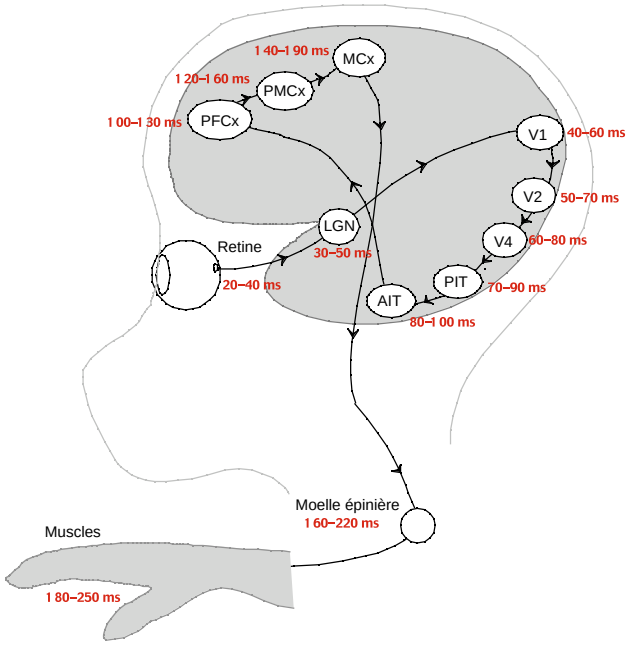


Figure 7.2 : représentation schématique du traitement visuel et de la réponse comportementale chez le singe dans la tâche de catégorisation. Le flèches indiquent les transferts d'information entre les différentes aires corticales, visuelles (V1, V2, V4, PIT, AIT), et frontales (PFCx, PMCx, MCx) avant d'être redirigées vers la moelle épinière et les muscles de la main qui permettent au sujet de relâcher le bouton. Pour la nomenclature des aires visuelles, se reporter à la figure I.2.3. Adapté de Thorpe (2000).

très grand nombre de neurones pour n'encoder qu'une seule valeur (Gautrais et Thorpe, 1998). Il est peu probable que les systèmes neuronaux favorisent un tel gaspillage d'énergie. Thorpe (1990) a imaginé un codage par population bien plus simple, dans lequel la latence relative de décharge des neurones encode le stimulus. La latence de décharge des neurones dépendant du contraste du stimulus présenté, il est possible de reconstruire l'image initiale à partir des latences de décharges des neurones de l'ensemble de la population (figure 7.3). Ce type de comportement correspond, dans une certaine mesure, à celui des neurones dans la rétine suite à un stimulus flashé. De cette façon, avec uniquement une seule décharge par neurone, il est possible d'encoder un stimulus complexe. Il est couramment admis que c'est la fréquence de décharge des neurones qui permet aux neurones de communiquer et cette hypothèse est donc novatrice. Elle ne signifie pas que la fréquence de décharge des neurones n'intervienne pas : simplement, les contraintes de la sélectivité rapide des neurones dans le système visuel font qu'elle n'est pas adaptée pour effectuer des traitements rapides suite à la présentation d'un stimulus. Les processus de codage de population que nous allons voir dans la partie suivante interviennent très probablement pour des stimuli apparaissant subitement, variant très rapidement ou suite à une saccade. Le modèle du traitement visuel que nous allons construire prendra en compte cette hypothèse de codage neuronal. Comme nous le verrons, un codage de population permet aux neurones de répondre très rapidement et de façon très sélective.

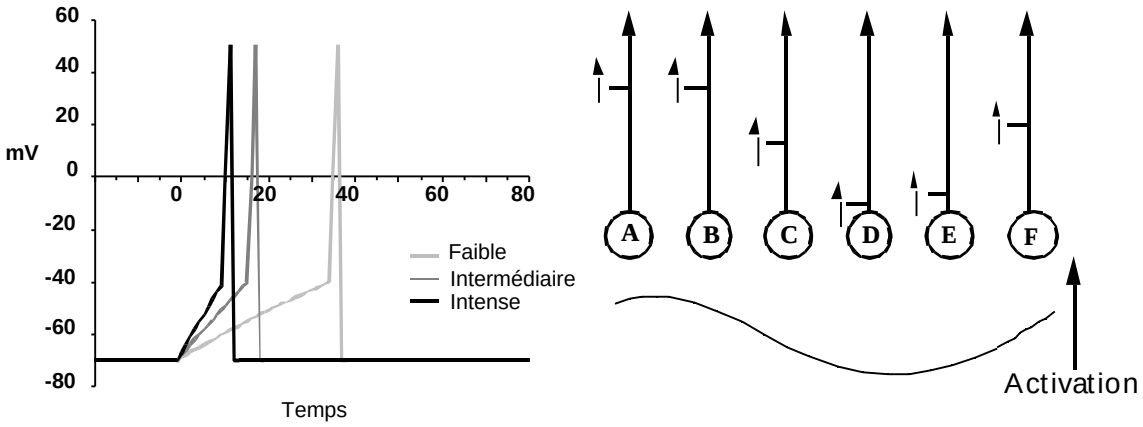


Figure 7.3 : à gauche, exemple de conversion d'une activation analogique en latence de décharge des neurones. La latence de décharge du neurone rend compte de l'intensité du stimulus, précoce pour les stimuli de forte intensité et tardive pour les stimuli de faible intensité. A droite, le délai d'activation de 6 neurones en fonction de l'intensité d'un stimulus à une dimension est représenté. La latence relative de décharge des neurones représente l'intensité du stimulus en chaque point. Un processus similaire a lieu dans la rétine quand une image est présentée : en chaque point de l'image, la latence de décharge des neurones dépend du contraste local dans l'image. Ce type d'information peut ensuite être intégré par des neurones cibles (non représentés). Adapté de Thorpe et Gautrais (1998).

Dans la seconde partie, je vais tenter de construire des modèles de propagation des décharges neuronales dans le système visuel compatibles avec ces contraintes. Je m'attacherai à la fois à réaliser des modèles biologiquement plausibles mais également computationnellement efficaces. Même si beaucoup de modèles du système visuel existent, très peu semblent pouvoir approcher les performances de celui-ci en terme de traitement d'images. Les 4 contraintes que j'ai indiquées sont excessivement restrictives et on voit mal comment on pourrait construire un modèle du système visuel efficace quand on sait quelles difficultés rencontrent les algorithmes de traitement d'images mathématiques et symboliques en intelligence artificielle. Cependant nous verrons que, malgré ces contraintes, l'efficacité des modèles que je vais présenter est redoutable, notamment en ce qui concerne la reconnaissance d'objets. Nous verrons comment, en implémentant des règles simples issues de la biologie, il est possible de développer un modèle artificiel du système visuel capable d'effectuer des tâches très complexes qui rivalisent avec les meilleurs systèmes experts de traitement d'image. Il semble que des milliers d'années d'évolution ont permis de contraindre et d'optimiser la rapidité et la précision de notre système visuel.

III

Modèles computationnels du traitement visuel rapide

Once I was a psychologist, with my own brain-centered view of the universe. Then, I was selected for the ACCN 1999 and I got the real scope of Neuroscience: to make converge, one day in the future, the Kandel & Schwartz "Principle of Neural Science" and the Matlab User's Guide into a single, bloody, book. I mean, human brain will have to be the best representation of Matlab or, at least, of one of its toolboxes!!!

Stefano Baldassi (1999)

Avant de présenter mon travail de modélisation, il convient de justifier mes choix en termes de dynamique des neurones simulés. Plusieurs niveaux de simulation sont possibles : principalement le niveau des neurones à décharge et celui des neurones à activation continue. Dans le cas de neurones à décharge, les potentiels d'action sont explicitement simulés. Il s'agit donc de simulations relativement proches du comportement des neurones réels. Pour les neurones à activation continue, plus classiques, on fait abstraction des décharges des neurones pour ne prendre en compte que leur niveau d'activation moyen - qui correspond en général à la fréquence de décharge de ces neurones. Bien que la première approche semble plus plausible du point de vue biologique, c'est la seconde qui a connu le plus grand succès. Les modèles à activation continue, plus facile à maîtriser en ce qui concerne leurs propriétés mathématiques, permettent de rendre compte de la majorité des comportements des réseaux de neurones réels, comme la sélectivité à l'orientation dans le cortex visuel primaire (Suarez et al, 1995; Adorjan et al, 1999; Carandini et al, 1997) et la sélectivité aux objets dans IT (Walis, 1994; Mel, 1997; Riesenhuber et Poggio, 1999). Il est possible de plus de construire des réseaux aux propriétés étonnantes, de mémorisation de formes ou de séquences et d'auto-organisation topologique (Kohonen, 1982; Durbin et Mitchison, 1990; Walis, 1994). Cependant, si les neurones présentent des niveaux d'activation continus, la dynamique du réseau reste globale et itérative. Le changement de niveau d'activation d'un neurone est automatiquement répercuté aux neurones sur lesquels il se projette ce qui, dans le cas de connexions réciproques, module en retour l'activité du neurone qui les a influencées. Ces processus de feedback pratiquement instantanés sont peu réalistes du point de vue de la biologie. Il paraît également difficile de simuler des réseaux *feedforward* à l'aide de ces

neurones, à moins qu'aucune connexion réciproque ne soit présente ce qui limiterait considérablement la dynamique de tels réseaux.

Les réseaux à décharge ne possèdent pas de telles limitations. Des connexions réciproques entre les neurones peuvent exister, mais, du fait des décharges discrètes des neurones, il est possible de limiter leur effet à une propagation *feedforward*. Bien que les réseaux classiques de neurones à activation continue soient encore très répandus, la récente découverte d'un type de plasticité synaptique va tendre à les marginaliser. Dans les réseaux classiques, la règle d'apprentissage synaptique est en général basée sur la décharge conjointe des neurones pré et postsynaptiques. Markram et al (1997) ont récemment montré que la plasticité synaptique ne pouvait se résumer à ce seul phénomène : elle semble également dépendre de la date exacte de décharge du neurone afférent par rapport au neurone efférent. Si le neurone afférent décharge avant le neurone efférent la connexion entre les deux neurones est renforcée et dans le cas contraire, elle est affaiblie. Étant donné cette nouvelle contrainte, il semble difficile de s'affranchir des décharges des neurones et donc d'utiliser des neurones à niveau d'activation continu.

Je me suis donc placé au niveau qui me paraissait le plus pertinent du point de vue de la simulation de neurones réels et de la propagation rapide, à savoir l'utilisation de neurones à décharge. De plus, comme on l'a vu dans la partie précédente, il paraît difficile d'implémenter un codage de population basé sur les décharges relatives des neurones dans un réseau où les neurones ont des niveaux d'activation continus. J'ai également tenté de me limiter à des processus *feedforward* : dans tous les modèles que je vais présenter, les neurones ne peuvent pas décharger plus d'une fois, de sorte que les processus itératifs de *feedback* sont rendus

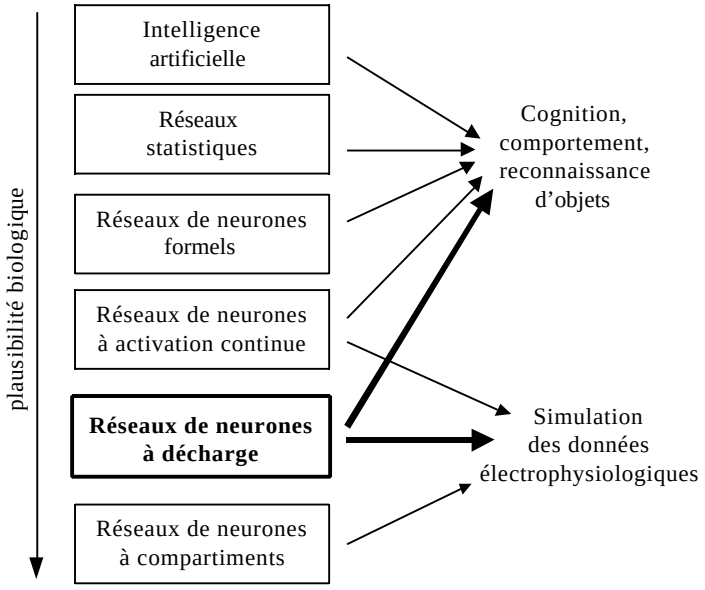


Figure 1 : hiérarchie des niveaux de simulation (à gauche) et correspondance avec le type de donnée que l'on tente de simuler (à droite). Le niveau de simulation que je me suis fixé est indiqué en gras. À ma connaissance, aucun réseau de neurones à décharge n'a jamais tenté d'atteindre un niveau comportemental.

totalement impossibles. Cette limitation, introduite volontairement, ne signifie pas que les neurones dans le système visuel se comportent de cette façon. Elle a pour unique but de montrer que le traitement de type *feedback* n'est pas indispensable pour que le système visuel puisse effectuer des opérations complexes. Parmi les différents types de neurones à décharge, à l'exception du premier modèle, plus détaillé, j'ai utilisé les neurones les plus simples possibles : des neurones "intègre et décharge" (IF) qui intègrent linéairement les décharges des neurones afférents et émettent un potentiel d'action dès qu'ils atteignent un certain seuil (annexe 2). Dans l'annexe 2, j'introduis un logiciel - SpikeNET - que j'ai programmé[1] pour pouvoir simuler un grand nombre de neurones de ce type et réaliser ainsi des traitements sur des images naturelles. J'indique également dans quelle mesure le comportement des neurones IFs est proche de celui des neurones réels et de celui de neurones modélisés de façon beaucoup plus détaillée comme ceux de Hodgkin et Huxley (1952). La figure 1 indique l'originalité de notre approche : en général, les réseaux à décharge sont utilisés pour rendre compte du comportement des neurones réels et ils ne sont pas utilisés pour réaliser des traitements plus complexes et tenter d'atteindre un niveau comportemental.

Dans cette partie, j'ai tenté de suivre une démarche ascendante : je commence par construire un modèle de sélectivité à l'orientation dans V1 dont la dynamique rapide permet de rendre compte de la sélectivité précoce de ces neurones. Dans un second modèle, j'introduis une règle d'apprentissage de type biologique et j'analyse dans quelle mesure la sélectivité des neurones de V1 est contrainte par cette règle et par les caractéristiques des stimuli présentés. Dans un dernier modèle, je tente, en me basant sur les mêmes principes, d'atteindre un niveau comportemental, en particulier celui de la catégorisation et de la reconnaissance d'objets dans les scènes naturelles. Je discute enfin les relations que ces modèles entretiennent avec d'autres modèles de traitement dans le système visuel et je tente d'inférer un modèle général de reconnaissance d'objets utilisant SpikeNET.

[1] SpikeNET est un terme générique, imaginé par Simon Thorpe, qui désigne un concept plutôt qu'un logiciel. L'idée est que les calculs neuronaux peuvent s'effectuer sur la base de l'organisation temporelle des décharges des neurones. Il existe plusieurs versions de SpikeNET et j'ai programmé ma propre version.

1

Sélectivité ultra-rapide à l'orientation dans V1

La sélectivité à l'orientation dans le cortex visuel primaire est le sujet de nombreuses études, tant en électrophysiologie qu'en modélisation, mais les mécanismes sous-jacents sont encore très controversés (Vidyasagar et al, 1996; Sompolinsky et Shapley, 1997). Certains auteurs prétendent, suivant par là l'idée originale de Hubel et Wiesel (1962), que les excitations en provenance du LGN sont suffisantes pour rendre compte de la sélectivité des neurones à l'orientation. Cependant ces modèles – dans leur version linéaire – ne permettent pas d'expliquer certaines propriétés des neurones comme la constance de la largeur de la courbe de sélectivité par rapport aux variations de contraste des stimuli. Dans ce type de modèles linéaires, l'activité des neurones étant proportionnelle au contraste, la largeur de la courbe de sélectivité des neurones dépend donc du contraste. D'autres modèles ont donc été proposés pour rendre compte de ces propriétés. La majorité de ces modèles se base sur des processus itératifs locaux qui permettraient de normaliser l'activité neuronale, c'est-à-dire en quelque sorte de rééquilibrer l'activité des neurones (Suarez et al, 1995; Carandini et Ferster, 1997a; Carandini et al, 1997b; Adorjan et al, 1999).

Ces simulations ne permettent cependant pas de comprendre la sélectivité rapide à l'orientation qui interviendrait dès les premières millisecondes après la présentation d'un stimulus. Si une barre orientée est présentée pendant 10 ms immédiatement suivie d'un masque perpendiculaire, les neurones enregistrés dans V1 émettent quelques décharges sélectives à l'orientation présentée, et ceci malgré la brièveté du stimulus, (Celebrini et al, 1993). Dans ce cours laps de temps, les processus itératifs - *feedback* horizontal - entre neurones voisins n'ont pas le temps de se mettre en place. À ce jour, aucun modèle ne peut expliquer cette sélectivité rapide, pas plus que la sélectivité en termes de latence de décharge des neurones. On observe en effet que la latence de décharge des neurones dépend du stimulus présenté, la latence étant très précoce pour des stimulus correspondant à l'orientation préférée du neurone et plus tardive pour des orientations voisines (Celebrini et al, 1993).

Dans ce chapitre, je présente le modèle d'un unique neurone biologiquement réaliste intégrant les décharges en provenance de la rétine et capable de tels comportements.

1.1 - Matériel et méthodes

La simulation d'un neurone cortical unique est effectuée à l'aide du simulateur NEURON (Hines, 1989). Le neurone modélisé est un neurone pyramidal à un seul compartiment. Cela signifie que la propagation du potentiel au sein de l'arbre dendritique n'est pas prise en compte. Un modèle très détaillé de neurones corticaux montre en effet que les stimulations reçues par l'arbre dendritique ont approximativement le même effet au niveau du soma (Jaffe et Carnevale, 1999), indépendamment de la position des synapses. La modélisation d'un neurone à un seul compartiment reste donc relativement acceptable du point de vue de la biologie.

Cette simulation est basée sur un ensemble de paramètres déterminés à partir de l'enregistrement de neurones réels (tableau 1.1). Le soma contient des canaux voltage-

Paramètre	**Valeur**
Surface	15000 μm^2

Tableau 1.1 : paramètres du modèle HH (Hodgkin-Huxley) à simple compartiment. Les valeurs utilisées ici ont été adaptées à partir de divers modèles (Suarez et al, 1995; Destexhe, 1997; Carandini et al, 1997a; Mell et al, 1998; Propatias et al, 1999), en particulier pour que la fuite du neurone (τ=9,5 ms) et son impédance d'entrée (35 MΩ) restent réalistes.

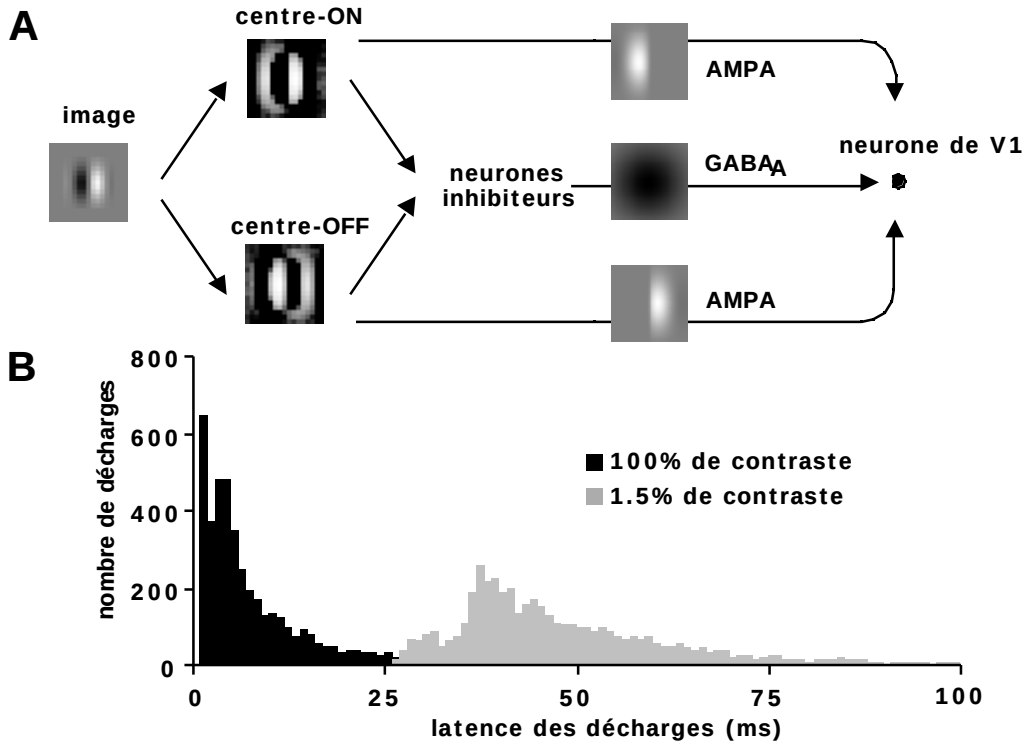


Figure 1.1 : A, image initiale contenant le stimulus, transformation en contrastes centre-ON et centre-OFF et projection en direction du neurone de V1. Pour les cellules centre-ON et centre-OFF, les niveaux de gris représentent la latence de décharge des neurones et les pixels en noir représentent les neurones non activés. La valeur des poids synaptiques est indiquée en dégradé de niveaux de gris, le gris moyen correspondant à des valeurs proches de 0, le gris clair à des poids synaptiques excitateurs AMPA et les points noirs à des poids synaptiques inhibiteurs $GABA_A$. B, latence de décharge des neurones centre-ON pour l'ensemble des 72 orientations des stimuli d'entrée à 100 % de contraste et à 1,5 % de contraste (pas de temps de 2 ms). La distribution pour les cellules centre-OFF est identique, au bruit près.

dépendants Na+ et K+ de type Hodgkin-Huxley (HH). L'activation provenant du LGN stimule l'activité de synapses excitatrices AMPA du neurone cible. Les canaux NMDA, dont la dynamique est très lente, ne sont pas pris en compte car la simulation se restreint aux quelques dizaines de millisecondes qui suivent la présentation du stimulus.

Un stimulus consiste en une image contenant un *grating*[1], à une orientation et un contraste donné[2]. Préalablement à la propagation, le stimulus est filtré par des filtres de contrastes centre-ON et centre-OFF[3]. Les valeurs d'activation résultantes sont ensuite utilisées pour

[1] Barres orientées dans une direction donnée. Les barres sont suggérées par une variation continue de luminance. Pour construire cette image, une fonction gaussienne bidimensionnelle (σ=2.3) est multipliée par une fonction sinus de période φ=25, l'unité étant le pixel.

[2] La taille de l'image est de 17x17 pixels. L'orientation varie par pas de 5° pour couvrir les 360° du cercle trigonométrique. Le contraste varie de 100 % - ce qui correspond à l'ensemble des niveaux de gris - à 1.5 %.

[3] Implémentée par une classique différence de gaussienne, la gaussienne la plus grande ayant une déviation standard correspond à 1/3 de la gaussienne centrale de signe opposé. Cette différence est normalisée afin que l'intégrale de la fonction résultante soit 0. Les contrastes centre-ON sont obtenus en appliquant directement le filtre sur l'image et les contrastes centre-OFF en appliquant le filtre sur l'image dont les pixels sont inversés.

calculer les latences de décharge des cellules ON et OFF en chaque point à l'aide d'un neurone intégrateur à fuite bruité[4]. Les latences de décharge les plus courtes correspondent donc aux zones de fort contraste – positif pour les ON et négatif pour les OFF - alors que les latences plus tardives correspondent à des niveaux de contraste moins élevés[5].

Pour chaque position "rétinienne" un stimulus ne peut induire qu'une seule décharge en provenance soit d'une cellule centre-ON, soit d'une cellule centre-OFF. Dans ces conditions, un code conventionnel basé sur la fréquence de décharge est totalement inopérant.

Chacun des neurones du LGN[6] est connecté par une seule synapse au soma du neurone cible. La répartition des poids synaptiques est déterminée à l'aide de la fonction suivante pour un neurone sélectif à des stimuli d'orientation 0° :

$$W_i = \left| \sin(A.d(r_i, n)) e^{\frac{-d(r_i,n)^2}{2\sigma^2}} \right|^+ + B$$

où W_i est le poids synaptique entre le neurone r_i centre-ON du LGN et le neurone n de V1 et $d(r_j,n)$ est la distance cartésienne séparant ces deux neurones[7]. Le paramètre B[8] a été ajouté pour vérifier que la sélectivité du neurone ne pouvait pas être due à l'activation d'une sous-population des synapses du neurone n.

Les neurones inhibiteurs ne sont pas modélisés explicitement. J'ai considéré ces neurones comme intégrant des informations à partir du LGN de façon suffisamment rapide pour que les IPSPs arrivent sur le neurone cortical avec un retard d'environ 2 à 5 ms par rapport aux EPSPs (Ferster et Jagadeesh, 1992; Hirsch et al, 1998). Cela correspond à peu près au temps que met un EPSP pour atteindre le soma, les IPSPs arrivant pour la plupart directement sur le soma ou sur la dendrite principale. Par souci de simplicité, dans le modèle, la date d'arrivée des IPSPs

[4] Le courant de fuite est fixé à 20 ms et Ll courant d'entrée dépend linéairement de la valeur d'activation résultant de l'application du filtre. Rapport signal sur bruit fixé à 1 pour le contraste le plus faible. Le bruit introduit correspond à une déviation standard de la latence des décharges des neurones de 3.6 ms au contraste le plus élevé et 12.6 ms au contraste le plus faible.

[5] Les latences des neurones sont générées à l'aide de SpikeNET (cf. annexe 2).

[6] 222 neurones, 11x11 centre-ON et 11x11 centre-OFF neurones en excluant les bords.

[7] A=0.75; B=0.1; σ=2.5, l'unité étant toujours le pixel

[8] Le paramètre B contribue pour la moitié du total des poids synaptiques du neurone cible.

correspond à celle d'arrivée des EPSPs sur le neurone cible. Un champ récepteur de synapses $GABA_A$ de répartition gaussienne a donc été défini[9] (figure 1.1).

1.2 - Résultats

Pour un niveau donné d'inhibition, les poids synaptiques excitateurs sont optimisés de telle façon que les orientations pour lesquelles le neurone décharge s'étendent sur 65° à 100 % de contraste[10].

Même en l'absence de *shunting inhibition*, le courant de fuite du neurone - reflet de sa

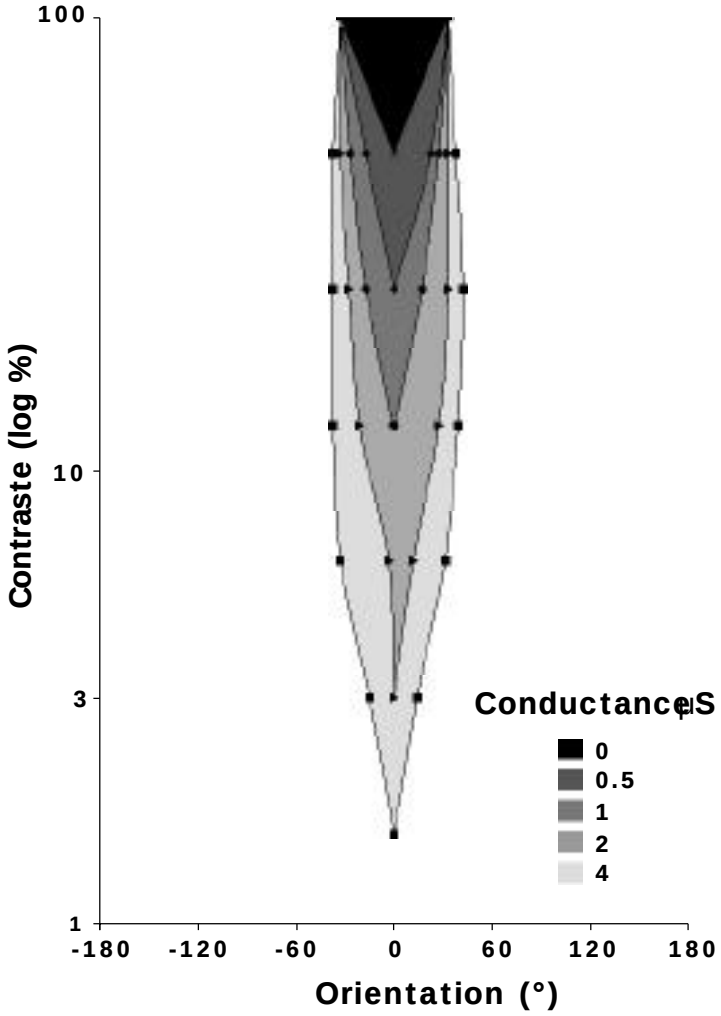


Figure 1.2 : orientations pour lesquelles le neurone décharge en fonction du contraste et de la *shunting inhibition*. Les surfaces en niveaux de gris représentent différentes conductances totales exprimées en µS (de 0 à 4 µS). La limite de ces surfaces est définie comme étant l'orientation pour laquelle le neurone décharge seulement dans la moitié des cas (en fonction du bruit). Les déviations standards pour ces points ne sont pas indiquées par souci de clarté (elles restent très faibles, de l'ordre de quelques degrés). Le niveau d'excitation du neurone est ajusté afin que la largeur de la courbe de sélectivité soit de 65° à 100 % de contraste. Sans inhibition (surface en noir), même dans le cas d'un stimulus à 50 % de contraste, le neurone n'est plus capable de décharger. Dans le cas de l'inhibition la plus forte, même avec 3 % de contraste résiduel, la largeur de la courbe de sélectivité du neurone est pratiquement identique à celle visible pour 100 % de contraste.

[9] A la fois pour les synapses excitatrices et les synapses inhibitrices, les délais de propagation synaptique ne sont pas pris en compte et la variation de la conductance est modélisée à l'aide d'une simple décroissance exponentielle.

[10] Du fait de la distribution des poids synaptiques, l'étendue de la décharge du neurone est centrée sur 0°.

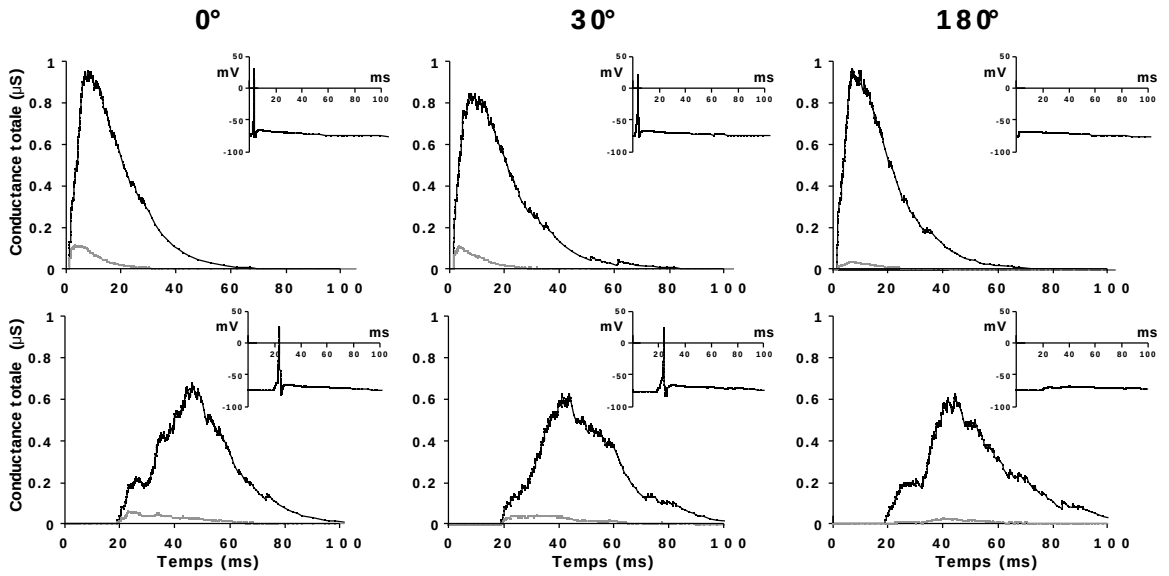


Figure 1.3 : conductances totales excitatrices (gris clair) et inhibitrices (gris foncé) du modèle d'un neurone de V1 pour trois orientations du *grating*, 0°, 30° et 180° et deux niveaux de contraste 100 % en haut et 3 % en bas. La conductance inhibitrice est identique quelle que soit l'orientation pour un niveau de contraste donné. Dans le coin supérieur droit de chaque courbe, le potentiel du neurone est représenté. Le neurone ne décharge que dans le cas de stimuli à 0° et 30° et la latence de la décharge est plus précoce à 0°. Cela signifie que la sélectivité du neurone s'exprime non seulement en termes de décharges – pas de décharge pour une orientation de 180° - mais également en termes de latence de décharge.

constante de temps membranaire[11] - lui permet d'être sélectif à l'orientation. Les EPSPs les plus amples doivent arriver suffisamment groupés pour que la décharge du neurone ne soit pas inhibée par le courant de fuite. En fixant les poids synaptiques à une valeur donnée, le neurone peut alors devenir sélectif à l'orientation sur une étendue de 65°. Cependant ce type de sélectivité est extrêmement fragile, car si le contraste du stimulus baisse, ne serait-ce que de moitié, les latences des neurones en entrée deviennent plus étalées et le neurone ne décharge plus (figure 1.2).

Cependant, si la *shunting inhibition* rapide est ajoutée au modèle et que les poids synaptiques excitateurs sont réajustés pour que le neurone conserve la même sélectivité, le neurone devient capable de répondre sur une grande plage de contrastes. Cet effet est parfaitement visible dans la figure 1.2, où, pour un niveau d'inhibition atteignant 4 µS[12], la sélectivité du neurone est pratiquement constante de 100 % de contraste à un niveau de contraste aussi faible que 3 %. Il est important de noter qu'il existe une relation linéaire[13] entre

[11] τ=9,5 ms dans notre modèle.

[12] Cette valeur représente la somme des conductances maximales de toutes les synapses inhibitrices.

[13] R=0.9996.

le total des poids synaptiques excitateurs et inhibiteurs et donc que l'équilibre entre excitation et inhibition est comparable dans tous les cas.

La sélectivité du neurone dépend principalement des latences relatives d'arrivée des entrées excitatrices par rapport à la *shunting inhibition* rapide. Si les poids synaptiques les plus forts sont activés de façon précoce, le neurone est capable de décharger avant que l'inhibition ne shunte ses entrées excitatrices (figure 1.3). Cependant, si les poids synaptiques activés de façon précoce sont de faible amplitude alors la balance entre l'excitation et l'inhibition jouera en faveur de l'inhibition et le neurone ne déchargera pas.

De plus, comme cela est visible sur la figure 1.3, la latence de décharge du neurone dépend de l'orientation du stimulus. La latence de décharge est plus faible pour l'orientation préférée du neurone et plus tardive pour des orientations voisines (figure 1.4). Ce modèle de neurone de V1 est donc sélectif en termes de latence de décharge. Des études en électrophysiologie ont montré que les neurones dans V1 déchargent de manière tout à fait similaire. Celebrini et al (1993) et Gawne et al (1996) ont tous deux montré une augmentation des latences de décharge des neurones entre leur orientation préférée et une orientation différant de 30° par rapport à celle-ci. Cela signifie également qu'à l'étape suivante du système visuel (V2 dans notre cas) les neurones peuvent également baser leur sélectivité sur la latence de décharge des neurones afférents.

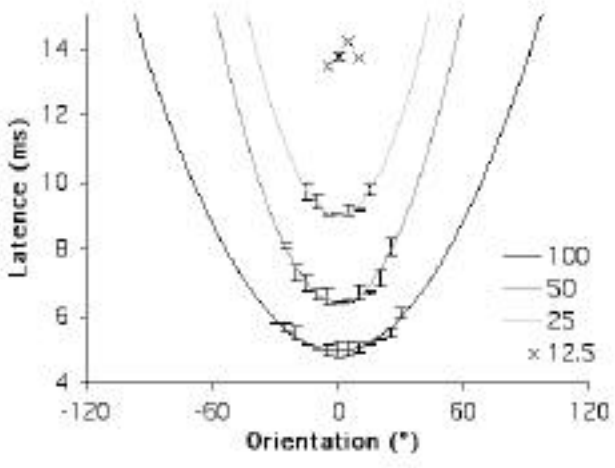


Figure 1.4 : latence de décharge du neurone dans V1 en fonction de l'orientation et du contraste du stimulus. L'inhibition est fixée à un niveau intermédiaire (somme des synapses inhibitrices équivalente à 1 µS). La déviation standard des latences de décharge est calculée pour 10 conditions de bruit des neurones centre-ON et centre-OFF. Dans tous les cas, la latence de décharge du neurone contient de l'information sur l'orientation du stimulus.

1.3 - Discussion

J'ai proposé ici un modèle hiérarchique à deux niveaux dans lequel les décharges des neurones dans le LGN excitent un neurone de V1. Je me suis focalisé en particulier sur les aspects temporels d'un tel processus - de l'ordre de quelques millisecondes – et sur

l'interaction entre les EPSPs et les IPSPs au sein de ce neurone de V1. J'ai proposé qu'une *shunting inhibition* rapide rende compte de l'invariance de la sélectivité des neurones de V1 par rapport au contraste du stimulus. Étant donné que l'inhibition dépend du nombre de décharges émises dans le LGN, elle filtre les excitations en fonction de leur ordre d'arrivée. Ce résultat démontre que, même dans un modèle biologique, la sélectivité à l'ordre des décharges des neurones afférents engendrée par la *shunting inhibition* fournit un mécanisme plausible de la sélectivité rapide à l'orientation invariante au contraste.

1.3.1 - Autres modèles de sélectivité à l'orientation

Les mécanismes de la sélectivité à l'orientation sont étudiés depuis maintenant plusieurs dizaines d'années mais le sujet reste encore très controversé (Vidyasagar et al, 1996; Sompolinsky et al, 1997). Pour Hubel et Wiesel (1962), la connectivité entre les cellules du LGN et les neurones de V1 suffirait à expliquer la sélectivité à l'orientation : des connexions excitatrices et inhibitrices orientées rendant le neurone sélectif à une orientation donnée. Cependant, il est maintenant avéré que l'inhibition intra-corticale joue également un rôle dans la sélectivité à l'orientation car la suppression des effets inhibiteurs par l'application d'un bloqueur des canaux GABAs, comme la bicuculine, fait disparaître ou atténue la sélectivité à l'orientation (Sillito, 1975; Nelson et al, 1994; Crook et al, 1997). De plus Hirsch et al (1998) ont montré que, si l'on présente dans le champ visuel des orientations perpendiculaires à l'orientation préférée du neurone, la diminution du nombre d'excitations en provenance du LGN est très faible : la connectivité du neurone ne peut donc pas rendre compte à elle seule de sa sélectivité à l'orientation.

La plupart des modèles de sélectivité à l'orientation postulent que seule la fréquence de décharge des neurones est importante. La comparaison avec notre modèle n'est donc pas aisée. Même dans le cas où les auteurs utilisent des neurones émettant des décharges, l'activité de ces neurones est simulée par un processus de Poisson (Maex et Orban, 1996), ce qui signifie que les latences de décharge des neurones sont décorrélées. Les simulations de ce type ont cependant montré que les modèles linéaires du type Hubel et Wiesel ne sont pas compatibles avec le comportement des neurones réels. En particulier ce modèle linéaire ne peut pas rendre compte de la réponse des neurones en présence de *grating* superposés ou de la dynamique complexe d'adaptation au contraste observée en psychophysique (Lorenceau, 1987). Pour rendre compte de ces phénomènes, les auteurs supposent qu'une amplification corticale (Suarez et al, 1995; Adorjan et al, 1999) et/ou une renormalisation (Carandini et al,

1997a; Carandini et al, 1997b) suive l'intégration linéaire initiale. D'autres modèles tentent de simuler de façon réaliste la dynamique d'activation des neurones de V1, et en particulier leurs interactions latérales (Seriès et al, 1999; Seriès et al, 2000), afin d'expliquer des résultats psychophysiques (Lorenceau, 1998). Cependant le problème avec toutes ces études est qu'elles ne prennent pas en compte la sélectivité à l'orientation et l'invariance au contraste dans la phase initiale de décharge des neurones.

Le mécanisme que je propose ici réconcilie les modèles linéaires et non linéaires. J'ai montré que la seule présence de connexions excitatrices suffit à obtenir une sélectivité à l'orientation mais que l'inhibition est nécessaire pour stabiliser ce type de sélectivité et la rendre invariante au contraste. Ferster et Jagadeesh (1992) ont en effet montré que l'inactivation de l'inhibition sur un seul neurone n'induisait pas une totale perte de sélectivité à l'orientation de ce neurone. Il aurait cependant été intéressant de faire varier le contraste du stimulus dans cette expérience. Nos résultats laissent penser que la sélectivité à l'orientation du neurone et principalement l'étendue des orientations sur lesquelles le neurone décharge de façon précoce auraient été grandement affectées.

1.3.2 - Pertinence biologique du modèle

Les décharges des neurones dans le modèle que j'ai présenté sont similaires à celles des neurones réels dans le LGN (Reich et al, 1997; Reich et al, 1998). Pour simuler de façon efficace les décharges de ces neurones, il semble que des modèles basés sur la fréquence de décharge se révèlent insuffisants et qu'il soit nécessaire d'utiliser un modèle d'intégrateur à fuite bruité comme nous l'avons fait. Le bruit introduit dans le modèle induisait des variations - compatibles avec les données électrophysiologiques (Reich et al, 1997) - d'environ 4 ms sur les décharges des neurones. Les différences entre la distribution des décharges des neurones du modèle et les données réelles sont principalement dues au fait que, dans le modèle, seule la première décharge des neurones est considérée.

Il a été montré par ailleurs que les neurones du LGN peuvent exciter directement les cellules inhibitrices présentes dans V1 (Sillito, 1975; Freund et al, 1989) et que des IPSPs peuvent arriver sur un neurone de V1 avec seulement 2 à 5 ms de retard par rapport aux premiers EPSPs induits par un flash lumineux (Hirsch et al, 1998). La présence d'une inhibition très rapide est principalement due au fait que les synapses excitatrices en provenance du LGN se forment principalement sur le soma de ces neurones inhibiteurs et les conduit donc à décharger très rapidement (McCormick et al, 1985). Les neurones excitateurs,

quant à eux, présentent une distribution différente de leurs synapses afférentes, les synapses excitatrices étant localisées principalement sur les dendrites distales du neurone et les synapses inhibitrices étant localisées au niveau du soma ou sur la dendrite principale des neurones. Les EPSPs doivent donc se propager le long de l'arbre dendritique pour atteindre le soma du neurone. Cela signifie que, *in vivo* dans V1, à la suite d'un flash, les IPSPs induits par des neurones corticaux inhibiteurs peuvent parfois masquer les EPSPs en provenance du LGN (Volgushev et al, 1993; Hirsch et al, 1998). Dans le modèle que j'ai présenté, j'ai donc pris le parti de faire arriver les EPSPs et les IPSPs simultanément au niveau du soma du neurone.

Notre modèle est similaire en bien des points au modèle de Shadlen et Newsome (Shadlen et Newsome, 1998). Ce modèle montre que, si l'entrée des neurones est totalement décorrélée dans le temps (bruit de Poisson), le décours des décharges des neurones ne contient aucune information. Les auteurs concluent que le seul rôle que l'on pourrait éventuellement attribuer à des neurones IF serait celui de détecteur de coïncidence, c'est-à-dire d'unité détectant des décharges synchronisées de leurs afférents. Cependant, j'ai déjà indiqué que la simulation des décharges des neurones dans le LGN par un processus de Poisson semblait être une simplification abusive (Reich et al, 1997; Reich et al, 1998). En introduisant une dynamique plus complexe, nous avons démontré la possibilité pour les neurones corticaux d'utiliser la *shunting inhibition* pour décoder l'activation temporelle de leurs afférents.

Bien que nous n'ayons pas exploré l'espace des paramètres en détail, la transformation du modèle de neurone HH en neurone IF ou la modification des paramètres membranaires n'induit pas de changements de comportement drastiques[14].

1.3.3 - La *shunting inhibition* rapide

La plupart des études publiées sur le sujet de la *shunting inhibition* sont directement ou indirectement en accord avec notre modèle. Je vais commencer par décrire les études qui renforcent les résultats que nous avons obtenus. Par la suite, je montrerai que même les études *a priori* contradictoires avec nos résultats ne sont pas si incompatibles qu'on pourrait le penser.

Des enregistrements *in vitro* de neurones corticaux montrent que, suite à la présentation d'un stimulus, ces neurones sont capables de produire des décharges très reproductibles

[14] J'obtiens des résultats comparables avec un modèle de neurone dans V1 présentant une surface totale plus réaliste S=4440 μm2; g_{NA}=450 nS; g_K=50 nS. Cependant, dans ce cas, la constante de temps membranaire et l'impédance d'entrée du neurone s'éloignent des données électrophysiologiques. Un modèle IF présente également un comportement similaire.

similaires à celles que l'on obtient dans notre modèle figure 1.4 (Mainen et Sejnowski, 1995). Les réponses des neurones dans V1 sont particulièrement reproductibles dans le cas de stimuli flashés alors qu'elles le sont moins pour des stimuli statiques (Mechler et al, 1998). En accord avec notre modèle, les auteurs soutiennent qu'un système de gain en fonction du contraste très rapide doit intervenir pour rendre compte de leurs résultats.

D'autre part, suite à un stimulus flashé, l'inhibition *feedforward* est très rapide et peut comme je l'ai déjà mentionné affecter le neurone avant qu'il ne décharge (Gabbott et al, 1988; Celebrini et al, 1993; Volgushev et al, 1995). Hirsch et al (1998) ont effectué des enregistrements intracellulaires de neurones dans V1 tout en présentant sur la rétine des stimuli flashés - des carrés noirs et blancs de différentes tailles et à différentes positions. Ils ont montré que pour un carré blanc, interagissant à la fois avec le champ récepteur ON et le champ récepteur OFF du neurone, les premières millisecondes d'intégration combinaient une composante excitatrice et une composante inhibitrice. Pour des stimuli non optimaux - un carré blanc dans la zone centre-OFF du neurone de V1 - les effets étaient une hyperpolarisation du neurone - c'est-à-dire une inhibition - et avaient lieu à des latences comparables aux effets excitateurs. Des effets similaires ont été obtenus avec des barres orientées flashées très rapidement (Volgushev et al, 1993), contrairement à notre modèle l'inhibition ne semblait pas dépendre de l'orientation du stimulus. Enfin, Celebrini et al (1993) ont montré que, pour un stimulus non optimal d'un neurone de V1 présenté pendant seulement 10 ms et suivi d'un masque, on observait une diminution de l'activité spontanée du neurone. Cette diminution pourrait être attribuée à des effets inhibiteurs très rapides.

Il existe en fait deux types d'inhibition, la première divisive - *shunting inhibition* - et la seconde soustractive par rapport aux courants excitateurs entrants. Ces deux types d'inhibition font en fait référence au même processus que l'on appelle en général IPSP. Lors de l'ouverture des canaux ioniques, c'est-à-dire pendant quelques millisecondes, l'inhibition est divisive - *shunting inhibition* - car la conductance de la membrane augmente. Cependant après la fermeture de ces canaux, l'inhibition n'est plus divisive et se propage sur l'ensemble du neurone du fait de la capacitance membranaire. Des expériences récentes ont montré que, suite à un flash, la conductance membranaire montait très rapidement - en quelques millisecondes - pour atteindre jusqu'à 3 fois la conductance de la cellule au repos (Berman et al, 1991; Celebrini et al, 1991; Borg-Graham et al, 1998). En total accord avec notre modèle, les auteurs interprètent ces changements rapides de conductance en termes de *shunting inhibition*.

Après avoir tenté de justifier notre modèle du point de vue de la biologie, nous allons maintenant analyser les conséquences des résultats que nous avons obtenus en terme de codage neuronal.

1.4 - Codage par ordre

Dans les paragraphes précédents, j'ai montré que la dynamique rapide de l'inhibition pouvait contrebalancer l'excitation et rendre le neurone sélectif à l'ordre d'activation de ses synapses afférentes. De cette façon, l'étendue de décharge du neurone n'est pas affectée par des variations de contraste.

Le codage par ordre est en fait un type de codage dans lequel le neurone est sélectif à l'ordre d'arrivée des décharges afférentes et le modèle précédent implémente donc exactement cette propriété. L'unique postulat du codage par ordre est que pour un neurone au repos qui reçoit une stimulation, l'effet inhibiteur augmente graduellement, interférant de plus en plus avec l'intégration des EPSPs. Les premiers EPSPs sont donc moins affectés par l'inhibition que les EPSPs tardifs.

1.4.1 - Les propriétés du codage par ordre

Le codage par ordre a avant tout une inspiration biologique. J'ai déjà passé en revue la majorité des raisons pour lesquelles une *shunting inhibition* rapide, provoquée par l'apparition d'un stimulus, est très plausible du point de vue électrophysiologique. J'ai également présenté dans la première partie, les données expérimentales en faveur d'un tel type de codage : la rapidité du traitement visuel est telle qu'à chaque étape du système visuel, chaque neurone n'a *a priori* pas le temps de décharger plus d'une fois.

Concernant l'aspect computationnel, le codage par ordre présente nombre de propriétés qui en font l'attrait (Thorpe, 1990; Thorpe et Gautrais, 1998). Un premier avantage est que, à la différence d'un code basé uniquement sur la latence de décharge des neurones (Hopfield, 1995), le codage par ordre fournit une normalisation automatique de l'activation des neurones. Seul l'ordre d'activation des afférents est pris en compte, de sorte que leur latence absolue est sans importance. Bien que des modifications d'intensité ou de contraste du stimulus provoquent des changements radicaux dans la latence de décharge des neurones, l'ordre dans lequel les neurones déchargent est très robuste à de telles variations.

Le second avantage du codage par ordre est qu'il est très facile de rendre un neurone sélectif à l'ordre dans lequel ses afférents déchargent. En principe, il est seulement nécessaire

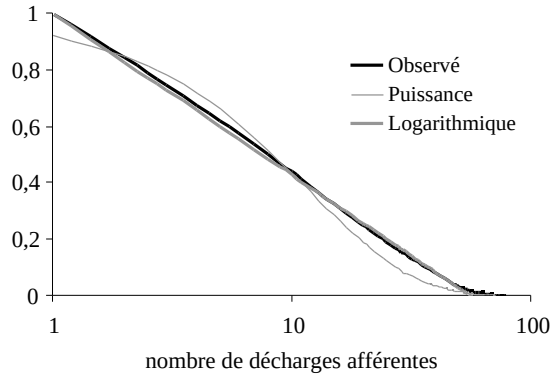


Figure 1.5 : différents types de modulation. En noir, la modulation nécessaire pour récupérer - de façon statistique - les informations de contraste. Un neurone possédant un champ récepteur de taille 11x11 observe les décharges des cellules centre-ON et centre-OFF (filtre de 3x3, différence de gaussienne normalisée à 0). Ce graphique présente les résultats pour la propagation d'environ 1 000 000 d'images naturelles 11x11 tirées de photographies commerciales. L'adéquation avec une courbe logarithmique est très bonne. Dans les simulations qui suivront nous avons préféré une loi puissance, qui reste relativement proche, du fait de la possibilité de calcul itératifs très rapides.

d'introduire un mécanisme de désensibilisation qui réduit la sensibilité du neurone à chaque fois qu'il est excité par un autre neurone. C'est donc uniquement dans le cas où ses neurones afférents déchargent dans l'ordre des poids synaptiques que le neurone cible est activé de façon maximale (figure 1.5).

Plus précisément soit $A = \{ a_1 , a_2 , a_3 \ldots a_{m-1} , a_m \}$ l'ensemble des afférents du neurone i, avec $W = \{ w_{1,i} , w_{2,i} , w_{3,i} \ldots w_{m-1,i} , w_{m,i} \}$ les poids synaptiques correspondant aux m neurones afférents et soit f une fonction arbitraire décroissante. Le niveau d'activation du neurone i au temps t est donné par

$$Activation(i,t) = \sum_{j\ [1,m]} f(ordre(a_i))\, w_{j,i}$$

où $ordre(a_j)$ est l'ordre de décharge du neurone a_j au sein de l'ensemble A[15]. Le neurone i déchargera au temps t si son niveau d'activation dépasse un certain seuil. L'analyse d'images naturelles a montré que la fonction f qui permettait de récupérer une partie des informations de contraste - à un ordre moyen, on peut assigner un contraste moyen - était une fonction logarithmique[16]. Dans les simulations qui suivront j'ai cependant préféré utiliser une fonction puissance du fait de ses propriétés algorithmiques[17]. De plus, une fonction logarithmique est relativement cohérente du point de vue de l'hypothèse de la *shunting inhibition*. Suite à une stimulation, la distribution de l'activation des neurones dans le LGN est décroissante de façon

[15] $f(ordre(a_j)) = 0$ si le neurone a_j n'a pas encore déchargé au temps t.

[16] $f(ordre(a_i))=1-log(ordre(a_i))*B$; B étant une constante.

[17] Il est en effet possible de calculer une fonction puissance de façon itérative, en multipliant la modulation courante par un facteur de désensibilisation à chaque fois que le neurone est excité par un neurone afférent.

grossièrement exponentielle (i.e. Gazeres et al, 1998). Il est raisonnable de penser que la distribution dans le temps de l'activation des neurones inhibiteurs serait de forme comparable. Si tel est le cas alors la modulation des neurones excitateurs suivrait une dynamique de type logarithmique. Ceci ne constitue absolument pas une démonstration, mais indique simplement qu'une modulation de type logarithmique s'accorde avec une vue intuitive de la dynamique d'activation des neurones dans le LGN et dans V1. Dans le modèle que j'ai présenté précédemment, l'inhibition était proportionnelle au nombre d'EPSPs sur le neurone cible, ce qui produit une inhibition excessivement forte et donc probablement exagérée. J'ai cependant choisi cette option afin de simplifier le modèle, chaque décharge dans la rétine évoquant à la fois un EPSP et un IPSP.

Malgré les arguments que j'ai pu donner, le codage par ordre est loin d'être accepté par la communauté scientifique. Les détracteurs d'un codage par ordre soutiennent que l'activité des neurones au sein de la population pourrait être utilisée pour calculer un ordre de décharge moyen pour des fenêtres temporelles très petites. Certains ont effectivement choisi cette voie (Gerstner, 2000) mais semblent avoir besoin d'un nombre de neurones redondants très important (Gautrais et Thorpe, 1998). Dans un codage par ordre, il est en effet possible de transmettre de grandes quantités d'informations pour un nombre de neurones réduit : une population de n neurones peut discriminer jusqu'à $n!$ différentes activations en entrée alors que ce chiffre n'est que de $n+1$ pour un codage en fréquence qui consisterait à compter le nombre de neurones activés. Il est cependant possible, et même probable, que plus tard dans le traitement d'un stimulus flashé, les neurones utilisent effectivement un codage par fréquence. Ces deux codes ne sont en fait pas contradictoires. Un moyen de calculer la fréquence d'un neurone à la suite d'une stimulation est de pondérer la fréquence en fonction du temps, ce qui signifie que l'on attribue plus d'importance aux décharges arrivant quelques millisecondes après la stimulation qu'aux décharges plus tardives (Theunissen et Miller, 1995). Dans le cas d'une population de neurones, il est possible d'appliquer la même technique mais en utilisant des fenêtres temporelles encore plus courtes. Ce type de calcul se rapproche du codage par ordre de décharge. La différence réside dans l'importance donnée aux latences exactes des décharges neuronales plutôt qu'à leur ordre.

Je tiens à souligner ici que les termes de "codage" et "d'information" que j'ai employés ne sont que des raccourcis syntaxiques pour décrire le comportement des neurones et non pas les dogmes associé généralement à ces mots. Comme je l'ai souligné dans l'introduction, les neurones ne seraient pas des unités optimisées pour l'encodage et le décodage de l'information. Ils réagiraient plutôt à leur environnement pour répondre en adéquation avec le

domaine au sein duquel ils évoluent. On peut cependant penser que la nature, en termes de minimisation de l'énergie produite et consommée, a su privilégier les mécanismes optimaux de ce point de vue. A cet égard, le codage par ordre est très clairement supérieur au codage en fréquence. On pourrait penser que la force du codage en fréquence réside dans sa redondance et sa résistance au bruit mais on verra que le codage par ordre offre également une exceptionnelle résistance au bruit.

Je vais maintenant passer en revue les autres types de codages neuronaux présents dans la littérature et tenter de déterminer leurs relations avec le codage par ordre.

1.4.2 - Codage par ordre et *bursts*

Je tiens à résoudre tout de suite le cas des *bursts* ou bouffées de décharge des neurones dans le système visuel. Les preuves s'accumulent en faveur des décharges en *burst* qui seraient beaucoup plus fiables pour rendre compte du stimulus que les décharges isolées (pour une revue, cf. Lisman, 1997). Dans le cortex visuel, les neurones émettant des *bursts* permettent clairement d'indiquer la position du stimulus alors que les décharges individuelles des neurones ne le permettent pas (Livingstone et al, 1996). On admet généralement que les *bursts* dans le LGN contiennent 3 à 5 fois plus d'informations sur le stimulus que les décharges individuelles ou qu'un train de décharge continu (Reinagel et al, 1999).

Du point de vue du neurone efférent, cela est cohérent avec le fait que la probabilité de libération d'une vésicule[18] au niveau d'une synapse est une fonction qui dépend du nombre de potentiels d'action ayant excité cette synapse sur un court laps de temps. Suite à un unique potentiel d'action, de nombreuses synapses ne libèrent aucune vésicule[19] et l'occurrence d'un *burst* permettrait donc de rendre plus fiable la transmission du message neuronal. Dans un modèle plus complexe, il serait tout à fait possible d'introduire des *bursts* et des synapses dynamiques, la latence du *burst* remplaçant alors la latence d'une décharge unique. Bien que des simulations soient nécessaires, je ne pense pas que cela dénaturerait le traitement effectué ni n'altérerait de façon dramatique les performances des réseaux. Le fait qu'un seul potentiel d'action, ou qu'un *burst* soit nécessaire est en fait sans grande conséquence pour notre modèle. En effet, comme je l'ai montré, à la fois dans ce modèle et dans ceux que je vais présenter par la suite, les résultats obtenus se basent sur la latence relative des "événement neuronaux" qui

[18] Bien que les manuels de physiologie représentent couramment une synapse comme libérant plusieurs vésicules suite à un potentiel d'action, il est tout à fait possible qu'elle n'en libère qu'une seule.

[19] Cependant un neurone afférent faisant contact en plusieurs points - une dizaine - avec un neurone efférent, la probabilité qu'aucune vésicule ne libère de neurotransmetteur est très faible.

encodent le stimulus. Que ces événements représentent la latence d'une unique décharge ou d'un *burst* a donc relativement peu de conséquence[20].

Cette hypothèse "d'événement neuronaux" semble de plus cohérente avec ce que l'on sait des *bursts*. Il a en effet été montré que seule la latence de la première décharge d'un *burst* était potentiellement importante et que le décours temporel ou le nombre de décharges au sein du *burst* encodait très peu d'informations sur le stimulus (Berry et al, 1997; Reinagel et al, 1999). De plus au niveau postsynaptique, il semble que la première décharge soit de la plus grande importance pour le traitement neuronal. Les synapses, notamment excitatrices, ont dans leur grande majorité tendance à s'inactiver très vite si le neurone afférent décharge à haute fréquence (Markram et Tsodyks, 1996). Cela signifie, que seules les premières décharges afférentes sur une synapse sont prises en compte. Dans le cas où les *bursts* seraient l'un des modes de communication privilégié des neurones, l'approximation que je fais ici en assimilant un *burst* à une décharge unique ne serait donc pas dépourvue de sens.

1.4.3 - Codage par différence de phases

De rares modèles basés sur la date exacte de décharge des neurones (Fukai, 1995; Hopfield, 1995; Maass et Natschläger, 1997; Ruf et Schmitt, 1997) sont également présents dans la littérature. Ces modèles sont cependant relativement théoriques. Leur point commun est d'utiliser, plutôt que la latence, la phase de la décharge neuronale par rapport à une oscillation. Pour décoder ces informations, certains utilisent des lignes à délais afin que les décharges au sein de la population de neurones, initialement décalées comme dans notre modèle, arrivent de façon synchrone sur le neurone cible. On peut alors assimiler cette approche à un codage par synchronie comme nous allons le voir après (Hopfield, 1995). Nakamura (1998) utilise une compétition entre les latences les plus rapides comparable à la nôtre et renforce les synapses activées de façon précoce. Il ne lui manque que la *shunting inhibition* pour avoir un codage par ordre. Nous verrons cependant dans la partie sur la reconnaissance des objets, qu'un tel code est relativement efficace et que la *shunting inhibition* n'est pas toujours nécessaire.

Le point commun de ces rares modèles est qu'ils n'ont pas su utiliser les règles d'apprentissage de type biologique pour rendre compte de la sélectivité des neurones. A ma connaissance aucun n'utilise de règle biologique du type de celle que je vais présenter dans le

[20] Dans le cas de *burst*, il n'aurait cependant plus été possible d'ignorer la dynamique des synapses en réponse à plusieurs décharges afférentes et le modèle aurait dû contenir des paramètres supplémentaires.

chapitre suivant pour renforcer les poids synaptiques. Aucun de ces modèles n'a non plus vocation à effectuer des simulations à grande échelle pour la reconnaissance des objets.

1.4.4 - Autres codes temporels

Je ne souhaite pas traiter le problème des codes neuronaux qui seraient basés sur les décharges multiples d'un même neurone, la date exacte de ses décharges, les unes par rapport aux autres, encodant une partie de l'information. Des analyses de résultats expérimentaux assez convaincants ont en effet montré que l'occurrence de triplets de décharge était corrélée avec celle de doublets (Lestienne et Strehler, 1988), et pouvaient donc difficilement constituer un code neuronal. De plus, il est possible d'expliquer les intervalles de décharge, ainsi que l'occurrence de doublets et de triplets dans le LGN et dans V1, sur la seule base de la fréquence de décharge d'un neurone (Oram et al, 1999).

Je souhaite plutôt me pencher sur la synchronie des décharges qui est le code couramment accepté dans le domaine des neurosciences computationnelles. Un nombre impressionnant de résultats expérimentaux concernent la synchronisation des réponses neuronales qui interviendrait dans le regroupement à la fois des contours d'un même stimulus (Gray et al, 1989) et des différentes caractéristiques d'un stimulus comme son mouvement et son contour (Engel et al, 1991)[21]. L'équipe de Singer dont sont issues les deux publications précédentes est pionnière dans ce domaine et montre que si deux neurones dans des aires corticales proches ou éloignées déchargent en même temps, c'est qu'ils traitent du même objet. De nombreux modèles utilisent effectivement ces propriétés dans des réseaux de reconnaissance d'objets (Hummel et Biederman, 1992; Hummel et Stankiewicz, 1996; Opara et al, 1996).

Sans entrer dans le détail de cette théorie et de ces modèles, je souhaite indiquer pourquoi les résultats expérimentaux ne s'opposent pas à un codage par latence ou par ordre. En fait la synchronie des décharges de deux neurones n'est jamais parfaite. Il existe toujours quelques millisecondes de décalage entre ces deux décharges (Alonso et al, 1996; Engel et al, 1991; Gray et al, 1989). Cependant, l'interprétation en termes de bruit dans la synchronisation des décharges n'est pas la seule possible. Avant les travaux de Singer, cette différence était interprétée en termes de connectivité et de distance, la synchronie diminuant avec la distance et les neurones activés tardivement recevant des entrées des neurones activés de façon précoce

[21] Ces premières expériences ont été réalisées chez le chat anesthésié. Il y a cependant des raisons de croire que l'anesthésique provoque des oscillations en lui-même. Cependant certains de ces résultats ont été reproduits chez le singe vigile et sont donc plus convaincants.

(Arnett, 1975). Une autre interprétation serait que ces faibles déviations implémentent un codage par ordre ou par latence. La déviation standard du déphasage entre deux décharges synchrones est d'environ 10 ms[22] (Aiple et Krüger, 1988) et parfaitement compatible avec de tels codes.

Le codage par synchronie est également très controversé. Bien que le débat soit toujours très vif, il a été montré que les synchronisations dépendent de la fréquence de décharge (Aiple et Krüger, 1988; Ghose et Freeman, 1992). Si les auteurs s'efforcent toujours de chercher des synchronisations qui ne seraient pas liées à la fréquence de décharge des neurones, le fait est que la plupart d'entre elles le sont. A ce propos, une expérience très connue de Rielhe et al (Riehle et al, 1997), semble montrer que la synchronisation de neurones dans le cortex moteur dépendrait du comportement du singe, il s'agit pourtant d'un exemple typique d'une possible interprétation erronée des résultats expérimentaux[23]. Les résultats obtenus avec d'autres espèces que les primates vont également à contresens d'un codage par synchronie. Par exemple, la complexité des oscillations de décharges neuronales chez les pigeons, loin de se simplifier, augmente en présence d'un stimulus visuel (Neuenschwander et al, 1993). L'oscillation, c'est-à-dire la synchronisation répétée, ne serait donc pas renforcée par un stimulus visuel. Chez les insectes, des oscillations sont également présentes dans différentes aires perceptives et notamment dans l'aire olfactive. La suppression de ces oscillations par des moyens artificiels n'affecte pas outre mesure la sélectivité des cellules répondant aux odeurs (MacLeod et Laurent, 1996). La synchronisation exacte des neurones ne serait donc pas obligatoirement nécessaire à la reconnaissance des stimuli.

La structure même des neurones ne semble pas en faveur d'un code par synchronie. Comme le disent si bien Ferster et Spruston (1995), il y a encore peu de raisons de croire que les conductances dendritiques normaliseraient les EPSPs en faveur d'une détection de la synchronie des décharges afférentes.

Enfin, la synchronie semble être liée à l'inhibition et nous allons voir que, loin d'aller dans le sens d'un codage par synchronie, l'inhibition permettrait de fournir une base de temps commune pour les neurones excitateurs. Les oscillations, et par voie de conséquence dans une certaine mesure la synchronie, semblent être contrôlées par l'inhibition. Quand les cellules

[22] Elle semble de plus comparable dans toutes les couches neuronales.

[23] Les auteurs tentent de normaliser la synchronie par rapport à la décharge moyenne des neurones en estimant la fréquence de décharge des neurones sur une fenêtre temporelle de 50 ms. Il ne leur est donc pas possible de prendre en compte des variations transitoires plus rapides de la fréquence de décharge qui pourrait être à l'origine de la synchronie.

inhibitrices déchargent de manière synchronisée, les cellules excitatrices sont inhibées. Du fait de l'activation périodique des cellules inhibitrices, les cellules excitatrices présentent donc, elles aussi, une activité rythmique (Chagnac-Amitai et Connors, 1989; Van Vreeswijk et al, 1994). Une suppression de l'inhibition par des moyens pharmacologiques fait disparaître ce phénomène. On peut facilement imaginer que ce type d'inhibition périodique fournirait non seulement une base temporelle au codage par latence ou par ordre, l'inhibition permettant en quelque sorte de réinitialiser une partie du système (Fukai, 1995; Hopfield, 1995; Maass et Natschläger, 1997; Ruf et Schmitt, 1997), mais également un mécanisme de désensibilisation pour le décodage de l'ordre[24].

La synchronie des décharges neuronales est le code neuronal le plus couramment accepté et bien que son rôle dans le traitement visuel rapide soit à mon avis discutable, il est à mon avis plus clair dans d'autres phénomènes. La synchronisation des aires corticales pourrait intervenir par exemple dans les phénomènes de mémoire à court terme ou dans la perception consciente (Rodriguez et al, 1999; Tallon-Baudry et Bertrand, 1999).

Ce chapitre compare la plausibilité biologique du codage par rang et celle des autres codes neuronaux et je pense avoir ici démontré qu'il se sort très bien de cette épreuve. Le codage par synchronie, son pire ennemi, est sur le déclin et les expériences à ce sujet ne sont en aucun cas incompatibles avec un codage par ordre. Les *bursts* sont également un phénomène tout à fait compatible avec le codage par ordre. Enfin le codage par latence ou par phase, ne constitue qu'un cas particulier du codage par ordre - il correspond a une absence de *shunting inhibition* - et c'est en ces termes que je le traiterai par la suite.

Dans le chapitre suivant, en nous basant sur les résultats obtenus avec ce premier modèle, nous allons tenter d'incorporer un mécanisme d'apprentissage. Comme je l'ai déjà indiqué, un nouveau type de plasticité synaptique, qui prend en compte la date de décharge des neurones, a été découvert récemment (Markram et al, 1997). Il nous a donc paru très intéressant de l'incorporer à notre modèle de V1 et d'étudier la dynamique d'un tel réseau.

[24] Si les oscillations inhibitrices ne dépendent pas des stimuli en entrée, il y a bien décodage de l'ordre mais ce type de décodage dépend alors du contraste et de la luminance du stimulus.

2

Émergence de la sélectivité à l'orientation dans V1

Connectionists argue that the supremely charasteristic features of human intelligence are, among other, associative thinking and the ability to learn from examples.

Anya Hurlbert and Tomaso Poggio (1988)

Dans le chapitre précédent, les résultats que nous avons obtenus montrent que la prise en compte de la latence des différents afférents d'un neurone est suffisante pour rendre compte de sa sélectivité à l'orientation.

Le présent chapitre a pour but d'introduire des règles d'apprentissage plausibles du point de vue biologique dans le modèle d'un neurone de V1 et d'étudier les effets de la propagation d'images naturelles au sein de ce modèle. Récemment une loi d'apprentissage basée sur l'ordre de décharge du neurone présynaptique et postsynaptique a été découverte en électrophysiologie (Markram et al, 1997; Bi et Poo, 1998). Si l'EPSP arrive avant la décharge postsynaptique alors la synapse est renforcée, dans le cas contraire elle est affaiblie. Les dates de décharge des neurones sont au premier plan dans ce type de loi et un décalage de quelques millisecondes peut inverser un effet de potentiation vers un effet de dépression synaptique.

Comme nous allons le voir, on peut difficilement trouver une règle plus adaptée au codage par ordre. Les résultats obtenus sont spectaculaires et nous analyserons dans quelle mesure ils sont cohérents avec les données neurophysiologiques.

2.1 - Matériel et méthodes

Comme dans le modèle précédent de sélectivité à l'orientation, le réseau est composé uniquement de deux couches de neurones, la première représentant les neurones dans la rétine ou/et le LGN (je considère ces deux étapes comme ne faisant qu'une) et la seconde les neurones dans V1.

Les neurones de la rétine/LGN sont organisés au sein de cartes rétinotopiques de la taille de l'image[1]. La rétine/LGN se décompose en deux cartes neuronales, centre-ON et centre-OFF, qui extraient les contrastes positifs et négatifs en chaque point de l'image. La valeur de contraste positif en un point détermine la latence de décharge du neurone[2].

Les neurones dans V1 possèdent tous le même champ récepteur centré sur l'image. Cela signifie que ces neurones intègrent la même zone de l'image (figure 2.1). Les neurones sont des unités "intègre et décharge" plus simples que celles implémentées dans le modèle de sélectivité à l'orientation précédent. Ils intègrent linéairement EPSPs et IPSPs et émettent un unique potentiel d'action lorsqu'ils atteignent leur seuil (cf. annexe 2).

En absence d'excitation, le potentiel des neurones réels converge vers leurs potentiels de repos, du fait de la présence des canaux passifs : c'est ce que l'on appelle le courant de fuite. En nous basant sur les résultats du modèle du chapitre précédent, nous avons décidé à ce stade de ne pas prendre en compte la fuite des neurones de V1 et ceci pour deux raisons. D'une part, les processus d'intégration synaptique sont de l'ordre de la milliseconde et interfèrent peu avec la dynamique du courant de fuite, plus lente[3]. D'autre part, cela permet de ne pas prendre en compte la distribution des décharges dans le LGN. Même quand les décharges de ces neurones sont simulées comme une fonction du contraste du stimulus, la relation exacte entre latence et contraste n'est pas clairement établie[4]. Sans courant de fuite, seul compte l'ordre de

---

[1] Les filtres de contraste appliqués sont de taille 3x3 pixels. Les filtres centre-ON et centre-OFF sont implémentés comme dans le modèle précédent par une classique différence de Gaussienne normalisée à 0.

[2] En chaque point, un seul un neurone (centre-ON ou centre-OFF) décharge.

[3] En général τ>10 ms, ce qui signifie qu'après 10 ms sans excitation, le potentiel du neurone est à mi-chemin du potentiel de repos.

[4] À ce propos, l'utilisation d'un intégrateur à fuite bruité comme au chapitre précédent semble être la meilleure approximation.

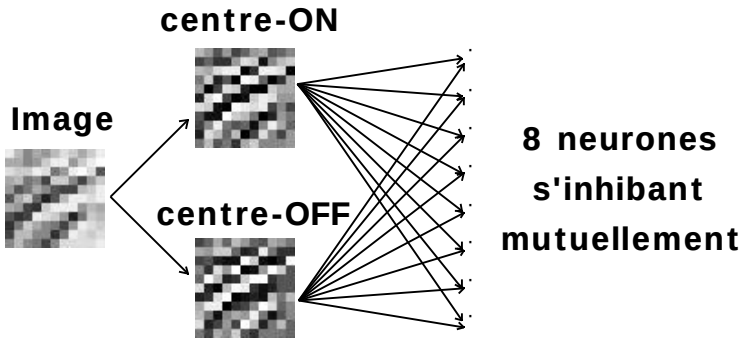


Figure 2.1 : architecture du réseau initial. Dans ce modèle, les décharges en provenance des neurones de la rétine centre-ON et centre-OFF sont intégrées par 8 neurones dont les champs récepteurs sont centrés sur l'image. Les niveaux de gris dans les cartes rétiniennes représentent la latence de décharge des neurones (les pixels clairs correspondent à des décharges précoces et les pixels foncés à des décharges tardives). Les *patterns* de connectivité afférente de ces neurones sont initialement constitués de bruit (cf. texte). De plus, les neurones s'inhibent mutuellement pour ne pas apprendre la même forme.

décharge des neurones[5]. Cela nous permet donc de choisir une fonction décroissante quelconque entre latence de décharge et contraste, la seule contrainte étant que les latences précoces correspondent aux contrastes élevés.

Au début de la simulation, les neurones dans V1 implémentent des patterns de connectivités afférentes, qui sont calculés à partir d'une équation gaussienne bidimensionnelle bruitée :

$$W_{i,j} = \frac{1+n_{x,y}}{2} e^{\frac{-d(n_i,n_j)^2}{2\sigma^2}}$$

où $W_{i,j}$ est le poids synaptique entre le neurone i dans la carte neuronale[6] (centre-ON ou centre-OFF) et j dans V1, $n_{x,y}$ représente un bruit uniforme (recalculé pour chaque synapse) et $d(n_1,n_2)$ est la distance Cartésienne séparant le neurone i et le neurone j[7]. Pour en terminer avec la connectivité, les neurones dans V1 s'inhibent mutuellement. La valeur du poids synaptique entre deux neurones est choisie afin que, lorsqu'un neurone décharge, les autres neurones ne soient pas capables de décharger (figure 2.1).

[5] Pour la même raison, les délais synaptiques ne sont pas pris en compte, les décharges des neurones étant propagées instantanément.

[6] La taille de la convolution est de 11x11 pixels. Les poids synaptiques inférieurs à 5 % de la valeur maximale théorique ont été supprimés.

[7] Les abscisses et les ordonnées des neurones sont des valeurs entières, l'unité étant le pixel. Le bruit est un paramètre variant aléatoirement entre 0 et 1. σ=2,0.

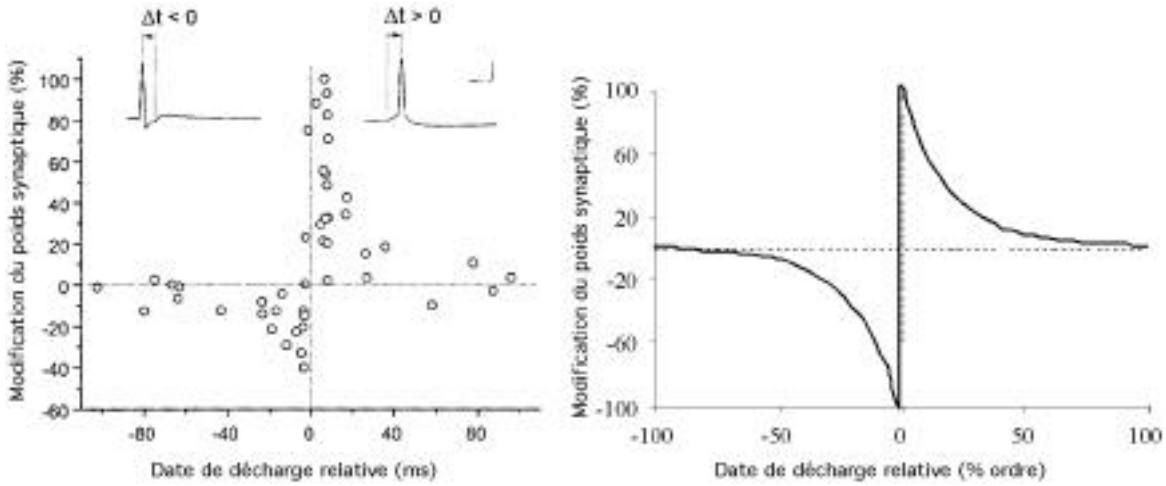


Figure 2.2 : modification des poids synaptiques en fonction de la date relative d'arrivée de l'EPSP et de la décharge du neurone postsynaptique, telle que l'ont enregistrée Bi et Poo (1998). La figure de gauche représente le niveau de potentiation de l'EPSP en fonction de la date relative entre cet EPSP et de la décharge du neurone. Un neurone en culture (cortex hippocampique) est stimulé toutes les secondes par un EPSP (stimulation d'un autre neurone connecté au neurone enregistré) et le neurone enregistré est contraint à décharger à différents intervalles de temps par rapport à l'EPSP. Suivant la date relative entre l'EPSP et la décharge du neurone, la connexion entre les deux neurones est renforcée ou affaiblie. À droite, une approximation de ce processus à l'aide de fonctions exponentiellement décroissantes. Les latences des neurones n'étant pas explicitement modélisées, il nous a fallu reformuler cette règle en termes de nombre de décharges afférentes plutôt qu'en unité de temps (correspondant à peu près à τ = 20 ms). Dans les deux graphiques, l'abscisse 0 correspond à la date d'occurrence de l'EPSP. Adapté de Bi et Poo (1998).

La règle d'apprentissage, dont j'ai déjà parlé au tout début de ce chapitre, s'inspire d'une règle de type Hebbienne découverte initialement par Markram puis vérifiée par Bi et Poo qui démontrent que la date d'arrivée d'une décharge afférente sur un neurone et de celle du potentiel d'action de ce neurone détermine le poids synaptique. Si la décharge afférente arrive avant le potentiel d'action du neurone, la connexion synaptique est renforcée, dans le cas contraire elle est affaiblie. Nous avons cependant adapté cette règle pour la rendre sensible au nombre de décharges afférentes plutôt qu'à leurs dates exactes car les latences de décharge des neurones ne sont pas simulées explicitement dans ce modèle (figure 2.2). Cette modification est en fait nécessaire à l'homogénéité du réseau. Du point de vue de la plausibilité biologique, je ne pense pas que cela pause un problème majeur : les deux phénomènes - ordre et latence - sont très proches et, les données électrophysiologiques ne sont pas incompatibles avec une loi d'apprentissage basée sur l'ordre. Nous reviendrons sur ce point précis dans la discussion.

Plus précisément, soit t_e la date d'activation de la synapse excitatrice et t_a la date de décharge du neurone

Si $t_e<t_a$ alors $dW = \alpha (1-W)e^{-|\Delta o|/\tau}$

Sinon $dW = -\alpha We^{-|\Delta o|/\tau}$

avec Δo l'ordre d'arrivée de l'EPSP sur le neurone par rapport à la décharge du neurone[8] et α la constante de potentiation et de dépression de la synapse[9]. Comme je l'ai déjà mentionné, dans ces équations, le nombre de décharges afférentes - l'ordre d'arrivée - est utilisé car les latences ne sont pas explicitement présentes dans notre modèle. Pour la potentiation, comme cela est visible sur les données expérimentales (figure 2.3), l'augmentation du poids synaptique dépend de la taille de la synapse, cela ne semble pas être le cas pour la dépression qui reste constante, en moyenne de 20 %.

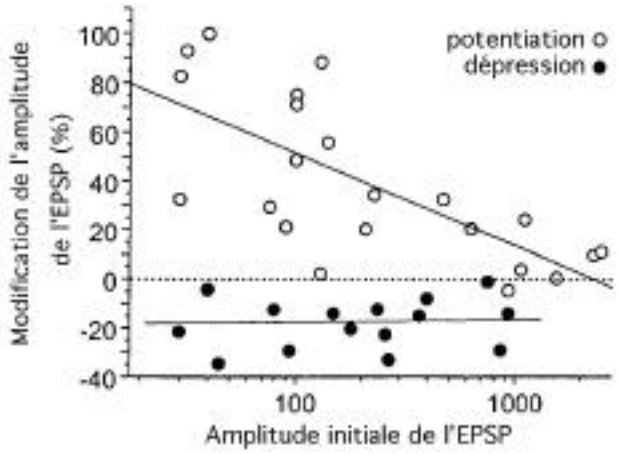


Figure 2.3 : plasticité synaptique en fonction du "poids" de la synapse (déduite à partir du maximum d'amplitude de l'EPSP). Les cercles blancs représentent les cas où l'EPSP précède la décharge du neurone. La potentiation dépend de la taille de la synapse et semble suivre une loi logarithmique : plus la synapse est grosse et moins elle est capable de se renforcer. Les cercles noirs représentent les cas ou l'EPSP est postérieur à la décharge du neurone. La dépression, de 20 % en moyenne, ne semble pas dépendre de la taille de la synapse. Adapté de Bi et Poo (1998).

Il reste cependant un dernier problème à régler, celui de la stabilité. Dans ce type de réseau, les poids synaptiques évoluant, le taux de décharge des neurones évolue également et la sélectivité peut alors disparaître. Pour éviter ce phénomène, il faut trouver un moyen de normaliser l'activité des neurones. Nous avons donc choisi de normaliser le total des poids synaptiques à chaque décharge, les seuils des neurones restant constants tout au long de la

[8] On peut fixer arbitrairement les EPSPs arrivant avant la décharge du neurone comme ayant des ordres négatifs et ceux arrivant après la décharge du neurone comme ayant des ordres positifs.

[9] τ=*20 ms*, α=*2 %,*. Les valeurs trouvées par Bi et Poo (1998) correspondent à peu près à 80% de potentiation après 60 stimulations pour les synapses les plus petites et 20% de dépression quelle que soit la taille de la synapse. Nous avons simplement divisé ces valeurs par le nombre de stimulation (60) afin d'obtenir les paramètres pour la convergence. Les estimations que j'ai pu faire à partir de l'article original (Markram et al., 1997) sont pratiquement identiques.

simulation. Cette procédure impose des contraintes relativement faibles sur la dynamique de réponse des neurones, à savoir que la probabilité de décharge des neurones et la répartition des poids synaptiques sont des paramètres libres qui ne dépendent que de la loi d'apprentissage. Une autre solution aurait été de contraindre la probabilité de décharge de l'ensemble des neurones à rester constante. Cette solution est à mon avis moins bonne car elle impose des contraintes plus fortes sur la dynamique d'activation neuronale : il est probable que les neurones réels présentent des probabilités de décharge différentes, et il est même possible que cette probabilité varie au sein d'un même neurone.

2.2 - Résultats

Nous avons propagé environ 7 millions d'images naturelles[10] 11x11 au sein d'un simple réseau de 32 neurones dans V1. Cependant les résultats sont décevants puisque pratiquement tous les neurones sont sélectifs à des orientations horizontales (figure 2.4). Quelques neurones deviennent cependant sélectifs à des orientations verticales ou obliques. On peut noter que du fait de l'inhibition entre les neurones, leur sélectivité n'est pas totalement identique, certains neurones étant sélectifs pour des orientations horizontales centrées dans le champ récepteur et d'autres étant sélectifs pour des orientations horizontales décalées vers le haut ou vers le bas.

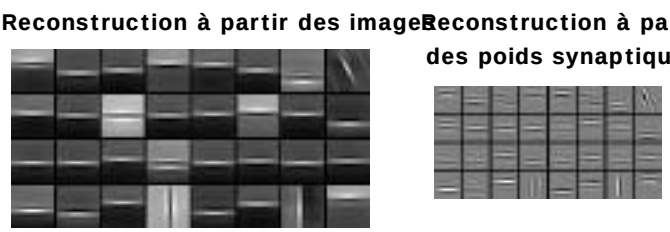


Figure 2.4 : sélectivité à l'orientation du premier réseau. Environ 7 millions d'images sont propagées. Les neurones sont presque tous sélectifs à des barres horizontales même si la position de cette barre varie d'un neurone à l'autre (voir la figure 2.6 pour des détails sur la reconstruction).

Le problème vient clairement du fait qu'un neurone sélectif à une certaine orientation ne l'est qu'à une certaine position dans son champ récepteur. S'il est sélectif à une barre horizontale centrée, il ne déchargera pas sur une barre horizontale décentrée et un autre neurone deviendra donc sélectif à ce type de stimulus. En fait, il est possible de résoudre ce problème en organisant les neurones au sein de cartes neuronales rétinotopiques.

Dans une extension mineure du modèle précédent, les neurones sont maintenant organisés au sein de cartes rétinotopiques, le reste de l'organisation du réseau restant la même (figure

[10] Il s'agit d'images 11x11 prises parmi des scènes naturelles 640x480 (base d'images Corel). Les scènes incluent des paysages, des prises de vue en ville, des objets...

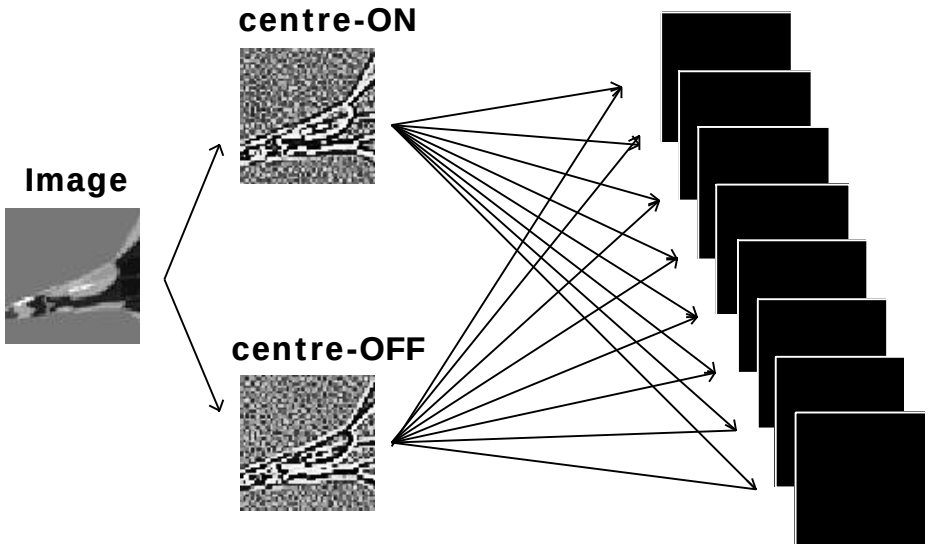


Figure 2.5 : extension du modèle de la figure 2.1 qui prend en compte l'organisation rétinotopique des neurones dans V1. Les neurones sont organisés en cartes neuronales homogènes où ils partagent tous les mêmes poids synaptiques. Si un neurone d'une carte neuronale décharge et voit ses poids modifiés par la règle d'apprentissage, alors cette modification est répercutée sur l'ensemble des neurones de la carte neuronale. Cette organisation permet d'ajouter une compétition entre les cartes neuronales, non plus seulement pour la même position mais également pour les positions voisines ce qui permet d'atteindre un éventail de sélectivité neuronale bien plus important que dans le cas du premier modèle.

2.5). Les neurones au sein d'une même carte ont les mêmes poids synaptiques que leurs voisins. Quand un neurone apprend, la modification des poids synaptiques se répercute donc sur l'ensemble des neurones de la carte. Les neurones d'une carte neuronale se projettent sur les autres cartes neuronales avec des connexions inhibitrices. L'inhibition n'est plus locale comme dans le réseau précédent mais s'étend latéralement (figure 2.5). Pour chaque neurone, des synapses inhibitrices projettent vers les autres cartes neuronales, implémentant une simple distribution gaussienne de poids synaptiques négatifs[11]. Cela signifie que si un neurone décharge à une position donnée alors les neurones au sein des autres cartes neuronales ne seront pas seulement inhibés à cette position mais aussi aux positions voisines. Du fait de l'organisation rétinotopique, si les neurones d'une carte sont sélectifs à une orientation horizontale, les neurones d'une autre carte seront inhibés et ne pourront pas devenir sélectifs à cette barre.

Trente deux cartes neuronales sont présentes dans le réseau, et les poids synaptiques étant

[11] Les simulations montrent que l'étendue des projections inhibitrices doit être à peu près le double de celle des champs récepteurs excitateurs (21x21 dans notre modèle; $\sigma=4$; les poids synaptiques inférieurs à 5 % du maximum sont supprimés). On peut peut-être rapprocher cela de données électrophysiologiques dans le cortex visuel primaire qui montrent que l'étendue des connexions latérales inhibitrices est plus large que celle des connexions excitatrices (à l'exception des connexions excitatrices à longue distance, cf. chapitre I.2).

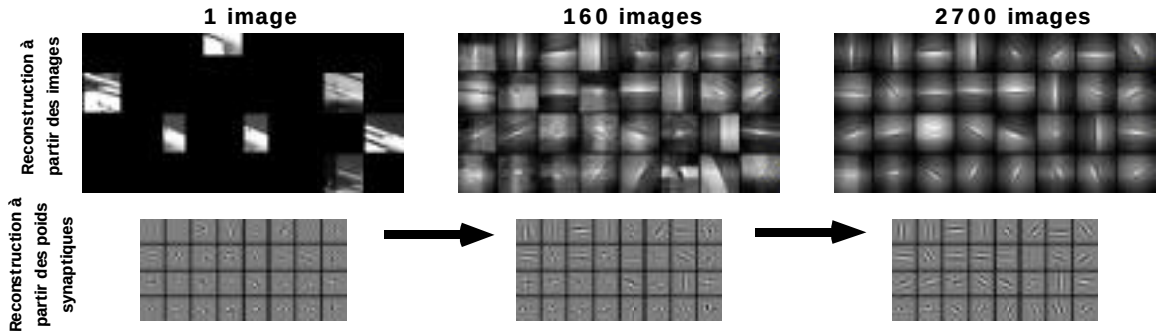


Figure 2.6 : résultats de la propagation d'images naturelles dans le second réseau à cartes neuronales rétinotopiquement organisées. L'évolution de l'apprentissage est représenté. Après la propagation d'une seule image, le profil de sélectivité des neurones est encore aléatoire. Après avoir propagé 160 images, on distingue une certaine organisation et, après 2700 images, les neurones présentent des sélectivités très variées. Sur la ligne supérieure, la reconstruction est effectuée à partir des images propagées. Pour chaque position où un neurone décharge, une image de taille 21x21 centrée sur cette position est extraite de l'image initiale 50x50 propagée dans le réseau. Pour chaque carte neuronale, cela permet de calculer la moyenne des images sur lesquelles les neurones de cette carte ont déchargé. Ce type de reconstruction, plus large que la zone d'intégration d'un neurone, permet de faire apparaître les effets de contexte présents dans les images naturelles. Sur la ligne inférieure, une reconstruction à partir des poids synaptiques en provenance des cellules centre-ON et centre-OFF de la rétine/LGN est représentée. Pour chaque poids synaptique, on applique un filtre (centre-ON ou centre-OFF suivant la provenance de la synapse) à la position du neurone dans la rétine/LGN duquel provient cette synapse. Ce filtre est ensuite pondéré par la valeur du poids synaptique. Le résultat final est normalisé, la valeur de gris moyenne correspondant toujours à 0. Ce type de reconstruction du profil de sélectivité des neurones, est également utilisé par Wallis (1994).

communs à tous les neurones au sein de chaque carte, on obtient 32 profils de sélectivité. Le résultat de la propagation de 2700 images naturelles[12] de taille 50x50 montre qu'il est possible d'obtenir dans de telles conditions des sélectivités très variées (figure 2.6). Elles vont de barres orientées, de toutes directions à des formes plus complexes qui s'apparentent à des blobs. Comme l'indique la figure 2.6, l'évolution de la sélectivité est relativement rapide au début de la simulation pour se ralentir et atteindre un état relativement stable à l'issue de la propagation des 2700 images.

2.3 - Discussion

Nos résultats montrent qu'à partir de simples hypothèses, il est possible d'extraire certaines statistiques des images naturelles. Les hypothèses sont que (1) les neurones se comportent comme des unités IF, (2) qu'une *shunting inhibition* rapide est présente, (3) qu'une inhibition compétitive existe entre des neurones de sélectivité différente, (4) que la plasticité synaptique dépend de la date d'arrivée de la décharge présynaptique et de la décharge postsynaptique.

[12] Les images 50x50 sont tirées de scènes naturelles 640x480 (même base d'images que précédemment).

Nous allons maintenant voir dans quelle mesure ces résultats s'accordent avec le comportement des neurones réels et quel est l'apport de ce modèle par rapport à ceux présents dans la littérature.

2.3.1 - Robustesse et plausibilité du modèle

Je ne reviendrai pas sur la plausibilité de la *shunting inhibition* rapide que j'ai traitée au chapitre précédent. Je ne discuterai pas non plus l'utilisation de neurone "intègre et décharge" (cf. annexe 2). On est cependant en droit de se demander pourquoi j'ai implémenté la *shunting inhibition* rapide, plutôt que de prendre de simples neurones IF. La raison en est très simple : la *shunting inhibition* rapide apporte une certaine stabilité au modèle. Sans la *shunting inhibition*, les profils de sélectivité des neurones divergent. Il arrive qu'une carte neuronale se mette à détecter une orientation durant quelques images, mais cette sélectivité est hautement instable[13]. Les problèmes de convergence existent également dans le cas où la *shunting inhibition* est présente. Si le seuil des neurones est trop élevé par exemple, les poids synaptiques des neurones ne sont jamais modifiés et ne convergent pas. Bien que le seuil des neurones intervienne dans la convergence de leurs sélectivités, ce ne semble pas être un paramètre critique, l'émergence d'un profil de sélectivité cohérent étant observée pour des valeurs de seuil variant du simple au double. Le facteur de convergence intervient également dans l'émergence de la sélectivité : s'il est trop faible, les neurones déchargent de façon non sélective et n'apprennent aucune forme structurée, s'il est trop élevé, les neurones changent constamment de sélectivité ou divergent, c'est-à-dire qu'ils perdent leur sélectivité car l'ensemble de leurs poids synaptiques devient pratiquement égal. Encore une fois, dans le cas où la *shunting inhibition* est présente, le réseau est relativement robuste à la variation du taux de convergence et ce paramètre peut varier du simple au quintuple sans modifier la dynamique de l'émergence de la sélectivité. De même, il est possible de modifier le taux de décharge des neurones dans la rétine/LGN : j'ai par exemple obtenu des résultats très similaires à ceux obtenus figure 2.6 avec seulement 10% de décharges dans la rétine/LGN.

Concernant mon postulat sur l'utilisation de l'ordre de décharge plutôt que la latence exacte, je tiens encore une fois à me justifier de ce choix. J'ai effectué de nombreuses simulations qui montrent que l'on peut obtenir des résultats similaires avec la latence exacte

[13] J'ai tenté d'appliquer de nombreuses variantes mais sans grand succès. J'ai contraint par exemple le taux de décharge des neurones à demeurer constant plutôt que de normaliser le total de leurs poids synaptiques. La dynamique dans tous les cas semble intrinsèquement instable.

de décharge. Le logiciel SpikeNET que j'utilise (cf. annexe 2) discrétise le temps, de sorte que les simulations que j'ai réalisées au départ utilisent les latences de décharge plutôt que leur ordre d'activation dans la dynamique d'apprentissage. J'ai obtenu des résultats comparables à ceux que je présente ici et cela signifie donc pour moi qu'il est tout à fait possible de trouver un ensemble de paramètres pour effectuer ces simulations dans le temps plutôt que dans l'espace de l'ordre[14]. En fait, j'ai simplement préféré construire un réseau homogène, où l'ordre était présent à tous les niveaux. J'ai donc été contraint de réinterpréter la règle de modification des poids synaptiques en fonction de l'ordre plutôt que de la latence afin de rester homogène avec le mécanisme d'intégration neuronale. Cependant ce type de mécanisme n'est pas dénué de sens : le facteur d'apprentissage au niveau d'une synapse pourrait dépendre du nombre d'EPSPs qui ont conduit le neurone à décharger plutôt que d'un facteur de temps absolu. Dans le même ordre d'idées, aucune expérience ne démontre que la latence est plus biologique que l'ordre d'activation des neurones. De plus, l'utilisation de l'ordre des décharges, en tant que code neuronal, semble plus logique *a priori*.

Nous allons voir également que les simulations dont l'apprentissage est basé sur l'ordre des décharges afférentes à un neurone, offrent l'avantage d'une explication simple pour la convergence des neurones et que la règle biologique que j'ai implémentée est compatible avec une règle d'apprentissage optimale pour le codage par ordre.

2.3.2 - Le mécanisme de convergence des neurones

Avant de me passionner pour la règle d'apprentissage Hebbienne dépendant de l'ordre d'occurrence entre l'EPSP et la décharge du neurone postsynaptique, j'ai utilisé au sein de ces réseaux une règle d'apprentissage optimale du point de vue de l'ordre des décharges des neurones afférents.

Cette règle fait converger les poids synaptiques des neurones vers la modulation - *shunting inhibition* - qui les affecte au moment de leur propagation. Cela signifie par exemple que si un EPSP arrive en $10^{ème}$ position sur un neurone, sa modulation, si elle suit une loi puissance, sera de X^{10} (X représentant le niveau de modulation, $1>X>0$) et le poids synaptique convergera donc vers cette valeur. Cette règle est optimale du point de vue du codage par ordre car la sélectivité d'un neurone est maximale si l'ordre d'activation des neurones afférents

[14] Du point de vue computationnel, il semble cependant plus efficace et plus stable de se baser sur l'ordre plutôt que la latence exacte.

est le même que celui des poids synaptiques (Veneau, 1996) [15].

Les résultats obtenus avec ce type de règle sont très similaires à ceux obtenus pour la règle d'apprentissage Hebbienne (figure 2.7). La règle d'apprentissage sur l'ordre peut en fait être ramenée à une règle binaire, renforçant les poids synaptiques arrivant avant la décharge du neurone et affaiblissant ceux qui arrivent après. Cette règle binaire est elle-même très proche de la règle d'apprentissage Hebbienne. Je vais tenter de détailler cette approche mais je tiens à

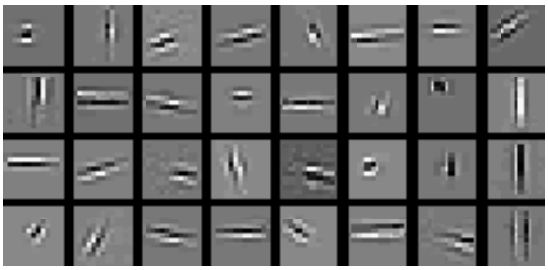


Figure 2.7 : exemple de propagation basée sur une règle d'apprentissage optimale pour un codage par l'ordre de décharge des neurones (cf. texte). Dans cet exemple, la modulation est relativement élevée et très peu de synapses activées suffisent à faire décharger un neurone. Avec des modulations plus faibles, on obtient des résultats similaires à ceux que j'ai présentés pour la règle d'apprentissage dépendant de la date relative de l'EPSP par rapport à la décharge du neurone.

souligner que je ne considère absolument pas ce qui suit comme une démonstration.

Tout d'abord, la loi d'apprentissage basée sur la date relative de décharge des neurones - optimale pour l'ordre - se rapproche de la loi d'apprentissage binaire. La variabilité contenue dans les images permet de linéariser en quelque sorte, cette loi binaire. Par exemple, lorsqu'un neurone décharge pour deux *patterns* qui ne sont pas totalement identiques, certains poids synaptiques seront renforcés dans un cas et affaiblis dans l'autre, d'autres seront soit renforcés soit affaiblis dans les deux cas. Il s'ensuit que la distribution des poids synaptiques, loin d'être binaire, sera continue et approchera celle de la loi d'apprentissage basée sur l'ordre, les synapses activées à des latences très précoces étant celles dont les poids synaptiques sont les plus élevés [16]. Déterminer le bruit que l'on peut attendre des images naturelles et dans quelle mesure la répartition des poids est affectée dans la loi d'apprentissage binaire est hors de propos ici et nécessite des analyses mathématiques plus poussées qui sont en cours (Perrinet, 1999).

---

[15] Je présente ici une démonstration que je pense bien plus simple pour le cas de l'activation de toutes les synapses afférentes d'un neurone. Supposons que l'ordre des poids synaptiques soit le même que celui de leur activation. Le potentiel atteint alors la valeur *A*. Soit deux synapses afférentes à un même neurone de poids w_1 et w_2 ($w_1>w_2$) activées dans un certain ordre. Pour *A*, on a $A=B+mw_1+m\alpha w_2$ avec *B* l'ensemble des autres activations synaptiques, *m* la modulation de la synapse de poids w_1 et $\alpha<1$. Dans le cas où on intervertit l'ordre d'activation des deux synapses, on obtient $A'=B+mw_2+m\alpha w_1$. Or $w_1-w_2>\alpha(w_1-w_2)>0$ et donc $w_1+\alpha w_2>w_2+\alpha w_1$, *soit* $B+mw_1+m\alpha w_2>B+mw_2+m\alpha w_1$. On en déduit que A>A' et donc que le niveau d'activation sera plus faible. La généralisation pour une distribution de poids synaptiques quelconque est triviale.

[16] J'ai également effectué des simulations avec cette loi d'apprentissage binaire. Les résultats sont qualitativement identiques à la règle d'apprentissage biologique et à celle optimale pour l'ordre.

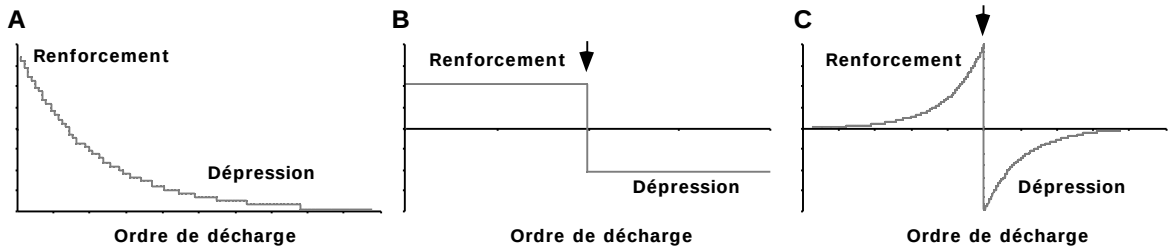


Figure 2.8 : représentation des trois règles d'apprentissage, optimale pour l'ordre, binaire et biologique. A, règle d'apprentissage faisant converger les poids synaptiques vers une valeur de modulation qui dépend de l'ordre. Cette règle est optimale du point de vue de la *shunting inhibition* rapide. B, règle d'apprentissage binaire, renforçant les poids synaptiques arrivés avant la décharge du neurone et affaiblissant les poids synaptiques arrivant après la décharge du neurone. La décharge du neurone est indiquée par une flèche. En présence de bruit dans les images, cette règle est très proche de la première (cf. texte). C, règle d'apprentissage biologiquese basant sur la date de l'EPSP par rapport à celle de la décharge du neurone. Cette règle est réinterprétée en utilisant l'ordre d'arrivée de l'EPSP par rapport à la décharge du neurone plutôt que sa latence exacte. Dans B et C, l'axe des abscisses coupe l'axe des ordonnés à 0 (les axes d'abscisses et d'ordonnées sont exprimés en unités arbitraires respectivement d'ordre de décharge et de renforcement/dépression synaptique).

La loi binaire elle-même est corrélée à la loi Hebbienne de type biologique que j'ai utilisée dans mon modèle. Les valeurs mesurées font état d'environ 50 ms pour la décroissance des courbes de potentiation/dépression (figure 2.2). Si l'on considère que le temps d'intégration synaptique qui conduit un neurone à décharger est d'environ 10 ms, les deux lois biologique et binaire sont très proches. La figure 2.8 indique dans quelle mesure les trois règles, biologique, binaire et optimale pour l'ordre, sont corrélées. Encore une fois, ces indications ne constituent en aucun cas une démonstration, et doivent être prise pour ce qu'elles sont, c'est-à-dire une approche intuitive des conséquences de la règle d'apprentissage biologique dans le cadre du codage par ordre.

On peut se demander cependant pourquoi les neurones utiliseraient la règle d'apprentissage décrite par Markram et al (1997) plutôt que celle basée sur l'ordre. En fait, la règle d'apprentissage de type biologique possède un avantage sur sa consœur basée sur l'ordre, à savoir que les poids synaptiques ne doivent pas nécessairement suivre la distribution des modulations. Dans la règle d'apprentissage où les poids synaptiques convergent vers la modulation, la distribution des poids synaptiques est nécessairement biaisée par les valeurs de modulation. La règle d'apprentissage de type biologique permet aux poids synaptiques de converger vers des valeurs qui dépendent de la statistique des images (bien qu'indirectement, du fait de l'intégration neuronale, elles dépendent aussi de la modulation). De plus la règle d'apprentissage biologique peut être implémentée en continue, sur un flot d'images, alors que la règle optimale pour l'ordre nécessite que l'on détermine un début à la propagation, début au niveau duquel la modulation sera minimale.

Très clairement, ces travaux ne couvrent pas l'ensemble des implications - notamment

mathématiques - de cette règle d'apprentissage. L'étendue de ces phénomènes et de leurs conséquences pour la modélisation est à mon avis majeure et nécessitera de nombreuses études. Nous allons voir maintenant la relation qu'entretient notre modèle avec ceux présents dans la littérature et quelles sont les extensions que l'on peut espérer.

2.3.3 - Apport des autres modèles

L'émergence de la sélectivité des neurones à partir de stimuli et de règles locales d'apprentissage est un problème central aux réseaux de neurones. En général, outre la règle d'apprentissage, ce sont les interactions latérales qui permettent aux neurones de s'organiser. Il y a bientôt 20 ans, le premier modèle de ce type est apparu et a fait grand bruit (Kohonen, 1982). Le réseau de Kohonen (1982) est intéressant dans le sens où il fixe des interactions locales entre les neurones. De cette façon et en utilisant une règle d'apprentissage Hebbienne, les neurones s'organisent en groupes pour devenir sélectifs aux stimuli présentés[17]. Des modèles plus évolués ont également montré que les interactions latérales étaient indispensables pour l'organisation topologique des neurones (Durbin et Mitchison, 1990). Dans le modèle que j'ai présenté, les interactions locales entre les neurones sont minoritaires[18] et interviennent principalement pour réguler l'apprentissage. Je vais donc me concentrer tout particulièrement sur la règle d'apprentissage qui est au centre du modèle.

Quelques travaux tentent d'étudier les propriétés de ce type de plasticité synaptique. Ces travaux se bornent cependant à analyser les réponses d'un neurone isolé. L'équipe de Abbott en particulier est précurseur dans ce domaine (Abbott et Song, 1999), elle a montré qu'en utilisant une règle d'apprentissage similaire à celle que j'ai utilisée, le changement de la fréquence de décharge des neurones afférents[19] ne modifiait pas les poids synaptiques (cf. également Kistler et van Hemmen, 2000). Par contre, la modification de la corrélation entre les décharges - le fait que deux synapses afférentes soient activées de façon synchrone - permet de renforcer ces synapses. Intuitivement, cela semble trivial : du fait du courant de fuite, si des synapses sont activées simultanément alors la probabilité que le neurone décharge est bien plus élevée que si leur activation n'est pas corrélée. Le problème avec ce type de modèle est que la distribution des poids synaptiques devient binaire, les synapses corrélées

[17] Par exemple des *patterns* sinusoïdaux à différentes fréquences. Les neurones sont des unités continues, il n'y a pas de décharges. Les interactions locales font que si un neurone apprend, ces voisins apprennent également.
[18] Cependant tous les neurones d'une carte apprennent le même *pattern*, ce qui peut en un sens être considéré comme une interaction locale. L'inhibition également peut-être considérée comme un type d'interaction latérale.
[19] Modélisé par un processus de Poisson.

ayant des valeurs de poids maximum alors que les synapses non corrélées ont des valeurs d'activation minimum. Cette distribution des poids synaptiques est loin de correspondre à celle des neurones réels.

Von Rossum (2000) résout ce problème en utilisant une règle d'apprentissage plus biologique encore - la même que nous avons utilisée[20]. Dans le modèle de Abbott, la potentiation/dépression ne dépend pas de la valeur du poids synaptique[21]. Cependant, d'après les données de Bi et Poo (1998), la dépression est proportionnelle au poids synaptique alors que la potentiation fait converger la synapse vers une valeur absolue. On obtient alors, dans les mêmes conditions que les modèles précédents, une distribution des poids synaptiques gaussienne cohérente avec les données neurophysiologiques.

Ces travaux, bien que très intéressants, restent relativement théoriques. Ils démontrent que la règle d'apprentissage de type biologique que j'ai décrite peut servir à détecter des EPSPs synchrones parmi les décharges afférentes à un neurone. Cependant, je pense que les implications du point de vue de l'auto-organisation de la sélectivité des neurones sont bien plus fortes dans mon modèle.

2.3.4 - Vers une plasticité synaptique plus complexe

L'apparente simplicité de la règle, de notre modèle et des modèles étudiant les propriétés de cette règle masque en fait sa véritable complexité. Par exemple, Roberts montre une loi d'apprentissage inverse chez le poisson électrique : si l'EPSP arrive avant la décharge du neurone cible, il y a dépression alors que la synapse est renforcée dans le cas inverse. Pour Roberts, cela permet aux neurones de ne pas répondre quand le poisson émet une décharge électrique. En effet, en construisant un modèle avec des neurones utilisant cette loi d'apprentissage, il montre que les neurones sensibles aux décharges électriques s'adaptent peu à peu à des décharges répétées pour ne plus du tout répondre sur ces stimuli.

Lorsqu'on analyse plus en détails les données de Markram et al (1997) et de Bi et Poo (1998), elles ne sont pas aussi claires qu'il peut y paraître au premier abord. Markram effectue des appariements EPSP/décharge à plusieurs fréquences - 2, 5, 10, 20, 30 et 40 Hz - et la potentiation/dépression est beaucoup plus importante dans les derniers cas, même si le nombre de décharge reste constant. On s'attendrait qu'à 40 Hz, potentiation et dépression

[20] J'avais pourtant imaginé cette règle avant d'avoir eu connaissance des travaux de Bi et Poo (1998) et de Von Rossum (2000). Cette règle me paraissait en effet la plus simple et la plus logique.

[21] Dans ce modèle, la dépression doit absolument être supérieure à la potentiation pour que la sélectivité des neurones se stabilise.

interagissent et que la potentiation synaptique soit moindre : or ce n'est pas le cas puisque c'est à 40 Hz que la potentiation est maximale. Cela montre que la fréquence, et pas seulement la corrélation temporelle EPSP/décharge postsynaptique, joue un rôle dans l'apprentissage. De même, dans l'expérience de Bi et Poo (1998), pour que la potentiation/dépression soit observée il faut que les stimulations soient répétées - 1 appariement EPSP/décharge postsynaptique par seconde répété 60 fois -. Bien que la durée séparant deux appariements soit très longue à l'échelle cellulaire, elle joue probablement un rôle dans la plasticité synaptique.

La potentiation/dépression semble en fait dépendre de la concentration de calcium au niveau post-synaptique. Si elle est élevée, on obtiendra une potentiation et si elle est faible, une dépression[22]. La clef réside probablement dans la dynamique différente des réservoirs de calcium (Frank, 2000) et pourrait expliquer la dépendance de la plasticité synaptique à la fréquence de décharge des neurones.

Bien que le problème ne se pose pas dans notre modèle puisque les neurones ne peuvent décharger qu'une seule fois, la question des décharges multiples sur une synapse est également primordiale. Si plusieurs décharges afférentes arrivent sur une même synapse, leur efficacité diffère grandement : le premier EPSP est généralement celui de plus grande amplitude mais cette dynamique est adaptative et le maximum d'EPSP peut parfois survenir à la deuxième, voire à la troisième décharge afférente (Markram et Tsodycks,1996; Tsodycks et al, 1998; Okatan et Grossberg, 2000). L'évolution de cette dynamique en fonction des dates relatives entre EPSP et décharge neuronale est le sujet d'un intense débat (Senn et al, 1997). Un modèle très intéressant (Stork et al, 2000) permet en fait d'adapter cette dynamique pour que des bouffées de décharge - *burst* - soient encodées de façon optimale. Comme je l'ai déjà mentionné au chapitre précédent, les *bursts* semblent être un mode de communication privilégié, une bouffée de décharge encodant bien plus d'informations par décharge qu'une décharge unique (Lisman, 1997; Reinagel et al, 1999). La latence des bouffées dépend en fait de l'intensité du stimulus en un point de la même façon que le stimulus est encodé en termes de décharges uniques dans notre modèle. Une loi d'apprentissage qui ferait évoluer la dynamique de la synapse pour que le maximum d'EPSP d'un *burst* coïncide avec la décharge du neurone (Stork et al, 2000), serait compatible à la fois avec la plasticité synaptique observée expérimentalement et avec le comportement de notre modèle - si le stimulus était

[22] Cela découlerait des propriétés de la molécule qui fixe le calcium, la calmoduline, qui n'est activée qu'en présence de hautes concentrations de calcium.

encodé par des *bursts* plutôt que par des décharges neuronales isolées -. Encore une fois, bien que des simulations soient nécessaires, je ne pense pas que ce type de mécanisme modifierait de façon drastique le comportement de notre réseau.

Nous avons vu dans ce chapitre la puissance d'une règle qui est basée sur la date relative d'un EPSP par rapport à la décharge d'un neurone et permet à la sélectivité des neurones de s'organiser de façon automatique sur des images naturelles. Cette propriété reproduit dans une certaine mesure celle des neurones de V1 dans le système visuel du primate. Nous avons vu également dans quelle mesure cette loi se rapprochait d'une loi optimale du point de vue du codage par ordre. Dans le chapitre suivant, je me bornerai à utiliser ce type de règle d'apprentissage pour tenter de simuler des phénomènes plus complexes encore, comme la reconnaissance des objets dans des scènes naturelles.

3

Modèles de reconnaissance des objets

The central connections, after all, determine the final result.

G.H. Bishop (1959)

Les propriétés des réseaux à *shunting inhibition* rapide, et celles des règles d'apprentissage que l'on peut y introduire, sont intéressantes en elles-mêmes. Cependant, malgré la forte composante biologique que j'ai tenté d'introduire dans mes simulations, ces modèles restent des modèles et il est nécessaire d'être très prudent quant à leur domaine d'application. Le meilleur moyen, à mon avis, de convaincre la communauté scientifique de l'utilité d'un modèle est d'en démontrer l'efficacité du point de vue computationnel. Le modèle doit être capable d'effectuer des traitements similaires à ceux réalisés dans le cerveau d'un primate et notamment dans le système visuel.

Cet objectif est très ambitieux car les systèmes artificiels de traitement d'image sont encore loin d'approcher les performances du système visuel, en particulier en ce qui concerne la reconnaissance des objets dans les scènes naturelles. Vouloir alors effectuer du traitement d'image à l'aide d'un modèle inspiré de la biologie, mais dont les paramètres sont difficilement contrôlables et dans lequel la dynamique n'est pas totalement connue, peut sembler utopique.

Cependant, de la même façon que l'aéronautique s'inspira du vol des oiseaux à ses débuts, je pense que tenter de simuler les stratégies de calcul des neurones dans le cerveau humain ne peut que conduire au succès. Si le système visuel est capable de détecter et de reconnaître des objets dans des scènes naturelles alors les modèles du système visuel doivent pouvoir atteindre ce niveau de performance.

Dans la première partie de ce chapitre, partant des résultats des modèles précédents, je présente un réseau volontairement très simple pour la reconnaissance de chiffres dactylographiés. Dans un second modèle, plus complet, un réseau similaire permet la reconnaissance de visages. Nous verrons que ces modèles sont très proches de ceux que j'ai présentés précédemment en particulier en ce qui concerne l'organisation topologique des neurones, leur sélectivité et la règle de plasticité synaptique. La principale différence réside dans le fait que l'apprentissage est dit supervisé, ce qui signifie que c'est l'expérimentateur qui détermine les neurones qui doivent apprendre. Dans le second modèle, nous verrons comment cette limitation peut être détournée. Enfin, la compatibilité de ces modèles avec les données neurophysiologiques sera discutée.

3.1 - Reconnaissance de chiffres

Le premier réseau de reconnaissance de chiffre se compose de neurones IFs organisés de façon rétinotopique. Les images sont initialement filtrées par des filtres de contraste[1]. Comme dans le modèle du chapitre précédent, les valeurs résultantes induisent des décharges neuronales en chaque position dans l'image dont la latence est inversement proportionnelle au niveau de contraste. En chaque point, soit un neurone centre-ON, soit un neurone centre-OFF décharge. Comme dans le modèle précédent, les dates exactes des décharges des neurones ne sont pas prises en compte car les neurones ne présentent pas de courant de fuite et seul l'ordre des décharges est important. Ces décharges sont intégrées à l'aide de filtres d'orientation à deux fréquences spatiales[2]. Ce type d'intégration est courant dans les modèles du système visuel pour rendre compte de la sélectivité à l'orientation. J'aurais évidemment pu utiliser les

[1] Filtre laplacien de gaussienne 3x3, positif au centre et négatif au pourtour (centre=-1,5*pourtour).

[2] Les filtres sont calculés à l'aide de fonctions gaussiennes bidimensionnelles asymétriques sur l'axe x et l'axe y (matrices 5x5 avec σ_x=3,2 et σ_y=0,5 à la fréquence spatiale élevée et matrices 11x11 avec σ_x=3,8 et σ_y=0,7 à la fréquence spatiale faible). Le total des poids contenus dans ces convolutions est normalisé et est donc égal dans tous les cas. Les filtres sont calculés pour 8 orientations différentes. Les poids synaptiques inférieurs à 10 % du maximum sont supprimés et les seuils des neurones dans cette couche sont fixés afin que la probabilité de décharge d'un neurone sur une image soit à peu près de 10 à 20 %.

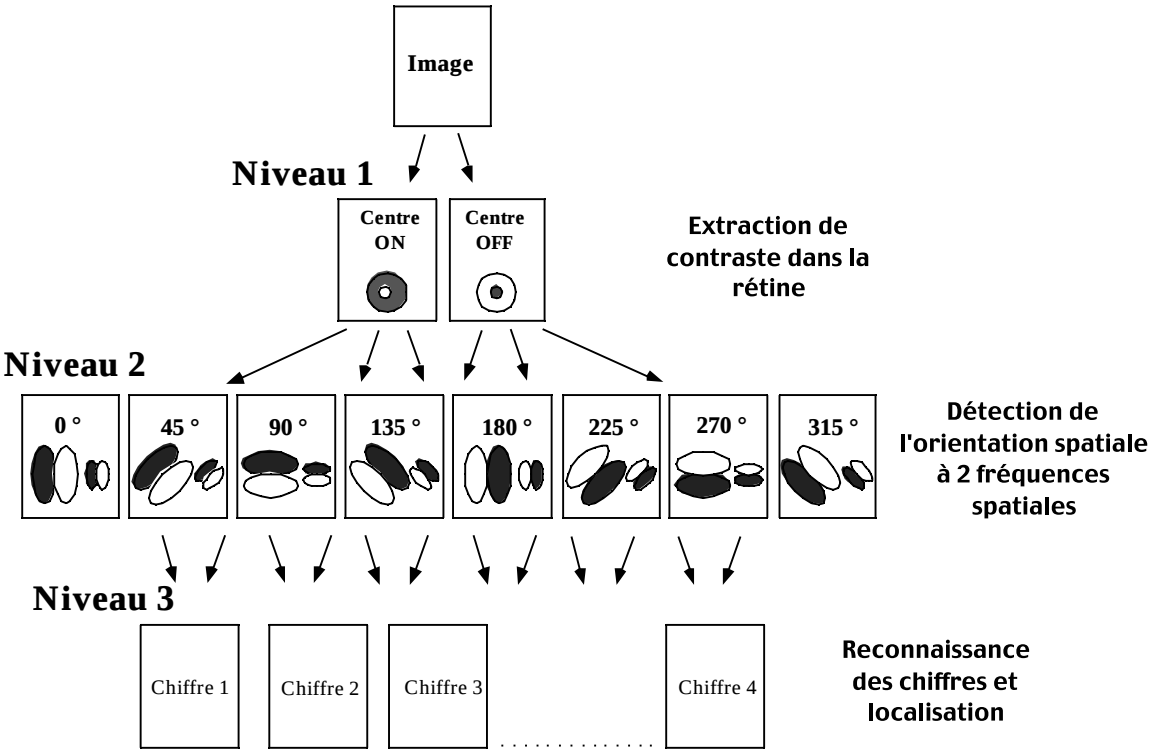


Figure 3.1 : architecture du réseau pour la reconnaissance des chiffres. L'image est tout d'abord décomposée en contrastes positifs et négatifs. Par la suite, ces informations sont propagées dans des cartes neuronales sélectives à différentes orientations et à différentes fréquences spatiales (8 orientations et 2 fréquences spatiales pour un total de 16 cartes). Les décharges en provenance de ces cartes sont enfin intégrées au sein de cartes chacune sélective à 1 chiffre. Les cartes neuronales de ce dernier niveau s'inhibent mutuellement afin qu'à une position donnée, 1 seul chiffre soit reconnu.

résultats du chapitre précédent et implémenter l'émergence à la sélectivité des neurones à partir de la présentation d'images naturelles. J'ai préféré cependant utiliser des filtres plus couramment acceptés. Au niveau supérieur, 10 cartes neuronales, sélectives aux dix chiffres sont introduites. Comme dans le réseau du chapitre précédent, ces cartes neuronales s'inhibent mutuellement : de cette façon, quand un neurone décharge à une position donnée, les neurones sélectifs à un autre chiffre ne déchargeront pas à cette position, ni à des positions voisines[3] (figure 3.1).

Pour chaque chiffre, 5 exemplaires, chacun dans une police de caractères spécifique et dans des conditions de bruit différentes, sont présentés au réseau[4]. L'apprentissage est dit supervisé dans le sens où l'expérimentateur indique aux cartes neuronales concernées où se trouvent les chiffres dans l'image présentée. Pour chaque carte neuronale, on définit donc les positions dans l'image pour lesquelles les neurones de la carte doivent répondre. Pour une

[3] Les convolutions inhibitrices sont modélisées par des fonctions gaussiennes de taille 7x7 et de σ=1,6. Les poids synaptiques inférieurs à 10 % du maximum sont supprimés.

[4] 10 % de bruit uniforme est introduit à l'aide du logiciel Photoshop.

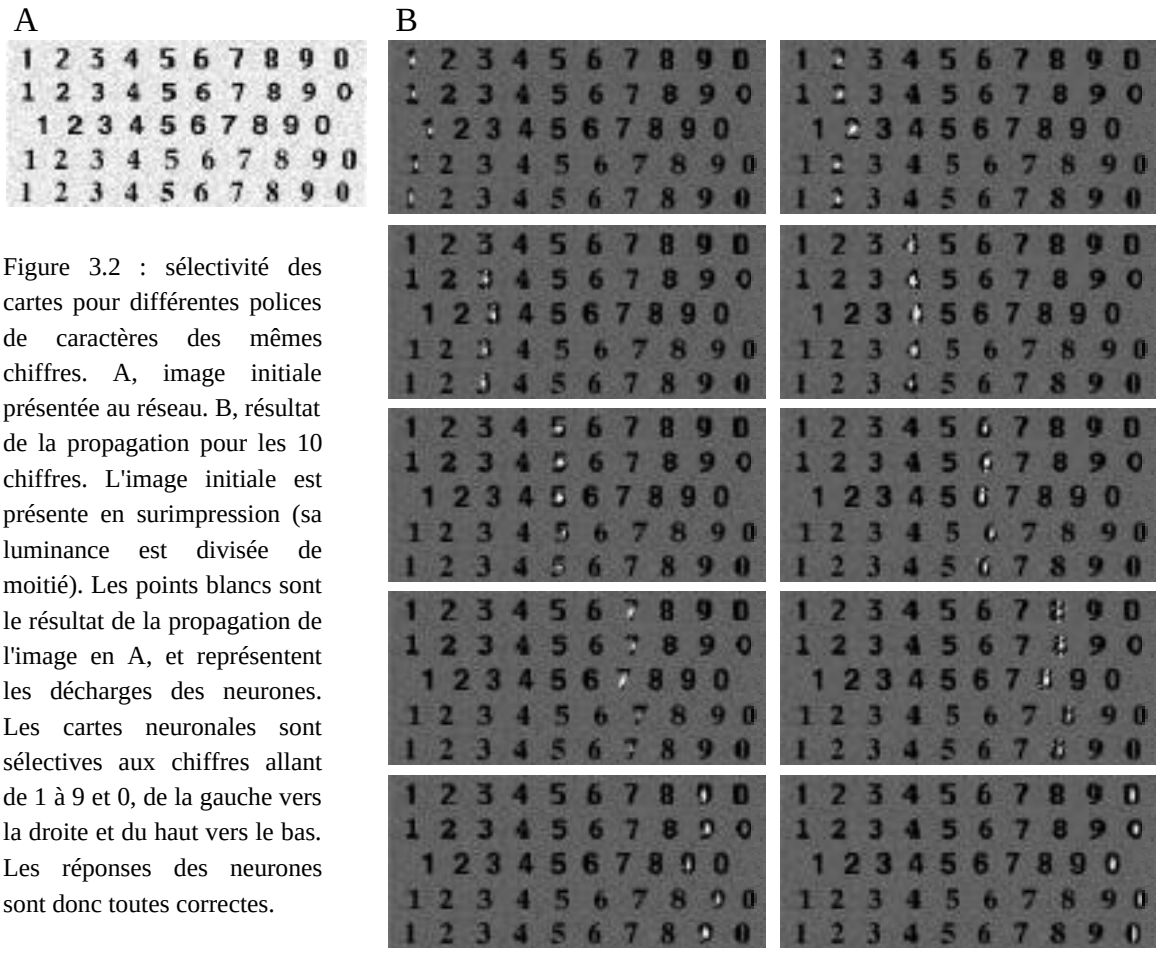


Figure 3.2 : sélectivité des cartes pour différentes polices de caractères des mêmes chiffres. A, image initiale présentée au réseau. B, résultat de la propagation pour les 10 chiffres. L'image initiale est présente en surimpression (sa luminance est divisée de moitié). Les points blancs sont le résultat de la propagation de l'image en A, et représentent les décharges des neurones. Les cartes neuronales sont sélectives aux chiffres allant de 1 à 9 et 0, de la gauche vers la droite et du haut vers le bas. Les réponses des neurones sont donc toutes correctes.

carte sélective au chiffre "1" par exemple, les neurones de la carte doivent répondre au centre des chiffres 1 présents dans l'image.

Une image contenant les 50 chiffres[5] (les 10 chiffres dans 5 polices de caractère) est propagée dans le réseau et les neurones des cartes neuronales sélectives à un chiffre donné apprennent aux positions où se trouvent ce chiffre[6]. Les neurones apprenant à reconnaître un chiffre voient en fait les poids synaptiques de leur champ récepteur converger vers la latence de décharge des neurones afférents. Comme dans le cas de la règle optimale du point de vue de l'ordre dans l'apprentissage non supervisé, les neurones afférents activés les premiers renforcent leurs connexions avec le neurone cible alors que les neurones activés tardivement affaiblissent leurs connexions avec le neurone cible. Concrètement, pour un neurone apprenant à une position donnée - position dans l'image où se trouve le chiffre auquel la carte

[5] La taille de l'image est de 256x118 pixels, et les différents chiffres ont une taille d'environ 17x11 pixels. L'image est présentée à la figure 3.2.

[6] Il est donc nécessaire de déterminer préalablement les coordonnées du centre de tous les chiffres présentés au réseau. J'ai effectué cette opération à la main. Malgré l'imprécision que cela peut engendrer, les résultats attestent que cette imprécision n'est pas critique pour la reconnaissance.

neuronale du neurone est associée - les poids synaptiques afférents de ce neurone convergent vers la latence de décharge des neurones sélectifs à l'orientation.

De cette façon, après la phase d'apprentissage, chaque carte neuronale est sélective à l'un des 10 chiffres. De plus, les cartes neuronales s'inhibent mutuellement : si un neurone, d'une carte neuronale sélective à un chiffre, décharge alors il inhibe les neurones sélectifs aux autres chiffres à cette position. De plus, les poids synaptiques étant partagés par tous les neurones au sein d'une carte neuronale, si un neurone de la carte apprend à reconnaître une forme à une position donnée, les autres neurones de la carte deviennent automatiquement sélectifs à cette même forme à d'autres positions dans l'image[7]. Enfin, le seuil de décharge des neurones sélectifs aux chiffres est ajusté manuellement afin que les neurones déchargent les premiers sur les chiffres auxquels ils sont associés.

Pour tester le réseau, la même image que celle utilisée pour l'apprentissage, contenant tous les exemplaires des chiffres en 5 polices de caractère, est présentée. Les résultats sont excellents, puisque le réseau ne commet aucune erreur (figure 3.2) : chaque carte neuronale sélective à un chiffre donné (0 et de 1 à 9) ne décharge que sur ce chiffre. Pour tester plus avant les capacités du réseau, l'apprentissage est effectué sur 4 chiffres très similaires par leur contour (les chiffres 0, 6, 8 et 9) à différentes tailles puis le réseau est testé. Là encore les performances sont très bonnes car les neurones sont sélectifs à plusieurs tailles de chiffres et cela malgré leur similarité (figure 3.3).

Malgré l'aspect rudimentaire du réseau, composé uniquement de neurones IF interconnectés, ces résultats montrent que si l'on utilise des règles d'apprentissage adaptées, il est possible d'atteindre un niveau de performance convenable du point de vue du traitement d'image. Cependant, bien que les niveaux de bruit soient différents d'une condition à l'autre, les chiffres utilisés pour l'apprentissage sont les mêmes que ceux utilisés pour tester le réseau. Il serait intéressant de déterminer si le réseau est capable de généraliser à de nouveaux objets. On peut d'ores et déjà être relativement confiant, les résultats préliminaires que nous avons obtenus sur la résistance à la variation de la taille sur les chiffres étant très prometteurs. Nous allons donc maintenant analyser les performances du réseau sur une tâche beaucoup plus complexe, celle de la reconnaissance de visages. En particulier nous nous intéresserons à la capacité de généralisation d'un tel réseau à de nouvelles images, qui n'ont pas été présentées pendant la phase d'apprentissage.

[7] Se référer à l'annexe 2 pour les détails du fonctionnement des carte neuronales à poids synaptiques partagés.

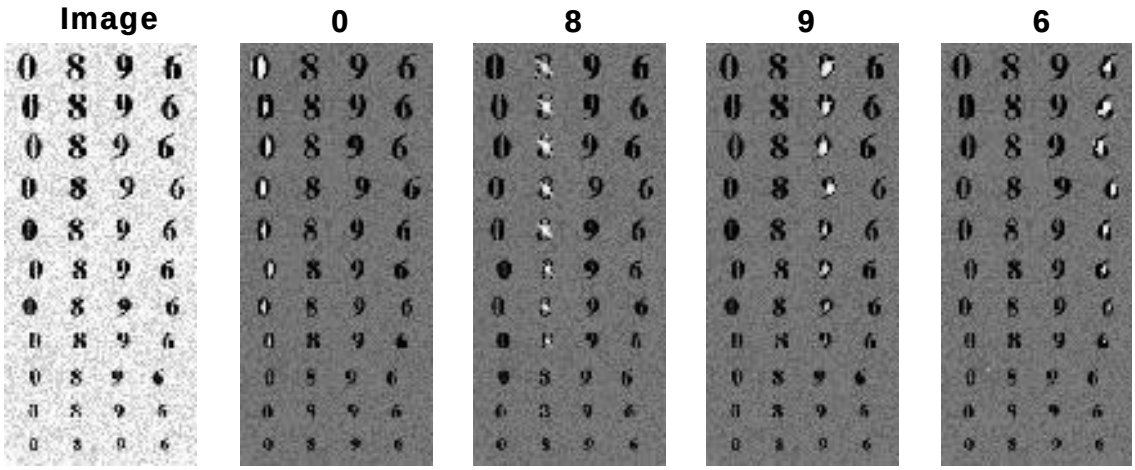


Figure 3.3 : résistance à la variation de tailles pour 4 cartes sélectives aux chiffres 0, 8, 9 et 6. Ces chiffres ont été choisis du fait de leur ressemblance (ils présentent tous une boucle). À gauche, l'image initiale est présentée et à droite le résultat de la propagation de cette image. On constate que la sélectivité des neurones est dans une certaine mesure invariante à la taille des chiffres.

3.2 - Reconnaissance des visages

Le réseau est en tout point similaire à celui utilisé pour la reconnaissance des chiffres. Cependant, dans la rétine/LGN, les filtres ont été modifiés pour que la somme totale des valeurs soit nulle[8]. Ce type de filtre est plus couramment utilisé en biologie et permet d'obtenir une invariance à la luminance moyenne de l'image. Si la luminance moyenne augmente par exemple de 10 niveaux de gris, l'augmentation de l'inhibition au pourtour sera entièrement compensée par l'augmentation de l'excitation au centre.

Au niveau supérieur, j'ai utilisé des fonctions Gabor[9] pour calculer la distribution des poids synaptiques. Comme dans le cas du premier niveau rétine/LGN, ce type de fonction est plus couramment utilisé pour simuler la sélectivité à l'orientation des neurones dans V1 que ceux du modèle précédent[10]. Comme dans le réseau précédent, les seuils des neurones dans V1 sont fixés afin que la probabilité de décharge des neurones sur une image soit d'environ 10 à 20 %. La dernière couche est également similaire à celle du modèle précédent, les neurones étant entraînés sur des visages et non plus sur des chiffres (figure 3.4). Le nombre d'objets à

[8] Dans le premier niveau rétine/LGN, à la différence du modèle précédent, le filtre 3x3 est une différence de gaussienne (centre = -1*pourtour). L'unité est le pixel.

[9] Les patterns de connectivité pour les orientations sont codés par des fonctions Gabor qui sont en fait le produit d'une fonction sinus unidimensionnelle sur un axe par une gaussienne bidimensionnelle. Pour la fonction sinus la phase est de 0.5 radian/unité de distance et pour la fonction gaussienne σ=1. Les valeurs inférieures à 5 % du maximum sont supprimées). L'unité est le pixel et la taille des fltres est réduite à 7x7 pixels.

[10] Il apparaît cependant que la différence de gaussienne utilisée dans le chapitre précédent est plus efficace du point de vue computationnel car les filtres sont plus indépendants les uns des autres (Wallis, 1994).

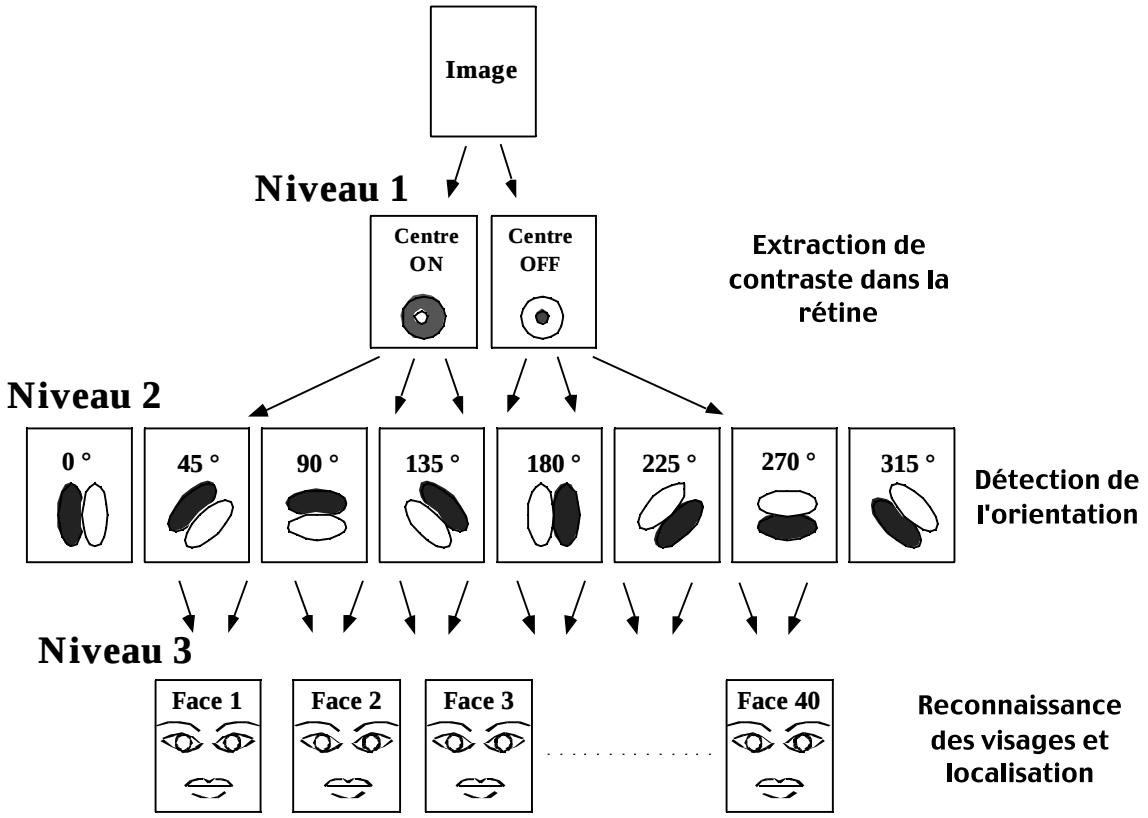


Figure 3.4 : architecture du réseau pour la reconnaissance des visages. Le réseau est similaire à celui utilisé pour la reconnaissance des chiffres, à l'exception près que les cartes du dernier niveau sont sélectives à des visages. Comme dans le cas des chiffres, les cartes neuronales de ce dernier niveau s'inhibent mutuellement afin qu'à une position donnée, 1 seul visage soit reconnu.

reconnaître est cependant bien plus important : le réseau devant être capable d'identifier 40 individus, 40 cartes neuronales, représentant chacune un individu, sont nécessaires. Dans ce réseau, j'ai également ajouté un mécanisme de *shunting inhibition* rapide qui module les poids synaptiques en fonction de leur ordre d'arrivée, les derniers poids synaptiques activés étant plus fortement modulés que les premiers[11].

La procédure d'apprentissage est identique à celle utilisée pour les chiffres. Les poids synaptiques cependant ne convergent plus vers la latence des décharges des neurones mais plutôt vers le niveau de modulation associé à cette décharge. Des analyses mathématiques ont en effet montré que l'activité du neurone est maximale uniquement dans le cas où l'ordre d'activation des neurones afférents est le même que celui des poids les associant au neurone

[11] La modulation est une fonction puissance du nombre d'afférents ayant touché un neurone modulation = $x^{nb\_afférent}$ (chaque neurone est considéré individuellement). Le chiffre x est déterminé comme suit : pour les neurones de V1, il correspond à la modulation nécessaire afin que 22 % des décharges afférentes à un neurone induisent une diminution de 50 % de l'efficacité des poids synaptiques. Pour les neurones sélectifs aux visages, ce facteur tombe à 10 % des poids afférents.

cible (Vennau, 1996). La modulation étant une fonction monotone de l'ordre et de la latence, faire converger les poids synaptiques vers la modulation plutôt que vers l'ordre est de peu d'importance. Dans le cas où l'on fait converger les poids synaptiques vers la modulation, on est cependant certain que la répartition des poids synaptiques d'un neurone sera optimale du point de vue de sa sélectivité à l'ordre. Nous verrons plus loin comment ce mécanisme peut s'accommoder des données électrophysiologiques.

La base d'images choisie est de 400 visages (10 vues de 40 personnes)[12]. Les individus sont de sexe et d'origine différente et sont très variés, certains ayant des lunettes, d'autres une barbe ou une moustache. Les 10 vues de chaque personne sont également assez variées, généralement de face ou de trois quarts (figure 3.5). Pour chaque personne, 2 vues sont sélectionnées aléatoirement et seront utilisées dans la base finale de test (vues inconnues des individus). Pour toutes les images restantes (8*40), 3 versions additionnelles sont réalisées, l'une où le contraste est diminué de moitié, les deux dernières où le contraste est diminué de moitié et où la luminance moyenne de l'image est soit augmentée, soit diminuée[13]. Parmi ces 4 versions, 2 sont sélectionnées aléatoirement et l'ensemble de ces images constitue la base d'apprentissage (soit 2*8*40= 640 images). Les deux versions restantes constituent une première base de test contenant des images avec des contrastes et des luminances modifiées. Enfin, la base finale de test est constituée des 2 images de chaque individu sélectionnées au départ (soit 2*40= 80 images) pour lesquelles on réalise également les 3 versions additionnelles. Il existe donc trois bases d'images, la base d'apprentissage, la base de test avec des vues connues d'individus présentées à des luminances et des contrastes différents, et la base de test finale contenant des vues inconnues des individus, présentées dans toutes les conditions de contraste et de luminance.

Avant de présenter les résultats que j'ai obtenus, il reste cependant à régler le problème de l'optimisation des seuils des neurones dans les cartes sélectives aux visages. Contrairement aux cartes sélectives aux chiffres dans le modèle précédent, du fait du grand nombre de cartes neuronales au dernier niveau (40 cartes chacune sélective à un individu), il n'est plus possible de les optimiser de façon manuelle. En effet, la sélectivité d'un neurone affecte celle des

[12] Base de visages ORL, AT&T Laboratories Cambridge. Je tiens à souligner qu'aucune image n'a été supprimée de la base d'image originale téléchargée sur Internet à l'adresse http://www.cam-orl.co.uk/facedatabase.html. Les images sont réduite à la taille de 23x28 pixels avant d'être propagées dans le réseau.

[13] Pour diminuer le contraste, on soustrait 128, le niveau de gris intermédiaire, à la valeur des pixels (entre 0 et 255), on divise par deux cette valeur et on ajoute 128. Augmenter (ou diminuer) le contraste signifie ajouter (ou retrancher) 64 à la valeur de tous les pixels.

A

B

	performance
Base d'apprentissage (BA)	100 %
Base de test (BT)	100 %
Base finale de test (BFT)	97,5%

Figure 3.5 : A, exemple de visages d'une même personne à différents niveaux de contraste et de luminance (cf. texte). B, performance du réseau pour la base de visages utilisée pour l'apprentissage (BA), pour le premier test effectué avec la même base mais avec des contrastes et des luminances différentes (BT) et pour le test final effectué avec des visages des mêmes personnes qui ne faisaient pas partie de la base d'apprentissage (BFT).

autres. Baisser le seuil d'un neurone va le conduire à décharger sur un plus grand nombre d'images et donc à inhiber les neurones sélectifs aux visages présentés. Des effets trompeurs apparaissent alors : les neurones d'une carte neuronale peuvent ne pas sembler sélectifs alors qu'en fait les neurones d'une autre carte neuronale, eux trop peu sélectifs, les empêchent de décharger. Pour résoudre ce problème complexe, j'ai donc imaginé un processus très simple. Pour 16 images de 40 individus (soit 640 images), les 40 cartes neuronales doivent en fait décharger dans 1/40=2,5 % des cas. Le taux de décharge des cartes est fixe mais elles ne sont pas assujetties à décharger sur les visages pour lesquels elles devraient être sélectives. Cela signifie que si une carte neuronale décharge sur 2,5 % des images de visage, l'algorithme de convergence est satisfait, que ces images contiennent ou non le visage auquel la carte neuronale est sélective. Ce processus se base sur l'hypothèse que les cartes neuronales sont effectivement sélectives et que si on les contraint à décharger sur 2,5 % des images, elles vont, selon toute vraisemblance, décharger sur les visages qu'on leur a appris. L'avantage de cette règle d'apprentissage est qu'une carte peu sélective ne peut pas décharger sur plus de 2,5 % des images présentées et ne pourra donc pas inhiber de façon incontrôlée les autres cartes neuronales. Existe-t-il une meilleure façon pour optimiser les seuils des neurones, notamment une règle prenant en compte le fait que la carte neuronale ait déchargé sur les visages qu'on lui a présentés pendant l'apprentissage ? C'est tout à fait possible, mais je discuterai pourquoi la règle que j'ai choisie est plus acceptable du point de vue de la biologie qu'une règle plus évoluée.

Les performances du réseau sont assez impressionnantes (figure 3.5) car le modèle est capable d'effectuer une reconnaissance avec 100 % de réponses correctes sur la base d'apprentissage et sur la première base de test avec des contrastes et des luminances

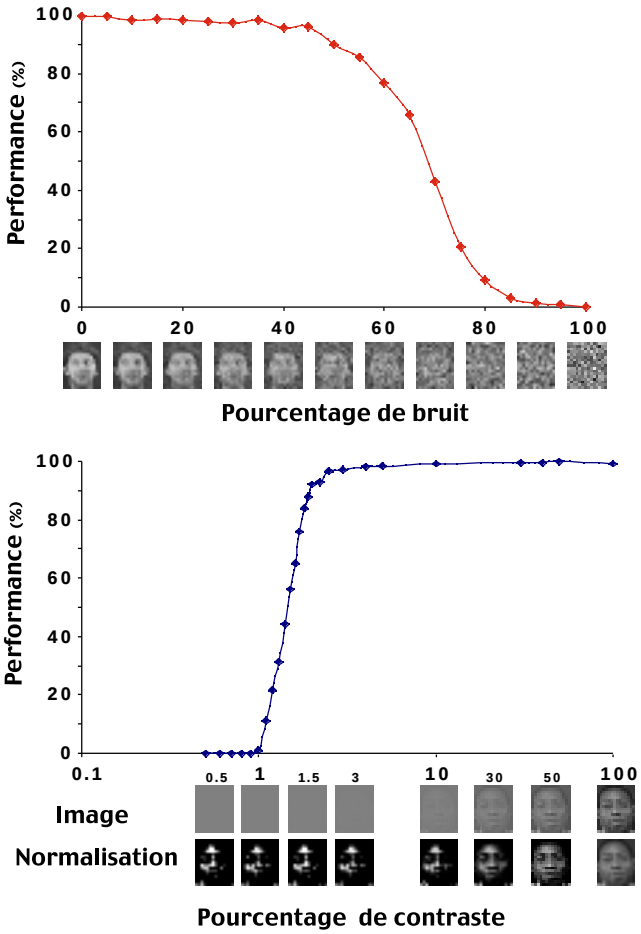


Figure 3.6 : résistance au bruit et à la diminution de contraste de la sélectivité aux visages. La résistance au bruit (en haut) est remarquable puisque la performance du réseau est de plus de 80 % d'images correctement catégorisées avec 50 % de bruit (cf. texte). Le visage représenté sous l'axe des abscisses illustre le niveau de bruit. Le visage présenté sous la courbe est le visage le mieux catégorisé par le réseau (le réseau reconnaît toujours ce visage avec plus de 90 % de bruit). En bas, représentation de la résistance aux variations de contraste. A 1,5% de contraste (ce qui représente de 1 à 4 niveaux de gris), la performance du réseau est toujours au-dessus de 90 % de réponses correctes. L'image sous l'axe des abscisses illustre les différents niveaux de contraste. Cette image est normalisée afin de se rendre compte de ce qui peut rester en terme d'information. Le visage représenté est le visage le mieux catégorisé (jusqu'à 1 % de contraste). Il est important de noter que le réseau a été entraîné avec des images non bruitées et de très faibles réductions de contraste (50 %). Les performances dans les conditions de bruit et de diminution de contraste présentées ici sont donc d'autant plus impressionnantes.

différentes. La luminance et le contraste des images ne semblent donc pas interférer fortement avec la reconnaissance. Les performances sur la base finale de test sont également très impressionnantes avec un taux de réponses correctes de 97,5 %. Le réseau est donc capable de reconnaître des nouvelles vues des individus présentés pendant la phase d'apprentissage.

Pour tester plus avant la résistance du modèle aux variations de contraste et au bruit, je réduis progressivement le contraste[14] et j'augmente le bruit[15] sur les images de la base d'apprentissage. On constate que la reconnaissance des visages est très résistante à la variation de ces deux paramètres (figure 3.6). Même avec 1,5 % de contraste résiduel dans les images, le taux de reconnaissance est supérieur à 90 %. De même, avec 50 % de bruit, le nombre de

[14] Pour diminuer le contraste, on soustrait 128, le niveau de gris intermédiaire, à la valeur des pixels (entre 0 et 255), on pondère cette valeur par le pourcentage de contraste désiré et on ajoute 128.

[15] L'addition de bruit se fait à l'aide d'une image contenant du bruit blanc (les valeurs des pixels sont situées aléatoirement entre 0 et 255). Pour obtenir l'image bruitée, on effectue une moyenne pondérée pour chaque pixel entre l'image initiale et l'image de bruit blanc. Pour 10 % de bruit par exemple, *pixel_bruité = 90%* pixel_initial+10%* pixel_bruit_blanc*.

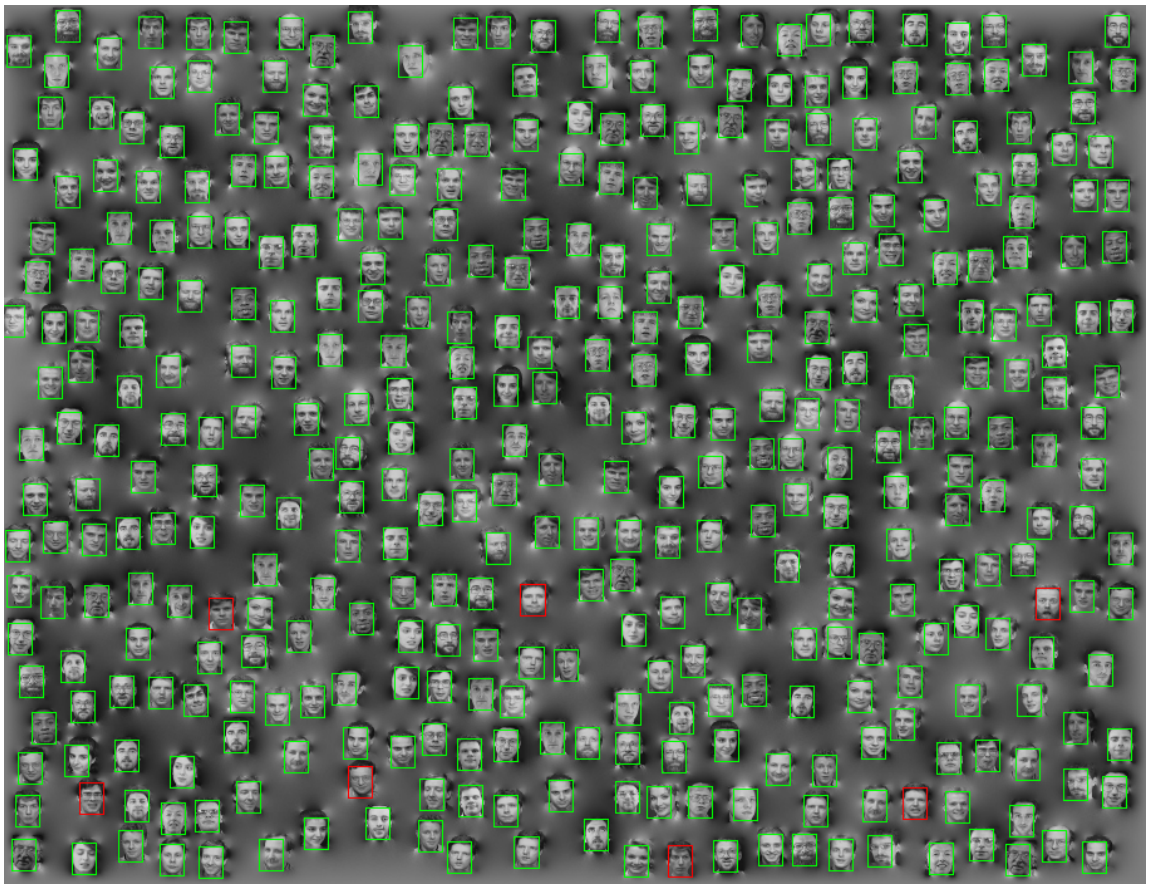


Figure 3.7 : propagation de l'ensemble des 10 vues des 40 individus (400 visages). La taille de l'image est de 910x700 pixels et la taille du réseau, qui dépend de la taille de l'image, atteint donc ici des chiffres considérables (32 millions de neurones et 245 milliards de connexions). Les visages entourés d'un rectangle vert indiquent une détection correcte et les visages entourés d'un cadre rouge indiquent une détection incorrecte. Les images des visages sont placées dans cette grande image et j'applique ensuite un algorithme de lancer de rayon pour atténuer les bords des images de visage et donner ainsi un aspect de foule.

reconnaissances correctes est encore supérieur à 90 %. Il est intéressant de noter que ces valeurs sont relativement proches des seuils perceptifs humains. De plus, étant donnée la base d'apprentissage utilisée et la relative simplicité du modèle, ces résultats sont à mon avis pour le moins inattendus.

Ces réseaux montrent la souplesse et la puissance des réseaux de neurones à décharge pour le traitement des images naturelles. En particulier, la résistance aux variations de bruit et de contraste est très impressionnante et aucun logiciel de traitement d'image à ma connaissance ne permet de telles performances.

Il est également important de noter que la durée de traitement des images est très courte. J'ai par exemple effectué la reconnaissance sur les 400 visages de façon simultanée (figure 3.7). La propagation de cette énorme image prend une trentaine de minutes sur un ordinateur

de bureau[16]. Pour simuler la reconnaissance de ces 400 visages, pas moins de 32 millions de neurones et 245 milliards de connexions sont nécessaires. Les raisons d'une telle vitesse de traitement sont indiquées dans l'annexe 2.

Nous allons voir maintenant dans quelle mesure le comportement du réseau est compatible avec les enregistrements neurophysiologiques et comparer les performances du réseau avec d'autres modèles s'inspirant de la biologie.

3.3 - Discussion

3.3.1 - Pertinence biologique

Très clairement, ces modèles sont les plus abstraits de ceux que j'ai pu présenter. Cependant cela ne signifie pas que les processus mis en jeu soient si éloignés de ceux présents dans le système visuel. Ces modèles reprennent bien des points des modèles précédents, en particulier en ce qui concerne la propagation des décharges neuronales. Je vais revenir brièvement sur ces points avant d'aborder ceux qui sont spécifiques aux modèles de reconnaissance d'objets présentés ici : l'organisation hiérarchique et l'apprentissage supervisé.

Tout d'abord, une attention particulière a été portée sur le choix des fonctions à utiliser pour les *patterns* de connectivité afférente des neurones du modèle de rétine/LGN et du modèle de V1. Ce type de profil de sélectivité est couramment utilisé dans les modèles du système visuel. La principale restriction de mes modèles reste que les neurones ne peuvent pas émettre plus d'un potentiel d'action, cela à la fois pour montrer qu'il est possible d'effectuer des traitements très complexes sur la base d'une seule décharge et pour supprimer tout processus itératif. Concernant ce dernier point, un argument à décharge pour la défense de mon modèle est que le fait d'utiliser une unique décharge par neurone ou un *burst* de potentiel d'action, qui semblent plus plausible biologiquement, ne modifie pas la dynamique de propagation du modèle. J'ai en effet déjà justifié pourquoi un *burst* pouvait être traité comme un événement unique.

Je ne peux cependant pas nier l'originalité du processus de *shunting inhibition* rapide et de la règle d'apprentissage. J'ai déjà indiqué dans quelle mesure la *shunting inhibition* est corrélée avec les données électrophysiologiques. Elle n'est cependant pas présente dans le premier réseau de reconnaissance de chiffres aux performances pourtant acceptables. De même, la règle d'apprentissage utilisée n'est pas si éloignée de la biologie comme nous l'avons

[16] Macintosh G3 à 266 Mhz.

montré dans le chapitre précédent sur l'apprentissage non supervisé. Je n'ai cependant pas testé la règle, plus biologique, basée sur la date relative d'arrivée entre l'EPSP et la décharge du neurone postsynaptique et c'est une extension directe que je pourrais apporter à ce modèle.

Concernant la convergence des poids synaptiques vers les valeurs des modulations plutôt que vers les latences de décharge des neurones, je ne pense pas que cela joue un rôle majeur dans la dynamique du réseau[17]. Je n'ai pas exploré ce problème en détail mais le premier modèle montre que même si les poids synaptiques convergent vers des valeurs inversement proportionnelles aux latences, la reconnaissance est également possible. Mon sentiment sur ce sujet est que tant que les valeurs des poids synaptiques sont une fonction monotone décroissante de la latence moyenne de décharge du neurone afférent, la dynamique du réseau sera peu affectée[18].

La règle d'apprentissage supervisée est évidemment la plus difficile à justifier. Il est à mon avis hors de propos d'invoquer un quelconque facteur extérieur, comme une sorte de mentor. Si l'on fait une analogie entre le réseau que j'ai construit et le cerveau d'un enfant, un facteur externe pourrait être un parent qui indique à son enfant la position des objets. Je ne crois pas trop à ce processus et je pencherais plutôt pour des signaux provenant de la voie dorsale qui indiqueraient la position des objets aux cartes neuronales sélectives aux objets. La carte neuronale, dont la sélectivité se rapproche le plus de l'objet présenté, déchargerait alors en premier. D'une part ces décharges inhiberaient les autres cartes neuronales à cette position et d'autre part elles déclencheraient le mécanisme de plasticité synaptique qui doit permettre à la carte neuronale de faire converger sa sélectivité vers l'objet présenté. Je reviendrai sur cette hypothèse dans le modèle général théorique que j'ai construit pour effectuer des reconnaissances sur des objets quelconques de façon non supervisée.

Dans le modèle de reconnaissance des visages que j'ai présenté, bien que l'apprentissage soit supervisé, l'optimisation des seuils des cartes ne dépend que du taux de décharge attendu des neurones d'une carte neuronale. Étant donné que la reconnaissance de nouvelles vues d'un visage par ces neurones est possible, si la présentation de nouvelles vues est équilibrée pour chaque visage, les neurones d'une carte peuvent continuer à apprendre tout en reconnaissant les visages. Les cartes rééquilibreraient de façon dynamique leur seuil, pour conserver la

[17] Nous sommes encore si loin d'approcher les performances du système visuel qu'il ne m'a pas semblé inutile de modifier la répartition des poids synaptiques pour qu'elle soit optimale du point de vue de la *shunting inhibition* rapide.

[18] L'optimisation des seuils des neurones peut cependant se révéler plus ardue si tous les poids synaptiques ont des valeurs voisines, ce qui peut être le cas quand ils convergent vers des valeurs inversement proportionnelles aux latences.

même sélectivité par rapport aux autres cartes neuronales. L'implémentation de ce type de mécanisme reste cependant à réaliser et il est possible qu'un apprentissage trop rapide de nouvelles vues déséquilibre l'ensemble du réseau.

Enfin, concernant la connectivité des neurones, le problème de la convergence des informations visuelles, qui n'est pas présente dans mes modèles, se pose également. Dans le système visuel, les différentes aires neuronales sont organisées de façon convergente, c'est-à-dire que les cartes de plus haut niveau ne présentent plus aucune rétinotopie, les neurones étant sélectifs à des objets quelle que soit la position à laquelle ils sont présentés dans le champ visuel. J'ai pour cette raison développé un algorithme complexe de convergence entre les cartes neuronales de différentes tailles (annexe 2) qui permet de conserver les connexions entre tous les neurones de la carte afférente et de la carte efférente. Les performances de cet algorithme restent à tester et cette prochaine étape permettra à mon avis de nous rapprocher du comportement des neurones dans le système visuel.

3.3.2 - Sélectivité des neurones

Outre les propriétés globales des modèles que j'ai présentés, il est également intéressant de déterminer, à un niveau plus local, dans quelle mesure la sélectivité des neurones du modèle est compatible avec celles des neurones biologiques, notamment des neurones sélectifs aux visages dans le cortex inféro-temporal. Tout d'abord à la différence des neurones réels, les neurones du modèle de reconnaissance des visages ne sont sélectifs qu'à une seule taille de visage. Les simulations que j'ai réalisées pour effectuer une reconnaissance à plusieurs tailles montrent que la performance du réseau baisse d'environ 5 % pour des variations de taille de plus ou moins 5 %[19]. Bien qu'il soit généralement admis que les neurones du cortex inféro-temporal (IT) sont sélectifs aux objets indépendamment de leur taille, cette propriété est largement exagérée. Certains neurones de IT ne sont sélectifs qu'à un objet de taille donnée et doubler ou réduire de moitié sa taille supprime complètement leur réponse, de sorte que les propriétés des neurones de mes modèles ne sont pas totalement irréalistes (Ito et al, 1995).

Dans V1, les neurones sélectifs à l'orientation le sont aussi à la disparité, à l'orientation du regard et à d'autres propriétés du stimulus (Trotter et Celebrini, 1999). La taille de leur champ récepteur est variable, de même que leur sélectivité au mouvement et tous ces paramètres semblent relativement indépendants (Schiller et al, 1976). De même dans IT, ce type de

[19] Ces résultats ne sont pas présentés car je n'ai pas effectué d'étude systématique de la variation de la taille sur la reconnaissance.

sélectivité multiple est parfois observé. Les neurones dans IT, que l'on considère comme des neurones à haut niveau d'intégration sont sélectifs à des images très complexes comme des visages, des mains ou des objets. Comme dans V1, les caractéristiques des objets sur lesquels répondent les neurones de IT sont *a priori* indépendants : la sélectivité à un objet par exemple ne correspond pas à la sélectivité attendue au niveau de la couleur. La couleur à laquelle sont sélectifs ces neurones est en effet très rarement la couleur prototypique des objets sur lesquels ils répondent de façon sélective (Komatsu et Ideura, 1993). Il est possible que les neurones de IT possèdent le même type de comportement mais sur un espace plus riche encore. On considère souvent comme un détail le fait que ces neurones sont souvent sélectifs à des stimuli totalement décorrélés. Par exemple, un neurone sélectif à la présentation d'un objet (i.e. une chaise) peut également être sélectif à un autre objet (i.e. un arbre), *a priori* sans rapport. Ce type de comportement est difficile à comprendre du point de vue du réseau connexionniste que j'ai présenté, mais à mon avis cela n'est pas incompatible. Une extension du modèle serait d'imposer à une carte neuronale de reconnaître plusieurs individus. En ajoutant des couches neuronales supplémentaires pour la reconnaissance des yeux comme on a pu le faire dans d'autres modèles de détection de visages (VanRullen et al, 1998), il sera, à mon avis, relativement facile d'atteindre ces résultats. Dans ce modèle, une couche intermédiaire de neurones est présente entre le modèle de V1 et celui de la carte neuronale sélective aux visages : ces neurones sont sélectifs à la présence d'une bouche, d'un oeil droit ou d'un oeil gauche dans l'image. La carte neuronale sélective à la présence d'un visage décharge si les neurones sélectifs à la présence d'une bouche et des yeux déchargent à des positions appropriées. On peut imaginer un modèle de reconnaissance basé sur le même principe, certains neurones intermédiaires étant sélectifs à la bouche ou aux yeux de certains individus. Je vais tenter d'exprimer dans ce cadre comment une représentation plus distribuée de la sélectivité pourrait émerger. Pour un neurone sélectif à deux individus par exemple, le neurone est excité par les neurones sélectifs à l'œil droit et ceux sélectifs à l'œil gauche des deux individus. Ces yeux doivent appartenir soit au premier individu, soit au second. Même s'il semble intolérable, pour certains modélisateurs, que le neurone réponde fortement quand l'œil droit du premier individu est présenté avec l'œil gauche du second, cette configuration a très peu de chance de se produire dans le monde réel. De plus en prenant une population de neurones répondant chacun sur plusieurs individus, si l'image d'un individu est présentée, de nombreux neurones répondront. Cependant, si un visage trafiqué est présenté très peu de neurones répondront, car peu de neurones combineront les caractéristiques des différents individus qui composent l'image. Une sélectivité des neurones de IT à plusieurs objets

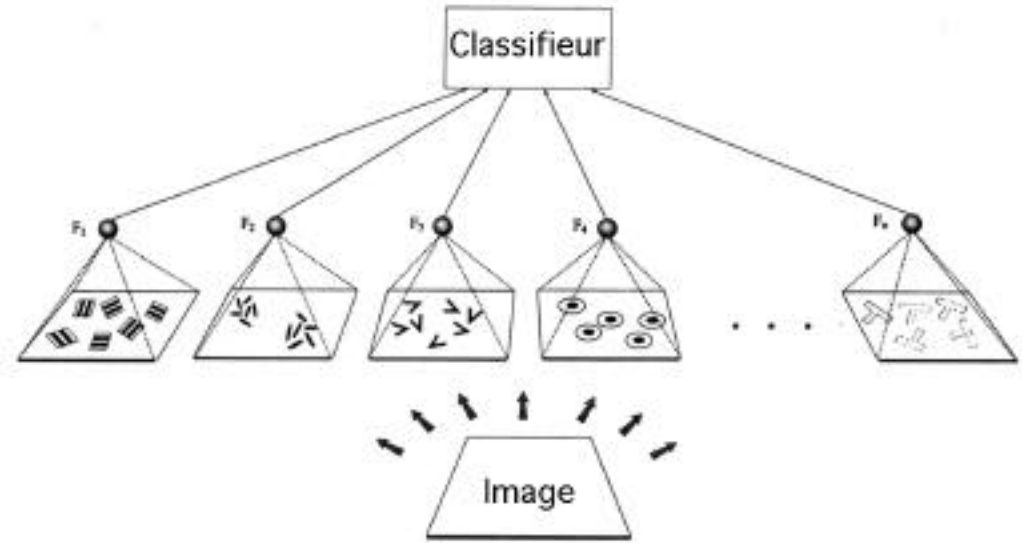


Figure 3.8 : modèle computationnel de Mel. Ce modèle est l'un des seuls à réellement poser la question de la performance. Il se décompose en deux couches, la première extrayant les attributs de l'image (Fi, orientation, coins, surfaces...) et la seconde les classant à l'aide d'un algorithme issu de l'ingénierie. Ce modèle est capable de reconnaître des images naturelles qu'il n'a jamais vues. Cependant en l'absence de couleur, ses performance sont assez faibles. Adapté de Mel (1997).

permettrait une grande résistance à la mort de certains neurones, ce comportement serait non redondant dans le sens où chaque neurone présente une sélectivité - à plusieurs objets - qui lui est propre. Cet exemple volontairement simple, illustre comment il serait possible d'obtenir une représentation plus distribuée de la sélectivité neuronale dans le modèle que j'ai présenté.

3.3.3 - Performances et aspects computationnels

La plupart des modèles de reconnaissance de visages se basent sur des outils mathématiques puissants. Je ne traiterai ici que des modèles d'inspiration biologique, bien que cette inspiration soit souvent très lointaine[20]. Il s'agit en général de modèles abstraits utilisant des réseaux probabilistes, des décompositions en composantes principales, ou des mémoires hétéro-associatives. Les performances de ces modèles ne sont cependant pas en rapport avec leur niveau d'abstraction. À titre de comparaison, sur la même base de visages, un modèle théorique utilisant un réseau probabiliste[21] commet 10 % d'erreurs de reconnaissance sur des images nouvelles (Samaria et Harter, 1994) alors que le nôtre n'en commet que 2,5 %. Cependant, dans ce modèle, la moitié des visages était utilisée pour l'apprentissage et l'autre moitié pour tester le réseau. Bien que nos résultats ne soient pas tout à fait comparables

[20] Les modèles purement symboliques en intelligence artificielle présentent des performances similaires.
[21] Chaînes de Markov Cachées.

puisque 8 vues sont utilisées pour l'apprentissage et 2 pour la phase de test, ils montrent cependant que la performance de notre approche rivalise et est même peut-être supérieure à celle utilisant des réseaux probabilistes.

Certains auteurs font de plus appel à un pré-traitement pour la reconnaissance des visages à l'aide de graphes élastiques (i.e. Würtz, 1997). Sur un visage, une matrice de points équidistants, qui correspondraient à des neurones, est appliquée. Ces points/neurones sont sélectifs à un *pattern* de contraste appris sur de nombreuses images de visages. Quand une image est présentée, les points/neurones optimisent leur position afin que le contraste local de l'image coïncide avec le pattern de contraste auquel ils sont sélectifs. Après ce prétraitement, un algorithme, comme une décomposition en composante principale, effectue la reconnaissance des visages sur ces points. Les auteurs pensent que ce prétraitement est nécessaire pour une reconnaissance efficace et soutiennent même que le comportement de ces neurones/points s'inspirerait du comportement des neurones réels. Cependant, il ne rapportent aucun argument vraiment convaincant à ce propos. De mon point de vue, les hypothèses que l'on a fait sur le comportement des neurones dans notre modèle semblent plus acceptables, de sorte qu'au regard des autres modèles, je ne pense pas prendre beaucoup de risques en affirmant que mon modèle de reconnaissance des visages est bien plus lié au comportement des neurones réels que ces modèles théoriques.

L'obtention d'une performance acceptable, la résistance au bruit et aux variations de contraste restent les problèmes les plus ardus à résoudre avec les modèles biologiques du système visuel. Les réseaux utilisant SpikeNET sont en fait bien seuls parmi ces modèles à atteindre un semblant de traitement d'image que l'on peut considérer comme efficace. Un modèle, Seemore, s'inspirant à la fois de la biologie et de l'intelligence artificielle est toutefois relativement performant (Mel, 1997). Dans ce modèle, les images sont décomposées en attributs simples - orientation, blobs, *end-stopped* - qui servent ensuite d'entrées à un puissant classifieur (figure 3.8). Le classifieur implémente des algorithmes optimisés du point de vue du traitement de l'information. Le modèle est entraîné sur 6 vues de 102 objets et testé sur 6 autres vues des mêmes objets (les objets sont découpés dans des photographies et présentés sur un fond noir). La précision de la reconnaissance est de 97 % dans le meilleur des cas et tombe à 80 % si les informations de couleur sont supprimées. Le taux de reconnaissance avec 30 % de bruit descend à 58 % et même jusqu'à 35 % si l'on supprime les informations chromatiques. À titre de comparaison, dans les mêmes conditions de bruit les performances de SpikeNET restent pratiquement inchangées. De plus, les performances de SpikeNET, à la différence de Seemore, se basent uniquement sur les contours des objets et l'information

chromatique n'est pas utilisée. Malgré la différence entre les bases d'images de Seemore et de SpikeNET[22], je ne pense donc pas que SpikeNET ait à avoir honte de ses performances, d'autant plus que du point de vue computationnel, la balance penche nettement du côté de SpikeNET[23].

Un dernier modèle relativement intéressant du point de vue de la physiologie est celui de Lee et Seung (1999). Il s'agit d'un sous-modèle de décomposition en composantes principales dans lequel les coefficients des filtres ne peuvent être négatifs[24]. Dans ce type de modèle, les images se décomposent sur un ensemble de filtres ou fonctions de base. Contrairement à une décomposition en composantes principales, les filtres générés à partir de visages semblent avoir une signification. Certains représentent par exemple une bouche et d'autres différents types d'yeux. Les algorithmes d'apprentissage utilisés sont de plus compatibles avec des règles d'apprentissage au niveau local dans des réseaux de neurones. Bien qu'il s'agisse ici d'optimisation de fonctions mathématiques et de neurones possédant des niveaux d'activation continus, ce type d'approche est à mon avis celle qui se rapproche le plus du comportement de notre réseau. Les capacités des réseaux utilisant des niveaux d'activation continus sont en fait très proches, sinon équivalentes, de celles utilisant la latence de décharge des neurones. Il faut en fait attribuer la force de SpikeNET, non seulement à l'originalité du traitement des latences, mais également à la quantité phénoménale de neurones qui peuvent être simulés en un temps record (cf. annexe 2).

Notre modèle est étonnement simple, et ses fondements mathématiques minimums, par rapport aux autres modèles de reconnaissance de visages qui se réclament des neurosciences computationnelles. Ses performances sont cependant comparables sinon supérieures à celles de modèles classiques. Les simulations que j'ai présentées ici démontrent, si cela était encore nécessaire, que l'intelligence artificielle a tout bénéfice à faire confiance à la biologie.

Même si la majorité des électrophysiologistes condamnerait bien volontiers les simplifications et les hypothèses que j'ai pu faire, je pense qu'elles sont minimes au regard de

---

[22] Par analogie avec les autres modèles de la littérature, je confondrai à partir de maintenant SpikeNET avec les réseaux que j'y ai implémentés.

[23] Les images étant approximativement de la même taille dans les deux modèles, à puissance de calcul équivalante, SpikeNET est au moins 10 fois plus rapide que Seemore pour reconnaître les visages (SpikeNET, 1 seconde/visage 92x112 sur un PowerPC à 266Mhz, Seemore, 2 minutes/image 120x120 sur un Sparc-20 à 60Mhz).

[24] Dans la décomposition en composantes principales, l'image se décompose sur un ensemble de filtres pondérés pour chaque image et qui s'ajoutent linéairement. Les coefficients sur les filtres peuvent cependant être négatifs, ce qui signifie que les filtres (qui résultent d'un calcul matriciel relativement simple) n'ont aucune signification du point de vue biologique.

celles présentes dans les modèles du même type. Nous verrons cependant au chapitre suivant des modèles théoriques de traitement dans le système visuel, peut-être plus plausibles du point de vue de la biologie mais malheureusement très peu efficaces pour le traitement des images naturelles. Je tenterai de synthétiser certaines de ces approches pour construire un modèle général théorique, computationnellement efficace, pour la reconnaissance des objets de façon non supervisée.

4

Modèle général de reconnaissance des objets

...when Sherrington posited the Pontifical Neuron as coding for an entire experiential memory, the latest neurophysiological evidence suggests that he was both wrong and guilty of this form of anthropomorphism. Barlow - though much maligned - did advocate a more distributed representation of sensory experience, extending Sherrington's ministerial metaphor to the new concept of Cardinal cells, as representing the culmination of a processing hierarchy. [...] both of the theories are misleading in the sense that they may perpetuate the myth of a central arbiter or executive [...] and in so doing postulate the existence of cells which, on current evidence, are excessively specialised and selective. However, their analogy may improve by taking one step further back. Continuing in the ecclesiastical vein, I propose a cellular Priesthood, or better still members of a monastic order, each specialised and more or less independent of the others but with similar goals, who together form the synergistic brotherhood.

Guy Wallis (1994)

La reconnaissance des objets, la perception des formes - appelez-la comme vous le voulez - la question de savoir comment nous identifions les objets a fourni matière à penser aux philosophes, psychologues, ingénieurs et neuroscientistes durant ces dernières décennies.

Nombre des modèles et de théories existant sur le sujet se sont donc développés. Je résume ici les principales tendances à travers quelques modèles. Je m'inspirerai de certains

pour construire mon modèle général de reconnaissance d'objets dans SpikeNET et, une fois n'est pas coutume, je partirai donc d'une revue des modèles existants du système visuel.

4.1 - Un peu d'histoire

Pour commencer, comment traiter de modèle de vision sans parler des modèles de Marr et Biederman qui font toujours référence dans le domaine des neurosciences computationnelles et de la psychologie expérimentale. Bien que ces modèles soient plus directement liés aux sciences de l'ingénieur qu'à l'activité des neurones dans le système visuel, je pense qu'il ne peuvent pas être ignorés

David Marr[1] a été un des précurseurs à la fois de la vision artificielle et des corrélats neurophysiologiques qui peuvent en rendre compte. Il a tout d'abord imaginé le système visuel comme une organisation hiérarchique capable d'extraire les objets 3D d'une scène naturelle. Le modèle de Marr (1982) postule 3 niveaux d'intégration. Tout d'abord un niveau purement bidimensionnel, où sont encodés les contours des objets à partir des différences de luminance. Ensuite, un niveau $2D_{1/2}$ qui décrit l'orientation des surfaces et donne une certaine invariance à la rotation (les normales aux surfaces conservant les mêmes relations). Enfin, les surfaces sont extrapolées à partir des contours et de lois très simples issues de la psychologie de la Gestalt[2]. Ce troisième et dernier niveau est celui de la représentation tridimensionnelle des objets. Ce modèle princeps a motivé la plupart des modèles qui ont suivi, notamment le celui de la reconnaissance d'objets, indépendante de leur orientation.

Toujours dans la lignée des modèles historiques incontournables, le modèle d'Hummel et Biederman (1992) fait encore aujourd'hui référence en psychologie. Il implémente la théorie des *geons* qui seraient à la base de la reconnaissance des objets. Pour Biederman (1986), les objets peuvent être décomposés en éléments géométriques tridimensionnels indépendants : les *geons*. Biederman en décrit environ 36 types (i.e. cônes, pavés, anses...), la plupart des objets étant composés de trois ou quatre *geons*. La description d'un objet en mémoire dépend alors des *geons* qui le composent, de leur dimension et de leur position relative. Un modèle de

[1] Le livre "Vision" de David Marr (1982) est mondialement connu. Il constitue une édition posthume de son œuvre, Marr étant mort d'un cancer à 45 ans en 1980.

[2] Les lois de la Gestalt sont un ensemble de règles auxquelles obéit la perception des objets. Les éléments d'un même objet sont en général à des profondeurs similaires, de couleurs similaires, leurs contours sont fermés, ils peuvent être symétriques... La loi de Prägnanzl par exemple, postule que s'il y a compétition entre deux formes pour être reconnues en tant qu'objet, celle qui est reconnue est celle qui regroupe le maximum de ces caractéristiques de bases.

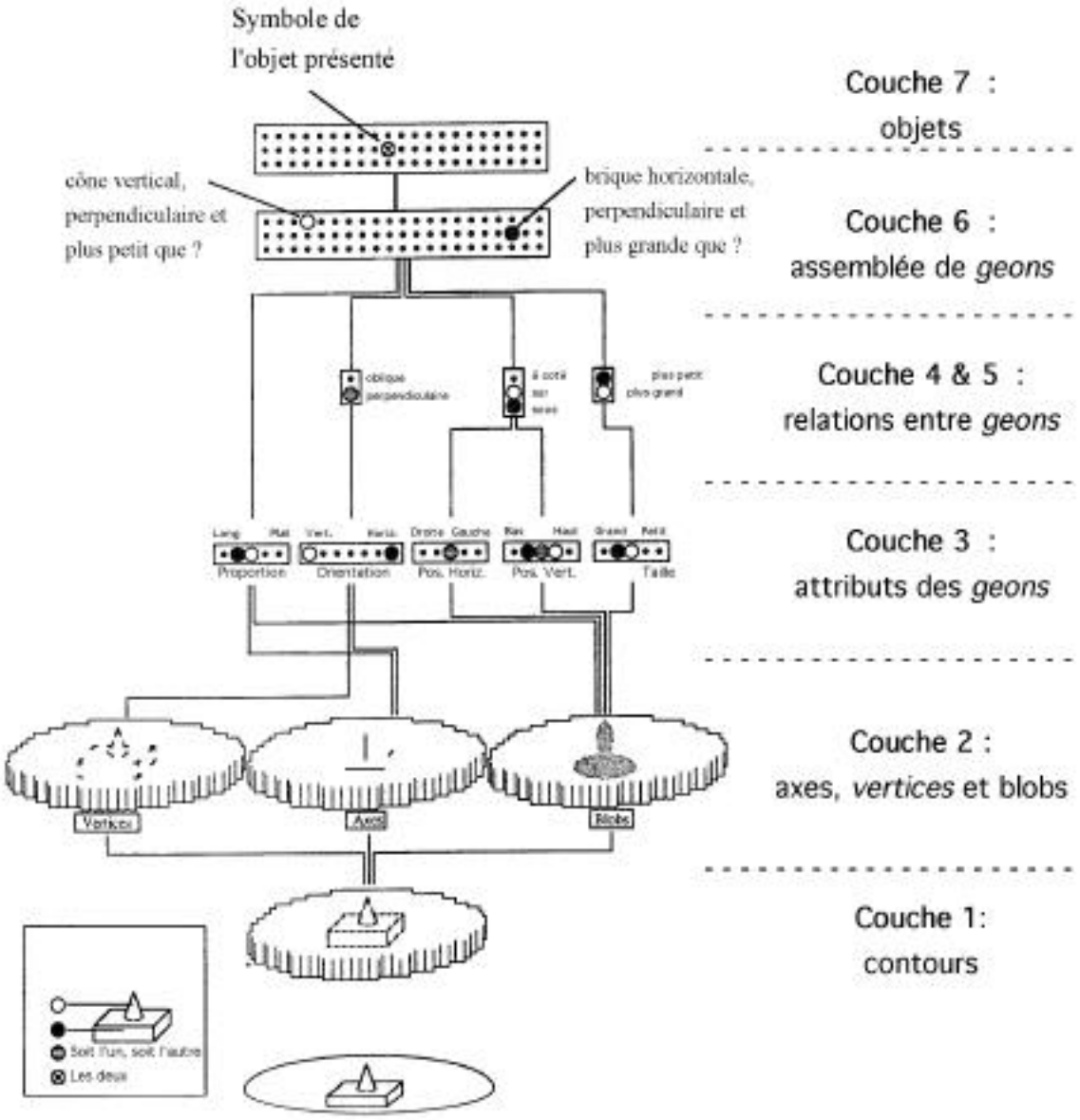


Figure 4.1 : modèle d'Hummel et Biederman. Le modèle est organisé hiérarchiquement : la première couche détecte les contours et la seconde des attributs plus complexes comme les coins, les axes et les surfaces (blobs). Les couches suivantes reconnaissent des objets de bases (*geons*), organisent leurs relations (orientation, proportion, position) et sont combinées pour détecter un objet. Adapté de Hummel et Biederman (1992).

reconnaissance a été construit sur ce principe (figure 4.1). Les deux premières couches extraient les contours des objets (contours, coins, axes...) et les autres couches tentent de construire des *geons* à partir de ces informations.

Dans ce modèle, les différentes propriétés d'un même *geon* sont liées par synchronie : les neurones encodant les attributs d'un même *geon* sont censés décharger en même temps[3]. Après la phase apprentissage, le modèle est capable de reconnaître les objets qui lui ont déjà

[3] Un signal externe aide également les neurones encodant les propriétés d'un même *geon* à devenir synchrones.

été présentés et des vues miroir de ces objets. Les réponses du modèles sont également peu affectées par des variations arbitraires de taille de l'objet et de rotations dans le plan ou dans l'espace. Pour rendre compte en électrophysiologie des réponses des neurones dans le système visuel qui dépendent de la vue de l'objet présenté (Rolls 1992; Wachsmuth et al, 1994; Booth et Rolls, 1998), Hummel et Stankiewicz (1996) ont complété ce modèle. En parallèle avec la décomposition en *geons,* ils ont ajouté une couche de neurones sélectifs à une vue des objets présentés. La dynamique du réseau reste cependant la même. Dans un modèle similaire, mais qui n'inclut pas la décomposition en *geons*, la synchronie des différents groupes de neurones est déterminée par la latence de décharge des neurones sur les objets : comme dans SpikeNET, les neurones ayant les latences les plus précoces - qui correspondent à des contrastes élevés - sont liés au même objet (Opara et al, 1996) et se synchronisent grâce aux interactions latérales.

Malgré leur intérêt historique, et leur connexion avec le comportement des individus, ces modèles présentent le désavantage de la non-homogénéité (dans le dernier modèle, la synchronisation ayant lieu uniquement dans les couches de haut-niveau). Non seulement leur connexion avec la biologie est discutable mais leur application au niveau du traitement d'image n'a jamais été réellement testée. Dans les paragraphes suivants, je présente des modèles plus récents et qui se basent de façon plus claire sur l'architecture du système visuel.

4.2 - Reconnaissance et traitements récurrents

Je vais commencer ma revue par trois récents modèles où le traitement récurrent est nécessaire à la reconnaissance des objets. Ces modèles à processus itératifs ou *feedback* font également intervenir l'organisation des neurones au niveau des colonnes corticales dont j'ai parlé dans l'introduction. Comme nous allons le voir, ces deux phénomènes sont intimement liés. Le modèle le plus tranché est sans aucun doute celui de Ullman (1995). Il présente un traitement bidirectionnel, ascendant pour la propagation des stimuli, et descendant pour celle des classes. Le stimulus est propagé de façon ascendante dans un modèle hiérarchique alors que les représentations des classes d'objets sélectionnées sont rétro-propagées - en fonction de l'attente du sujet - pour, au fur et à mesure de leur cheminement vers les niveaux hiérarchiques inférieurs, devenir invariantes à leur taille et à leur position. Les informations ascendantes et descendantes suivent des chemins différents mais interagissent au niveau des colonnes corticales dans des boucles amplificatrices (figure 4.2). Quand l'algorithme atteint un état stable, les neurones qui restent activés définissent l'objet qui serait présent dans

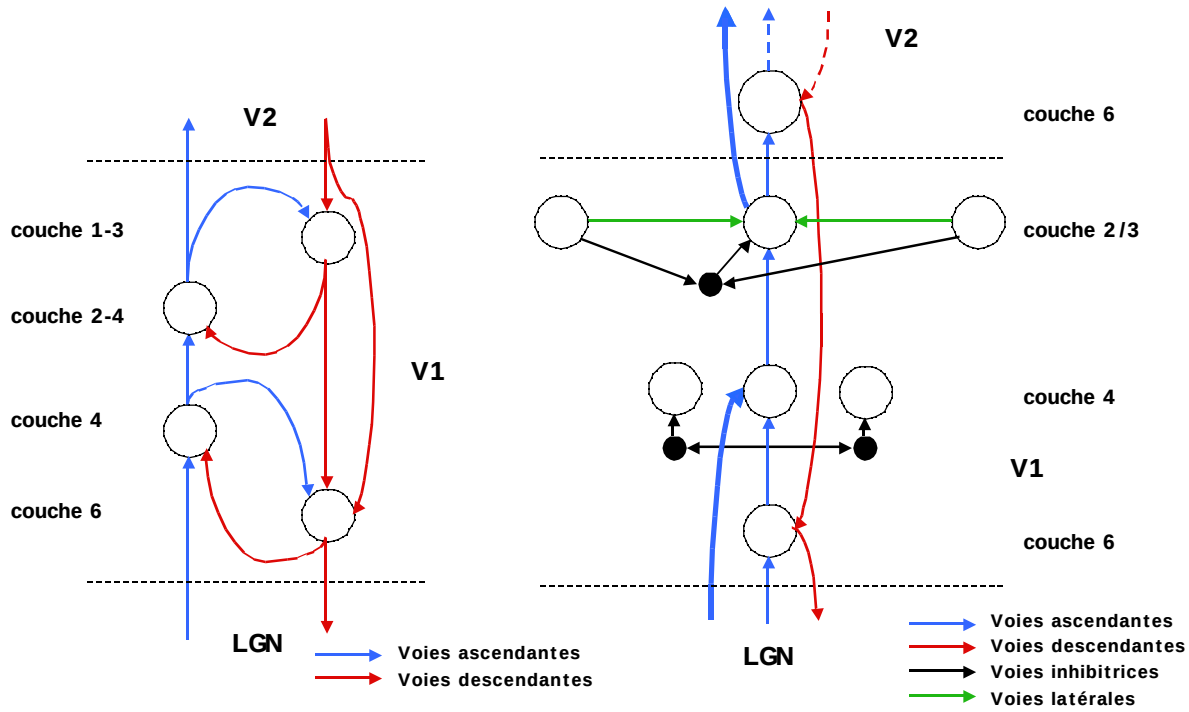


Figure 4.2 : à gauche, le modèle bidirectionnel de Ullman. Les informations ascendantes (stimuli) interagissent avec les voies descendantes (classes). Quand une voie est activée par les deux, elle est renforcée. La voie la plus active à l'issue de cette interaction permettrait de sélectionner l'objet reconnu. À droite, le modèle de Grossberg. La voie descendante sert à la fois de renforcement des couches supérieures vers les couches inférieures (par exemple quand un neurone a été activé dans V2 dans la couche 6, il tente de renforcer ses entrées). L'attention spatiale passerait par la même voie, préactivant les neurones de la couche 6 en cascade à partir des aires visuelles de plus haut niveau. Ce modèle a l'avantage d'être dynamique et d'intégrer l'inhibition alors que le modèle de Ullman ne la prend pas du tout en compte (cf. texte). Adapté d'Ullman (1995) et de Grossberg (1999).

l'image et la voie qu'il a fallu suivre pour détecter cet objet. L'interprétation en termes de propriétés des neurones au sein d'une colonne corticale est intéressante. Comme on l'a vu dans l'introduction, le cortex est une structure laminaire composée de 6 couches. Ullman propose un rôle pour chaque couche, certaines intervenant dans la voie ascendante et d'autres dans la voie descendante. En accord avec certaines données neurophysiologiques, les couches 2 et 4 seraient responsables de la propagation *feedforward* et les couches 1, 3 et 6 de la propagation *feedback*. Cependant, ce modèle ne prend pas en compte les interactions locales entre les colonnes, notamment les interactions inhibitrices - dont on a vu dans l'introduction et dans nos modèles qu'elles pourraient jouer un rôle primordial dans la compétition pour la représentation d'un objet. De plus, le type d'approche qu'il propose semble par trop restrictif car (i) il n'est pas possible *a priori* de détecter les objets qui n'ont pas été présélectionnés et (ii) une préactivation massive des voies descendantes, à toutes les échelles pour un ensemble de classe, résulterait probablement en une totale confusion. Le modèle est en fait avant tout

théorique et la simulation présentée sur la reconnaissance d'un unique objet est loin d'être convaincante.

Le modèle de Grossberg (1999) prend en compte les interactions locales ignorées dans le modèle de Ullman. Les couches 2 et 4 sont toujours responsables du traitement *feedforward* et la couche 6 du *feedback*. La dynamique semble cependant bien différente du fait des interactions locales au sein de chaque couche. Au sein d'une colonne corticale, les neurones des différentes couches interagissent pour supprimer les signaux qui ne sont pas suffisamment importants. Il existe deux systèmes d'inhibitions. Tout d'abord, les décharges neuronales se propagent vers les couches 2 et 3 où les neurones sont en compétition directe. Cependant la compétition serait de type normalisante, de sorte qu'un neurone doit coopérer avec ses voisins pour décharger. Si le neurone est excité uniquement par ses afférents de la couche 4, son activation sera très faible du fait de l'inhibition des autres neurones. Si cependant deux neurones des couches 2 et 3 s'activent mutuellement, ils peuvent dépasser cette inhibition et propager leur activation à l'étape suivante, dans les couches 4 et 6 de l'aire corticale de niveau supérieur. Il y a alors feedback vers la couche 6 du niveau inférieur pour amplifier le signal transmis par ces neurones. Ce feedback permet de renforcer l'activation de la voie *feedforward* concernée en supprimant les signaux dans les voies *feedforward* voisines, moins sélectives. Du point de vue computationnel, cela permet de filtrer de façon efficace les voies les plus actives et éventuellement d'implémenter l'algorithme MAX (Riesenhuber et Poggio, 1999) que l'on verra par la suite. Ce modèle présente également l'avantage d'expliquer l'attention spatiale qui passerait par une activation en cascade de la couche 6 - couche 6 de V4 vers couche 6 de V2 vers la couche 6 de V1 - pour renforcer les signaux visuels dans certaines zones et donc favoriser le traitement des objets à ces positions.

La couche 5 n'est présente ni dans le modèle de Ullman (1995), ni dans celui de Grossberg (1999). Les neurones de la couche 5 seraient en fait impliqués dans la régulation et éventuellement dans la génération d'oscillations qui permettrait de séquencer les activations continues en un traitement discret des informations rétiniennes (Silva et al, 1991). Ce type de mécanisme pourrait en effet être responsable du traitement discret des images que j'ai implémenté dans SpikeNET. Le modèle de Rodemann et Körner (Rodemann et Körner, 2000; Körner et al, 1999), reprend celui de Grossberg tout en intégrant cette propriété. La décomposition en colonnes corticales est similaire à celle de Grossberg mais plus complexe et réellement orientée vers les sciences de l'ingénieur (dans le sens où chaque type de neurone dans chaque couche a une fonction bien précise). Le modèle de Rodemann et Körner est un modèle de synthèse intéressant qui reprend à la fois la propagation dans SpikeNET, et la

fonction MAX du modèle de Riesenhuber et Poggio que nous verrons plus loin. C'est un modèle hiérarchique du même type que SpikeNET : les neurones déchargent à des latences qui sont traitées par les neurones postsynaptiques. La décharge à la latence la plus précoce est sélectionnée et les autres sont ignorées, conformément à l'algorithme MAX. De plus, les noyaux intralaminaires du thalamus (ILN) fourniraient une base temporelle - peut-être par l'intermédiaire de la couche 5 ? - activant ainsi de façon graduelle l'ensemble des aires visuelles et permettant aux neurones de décharger, la latence de leur décharge dépendant alors de leur activation par l'aire visuelle de niveau inférieur[4]. La latence de décharge des neurones dépend donc à la fois de l'adéquation du stimulus avec la sélectivité du neurone et du niveau d'activation de l'ILN. Ce mécanisme complexe de génération de décharges possède l'avantage de fournir une base temporelle discrète, les oscillations en provenance de l'ILN permettant en fait de segmenter le flux visuel (figure 4.3).

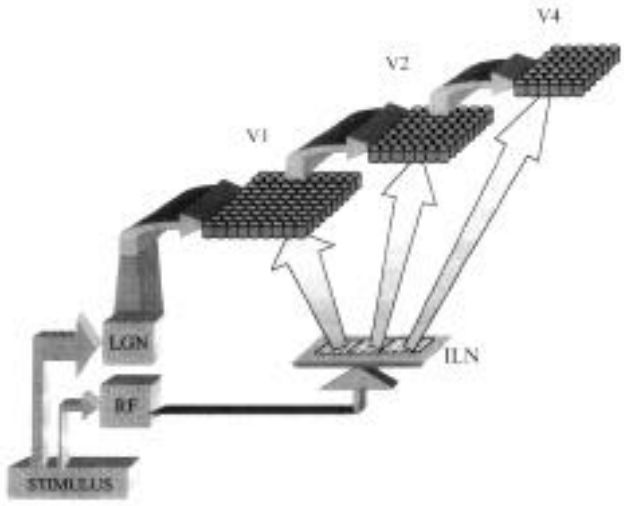


Figure 4.3 : modèle de Rodemann et Körner. Les neurones sont organisés de façon hiérarchique comme dans le système visuel. L'excitation suit deux voies principales, la voie classique LGN-V1-V2-V4 pour le traitement des caractéristiques du stimulus et une voie indirecte de régulation qui passe par la formation réticulée (RF) et l'ILN. La voie indirecte permet de synchroniser les neurones pour initier une propagation dans la voie classique (cf. texte). Adapté de Körner et al (1999).

Dans le modèle de Rodemann et Körner (2000), le mécanisme de *feedback* inspiré de celui de Grossberg est à mon avis le point le plus interessant : des informations en retour provenant de la couche détectant les objets peuvent biaiser les neurones des couches inférieures et modifier la sélection du MAX à ce niveau. Par exemple, si la forme d'un marteau active à la fois les neurones sélectifs aux marteaux et ceux sélectifs aux pioches, ces deux objets réactivent les couches de bas niveaux codant les orientations. Cette réactivation peut biaiser la décharge dans cette couche : si, à une position donnée, un ensemble de neurones reconnaît

[4] Il semble qu'en fait les stimulations de la couche inférieure ne soient pas suffisantes pour faire décharger le neurone. A la différence de SpikeNET, les neurones commencent par intégrer tous les EPSPs en provenance de la couche inférieure. Ensuite, une activation graduelle de l'ILN leur permettrait de décharger, la phase de leur décharge par rapport à l'ILN rendant compte de leur sélectivité.

une orientation et que ces neurones sont biaisés par l'activité des couches supérieures, dans la mesure où cela est compatible avec le stimulus, l'algorithme MAX choisira l'orientation la plus proche à la fois du stimulus et de l'un des objets à cette position. De cette façon, les neurones sélectifs aux marteaux seront activés plus fortement que ceux sélectifs aux pioches.

Malgré l'indéniable inspiration biologique de ce modèle, son intérêt reste limité. Du point de vue de la propagation *feedforward*, ce modèle, clairement inspiré de SpikeNET[5] (VanRullen et al, 1998), n'apporte à mon avis rien de nouveau. Toutefois, la rétroaction des couches supérieures sur les couches inférieures est très séduisante et pourrait facilement être intégrée dans le modèle de reconnaissance du chapitre III.3 si cela se révélait nécessaire[6]. La principale critique que l'on peut faire à ce modèle est que ses performances restent très faibles: le modèle n'a pas été testé avec des images naturelles et se limite à catégoriser 4 ou 5 images d'objets dessinés aux traits[7]. Les auteurs semblent vouloir synthétiser trop de techniques à la fois : partant de postulats biologiques, ils en arrivent finalement à utiliser des processus difficilement justifiables - i.e. le fait que le début de la propagation des latences soit synchrone avec le début de l'activation de l'ILN. Si dans SpikeNET, les processus mis en jeu font appel à moins de structures biologiques - couches neuronales, ILN... - ce modèle préserve ainsi une simplicité et une homogénéité qui font à mon avis sa force.

Dans le reste de cette discussion, je vais me concentrer sur les modèles de reconnaissance *feedforward*, les processus de feedback, comme je l'ai longuement discuté au chapitre II.7 et III.3, n'étant probablement pas indispensables au traitement visuel rapide.

4.3 - Propagation *feedforward* et attention spatiale

Si les modèles que je vais présenter ici considèrent que le traitement visuel peut s'effectuer de façon *feedforward*, ils partent également du principe qu'il serait nécessaire de préactiver certaines positions dans le champ visuel. En quelque sorte, le système visuel ne serait pas assez puissant pour effectuer l'ensemble du traitement relatif à une image rétinienne et l'attention spatiale, en focalisant le traitement visuel, permettrait de résoudre ce problème (Rensink et al, 1997; O'Regan et al, 1999). Olshausen et al (1993) proposent donc un modèle

[5] Les discussions que j'ai pu avoir avec l'auteur indiquent qu'il s'est inspiré de SpikeNET pour réaliser son modèle.

[6] Dans le modèle de Rodemann, les images sont propagées plusieurs fois, la rétroaction des couches supérieures sur les couches inférieures n'intervenant qu'à partir de la seconde propagation.

[7] Les auteurs travaillent cependant pour Honda industrie et l'on aurait pu croire que leur principal souci aurait été la performance du modèle.

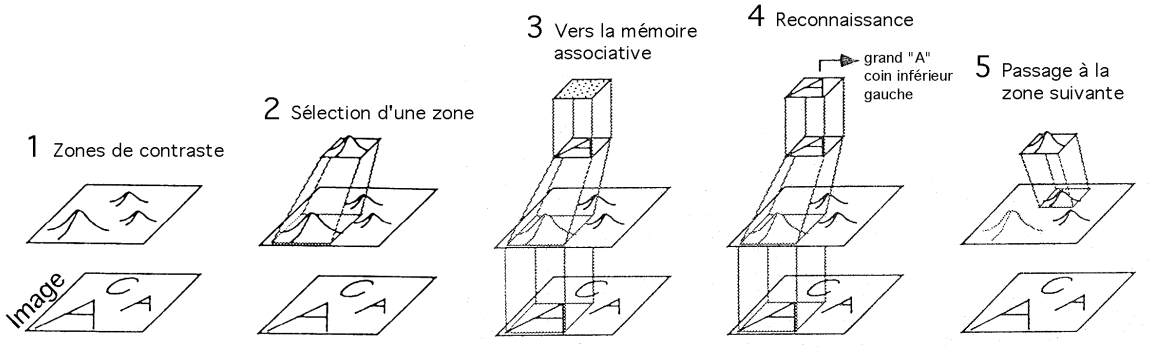


Figure 4.4 : modèle préattentif de Olshausen et al. Au sein de l'image, les zones de plus fort contraste sont sélectionnées (1 et 2) pour être traitées en premier (3). Ces zones sont dirigées dynamiquement vers les zones permettant de reconnaître les formes et les objets (4). Quand la zone de plus fort contraste est propagée, c'est au tour de la seconde zone de contraste élevé d'être traitée (5). Adapté d'Olshausen et al (1993).

de routage dynamique des informations visuelles vers les aires de haut niveau. Des zones d'intérêt doivent donc être définies sur lesquelles le traitement se fera en priorité. Les différentes zones d'intérêt sont traitées de façon séquentielle en fonction de leur saillance. Cette idée semble initialement avoir été développée par Koch et Ullman (1985) mais Olshausen et al (1993) ont été les premiers à effectuer réellement des simulations pour la reconnaissance de caractères.

Dans le modèle d'Olshausen et al (1993), un filtre grossier détecte les zones de contraste les plus élevées. La zone de plus fort contraste est traitée en premier, ce qui correspondrait à un phénomène préattentif dans le système visuel. L'intérêt du modèle réside en fait dans le routage des informations : quelle que soit la taille de la fenêtre attentionnelle, un mécanisme de contrôle permet de ramener l'objet à une taille constante qui sera utilisée par la suite pour la reconnaissance. Ce mécanisme de routage est en fait très simple : en fonction de la taille de la fenêtre attentionnelle, certaines connexions synaptiques sont activées et d'autres sont inhibées. En fonction des connexions synaptiques activées, le zoom et le recentrage sur la carte d'entrée seront différents (figure 4.4). Dans une version plus élaborée de ce modèle (Olshausen et al, 1995), l'invariance à la taille est réalisée avant le routage dynamique pour rendre compte de la perte de résolution du système visuel à la périphérie. Pour 3 échelles par exemple, une petite fenêtre centrée sur l'image est propagée directement - cela correspond à la fenêtre fovéale - alors que des fenêtres de tailles double et quadruple sont d'abord ramenées à la taille de la fenêtre fovéale avant d'être propagées au niveau suivant. Les formes à la périphérie sont en effet de résolution réduite dans la rétine - car la densité en cellules ganglionnaires est très faible dans ces zones - et il n'est donc pas nécessaire, ni logique, de les

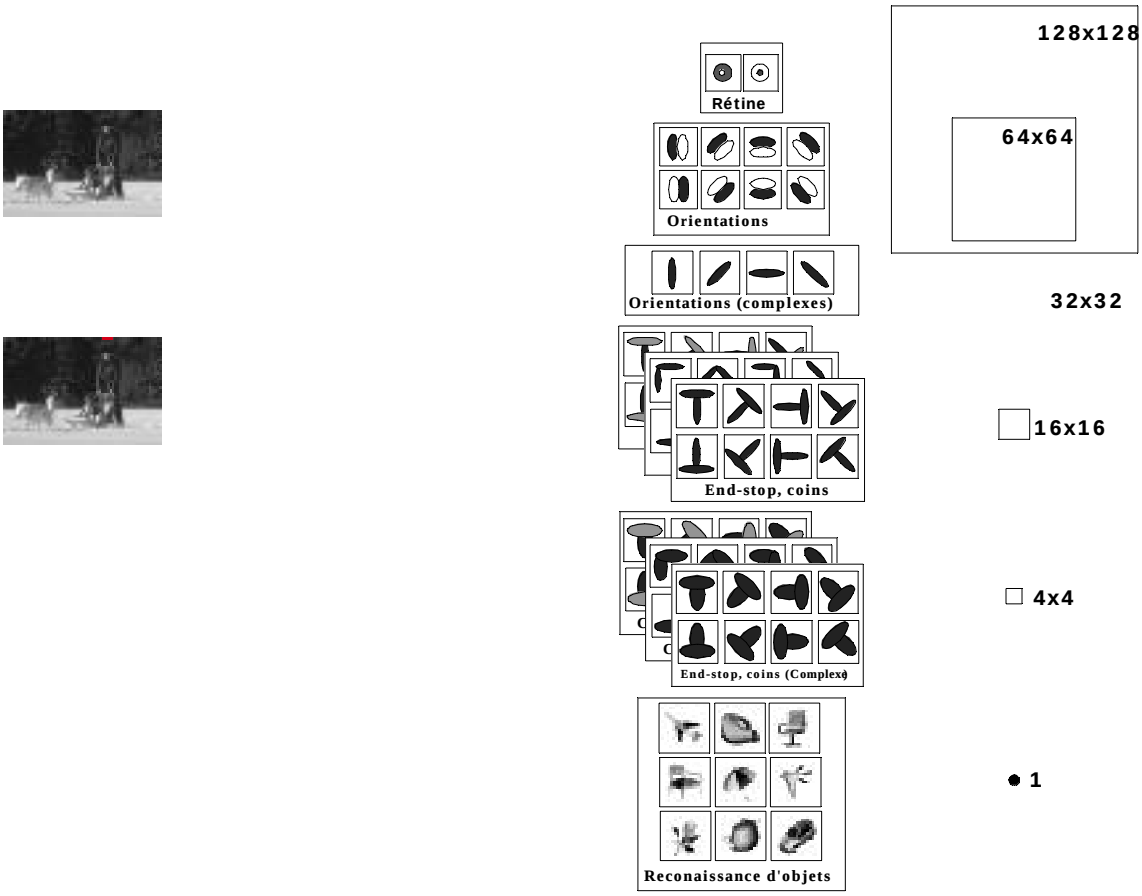


Figure 4.5 : A, reconstruction de l'image en fonction de la latence relative des décharges des neurones (les premiers 1 % des décharges). En favorisant certains neurones à certaines positions (attention 1, 2 et 3), les latences de décharge de ces neurones sont plus précoces. Cela permet de favoriser le traitement à ces positions. B, implémentation du principe exposé en 1 dans SpikeNET. La taille des cartes neuronales diminue avec le niveau hiérarchique pour atteindre 1 seul neurone par objet dans la dernière couche. Le réseau est entraîné sur 9 images d'objets de taille 64x64 (représentées dans la dernière couche). Si l'on présente plusieurs objets dans la rétine (de taille 128x128), l'activation induite par les différents objets présentés se mélange et les performances du réseau sont très faibles. Si l'attention est focalisée sur un objet cependant, son traitement est favorisé et le réseau le catégorise correctement. Adapté de VanRullen et al (2000c).

traiter à la résolution la plus élevée. Ces trois échelles subissent ensuite un routage dynamique en fonction des zones de saillance qu'elles contiennent comme dans le premier modèle[8].

Bien que le routage dynamique et la reconnaissance des objets en soient absents, le modèle de Koch et al (Koch et Ullman, 1985; Itti et al, 2000) est plus élaboré pour la sélection de la zone de traitement. Il prend en compte la couleur, l'intensité et l'orientation, rendant ainsi compte du comportement humain de recherche de cible dans une scène complexe. Dans ces

[8] Le routage dynamique prend alors uniquement en compte le recentrage et non plus le recentrage et le zoom. Les auteurs affirment que le routage des trois fenêtres pourrait être réalisé en parallèle (ils ne l'ont cependant pas implémenté).

modèles, le passage d'une zone à l'autre dépend non seulement de la saillance de cette zone mais également de sa distance par rapport à la zone en cours de traitement, un comportement plus en rapport avec le comportement humain (Reinagel et Zador, 1999).

Les corrélats biologiques de ce type de processus, notamment pour le routage dynamique, restent cependant limités. Les auteurs soutiennent que le Pulvinar pourrait être responsable de ce phénomène mais les arguments qu'ils apportent sont loin d'être totalement convaincants. De plus, la réallocation de l'attention pour le traitement séquentiel des objets prend un certain temps - environ 50 ms - et semble difficile à accorder avec les expériences de catégorisation rapides que nous avons montrées, où la cible n'est pas systématiquement l'objet de plus fort contraste présent dans l'image. Ces processus interviendraient donc uniquement quand une recherche visuelle est nécessaire pour localiser un objet.

VanRullen et al (2000c) ont également montré, à l'aide d'un réseau très simple dans SpikeNET, que l'attention spatiale pouvait biaiser la propagation de tel ou tel objet et ainsi résoudre le problème de la convergence des aires visuelles. Le modèle est totalement convergent et chaque neurone de la couche supérieure est sélectif à un objet (figure 4.5), la rétinotopie étant complètement absente à ce dernier niveau. Des expériences montrent que l'attention spatiale aurait tendance à privilégier des positions spatiales dès V1 (Kastner et al, 1998; Chelazzi et al, 1999). En implémentant un tel mécanisme dans SpikeNET, cela permet d'obtenir, lors de la présentation simultanée de 2 stimuli, une réponse du modèle sur un des deux objets en fonction de la position de l'attention spatiale (VanRullen et al, 2000c). Sans attention spatiale, du fait de la convergence totale, les traits des différents objets présentés simultanément se mélangeraient. Malgré la simplicité du modèle, il démontre la faisabilité de ce type d'approche dans SpikeNET[9]. Les projections entre cartes neuronales de différentes tailles que j'ai depuis lors implémentées (cf. annexe 2) permettront, je l'espère, de donner une dimension plus réaliste à ce type de modèle en étendant ses capacités au traitement des images naturelles.

[9] Du point de vue computationnel, étant donné le nombre de neurones (80 000), les performances du modèle restent relativement faibles, les 9 images d'objets ne pouvant être reconnues qu'à 4 positions dans l'image. La principale raison est probablement que je n'avais initialement implémenté que des projections entre les cartes de différentes tailles très rudimentaires (cf. annexe 2 sur le zoom dit naïf).

4.4 - Emergence de représentations neuronales

Ce dernier paragraphe introduit des modèles comparables à mon modèle de reconnaissance des objets dans la mesure où la dynamique en est totalement *feedforward*. Commençons par le modèle de Riesenhuber et Poggio (1999). L'architecture du modèle et les couches neuronales correspondent, dans une certaine mesure, à celles observées dans le système visuel : les neurones sont organisés hiérarchiquement et la convergence totale comme dans le modèle de VanRullen (2000c). Cela signifie que partant de la rétine et traversant les différentes couches, la rétinotopie n'est plus présente au plus haut niveau, les neurones encodant chacun une vue d'un objet quelle que soit sa position dans l'image fournie en entrée au réseau. La dynamique de propagation n'est pas simulée, l'activité des neurones étant continue et proportionnelle au niveau de contraste dans la rétine. A chaque étape, seul le maximum de sortie au sein d'une couche neuronale est pris en compte à un point donné. Ce modèle est capable de reconnaître des *paperclips* de différentes tailles et dans diverses positions (figure 4.6). La sélectivité des neurones en termes d'invariance à la taille et à l'angle de vue des objets semble reproduire assez fidèlement le niveau d'activation des neurones de IT chez le singe sur les mêmes stimuli. Il serait intéressant de tester ce type de modèle sur des images naturelles[10], le principal défaut étant à mon avis que les connexions entre les neurones sont fixées à l'aide de fonctions mathématiques *ad hoc* et que la dynamique de décharge des neurones n'est pas prise en compte. Dans une certaine mesure ce réseau se rapproche du réseau décrit par Vetter et al (1995) utilisant des *radial basis function*[11] (RBF) pour reconnaître le même type de *paperclips*. Ces fonctions RBF tentent en fait de faire correspondre un stimulus en entrée à une forme grossière définie par un ensemble de fonctions gaussiennes. Etant donné la simplicité du processus, les performances du modèle sont très impressionnantes et pourraient rendre compte d'une partie de la sélectivité des neurones réels.

Les modèles que je viens de présenter, pas plus que SpikeNET, ne prennent en compte la séquence de présentation des images : les réseaux de neurones reconnaissent des stimuli présentés de façon ponctuelle et l'ordre de présentation des stimuli importe peu. Il est cependant probable que, dans IT, la séquence de présentation soit responsable en partie des

[10] Des performances similaires à celles observées sur les *paperclips* ont été obtenues sur des images de synthèse de voitures (Riesenhuber, communication personnelle).

[11] Une vue d'un objet est décrite en termes d'un ensemble de gaussiennes à des positions précises dans l'image et ayant chacune une influence différente sur le niveau d'activation d'une unité.

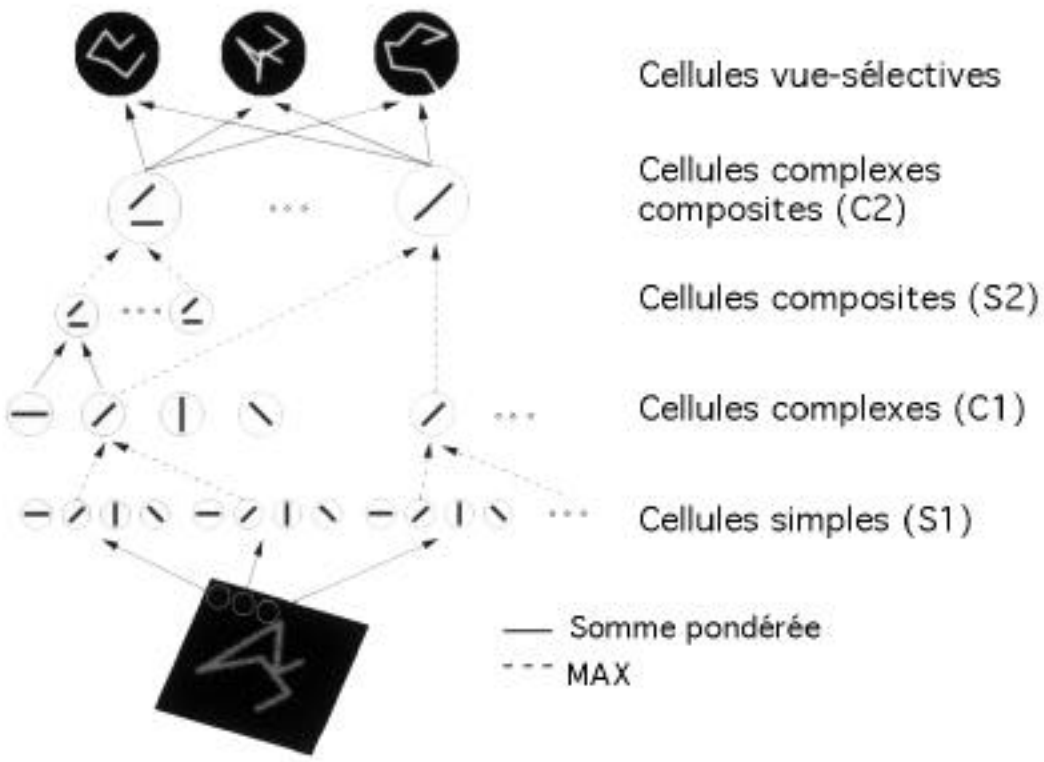


Figure 4.6 : modèle de Riesenhuberg et Poggio (1999) pour la reconnaissance de *paperclips* (formes géométriques 3D définies par des barres interconnectées). Ce modèle est totalement convergent au sens où chaque cellule de plus haut niveau ne reconnaît qu'un *paperclip*. Les neurones effectuent soit une somme pondérée soit un MAX sur leurs entrées (cf. texte). Outre les performances intéressantes du réseau, le comportement des neurones est similaire à celui de neurones réels sur les mêmes stimuli. Adapté de Riesenhuberg et Poggio (1999).

réponses sélectives des neurones invariantes à la position et à la taille des objets présentés. En effet quand le neurone décharge sur un objet présenté dans le champ visuel, il y a toutes les chances qu'il décharge encore quelques centaines de millisecondes plus tard même si cet objet a un peu bougé. Le neurone serait alors capable d'associer ces deux représentations et le fait que leurs occurrences soit très rapprochées dans le temps ne serait donc pas anodin. Le modèle de Wallis et Rolls (Wallis, 1994; Wallis et Rolls, 1997) pose clairement cette question et implémente une règle d'apprentissage qui prend en compte la séquence de stimuli présentés pour renforcer/affaiblir les connexions entre les neurones.

Le modèle est aussi simple que possible et reprend l'organisation hiérarchique exprimée précédemment (figure 4.7). Il se base sur 3 postulats. Tout d'abord, le champ récepteur des neurones augmente avec le niveau hiérarchique. Ensuite, des interactions locales inhibitrices au sein des cartes corticales implémentent une compétition "douce" pour faire la balance entre sélectivité et représentation distribuée. Enfin, une règle d'apprentissage qui prend en compte le temps est implémentée. Cela signifie que, si un neurone décharge sur un premier stimulus

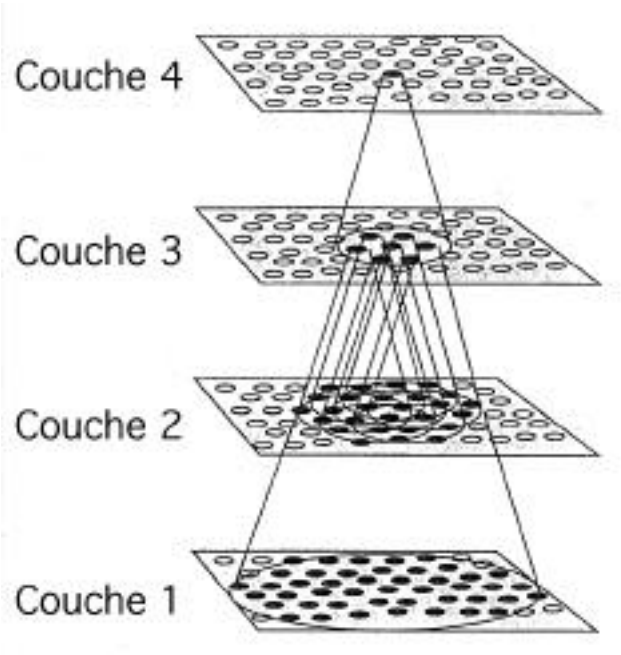


Figure 4.7 : modèle de Wallis et Rolls. Ce modèle est totalement convergent : les premières couches encodent les attributs de l'image (équivalent de V1) et les dernières couches les propriétés des objets (équivalent de V4-IT). L'originalité du réseau réside dans l'apprentissage qui dépend en fait de la séquence d'images présentées. Un même objet est présenté à plusieurs positions voisines dans le champ visuel. Les neurones dans les couches les plus élevées (2,3 et 4), du fait qu'ils restent actifs entre deux présentations, s'organisent automatiquement pour reconnaître les différentes vues des objets présentés. Les interactions latérales permettent également aux neurones de se "regrouper" s'ils encodent des attributs des objets communs. Adapté de Wallis et Rolls (1997).

et qu'un second stimulus présenté juste après conduit également le neurone à décharger, les connexions synaptiques activées dans le second cas seront renforcées. Dans ce modèle, au sein de 4 couches, seuls les neurones de la première couche - représentant V1 - ont des champs récepteurs fixés *a priori*[12]. Après apprentissage sur des visages à différentes positions, chaque photographie de visage étant présenté séquentiellement et se déplaçant sur l'ensemble du champ visuel du modèle, les neurones du dernier niveau deviennent sélectifs à ces visages. Du fait de la dynamique de l'apprentissage, leurs réponses est plus ou moins invariante à la translation et/ou à la position de ces visages. Certains problèmes restent cependant posés comme celui de la présentation simultanée de plusieurs formes. De plus, lors de l'apprentissage, si deux objets sont présentés en même temps, on a tout lieu de croire que les neurones mélangeront ces deux stimuli.

Bien que les performances du réseau soient sans commune mesure avec celles de SpikeNET[13], il présente pourtant l'avantage de montrer dans quelle mesure l'invariance à la vue ou à la position peut émerger car l'algorithme implémenté est totalement non supervisé. Je vais donc tenter de voir comment je pourrais m'en inspirer dans un modèle général de reconnaissance visuelle utilisant SpikeNET.

[12] Par des différences de gaussiennes, qui peuvent être à 4 fréquences spatiales différentes.

[13] Le nombre d'images de visage utilisées (7 dans le cas de l'invariance à la position et 21 dans le cas de l'invariance à la vue) reste très faible et on ne peut pas réellement parler de traitement d'image. De plus le réseau n'est pas testé avec de nouveaux stimuli.

4.5 - Vers un modèle général de reconnaissance

Le bref aperçu qui précède montre que les modèles de reconnaissance d'objets conçu pour expliquer le comportement du système visuel sont relativement différents les uns des autres sans qu'aucun d'eux ne soit réellement efficace pour le traitement des images. Parallèlement, les systèmes de reconnaissance d'objets en intelligence artificielle tentent de développer des applications efficaces, sans réellement y parvenir. Cependant les deux catégories de modèles, que l'on aurait pu croire confondues, ont très peu d'interactions. Les modèles du système visuel permettent rarement de développer des applications efficaces et les systèmes conçus en intelligence artificielle s'inspirent relativement peu de la biologie.

4.5.1 - Intérêts des autres modèles pour la détection d'objets

Le modèle de reconnaissance d'objet que j'ai développé dans le chapitre précédent avec SpikeNET se trouve étonnamment seul à constituer à la fois un modèle du système visuel et une application efficace pour la reconnaissance des objets. De plus les hypothèses sont minimales. La topologie du réseau reprend celle simplifiée du système visuel. Les neurones sont de simples neurones IFs couramment utilisés pour modéliser le comportement des neurones réels (Reich et al, 1997; Reich et al, 1998). Le fait que chaque neurone ne puisse décharger qu'une seule fois semble plutôt une contrainte qu'un avantage et la *shunting inhibition* rapide, relativement plausible du point de la biologie, n'est de toute façon pas nécessaire pour la reconnaissance des chiffres. Seule la rétinotopie des cartes neuronales de haut niveau et la règle d'apprentissage utilisée peuvent éventuellement poser problème à l'électrophysiologiste.

Je souhaiterais donc étendre les résultats que j'ai présentés à un modèle de reconnaissance générale des objets à la fois plus proche de la biologie et plus efficace. Concernant, la revue des modèles que j'ai pu présenter ici, je pense que certains peuvent nous être utiles. Je ne souhaite pas discuter des premiers modèles pour des raisons que l'on peut qualifier d'ontologiques : je ne suis pas persuadé que le feedback soit nécessaire à la fois pour la reconnaissance des objets par les sujets humains et dans les réseaux de neurones modélisant ces propriétés. Les expériences électrophysiologiques et psychophysiques vont également dans ce sens. De plus, les modèles que j'ai présentés ne nécessitent aucun *feedbak* et sont pourtant très efficaces, même en présence de bruit. Il serait inutile de compliquer le modèle pour y rajouter des processus qui, bien que probablement utiles, ne sont pas vitaux dans un premier temps pour obtenir un niveau de performance acceptable avec SpikeNET.

Concernant l'attention spatiale, ma préférence va bien évidemment au modèle de VanRullen et al (2000c) que je vais tenter d'intégrer en le modifiant dans mon modèle général. Le modèle de Olshausen et al (1993) de routage dynamique nécessite un système de contrôle très complexe qui n'a encore jamais été observé. Le processus préattentionnel utilisé, et auquel les modèles de Koch (Koch et Ullman, 1985; Itti et al, 2000) font écho, est très intéressant mais ne semble intervenir que dans une approche exploratoire des images et non dans leur catégorisation rapide. Dans l'optique de construire un système efficace de reconnaissance rapide, la simplicité est de mise et si des processus ne sont pas absolument nécessaires, il ne semble pas utile de les prendre en compte dans un premier temps.

Pour en terminer sur les modèles de représentation dans le système visuel, le modèle de Riesenhuber et Poggio indique que le processus MAX a probablement une inspiration biologique. Ce processus cependant correspond à un mécanisme de *winner take all*[14] au niveau local. Les neurones les plus fortement activés sont les seuls à répondre au bout d'un certain moment. Ces mécanismes font probablement intervenir une forte inhibition latérale comme celle décrite par Grossberg (1999) et comme celle, - je tiens ici à le rappeler – qui est présente dans mes modèles. De plus, en présence de *shunting inhibition*, les neurones de SpikeNET, sélectionnent le maximum d'activation de la couche précédente, c'est-à-dire les latences les plus courtes. Le processus MAX décrit par Riesenhuber et Poggio est donc, pour moi, déjà présent dans SpikeNET.

Le modèle de Wallis et Rolls me semble d'un bien plus grand intérêt. La loi d'apprentissage non supervisé offre en effet un mécanisme pour l'organisation topologique des cartes neuronales pour la reconnaissance des objets. Les implications du point de vue computationnel par rapport à l'approche que nous avons utilisée restent, à mon avis, encore à découvrir. Dans mon modèle général de reconnaissance d'objets, j'intègre ce principe mais d'une façon détournée. Ce n'est pas en effet la répétition du stimulus et ses variations au cours du temps - en position, en orientation... - qui permettront à la sélectivité d'émerger. Il s'agira, comme nous allons le voir, d'un signal extérieur fourni par la voie dorsale.

4.5.2 - Tentative de synthèse

A partir du paragraphe précédent et des modèles de reconnaissance de visages et de chiffres que j'ai présentés au chapitre III.3, je vais tenter de définir les grandes lignes d'un

[14] Processus bien connu en modélisation où le neurone le plus activé inhibe les autres pour se retrouver seul à décharger.

modèle de reconnaissance général des objets. Ce modèle général serait de type non supervisé et reprendrait l'architecture du modèle du chapitre III.3. Sa première couche représenterait la rétine, telle que je l'ai modélisée jusqu'à maintenant. De même, les cartes neuronales représentant les différentes orientations seraient fixées comme elle l'étaient déjà. Dans la voie ventrale, deux couches représentant des neurones dans V4 et dans IT seraient capables d'apprendre à reconnaître des stimuli, la taille du champ récepteur des neurones et leur niveau hiérarchique déterminant le type de sélectivité qu'ils peuvent espérer obtenir.

L'originalité du modèle viendrait du fait que l'apprentissage de type non supervisé pourrait être dirigé par les neurones de la voie dorsale (figure 7.8). Un modèle de la voie dorsale, où les neurones sont hautement sélectifs à des mouvements dans différentes directions et à différentes vitesses, a en effet déjà été implémenté dans un modèle utilisant SpikeNET (Paquier et al, 2000). Certains neurones des couches supérieures du modèle de Paquier et al. sont sélectives à des mouvements très complexes comme des rotations ou des expansions. Lors de la présentation d'une image flashée, la voie dorsale étant activée de 20 à 30 ms plus rapidement que la voie ventrale, elle pourrait tout à fait biaiser l'activité des neurones de la voie ventrale avant même que l'information visuelle ascendante sur l'objet n'atteigne ces neurones (Nowak et Bullier, 1997). Cette idée n'est pas nouvelle (Vidyasagar, 1999) mais elle ne semble pas encore avoir été intégrée à un modèle. Cela semble de plus compatible avec la dynamique observée dans le cortex visuel : la voie dorsale - majoritairement magnocellulaire - intervient fortement dans l'activité des neurones de la voie ventrale (Ferrera et al, 1994). Une hypothèse serait que cette action indirecte modulerait l'activité des neurones de la voie ventrale, notamment en rendant plus saillant le stimulus à ces positions[15].

Tout se passerait alors comme si le contraste des stimuli à ces positions était augmenté, processus qui peut être assimilé à un effet préattentionnel. Des expériences très convaincantes dans l'équipe de Desimone montrent que la modulation de la décharge d'un neurone (Reynolds et al, 2000) - ou le degré d'activation d'une aire cérébrale en IRMf (Desimone, 2000) - sur un stimulus dépend du contraste relatif de ce stimulus avec les objets alentour. L'attention spatiale a pour effet de supprimer partiellement l'effet des stimuli alentour, et donc en quelque sorte d'augmenter le contraste du stimulus dans la zone concernée par l'attention (Reynolds et al, 2000). L'effet d'augmenter le contraste à une position donnée est en fait le

[15] Attention, il ne faut pas confondre les activations directes de la voie ventrale par le système magnocellulaire et les activations indirectes, passant par la voie dorsale. Dans la partie expérimentale j'ai indiqué comment la voie magnocellulaire pouvait intervenir directement – au sein de la voie ventrale - dans la reconnaissance des objets. Elle peut également intervenir par des interactions de la voie dorsale sur la voie ventrale.

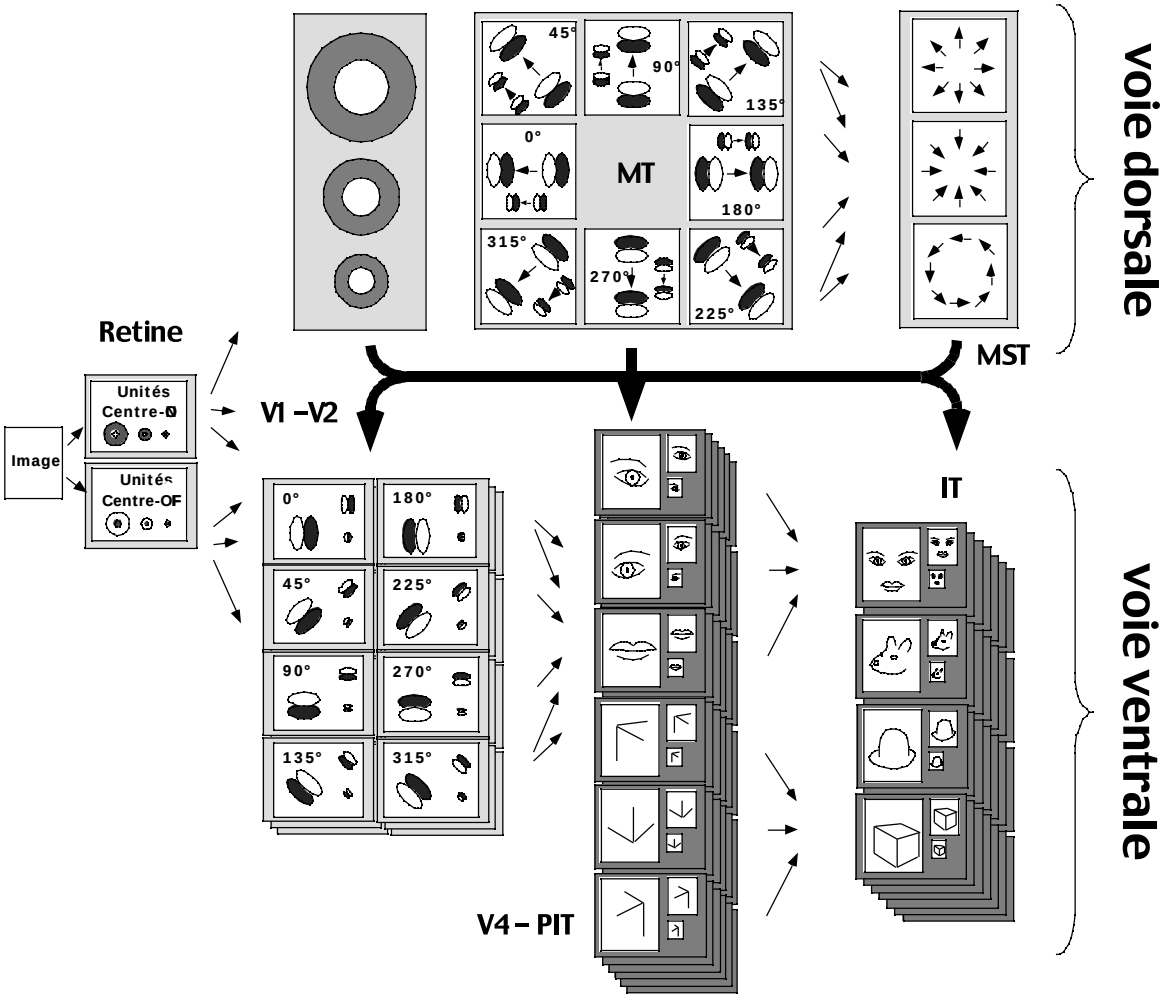


Figure 4.8 : modèle général de reconnaissance des objets. Ce modèle théorique combine apprentissage supervisé et apprentissage non supervisé dans le sens où la voie dorsale induit l'apprentissage dans la voie ventrale. Les abbréviations utilisées représentent la localisation probable des aires neuronales présentes chez le singe : pour la nomenclature des aires corticales, se reporter à la figure I.2.3 du chapitre I.2. Les cartes neuronales en gris foncé (dans V4-PIT et dans IT) contiennent des neurones dont les connexions afférentes sont plastiques. L'activité de ces neurones est en fait biaisée par les cartes neuronales de la voie dorsale : les neurones sont par exemple pré-excités pour des positions où un objet est en mouvement et renforcent leurs connexions avec les neurones des cartes neuronales du niveau précédent dans la voie ventrale pour devenir sélectifs à l'objet présenté. Ce modèle est non supervisé dans le sens où le réseau n'a aucune information *a priori* sur le stimulus présenté. Il est supervisé dans le sens où la voie dorsale indique à la voie ventrale quels sont les endroits dignes d'intérêt dans l'image présentée.

même que celui de raccourcir les latences des neurones dans la zone concernée. Le stimulus sera donc propagé en premier et les neurones sélectifs à ce stimulus inhiberont les neurones sélectifs aux autres objets. Dans le modèle de VanRullen et al (2000c), l'attention spatiale consistait à baisser les seuils des neurones - ce qui revient en fait à une excitation externe des neurones – ce qui aurait pu provoquer une baisse de leur sélectivité. De plus, ce type de mécanisme ne semble pas être cohérent avec les expériences qui montrent que la décharge spontanée des neurones n'augmente pas avec l'attention (Reynolds et al, 2000). Tout en

gardant l'idée de favoriser le traitement dans certaines zones du champ visuel, l'augmentation locale du contraste paraît donc être une meilleure solution[16] car elle ne provoquera pas d'augmentation de la décharge spontanée du neurone.

La préactivation de la voie ventrale par la voie dorsale serait un processus préattentif et non un phénomène attentif proprement dit. Toutefois, à la fois les expériences de l'équipe de Desimone (2000) sur l'interaction entre la saillance du stimulus et l'attention et les modèles détaillés comme celui de Grossberg (1999), indiquent que les deux phénomènes, attentif et préattentif, pourraient faire intervenir les mêmes mécanismes neuronaux. Dans le modèle de reconnaissance général des objets, plus le contraste est élevé dans une zone donnée et plus la latence de décharge des neurones sera courte. Les objets de fort contraste, ou ceux sur lesquels les effets préattentifs induits par la voie dorsale se focalisent, voient donc leurs traitements favorisés par rapport à celui d'autres objets. Dans le cas présent, la sélection de zones d'intérêt sert à favoriser la décharge de certains neurones, mais peut aussi, dans une certaine mesure, résoudre le problème de la convergence. Les deux mécanismes ne sont pas contradictoires, et l'on pourra tester dans ce modèle les effets d'un réseau plus ou moins convergeant à l'aide des algorithmes de convergence que j'ai mis au point (annexe 2).

Suite à cette présélection spatiale, les neurones les plus sélectifs aux stimuli présentés déchargeront. Par le même mécanisme d'apprentissage non supervisé que celui présenté au chapitre III.2, les poids synaptiques des neurones afférents convergeront afin d'augmenter la sélectivité des neurones sélectionnés. Ce n'est plus l'expérimentateur qui indiquera la position de l'objet mais le biais préattentif, qui favorisera la décharge de certains neurones et donc le fait qu'ils apprennent, de façon non supervisée, à reconnaître un ou plusieurs objets. Pour atteindre ce but, il reste à surmonter des problèmes techniques comme le choix de la base d'apprentissage, qui doit être suffisamment riche en objets, le but ultime étant de présenter un film au réseau et que le réseau apprenne seul à y reconnaître les objets présents. L'optimisation des paramètres de l'apprentissage ne sera pas non plus une chose aisée. Un apprentissage trop rapide rendra probablement les neurones instables alors qu'un apprentissage trop lent ne permettra à aucun neurone de devenir sélectif. Enfin le dernier problème reste le décodage de l'activité des neurones de plus haut niveau, afin de pouvoir déterminer si le réseau est en effet sélectif à certains objets. Dans le cas d'une représentation

[16] Du point de vue de l'implémentation biologique, on peut facilement imaginer un processus qui modifie le courant de fuite du neurone de sorte que les courants entrants se trouvent amplifiés et leurs latences plus courtes. En l'absence de stimulation, la décharge spontanée des neurones n'augmente pas.

distribuée des objets, il sera nécessaire de déterminer le *pattern* d'activation qui correspond à chaque objet présenté.

Un tel système ne constituera bien évidemment pas l'ultime modélisation du système visuel (en existe-t-il une ?). Il est en effet *a priori* difficile d'imaginer de ne pas prendre en compte les séquences de présentation (cooccurrence des objets, occurrences successives...) ni la sémantique des objets. Dans une seconde étape, je pense qu'il faudra ajouter une dimension temporelle à la dynamique d'apprentissage comme pour les neurones réels. Les poids synaptiques ne sont pas modifiés instantanément et il faut environ quelques minutes pour que l'effet d'une potentiation devienne visible. De cette façon, et en modifiant la règle d'apprentissage non supervisée pour qu'elle prenne en compte la fréquence d'activation des neurones comme dans le modèle de Wallis et Rolls, les neurones pourront apprendre en continu sur flot d'images continue[17].

Concernant la prise en compte de la sémantique des objets, dans la conclusion, je montre comment ce phénomène peut émerger en relation avec les autres modalités sensorielles, les sentiments et la conscience visuelle. A partir des données bibliographiques sur le sujet, je tente de déterminer les éléments nécessaires à intégrer à un modèle pour voir émerger de tels phénomènes.

[17] Comme on l'a vu au chapitre concernant l'apprentissage non supervisé, la fréquence de décharge des neurones semble également intervenir dans la loi de plasticité synaptique qui dépend de l'ordre d'activation des neurones pré- et postsynaptiques.

IV

Perspectives et Conclusion

"A further knowledge of facts is necessary before I would venture to give a final and definite opinion."

Sherlock Holmes

Avant toute chose, je souhaite revenir sur le cadre conceptuel que je m'étais fixé, à savoir les systèmes autonomes. Il semble que toutes les expériences et les modèles sur lesquels j'ai travaillé soient en fait très éloignés d'un processus quelconque de conservation de l'autonomie. Le système visuel tel que je l'ai décrit peut sembler correspondre à une machinerie sans vie dont le seul but serait de traiter les informations contenues dans les images. Je vais brièvement montrer qu'il n'en est rien et tenter de resituer mes travaux dans le cadre des systèmes autopoïétiques (Varela, 1989). A mon avis, l'autonomie est présente à tous les niveaux : depuis le neurone vers l'organisme tout entier. L'organisme et notamment le cerveau serait une pyramide de systèmes autonomes imbriqués, chacun ayant un rôle différent dans la cognition (Burnod, 1991).

Au niveau *a priori* le plus bas dans le système visuel, comme je l'ai souligné dans l'introduction, les interactions locales entre les neurones doivent permettre l'émergence de traitements complexes. Dans les modèles que j'ai présentés, même si l'autonomie des neurones, et les comportements propres qu'ils peuvent avoir pour préserver leur identité, ne sont pas explicités, il sont présents sous la forme des contraintes que j'ai introduites. Tout comme les neurones réels, les neurones de mes modèles intègrent les décharges de leurs afférents et émettent des potentiels d'activation. La plasticité synaptique, qui permet aux neurones d'organiser leur profil de sélectivité, est de plus comparable avec celle des neurones réels. Enfin la structure hiérarchique du système visuel est grossièrement implémentée. Ces contraintes sont en partie le reflet de la préservation de l'autonomie des neurones au sein du système et j'ai montré, qu'à partir de ce type d'organisation, des comportements complexes comme la sélectivité à l'orientation pouvait émerger.

Au niveau supérieur, il est également possible d'imaginer le système visuel comme une entité autonome, un système modulaire (Fodor, 1983) qui serait lui-même composé de sous-systèmes autonomes comme les différentes aires visuelles (Zeki et Shipp, 1988). Si le système visuel est un système autonome ou autopoiétique, il n'est pas nécessairement orienté vers le traitement d'images. Il se situe au sein d'un ensemble d'autres systèmes - auditif, moteur... - et on peut imaginer que des processus homéostatiques le contraignent à maintenir son activité

pour conserver son identité. Il est possible que les systèmes perceptifs tentent de s'adapter à leur environnement - c'est-à-dire aux autres systèmes - en minimisant leurs interactions. Si le système visuel tente par exemple de conserver son identité en minimisant ses interactions avec le système auditif, seuls les neurones qui déchargent de façon corrélée dans les deux systèmes interagiront. On peut assimiler ce type de processus à une descente de gradient ou la minimisation de l'énergie contraint le système à s'organiser pour coder de façon optimale les stimuli qui le perturbent. Cette hypothèse est gratuite et probablement fausse mais elle illustre de quelle manière, à différents niveaux, les systèmes peuvent à la fois être dans une dynamique de conservation de leur identité et de traitement de l'information.

Enfin au niveau de l'organisme, les études que j'ai présentées ont également un sens du point de vue de la conservation de l'identité. Je me suis intéressé tout particulièrement aux processus visuels rapides. Chez les mammifères, la réaction rapide à la modification de leur environnement est le principal garant de leur survie. Un prédateur doit pouvoir réagir rapidement aux mouvements de sa proie s'il veut pouvoir se nourrir. De même, la proie doit être capable de répondre très rapidement à toute attaque. Dans le cadre de la préservation de l'organisme, la vitesse du système visuel a donc un sens.

Après cette brève mise au point, je pense qu'il serait dommage de terminer ma thèse sans parler des implications éventuelles de l'approche que j'ai utilisée sur des processus perceptifs de plus haut niveau, comme l'accès à la sémantique des objets, aux émotions et à la conscience visuelle. Dans le paragraphe suivant, je présente deux modèles qui, malgré leurs imperfections, donnent une explication biologiquement plausible pour l'implémentation de ce type de processus. Le premier modèle tente de resituer le système visuel dans son interaction avec les autres processus neuronaux. Il traite en particulier des interactions entre les différentes modalités sensorielles et de la façon dont elles pourraient interagir pour faire émerger la sémantique des objets. Le second modèle traite de processus de plus haut niveau encore comme les phénomènes émotionnels et la conscience. J'aborderai ce problème dans le cadre des travaux de Damasio (1999), et je discuterai de leur émergence au sein d'un système artificiel.

1 - Modèle perceptif multimodal

Aucun des modèles que j'ai présentés n'aborde le problème de la sémantique des objets. Existe-t-il un ou plusieurs systèmes sémantiques, et s'il en existe plusieurs, quelles sont leurs

relations ? Concernant l'existence de plusieurs systèmes sémantiques, il semble tout d'abord y avoir une forte distinction entre deux catégories principales d'objets : les êtres vivants et les objets manufacturés. On observe par exemple des patients cérébro-lésés qui ne sont capables de reconnaître que les objets animés alors que d'autres ne sont capables de reconnaître que les objets inanimés. En IRMf, ces deux catégories d'objets semblent activer des zones cérébrales distinctes dans l'hémisphère gauche (Moore et Price, 1999) et la dissociation est encore plus importante quand on demande aux sujets de nommer les objets qui leurs sont présentés (Martin et al, 1996).

D'aucuns pensent qu'il s'agit d'une distinction entre les caractéristiques des objets. Les êtres vivants partagent en général de nombreuses caractéristiques, des éléments diagnostiques comme les pattes et les yeux, et il est nécessaire d'effectuer un traitement très fin de leur caractéristiques pour déterminer leur identité. Les objets manufacturés, par contre, ont des formes très diverses et ne semblent pas partager beaucoup d'attributs visuels. En général, leur forme dépend de plus de leur fonction. Il n'y aurait qu'un pas à faire pour émettre l'hypothèse que les êtres vivants et les objets manufacturés sont traités par deux systèmes sémantiques différents, l'un basé sur la forme et l'autre sur la fonction.

A un niveau plus fin, il semble même exister trois principales catégories d'objets, les êtres vivants animés, les plantes et les objets manufacturés. On observe par exemple des sujets sélectivement déficients soit aux fruits et légumes soit aux animaux. Le modèle HIT[1] de Humphreys et Forde permet d'expliquer à la fois les déficits spécifiques aux catégories et la distinction entre caractéristiques fonctionnelles des objets manufacturés et caractéristiques de forme des êtres vivants (figure 1). Humphreys et Forde soutiennent que pour déterminer les caractéristiques d'un objet - par exemple les nommer - il est nécessaire de réactiver les zones concernées par leur traitement, qui dépend de la catégorie d'objets traitée. Quand l'objet est présenté, il active tout d'abord l'aire neuronale visuelle puis les aires auxquelles il est plus particulièrement associé. Par exemple, un objet manufacturé activera principalement le cortex moteur, en relation avec ses caractéristiques fonctionnelles. Par contre, un être animé réactivera le système visuel pour déterminer ses caractéristiques fines de contour. Une plante activera en partie le cortex olfactif. Dans ce modèle, c'est l'interaction entre les différentes modalités sensorielles et motrice qui détermine la voie que doivent suivre les informations visuelles. Ce seraient en fait les caractéristiques des objets, associées au rôle qu'ils ont pour nous, qui détermineraient le cheminement des informations et par conséquent le regroupement

[1] Hierarchical Integrative Theory.

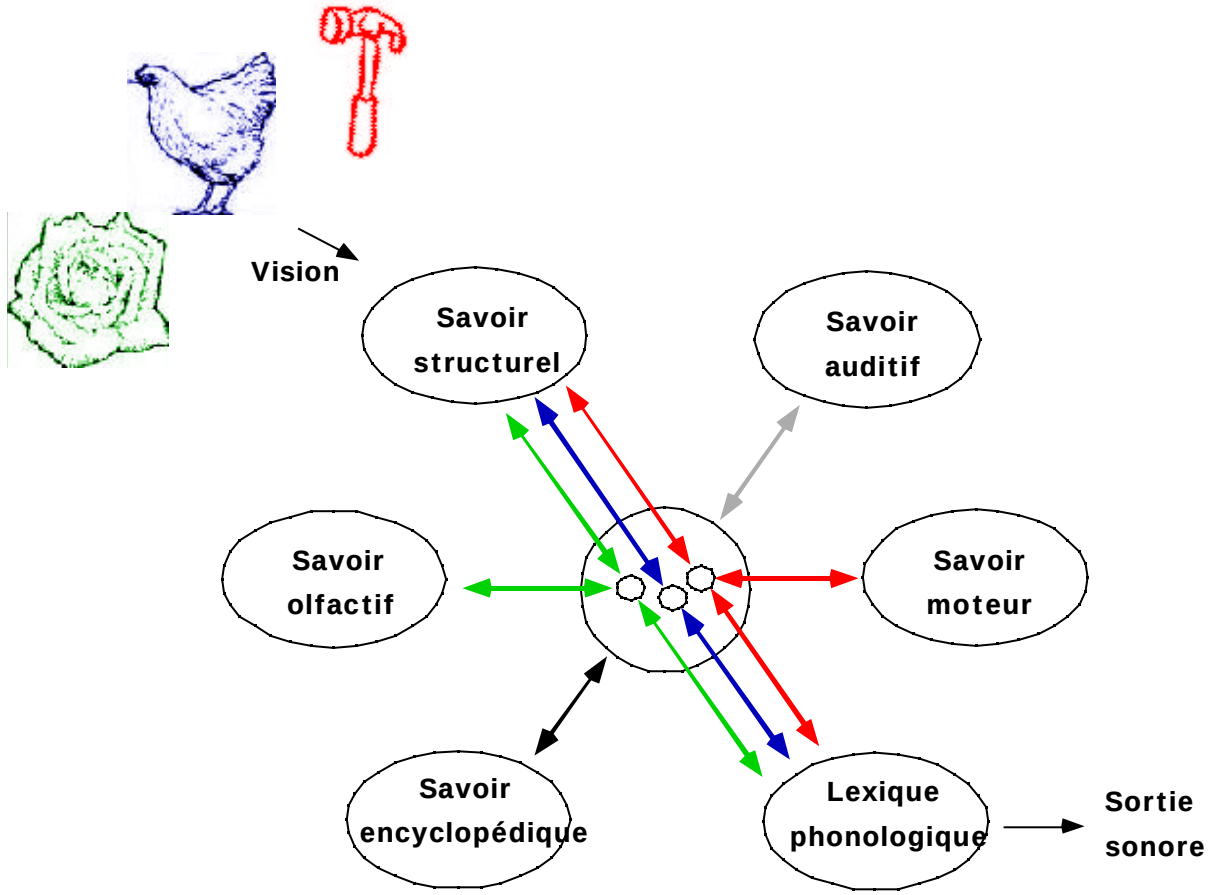


Figure 1 : modèle de Humphrey d'interaction entre les différents systèmes sensoriels et moteur. Suivant l'objet présenté, les voie et les aires neuronales impliquées sont différentes. Ce serait l'interaction de ces différents systèmes qui permettrait à un objet d'acquérir un sens. Adapté de Humphreys et Forde (2000).

en des zones de traitement localisées, comme une sorte d'autoroute entre deux systèmes sensoriels ou moteur. On peut également imaginer la dynamique sensori-motrice comme un processus complexe de changement de référentiel (Burnod et al, 1999).

Ce qui est intéressant, c'est que du sens peut émerger de ces interactions : on imagine bien qu'une activation des aires motrices n'a pas la même conséquence du point de vue de la réaction qu'un individu peut avoir face à un objet, qu'une activation des aires sensorielles. Même sans postuler des zones d'intégration plus complexes, les objets acquièrent un certain sens : les objets que l'on peut manipuler activent les voies motrices et sont par là-même associés à une action, alors que les autres objets ne le sont pas.

Cette vue est très schématique mais reflète bien l'idée générale d'un traitement partagé en fonction des caractéristiques de l'objet. Elle est de plus compatible avec la majorité des lésions observées chez les patients humains (Humphreys et Forde, 2000). Nous avons vu comment la sémantique des objets pourrait émerger à partir des voies neuronales qu'ils activent entre les aires sensorielles et motrices. Nous allons maintenant tenter d'aller encore plus loin et d'aborder l'aspect émotionnel et conscient des processus visuels.

2 - Conscience visuelle

L'étude de la conscience visuelle est très difficile car il n'est pas possible de l'aborder directement, ce phénomène n'étant accessible que par le sujet lui-même. De nombreuses études, certaines très sérieuses, ont cependant été menées sur ce sujet. On a cherché en particulier à déterminer le corrélat neuronal de la conscience visuelle : quelles conditions doivent remplir les décharges des neurones pour avoir une chance d'intervenir sur la perception consciente du sujet ?

Il a tout d'abord été montré que l'activité des neurones peut biaiser la perception du sujet. Par exemple, si les neurones de certaines aires sont stimulés, on peut biaiser la réponse du singe (Salzman et al, 1990; Celebrini et Newsome, 1995). Dans cette étude le singe devait répondre quand des points présentés en mouvement sur un écran étaient perçus dans une certaine direction. Quand les neurones de l'aire MT qui sont sélectifs aux mouvements des objets sont stimulés électriquement, on peut biaiser la réponse du singe. Cela montre clairement qu'il existe une relation entre ce que le sujet perçoit et l'activité des neurones.

Cependant toutes les aires visuelles ne semblent pas entretenir une relation privilégiée avec ce que le sujet perçoit : cela semble être le cas en particulier des aires visuelles de bas niveau comme V1, V2 ou V4 (pour une revue cf. Crick et Koch, 1998). Des expériences chez le singe montrent que seuls les neurones de IT sont majoritairement affectés par ce que perçoit le sujet (Sheinberg et Logothetis, 1997; Logothetis, 1998). Les auteurs ont présenté deux images différentes à chaque œil d'un singe. Dans ces conditions, l'homme comme le singe n'en perçoit qu'une des deux mais celle qu'il perçoit change aléatoirement au cours du temps (Kovacs et al, 1996). Si l'on demande à un singe de répondre uniquement sur l'une des deux images, on peut déterminer quand il la perçoit[2]. Les auteurs remarquent que les décharges des neurones dans IT sont biaisées par ce que le singe perçoit alors que ce n'est pratiquement pas le cas pour les neurones dans V4.

Chez l'homme également, il semble que l'activité dans le système visuel ne soit pas nécessairement consciente. Les patients présentant une "vision aveugle" ne sont pas conscients de leur vision résiduelle dans leur hémichamp visuel atteint (Sahraie et al, 1997). Cependant dans cet hémichamp, ils peuvent percevoir inconsciemment en quelque sorte les mouvements. Si on leur demande d'indiquer la direction d'un objet en mouvement dans cet hémichamp, ils nieront avoir vu l'objet. Cependant dans une situation de choix forcé (vers la

[2] De plus en flashant brièvement l'une des deux images présentées simultanément, on contraint le système visuel à percevoir cette image.

gauche ou vers la droite), leur performance dépasse 80 % de réponses correctes. D'autres expériences montrent que l'on peut avoir un traitement très complexe de l'information visuelle sans pour autant que le stimulus devienne conscient. Par exemple, quand deux images sont présentées simultanément, une à chaque œil, pendant moins de 250 ms, les sujets sont incapables de désambiguïser l'information en provenance de chaque œil. Ils réussissent à deviner où se trouve la cible mais n'ont pas conscience de l'avoir vue (Kolb et Braun, 1995). De même l'expérience sur la perception inconsciente de chiffres ou d'images présentés très brièvement et suivies d'un masque renforce ce résultat (Dehaene et al, 1998; Bar et Biederman, 1999).

Comme on l'a vu, des stimuli peuvent être traités par le système visuel sans pour autant devenir conscients. L'extrême brièveté des présentations implique que les neurones ne doivent pas être activés très fortement. Cela a conduit certains auteurs à émettre l'hypothèse que seules les décharges soutenues des neurones pourraient permettre d'accéder à un niveau conscient (O'Brien et Opie, 1999). Toutefois, comme le dit si bien O'Regan (2000), cela ne résout absolument pas le problème de la conscience : quand bien même saurions-nous exactement quelles sont les conditions que doivent remplir les décharges neuronales pour influencer la perception consciente du sujet, cela ne nous indiquerait en aucune manière ce qu'est la conscience visuelle.

Ces études présentent un défaut inhérent à l'approche utilisée : elles sont à mon avis trop attachées aux mécanismes neuronaux et voudraient absolument tout ramener à l'activité neuronale. Je crois personnellement que l'activité neuronale en elle-même ne signifie que ce qu'elle véhicule. Si par exemple un neurone du cortex moteur interagit avec un neurone sélectif à certains attributs d'un objet alors cela signifie que l'objet présenté a une quelconque relation avec un comportement moteur : du sens émerge de cette activation conjointe, mais la décharge individuelle de l'un des deux neurones ne signifie rien. Dans le paragraphe suivant, nous allons voir comment dans ce cadre, la conscience peut émerger. Il nous faudra cependant sortir du cadre de la conscience visuelle, pour une raison très simple, c'est que la vision n'est pas nécessaire à la conscience.

3 - Sentiments et conscience étendue

A la différence de ce que l'on pourrait penser, je ne vais pas ici me lancer dans des considérations philosophiques, mais plutôt retourner vers la biologie. Je vais présenter le

modèle de Damasio qui est le plus plausible de mon point de vue, justement parce qu'il se base sur la neurophysiologie.

En continuité avec l'approche consistant à considérer les décharges neuronales comme véhiculant un certain sens par rapport au circuit dans lequel elles sont engagées, Damasio établit tout d'abord une distinction entre représentation mentale et image mentale. L'image mentale est un terme issu de la philosophie qui désigne le résultat de la construction mentale suscitée par un objet qui n'est accessible qu'à la première personne. La pensée pour Damasio est un flux d'images mentales logiquement corrélées. Une représentation[3], en terme d'activité neuronale, par opposition est le résultat d'une observation à la 3ème personne : un individu observe le système et établit des relations entre activité des neurones et la présentation d'objets.

Ces considérations théoriques faites, pour Damasio, il existe au moins deux niveaux de conscience : la conscience noyau et la conscience étendue. La première constituerait la base inconsciente de notre interaction avec le monde et la seconde, une perception plus évoluée dans le sens où elle serait à la base de ce phénomène si mystérieux : la conscience. La conscience noyau serait en fait la régulation de nos processus homéostatiques, rythme cardiaque, veille/sommeil, faim. Cependant elle serait également basée sur la présence de la somesthésie. En fait la conscience noyau intègrerait à la fois, les processus de régulation automatique de notre corps et les entrées somesthésiques. Il est d'ailleurs intéressant de noter qu'au niveau du tronc cérébral, les deux phénomènes se réunissent. L'étude des patients lésés renforce également l'hypothèse d'une localisation de la conscience noyau dans le tronc cérébral. Des lésions du tronc cérébral provoquent en général le coma. Cependant certaines des lésions du tronc provoque une paralysie globale où seuls les yeux peuvent encore bouger. La particularité de ce type de lésions est qu'elles se situent au-dessous du dernier nerf crânien sensoriel. Une lésion au-dessus de ce nerf, qui signifierait que le cerveau n'a plus aucune entrée somesthésique, provoque un coma irrémédiable. Sans somesthésie pas de conscience possible. De même des lésions du cortex somatosensoriel provoquent des comas ou des altérations marquées de la conscience, dans lesquels les patients ne se rendent pas compte de la paralysie de l'un de leur membre. Par opposition, les lésions d'autres cortex, visuel

[3] J'ai souligné dans le chapitre introductif que le terme représentation était biaisé : l'activité des neurones n'est pas obligatoirement liée à la présentation d'un stimulus qu'ils seraient censés représenter. On ne peut cependant nier que l'activité de certains neurones est corrélée avec l'objet présenté et c'est dans ce cadre que j'utilise le terme de représentation.

sensoriel, auditif, olfactif... ne provoquent pas le coma. Des lésions du système visuel par exemple peuvent provoquer une cécité mais en aucun cas altérer la perception consciente (non visuelle) du sujet.

De façon surprenante, bien que la sensibilité des organes internes soit très développée, c'est la seule que nous ne puissions ressentir consciemment. C'est peut-être justement parce qu'elle est à la base de notre individu que nous ne pouvons pas y accéder. La conscience noyau, mettant en relation somesthésie et homéostasie, définirait un état émotionnel basique, de fond. Le fait que nous soyons fatigué, motivés... serait déterminé par la conscience noyau. La conscience noyau interagit également avec les objets présentés. Un objet entre en étroite collaboration avec les états émotionnels et les configurations motrices qui lui sont associés. Dans ce cadre, pour compléter le graphique de Humphreys et Forde, il faudrait ajouter une case "état mental" qui serait gouvernée par notre conscience noyau.

Au second niveau, se trouve le soi central et la conscience étendue. Le soi central est en fait une carte de second niveau qui observe en quelque sorte la conscience noyau interagir avec le monde, une représentation de second ordre. Dans le cadre de l'étude de la conscience visuelle, les cartes de second ordre - dont certaines seraient probablement situées dans le cortex cingulaire - observeraient le système visuel interagir avec la conscience noyau. Cependant la présence de la conscience étendue, comme dans le cas de la conscience noyau, ne dépend pas d'une modalité sensorielle particulière[4].

La conscience étendue en elle-même n'aurait qu'une portée limitée. Elle devrait pour déployer ses capacités de planification à long terme se reposer sur le soi autobiographique, qui serait en quelque sorte l'histoire d'un individu. L'étude des patients cérébro-lésés est ici d'un intérêt considérable. Certains de ces patients semblent en fait souffrir d'un accès à ce soi autobiographique. Les individus sont parfaitement conscients mais ne peuvent se rappeler les événements qui ont eu lieu plus de 2 minutes auparavant[5]. Ils possèdent une connaissance sur les objets, leur fonctions et leurs relations, ils peuvent ressentir, exprimer ou mimer tout type de sentiment mais ils ne sont pas capables de planifier leur comportement au-delà de 2 minutes. Une expérience concernant les effets d'amorçage renforce ces résultats (Subramaniam et al, 2000). On présente au sujet des séquences rapides d'images - une

[4] La présence du langage, de la vision ou d'une quelconque modalité sensorielle n'est pas nécessaire à la conscience.

[5] L'analyse de patients cérébro-lésés indique que les structures impliquées dans ce type de conscience se situent toutes dans des zones anciennes du point de vue de la phylogénie, en particulier dans le cortex cingulaire.

nouvelle image toutes les 70 ms. On demande ensuite au sujet de déterminer laquelle de deux images était présente dans la séquence et le sujet en est totalement incapable. Même dans une tâche de choix forcé, sa précision est au niveau de la chance. Ce qui est très intéressant dans cette expérience est que les sujets ont parfaitement conscience de voir les images qui leur sont présentées. Cette expérience indique que la mémoire autobiographique ou à court terme, même inconsciente, est dissociable du fait de percevoir consciemment l'objet. Mon interprétation de ces résultats est qu'il est probable que certains animaux, bien qu'ils aient une conscience étendue, présentent une mémoire autobiographique réduite. Cette mémoire ne leur permettrait donc pas de planifier des opérations complexes. Par opposition, l'homme aurait une mémoire autobiographique bien supérieure, ce qui lui permettrait de planifier son comportement à beaucoup plus long terme.

La distinction entre conscience noyau et conscience étendue permettrait également d'expliquer la différence entre émotions et sentiments. Les émotions, probablement contrôlées par la conscience noyau, sont des phénomènes dirigés vers l'extérieur et provoquent des altérations physiologiques. Les sentiments par contre sont dirigés vers l'intérieur et seraient en quelque sorte le ressenti de l'émotion. Alors que les émotions se situeraient au niveau de la conscience noyau, les sentiments seraient en fait la représentation des émotions au niveau de la conscience étendue. Dans certains cas, comme dans celui des modifications physiologiques induite par la peur, le sentiment (de peur) serait en fait la perception - au sens propre du terme - de l'émotion.

Nous avons traité des émotions et des sentiments, qu'en est-il de la conscience de soi ? La conscience de soi serait en fait un sentiment particulier, le sentiment de soi. A la différence des autres sentiments, le sentiment de soi serait omniprésent à l'état de veille. On peut se demander pourquoi ce type de sentiment serait présent chez les animaux et quelle en est l'utilité ? Une hypothèse qui me plait beaucoup, car elle rejoint celle de la conservation de l'identité des systèmes autonomes, serait que, étant donné que nous simulons en permanence les pensées de nos congénères, sans le soi, l'organisme ne serait pas capable de déterminer à qui ces pensées appartiennent. Le soi est le garant de la conservation de l'identité de l'individu.

La question qui m'intéresse maintenant est "dans quelle mesure est-il possible de simuler la conscience (visuelle en particulier) ?". C'était également le but de la cybernétique que de tenter de répondre à cette question. Cependant, le débat est resté centré sur des considérations théoriques et mathématiques. Il n'y avait *a priori* aucun obstacle théorique à ce projet, Turing

ayant montré que sa machine théorique[6] pouvait implémenter n'importe quel phénomène déterministe. De plus, même dans un espace déterministe, des processus complexes, souvent non solubles dans cet espace, peuvent émerger[7]. L'émergence de nouveaux comportements, non déterministes, à partir de bases déterministes est donc tout à fait possible. Sans entrer plus avant dans les considérations théoriques qui avaient lieu à l'époque je tiens à citer cette phrase qui résume bien à la fois l'approche cybernétique et celle à laquelle j'adhère sur ce sujet :

"La cybernétique [...] libère l'homme de la fermeture contraignante de l'organisation en le rendant capable de juger cette organisation, au lieu de la subir en la vénérant et la respectant parce qu'il n'est pas capable de la penser ..."

Gilbert Simondon (1969)

Pour paraphraser Simondon, ce n'est pas parce que l'on n'est pas capable de penser les sentiments et la conscience, que l'on ne peut pas les simuler. La mauvaise nouvelle est que nous ne pourrons jamais comprendre les processus conscients : *"Car comme il doit être plus qu'évident désormais, rien à l'intérieur d'un cadre ne permet de formuler quelque chose, ou même de poser des questions sur ce cadre*" (Watzlawick et al, 1972). La bonne nouvelle cependant est qu'il ne nous sera pas *a priori* impossible de simuler ce phénomène : si l'on en croit Damasio, ces travaux constituant à mon avis l'approche la plus aboutie du point de vue scientifique, il est nécessaire d'avoir un corps, de le sentir pour devenir conscient. La conscience émerge en quelque sorte de l'interaction des systèmes somatosensoriels et de régulation de l'organisme avec les systèmes sensoriels. Pour voir émerger une conscience - même rudimentaire - il faudrait donc que le système artificiel possède un corps sensible en tant que robot par exemple[8], et que l'état de ses organes internes - qui définissent son état émotionnel de base - modifie les interactions qu'il peut avoir avec les objets. En complétant le modèle d'Humphreys et Forde, en plus de la case état mental de la conscience noyau, une petite case de représentation de l'interaction de ces états mentaux avec la représentation sensorielle du monde extérieur suffirait peut-être pour voir émerger ces phénomènes. Il manque encore "quelques cases" à nos modèles, mais....

[6] Machine de Turing. Cette machine théorique hyper simple est une sorte d'ordinateur qui possède une mémoire infinie et des opérations de base très simples ainsi qu'une entrée et une sortie.

[7] Aucun processus, aucune machine de Turing, ne peut déterminer si une autre machine de Turing va s'arrêter.

[8] Ou virtuel dans un environnement informatique.

Annexes

Annexe 1

Méthode expérimentale

Cette annexe présente en détail la tâche go-nogo et le dispositif expérimental utilisé : la tâche étant commune à toutes les expériences, il a paru utile de regrouper ces informations. Cette annexe indique également de quelle façon les potentiels évoqués ont été enregistrés et traités. Une dernière partie a trait à l'utilisation et l'interprétation des courbes de d' qui permettent d'étudier l'évolution de la précision des réponses au cours du temps.

1 - Tâche go-nogo

Les expériences que j'ai réalisées sont toutes basées sur une tâche go-nogo. Les sujets doivent relâcher un bouton (en moins de 1s) quand une cible leur est présentée et maintenir l'appui sur le bouton pendant au moins 1 s sur un distracteur. La répartition des cibles et des distracteurs au sein d'une série est aléatoire. Cependant chaque série est composée de 50 % de cibles et 50 % de distracteurs. Chaque homme travaille sur des séries de 100 images qu'il n'a jamais vues auparavant. Il n'y a pas de *feedback* dans les expériences où nous avons associé chez l'homme étude comportementale et étude EEG.

Dans le poste singe cependant, où certaines des expériences que je présente chez l'homme ont également été menées[1], les sujets doivent, en plus de relâcher le bouton sur les cibles, aller toucher l'écran à la position où la cible est apparue. Chaque réponse correcte (go après une cible ou no-go après un distracteur) est signalée par un son et pour le singe récompensée par une gorgée de jus de fruit; en revanche chaque erreur est punie par le réaffichage de l'image incorrectement classée pendant 3 s sur l'écran. Le singe travaille quotidiennement sur une série de 100 images qui peuvent lui être présentées plusieurs fois mais toujours dans un ordre

[1] Afin de faciliter la comparaison entre le comportement des deux espèces.

différent pour éviter tout apprentissage de séquence. Parmi ces 100 stimuli, 10 n'ont jamais été vus par l'animal (5 cibles et 5 distracteurs).

Dans toutes les tâches où l'on enregistre les TRs, les réponses d'un même sujet varient de façon très importante. Les TRs varient également beaucoup d'un sujet à l'autre. Étant intéressé en particulier par la catégorisation rapide, il était important pour nous d'obtenir les TRs les plus courts possibles. Dans des tâches de *naming* où il faut nommer les objets qui sont présentés, les temps de réaction sont très longs, de l'ordre de la seconde (i.e. Boucart et Humphreys, 1992). De même, dans les tâches où le sujet doit appuyer sur un bouton pour une cible et appuyer sur un autre bouton pour un distracteur, les TRs sont de l'ordre de 600 ms chez l'homme. Une tâche go-nogo permet d'obtenir les TRs les plus rapides - 400 ms en moyenne, cf. chapitre ? - et donc *a priori* de minimiser les effets parasites qui ne sont pas directement liés au processus de catégorisation.

Les images utilisées sont des photographies de scènes naturelles sélectionnées dans une très riche base commerciale d'images (Corel photo). Les cibles incluent des mammifères, des reptiles, des poissons, des oiseaux, des insectes[2], … dans leur environnement naturel ou dans des environnements façonnés par l'homme. Ces animaux peuvent êtres partiellement visibles, dans des positions très diverses, en gros plan ou en plan large. La photographie peut contenir plusieurs animaux, la plupart du temps appartenant à la même espèce. Nous avons préféré cependant ne pas utiliser de photographies représentant des hommes[3]. Au chapitre ?, les images sont classées selon leur contenu et l'on peut appréhender la variété des images cibles présentées. Les distracteurs incluent des paysages naturels et artificiels très divers, des fleurs, des fruits, des arbres, des monuments, des objets manufacturés[4]…

Le temps de présentation des images est également critique dans ce type d'expérience. Des effets adaptatifs très rapides peuvent intervenir au niveau oculaire seulement 80 ms après la présentation de l'image chez l'homme - respectivement 50 ms chez le singe. Des saccades oculaires peuvent également être initiées à des latences similaires. Il était donc nécessaire de s'en affranchir en présentant les images pendant un très court laps de temps, 20 ms chez

[2] Seulement dans les deux expériences réalisées chez l'homme où les potentiels évoqués étaient enregistrés.

[3] Dans les deux expériences réalisées chez l'homme où les EEGs sont enregistrés, certaines images cibles représentaient des animaux en présence d'une ou plusieurs personnes (cavalier sur son cheval, dresseur).

[4] Dans les deux expériences réalisées chez l'homme où les EEGs sont enregistrés, certains distracteurs présentaient des êtres humains et d'autres présentaient des statues, des dessins ou des peintures d'animaux face auxquels les sujets avaient pour consigne de ne pas répondre.

l'homme et 30 ms chez le singe[5]. De plus, l'utilisation de temps de présentation très courts permet probablement d'obtenir des enregistrements électroencéphalographiques plus propres[6].

2 - Dispositif expérimental

Cette partie, essentiellement d'ordre technique, présente le dispositif expérimental à la programmation duquel j'ai considérablement contribué et que j'ai utilisé pour réaliser les expériences chez le singe et chez l'homme.

Le poste expérimental est composé de deux ordinateurs. Le premier ordinateur est utilisé pour présenter les stimuli. Il comprend une carte VSG (Cambridge Research System) qui permet l'affichage des images sur un écran supplémentaire situé dans la pièce d'enregistrement. Les images sont chargées dans une partie de la mémoire de la carte qui n'est pas visible à l'écran. Quand l'image doit être affichée pendant 20 ms (2 frames à 100Hz, l'image correspondant à un angle visuel de 4,5°x6,7° pour une distance d'environ 1m-1,1m entre le sujet et l'écran), on décale simplement la zone de mémoire de la carte qui sert à l'affichage. De cette façon, il est possible de contrôler très strictement les temps de présentation. La carte VSG permet également de connaître la position du faisceau d'électrons sur l'écran et de synchroniser l'affichage par rapport à la fréquence de l'écran (figure 1).

Une autre carte ISA permet de mesurer le temps. Il est ainsi possible de déterminer le temps avec une précision de l'ordre de la milliseconde tant pour la durée de présentation des images que pour les temps de réaction des sujets et le temps séparant deux images. Les temps de réaction des sujets étaient initialement enregistrés à partir d'une souris d'ordinateur. Le sujet devait appuyer de façon continue sur le bouton droit de la souris et, quand il relâchait ce bouton, le temps de réaction était recueilli par l'ordinateur. Des tests effectués par Simon Thorpe lors de la mise en place du poste expérimental ont montré que les temps de réaction des sujets étaient plus rapides quand ils relâchaient le bouton que quand ils devaient appuyer dessus. Nous avons par la suite été amenés à améliorer ce dispositif pour deux raisons. Tout d'abord, il semblait que les temps de réaction des sujets humains étaient plus rapides lorsqu'ils

[5] Dans les deux cas, cela correspond à 2 frames, les fréquences de rafraîchissement étant différentes sur l'écran du poste "homme" (100 Hz) et sur l'écran du poste singe (60 Hz).

[6] Cette affirmation n'engage que moi. Les enregistrements EEGs décrits dans la littérature, suite à la présentation d'une image pendant plusieurs centaines de ms, semblent très bruités, notamment au niveau des parties initiales des potentiels évoqués moyens pour une tâche similaire (Antal et al, 2000). Il peut s'agir du dispositif expérimental, mais j'attribuerais plutôt cette différence au fait que l'image continue à être intensivement traitée par les aires de bas niveau du système visuel.

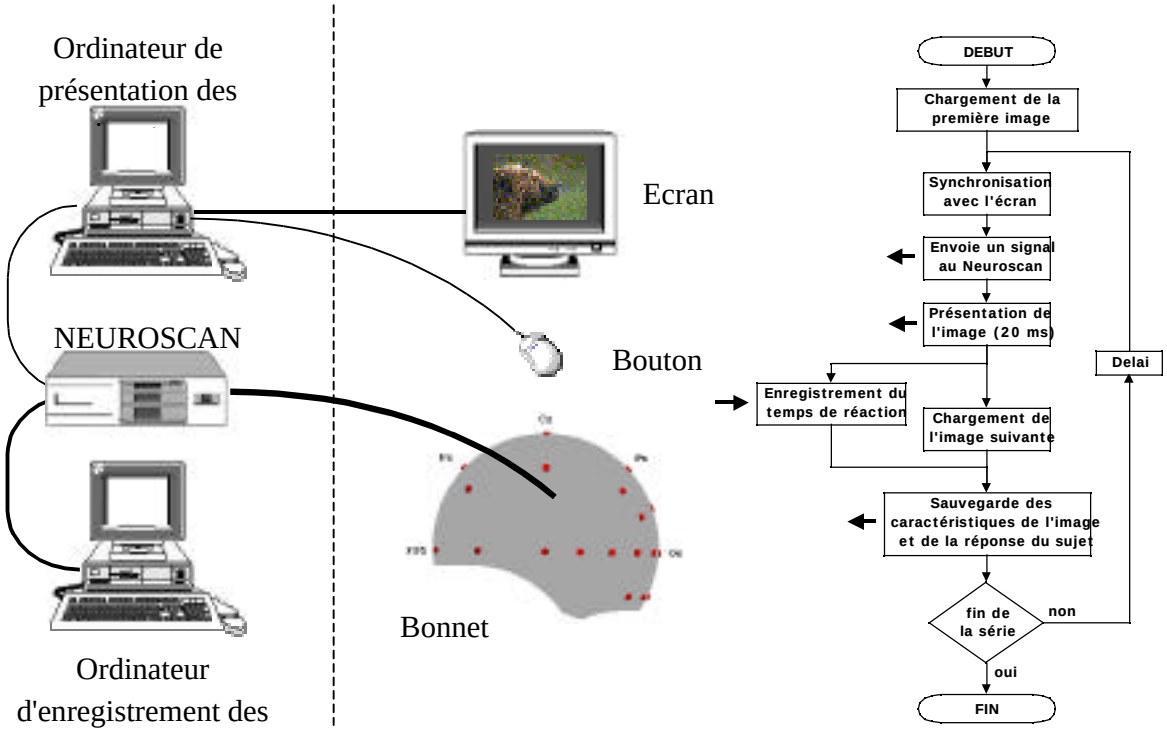


Figure 1 : À gauche, le dispositif expérimental est représenté. Un ordinateur présente les images et enregistre les temps de réaction des sujets. Il envoie un signal au Neuroscan à chaque présentation d'image. Un autre ordinateur enregistre les données brutes renvoyées par le Neuroscan . À droite, l'algorithme de présentation des images est indiqué (il est implémenté sur l'ordinateur présentant les images). Les temps de réaction sont recueillis en parallèle du déroulement du programme par une carte d'acquisition asynchrone.

étaient testés dans le poste expérimental dédié aux singes : la principale différence entre les deux postes étant que dans le poste "singes", les temps de réaction était recueillis à l'aide d'une plaque conductrice. Quand le sujet lève son doigt de la plaque, les propriétés électriques de la plaque sont modifiées et l'ordinateur enregistre la réponse du sujet. La seconde raison était que le système d'exploitation multitâche de l'ordinateur (qui gère la souris) pouvait biaiser l'enregistrement des temps de réaction.

Nous avons équipé le poste expérimental "homme" avec un bouton capacitif identique à celui du poste "singe". Ce bouton était branché directement sur la carte ISA permettant de mesurer le temps. Ainsi les temps de réaction sont recueillis de manière asynchrone par rapport au déroulement du programme de présentation (figure 1). L'avantage est double, puisque le programme, au lieu de tester en permanence l'état du bouton (ou de la souris), pouvait effectuer d'autres tâches comme le chargement d'image dans la mémoire de la carte vidéo. Cela nous a donc permis d'augmenter la fréquence de présentation des images puisque le chargement de ces images avait lieu pendant que le sujet répondait et non après.

Il est apparu cependant que ce bouton capacitif interagissait fortement avec l'enregistrement des potentiels évoqués du fait de ses propriétés électriques[7]. Les améliorations du poste prévues, actuellement en cours de mise au point, incluent donc l'utilisation d'un bouton équipé d'un capteur infrarouge. Quand le sujet lève le doigt, le faisceau lumineux est rétabli et le temps de réaction communiqué à l'ordinateur. Une autre amélioration est la suppression de la carte VSG, trop lente pour des présentations rapides de séquence d'images. On lui a préféré des cartes graphiques standard de jeux qui permettent d'atteindre des fréquences d'affichage très élevées.

D'autre part, pour chaque image flashée, l'ordinateur présentant les images envoie un signal au Neuroscan au moment de l'affichage de l'image. Ce signal est ensuite retransmis à l'ordinateur enregistrant les potentiels évoqués. Un oscilloscope et un photomètre ont permis de déterminer le temps exact séparant l'apparition de l'image et l'envoi du signal.

3 - Traitement des données EEG

Les images sont présentées par série de 100 et les potentiels cérébraux sont enregistrés en continu durant cette période à 1000 Hz sans filtrage. La position exacte du haut du bonnet, estimée par rapport au milieu des deux électrodes les plus frontales, doit être 10 % de la distance nasion-ignon[8]. Lors de la pose du bonnet, l'impédance de chaque électrode avec la surface du scalp est réglée pour être inférieure à 5 kOhm. La configuration du bonnet suit la nomenclature 10-20 (figure 2), mais le bonnet standard a été modifié afin d'inclure plus d'électrodes occipitales en vis-à-vis du système visuel qui nous intéressent tout particulièrement.

Dans la première des deux expériences EEG présentées (chapitre III.5), l'électrode de masse était FPZ et la référence était enregistrée front et oreilles liées. Il est apparu cependant que la technique consistant à prendre CZ comme référence et la moyenne du signal des électrodes pour reconstruire l'activité en CZ[9] permettait de minimiser le bruit et cette technique a donc été utilisée dans la seconde expérience (chapitre III.6).

---

[7] Nous avons isolé le doigt des sujets à l'aide d'un morceau de caoutchouc dont la face externe était solidaire d'une pièce métallique reliée à la terre. De cette façon, les temps de réaction pouvaient être enregistrés par le bouton capacitif et les effets indésirables sur l'électroencéphalogramme disparaissaient.

[8] Le nasion correspond à la limite entre le nez et le front et l'ignon au creux que l'on peut sentir à l'arrière du crâne.

[9] CZ est la référence, donc chaque électrode enregistrée par rapport à CZ. Quand on fait la moyenne des électrodes, on a *potentiel_moyen = moyenne(électrode) - 32 * CZ*. Idéalement *moyenne(électrode) = 0* et donc *CZ = potentiel_moyen / -32.*

		théta	phi
1	FP1	-92	-72
2	FP2	92	72
3	F3	-60	-51
4	F4	60	51
5	C3	-46	0
6	C4	46	0
7	P3	-60	51
8	P4	60	-51
9	O1	-92	72
10	O2	92	-72
11	F7	-92	-36
12	F8	92	36
13	T7	-92	0
14	T8	92	0
15	P7	-92	36
16	P8	92	-36
17	Cz	0	0
18	Fz	46	90
19	Pz	46	-90
20	IF7	-92	18
21	IF8	92	-18
22	PO7	-92	54
23	PO8	92	-54
24	PO3	-74	61
25	PO4	74	-61
26	POz	69	-90
27	Oz	92	-90
28	Iz	115	-90
29		-115	72
30		115	-72
31	PO9	-115	54
32	PO10	115	-54

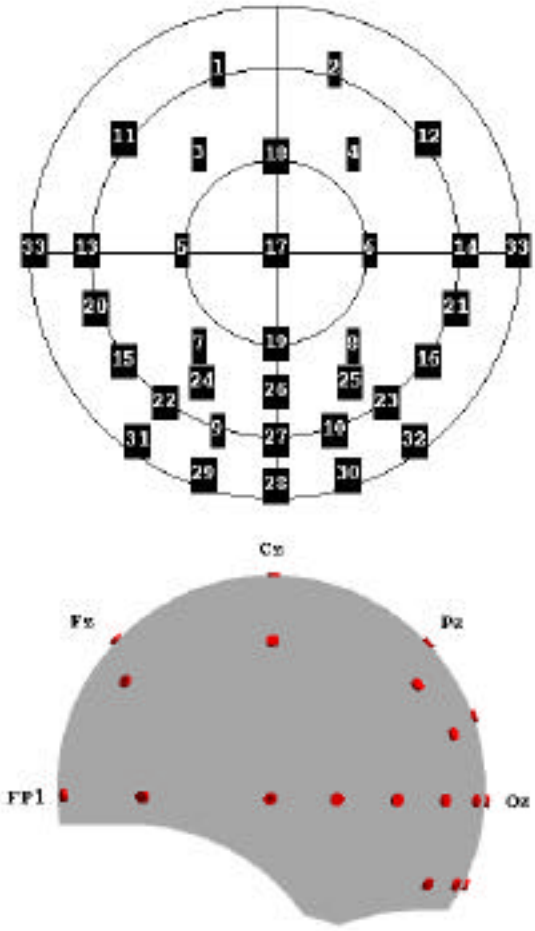


Figure 2 : Bonnet à 32 électrodes utilisé lors des deux expériences chez l'homme. La première colonne indique le nom de l'électrode (système 10-20 étendu) et sa position dans un système de coordonnées sphériques standardisé (T5 et T6 au chapitre III.6 correspondent respectivement à P7 et P8)

Les données continues correspondant à chaque série d'images sont ensuite retraitées à l'aide du logiciel Neuroscan. Les potentiels correspondants à chaque image sont découpés (de 100 ms avant la présentation de l'image à 600 ms après la présentation de l'image). On utilise ensuite des filtres pour supprimer les essais où le sujet a effectué un mouvement des yeux (de –50 à 50 ms par rapport à la présentation de l'image, le signal EEG ne doit pas dépasser 50 µV) et ceux où le sujet présentait des ondes alpha de grande amplitude (de –100 à 400 ms par rapport à la présentation de l'image, le signal EEG ne doit pas dépasser 30 µV). Il est très important de ne pas supprimer les ondes alpha sur l'ensemble des 700 ms correspondant à un essai car l'amplitude des signaux EEG à 500 ms dépend très fortement du type d'image présentée et/ou de la réponse produite par le sujet. Pour chaque sujet, ces potentiels sont ensuite moyennés en fonction du type d'image (par exemple les images contenant des animaux et les images ne contenant pas d'animaux). Suivant la superposition des courbes pour chaque groupe d'images, les données peuvent subir différents types de filtrages. Le

traitement est adapté au sujet afin de déterminer le nombre optimal d'essais permettant d'obtenir les données les plus propres.

Les potentiels correspondants à un type d'image donné sont ensuite regroupés pour l'ensemble des sujets. Chaque moyenne d'un sujet est considérée individuellement ce qui signifie que, dans la moyenne globale, la moyenne individuelle de chaque sujet prend une part égale. Cela permet, par la suite, d'effectuer un t-test à chaque pas de temps pour l'ensemble des sujets et de déterminer à quelles latences, par rapport au début de la présentation des images, le signal EEG diffère significativement de la ligne de base.

La plupart des traitements que nous avons effectués se basent sur la différence entre des moyennes de signaux EEG (par exemple entre les signaux moyennés sur les essais animaux et les signaux moyennés sur les essais distracteurs). Une différence permet en effet d'isoler spécifiquement les variations en termes de traitement visuel entre deux catégories d'images en calculant la latence à laquelle les signaux EEGs divergent significativement. En général, à une onde positive correspond toujours une onde négative sur une autre électrode (les courants circulant dans le cerveau peuvent en effet être vus comme des dipôles ayant un pôle positif et un pôle négatif). Lors des tests de significativité, je me suis donc appliqué à faire correspondre à une différence négative sur certaines électrodes une différence positive sur d'autres.

Des tests statistiques de Student (t-tests) sont réalisés sur les données non-filtrées des activités différentielles. Pour évaluer le début de la différence entre les potentiels évoqués moyens obtenus pour les stimuli cibles et non-cibles, j'ai considéré l'effet comme significatif si 15 valeurs de *t* consécutives sont effectivement significatives (Rugg et al, 1995). Comme je l'ai indiqué, j'applique une contrainte supplémentaire en m'assurant que le signal est symétrique, chaque différence significative positive devant être associée à une différence significative négative. Deux groupes d'électrodes ont particulièrement retenu mon attention : les électrodes frontales (FZ, FP1, FP2, F3, F4, F7, F8, en nomenclature 10-20) et les électrodes occipitales (OZ, I, O1, O2, O1', O2', CB1, CB2, CB1'', CB2''). J'ai également contrôlé que les autres électrodes n'atteignaient pas le seuil de significativité avant celles-ci et si cela était le cas, j'ai vérifié que la différence entre le début de la significativité pour cette électrode ne précédait pas de plus de 10 ms celle du groupe fronto-occipital.

Pour finir, lors de la présentation des potentiels, il est préférable de filtrer les signaux pour améliorer la lisibilité des données[10].

[10] Filtre passe-bas à 35 Hz, 12db.

4 - Analyse de sources

L'analyse des sources a été effectuée à l'aide du logiciel BESA[11] 99 (Miltner et al, 1994). Le logiciel utilise un fichier de moyenne des EEG. Ce fichier est filtré (filtre passe-bas à 35 Hz, 12db) et les deux électrodes temporales T3 et T4 sont systématiquement supprimées[12]. L'expérimentateur choisit la fenêtre temporelle sur laquelle il désire modéliser le signal et place un certain nombre de dipôles. Un algorithme optimise ensuite la position et l'amplitude au cours du temps de ces dipôles afin qu'ils reflètent au mieux le signal EEG enregistré dans un modèle ellipsoïdal[13] (algorithme simplex ou algorithme génétique). La solution est unique pour un dipôle mais pour un nombre plus élevé de dipôles, plusieurs solutions acceptables sont possibles. Je n'ai effectué de modélisation de dipôles que sur les différences de moyennes (entre les animaux et les distracteurs par exemple). Il s'agit alors de déterminer les sources neuronales qui permettent d'expliquer au mieux la différence entre les deux catégories d'images. Deux dipôles sont introduits dans le modèle et sont contraints pour avoir une position symétrique. Il est en effet très probable que si une zone donnée du cerveau s'active spécifiquement pour une catégorie, la zone correspondant dans l'autre hémisphère soit également active. Je n'ai personnellement pas tenté d'interpréter directement la position des signaux. J'ai préféré analyser la position des sources d'une condition à l'autre (cf. chapitre III.6). L'interprétation de la position des signaux est en fait très difficile car très peu précise : la précision spatiale est de l'ordre du centimètre ce qui peut conduire à des conclusions erronées quant aux aires cérébrales impliquées.

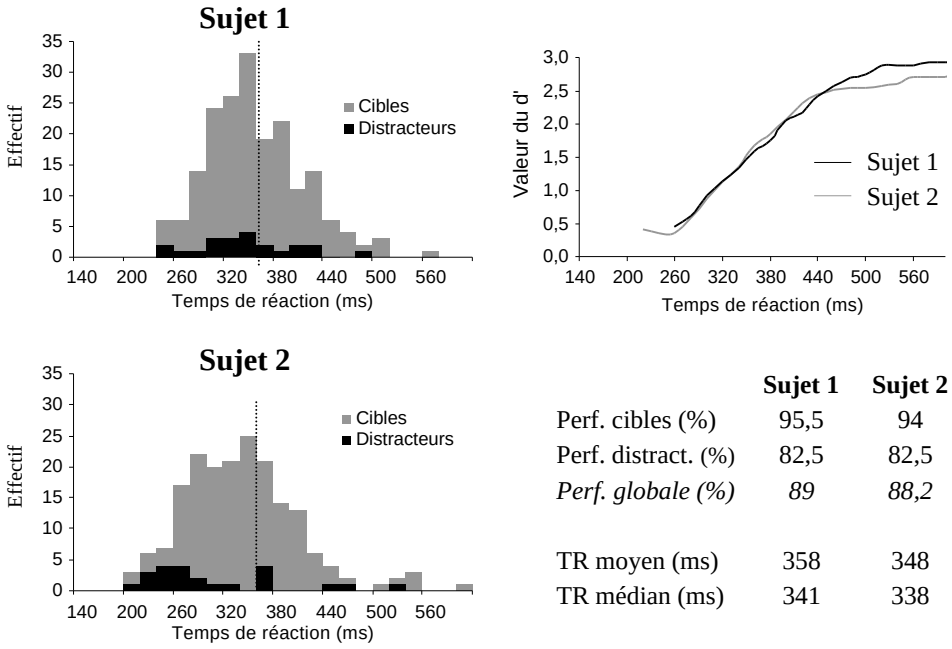
5 - Calcul de la précision en fonction du temps : d'

La précision d'un sujet s'évalue en calculant et en comparant le nombre de cibles et de distracteurs correctement catégorisés. Cependant, comment estimer la performance sur les temps de réaction inférieurs à 500 ms? Il n'est alors plus possible de comparer le pourcentage de réussite des sujets sur les cibles et les distracteurs. J'ai donc utilisé une solution alternative : le d'. Le d' est un outil statistique plus puissant que l'analyse du simple pourcentage et il permet notamment d'estimer la précision d'un sujet au cours du temps dans

[11] Brain Electrical Source Analysis.

[12] Ces électrodes sont en effet très bruitées par l'activité musculaire.

[13] Pour des raisons purement algorithmiques, modéliser le cerveau comme une ellipsoïde est beaucoup plus rapide et plus simple que de prendre en compte sa géométrie exacte.



	Sujet 1	Sujet 2
Perf. cibles (%)	95,5	94
Perf. distract. (%)	82,5	82,5
Perf. globale (%)	*89*	*88,2*
TR moyen (ms)	358	348
TR médian (ms)	341	338

Figure 3 : Illustration de l'intérêt de l'utilisation du d' pour deux sujets. À gauche, histogrammes des TRs des deux sujets. Le second sujet est plus rapide que le premier et présente des TRs entre 200 et 240ms. Cependant il commet beaucoup d'erreurs sur ces réponses très rapides. Le d', à droite, permet d'estimer la différence de performance au cours du temps et elle est identique pour les deux sujets. La valeur finale du d' reflète la performance globale des sujets (visible dans le tableau). La barre verticale en pointillé représente la position du TR médian.

une tâche go-nogo[14]. Imaginons, par exemple, deux sujets, l'un ayant des temps de réaction très lents et commettant peu d'erreurs et l'autre présentant des temps de réaction très rapides et commettant beaucoup d'erreurs (figure 3). Bien que le sujet 2 ait des temps de réaction très rapides, étant donné qu'il commet beaucoup d'erreurs lors de réponses précoces, ses performances sont donc faibles pour les temps de réaction rapides. Par opposition le sujet 1 a des temps de réaction plus lents, mais commet moins d'erreurs pour ses temps de réaction très rapides. Bien que le comportement des deux sujets diffère, ils ont cependant des courbes de performance tout à fait similaires tout du moins dans leur partie initiale. Le niveau du plateau des courbes de performance dépend directement de la précision des sujets dans la tâche.

J'ai, dans l'exemple précédent, cumulé la précision des sujets en fonction du temps, mais il aurait également été possible de calculer un d' pour chaque intervalle de temps afin d'obtenir

[14] À partir du pourcentage de réponse sur l'ensemble des cibles, on détermine à l'aide d'une table de correspondance un Z-score à un pas de temps donné : *Zs*. On effectue de même pour les distracteurs : *Zc*. La valeur du d' est la soustraction de ces deux valeurs : $d' = Zs - Zc$

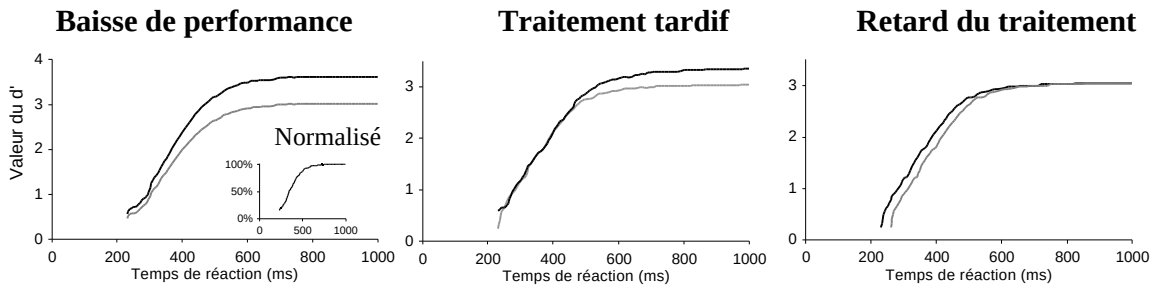


Figure 4 : Interprétations pour différentes courbes de précision cumulées en fonction du temps (d'). À gauche, les deux courbes sont proportionnelles et donc seule la performance globale est affectée, la performance étant plus faible pour la courbe grise et cela indépendamment du temps puisque les courbes normalisées sont identiques. Au milieu, seule la performance tardive est affectée (après 500ms sur ce graphique), ce qui signifie que les réponses précoces sont identiques et les réponses tardives diffèrent. À droite, la courbe grise est décalée par rapport à la courbe noire. Cela signifie que dans le cas de la courbe grise, le traitement est globalement plus lent mais que cela n'affecte pas la précision car les deux courbes convergent vers un même plateau.

la précision - non cumulée - des sujets en fonction du temps[15]. Cette approche permet de rendre les intervalles de temps indépendants les uns des autres. Dans la pratique, on se heurte cependant à un problème d'ordre technique, car, malgré la quantité de données recueillies, le nombre de réponses par intervalle de temps de 10 ms ou de 20 ms reste très faible et ne permet pas d'obtenir des courbes interprétables. De façon épisodique, dans les cas où cela est possible, nous utiliserons le d' non cumulé. Néanmoins, dans l'usage que j'en ai fait, l'utilisation de la précision cumulée plutôt que la précision non cumulée en fonction du temps est sans grandes conséquences. Je me suis en effet intéressé principalement au traitement ultra-rapide, pour lequel les deux types de courbes de d' sont très similaires. Dans ce genre d'analyse, il est également primordial de noter où se situe la médiane des temps de réaction. La figure 3 illustre le fait que la partie initiale de la courbe de d' est la partie la plus importante puisqu'elle correspond aux 50 % des réponses les plus rapides.

La figure 4 illustre trois différentes configurations dans la comparaison de deux courbes de d' et la façon dont il faut les interpréter. Tout d'abord, si les courbes sont proportionnelles, c'est-à-dire si les courbes normalisées sont confondues, alors il s'agit d'une simple variation de précision indépendante du temps de réaction : les précisions cumulées sont en effet proportionnelles à chaque pas de temps[16]. Ensuite, si seul le plateau de précision terminal diffère alors cela signifie que le traitement rapide est identique pour les deux catégories -

[15] La différence avec le d' cumulé réside uniquement dans la sélection des données. Pour un d' cumulé à 200 ms, toutes les réponses de 0 à 200 ms sont considérées. Dans le cas d'un d' non cumulé à 200 ms, seules les réponses entre 180 et 200 ms sont prises en compte (pour un pas de temps de 20 ms).

[16] Pour une variation de la performance indépendante en fonction du temps, les performances sont proportionnelles à chaque pas de temps et les performances cumulées le sont donc également. Si on normalise les courbes de performance alors elles sont confondues.

parties initiales des courbes - et diverge pour les réponses tardives, la catégorisation devenant plus précise pour l'une des deux catégories. Enfin, si l'une des courbes semble retardée par rapport à l'autre, cela signifie que l'une des catégories est traitée plus rapidement que l'autre et que ce gain est indépendant du temps de réaction. Dans ce cas, on peut observer également des distributions de TR similaires mais décalées l'une par rapport à l'autre. Dans les analyses présentées, les cas sont rares où l'interprétation est aussi claire. Souvent, les différents types d'effets se mélangent et compliquent l'interprétation des résultats.

Un type de problème courant avec ce type d'analyse est dû au fait que les intervalles utilisés pour calculer le d' sont de 10 ms - éventuellement de 20 ms. Étant donné la rapidité du traitement dans le système visuel, suite à l'absence ou à la présence d'une caractéristique dans l'image, il est très probable que les effets sur les TRs seront de l'ordre de quelques ms. Ces différences si elles existent ne seront donc pas détectables avec la résolution temporelle de ces analyses.

Annexe 2

SpikeNET

Il existe actuellement un grand nombre de logiciels permettant d'effectuer des simulations de réseaux de neurones. Beaucoup d'entre eux ont été conçus pour simuler des réseaux de neurones artificiels et sont inadaptés à la modélisation des propriétés biophysiques des neurones. Ces unités ne possèdent aucune structure et leur sortie consiste généralement en une valeur continue (souvent entre 0 et 1 ou entre -1 et 1). Alors que de tels systèmes ont été largement utilisés dans des domaines aussi variés que l'ingénierie ou la finance, il semblerait pour le moins impensable de les utiliser pour simuler des réseaux de neurones réels.

À l'opposé, il existe des programmes sophistiqués comme GENESIS et NEURON qui sont adaptés à la simulation détaillée de neurones réels et prennent en compte la structure de l'arbre dendritique et la dynamique complexe des canaux ioniques (Hines et Carnevale, 1997; Bower et Beeman, 1998). Cependant dans le cas de ces logiciels, la finesse de la simulation et le nombre considérable de paramètres en font des outils inadaptés à la simulation de gros réseaux de neurones.

Dans cette annexe, je décris un logiciel de simulation de réseaux de neurones intermédiaire entre ces deux extrêmes : SpikeNET. Les neurones de SpikeNET sont suffisamment proches de la biologie pour étudier la dynamique temporelle des décharges des neurones comme la synchronie ou l'asynchronie mais encore suffisamment simples pour permettre la simulation de très larges réseaux de neurones en temps réel et l'application à des problèmes concrets comme le traitement d'image. SpikeNET ne se restreint pas à la modélisation du système

visuel mais les capacités computationnelles exceptionnelles du système visuel pour le traitement des images en fond un candidat idéal pour démontrer les capacités de ce logiciel.

Je vais dans un premier temps décrire l'implémentation de la propagation des décharges des neurones dans SpikeNET. Ensuite, j'analyserai dans quelle mesure SpikeNET est adapté à la simulation de très gros réseaux de neurones et je m'attacherai tout particulièrement à l'implémentation de connexions entre des groupes de neurones rétinotopiquement organisés.

1 - SpikeNET, simulateur de réseaux de neurones

1.1 - Historique du développement

SpikeNET a été initialement conçu par Jacques Gautrais et Simon Thorpe pour traiter des scènes naturelles en utilisant de très gros réseaux de neurones "intègre et décharge" (IF). Ils en ont ainsi posé les bases de fonctionnement. Le postulat le plus important, et qui fait à mon avis la force de SpikeNET, est le suivant : le temps doit être discrétisé et les décharges des neurones doivent se propager de manière événementielle[1]. Cela implique que les neurones qui ne sont pas activés ne sont pas traités. Les neurones sont regroupés au sein de cartes rétinotopiquement organisées. On suppose en général que, au sein d'une carte neuronale, le champ récepteur des neurones est identique (seule la position de ce champ récepteur change d'un neurone à l'autre). La rapidité des simulations dans SpikeNET est due à ce type d'implémentations : comme on le verra, il est possible de simuler plusieurs dizaines de millions de neurones et plusieurs centaines de milliards de connexions synaptiques sur un ordinateur standard.

Ma contribution au développement de SpikeNET a été principalement centrée sur l'apprentissage supervisé et non supervisé pour le traitement des images naturelles. J'ai tout d'abord implémenté un algorithme pour la reconnaissance des caractères (Chapitre III.3) que Rufin VanRullen a étendu à la détection des visages (VanRullen et al, 1998). J'ai ensuite implémenté un modèle de reconnaissance d'objet que j'ai testé pour la reconnaissance des visages (Chapitre III.3). Concernant le logiciel SpikeNET même, j'ai étendu ses capacités au

[1] Cela signifie que ce sont les événements, c'est-à-dire les décharges des neurones, qui sont au centre de l'algorithme de propagation.

traitement des images à plusieurs échelles, à la connexion de cartes neuronales rétinotopiquement organisées de différentes tailles et au calcul réseau. Programmé initialement en C, j'ai également dû reprogrammer SpikeNET en C++.

Les développements futurs de SpikeNET sont axés sur le traitement du mouvement dans les images et sur la prise en compte des informations stéréoscopiques. William Paquier (2000) sera chargé de ces nouveaux développements ainsi que de l'optimisation du code de SpikeNET pour le passage de message au sein du réseau.

1.2 - Organisation de base

Dans SpikeNET, les objets sont des matrices de neurones IF - que j'appellerai cartes à partir de maintenant -. Chaque unité neuronale est caractérisée par un nombre relativement faible de paramètres : un potentiel membranaire, un seuil, éventuellement un niveau de désensibilisation et une constante de fuite. Quand un neurone afférent décharge, le poids synaptique d'une connexion reliant ce neurone à ses cibles est ajouté au potentiel membranaire des neurone-cibles. Le programme teste si ce potentiel dépasse un certain seuil, si tel est le cas, le neurone cible voit sa référence ajoutée à la liste des neurones ayant déchargé à ce pas de temps et son potentiel remis à zéro. La propagation de l'activité des neurones dans SpikeNET

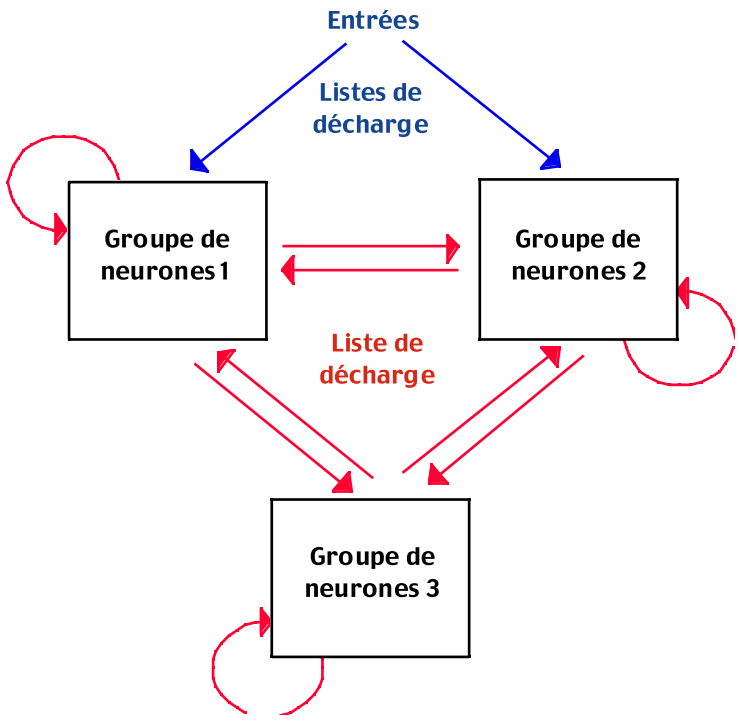


Figure 1 : Organisation des cartes de neurones dans SpikeNET. SpikeNET redirige des listes de neurones entre différentes cartes neuronales organisées de façon rétinotopique. Étant donné qu'un pourcentage très faible de neurone décharge à chaque pas de temps, les communications entre les groupes neuronaux restent limitées.

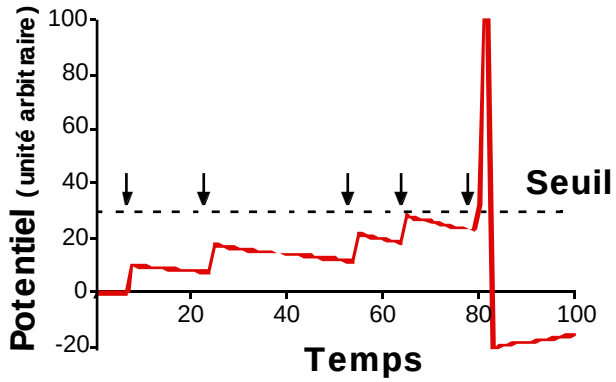


Figure 2 : Comportement d'un neurone IF. Le neurone intègre des décharges afférentes jusqu'à ce que son potentiel atteigne un certain seuil. Quand le potentiel dépasse le seuil, le neurone décharge. La latence de la décharge du neurone dépend donc de l'intensité des stimulations.

consiste à communiquer aux différentes cartes neuronales les listes des neurones ayant déchargé (figure 1). La nature événementielle de ce type de message fait que l'augmentation du nombre de pas de temps n'a pratiquement aucun effet sur le temps de calcul : en effet augmenter le nombre de pas de temps n'augmente pas le nombre de neurones ayant déchargé, paramètre critique dont dépend la vitesse de calcul.

1.3 - Entrées du réseau

La plupart des neurones dans SpikeNET sont affectés uniquement par les décharges d'autres neurones. Cependant, pour certains neurones d'entrée, qui correspondent par exemple aux neurones ganglionnaires de la rétine, la date de décharge du neurone dépend directement de l'image présentée au réseau. Pour simuler le comportement d'une cellule ganglionnaire, on applique généralement un filtre de type chapeau mexicain sur l'image, la valeur renvoyée en chaque point indiquant le contraste local. Les contrastes locaux sont ensuite convertis en latences de décharge : les contrastes les plus élevés correspondent à de courtes latences de décharge et les contrastes très faibles, à de longues latences de décharge (figure 2). Des études en électrophysiologie ont montré que le comportement des cellules dans le LGN[2], c'est-à-dire au point d'entrée du cortex visuel, était bien expliqué par un simple modèle intègre et décharge pour les neurones, dont le courant d'entrée dépend du stimulus (Reich et al, 1998). L'implémentation exacte de ce type de mécanisme est indiquée au chapitre III.1.

[2] Dans le LGN, il existe 3 types de neurones dont deux principaux, les neurones magnocellulaires et les neurones parvocellulaires. Le type de comportement décrit ici correspond aux neurones parvocellulaires dont on pense qu'ils véhiculent une grande partie de l'information visuelle nécessaire à la reconnaissance des objets.

1.4 - Propagation des décharges et application réseau

Les neurones sont regroupés au sein de cartes rétinotopiquement organisées, cartes qui sont placées à chaque pas de temps dans une boucle. Le programme parcourt ensuite la boucle, traitant séquentiellement chaque carte neuronale et vérifiant si des décharges ont été émises par les afférents de cette carte au pas de temps précédent. Si tel n'est pas le cas, la carte est supprimée de la boucle et le traitement la concernant est considéré comme étant terminé pour le pas de temps courant. Cependant, si des décharges ont été émises au sein des cartes afférentes au pas de temps précédent, elles sont propagées dans la carte en cours de traitement. Les poids synaptiques correspondant à une connexion entre un neurone présynaptique et un neurone postsynaptique sont simplement ajoutés au potentiel du neurone postsynaptique[3]. Le potentiel de chaque neurone touché est ensuite comparé à son seuil : dans le cas où il le dépasse, les références du neurone sont ajoutées à la liste des neurones ayant déchargé au sein de la carte pour ce pas de temps. Quand une carte a terminé son traitement, elle est supprimée de la boucle. La propagation d'un pas de temps se termine avec le traitement de la dernière carte neuronale encore présente dans la boucle.

Dans le cas où toutes les cartes sont traitées sur le même ordinateur, l'agencement des cartes neuronales au sein d'une boucle est inutile. Cependant, ce type d'organisation montre son intérêt dans le cas de traitements au sein d'un réseau d'ordinateurs. La figure 3 illustre la propagation de l'activité dans un modèle simple quand l'une des cartes neuronales est traitée sur un ordinateur distant. La propagation est centrée sur les cartes de neurones dans le sens où ces cartes récupèrent l'information dont elles ont besoin pour effectuer des calculs. Les cartes neuronales d'entrées et les cartes neuronales de réseau (représentation locale d'une carte se trouvant sur un autre ordinateur) sont purement passives : elle stockent uniquement des listes de neurones ayant déchargé dans un pas de temps donné (calculées à partir du stimulus ou fournies par le réseau) et les rend disponibles aux autres cartes neuronales effectuant des traitements. Si la liste de décharge d'une carte afférente n'est pas disponible - par exemple, le réseau n'a pas encore envoyé cette information - alors la carte neuronale en cours de traitement est mise en attente : la carte reste dans la boucle et le processeur passe à la carte suivante. Les

[3] Nous verrons que l'intégration de ces décharges peut prendre des formes bien plus complexes.

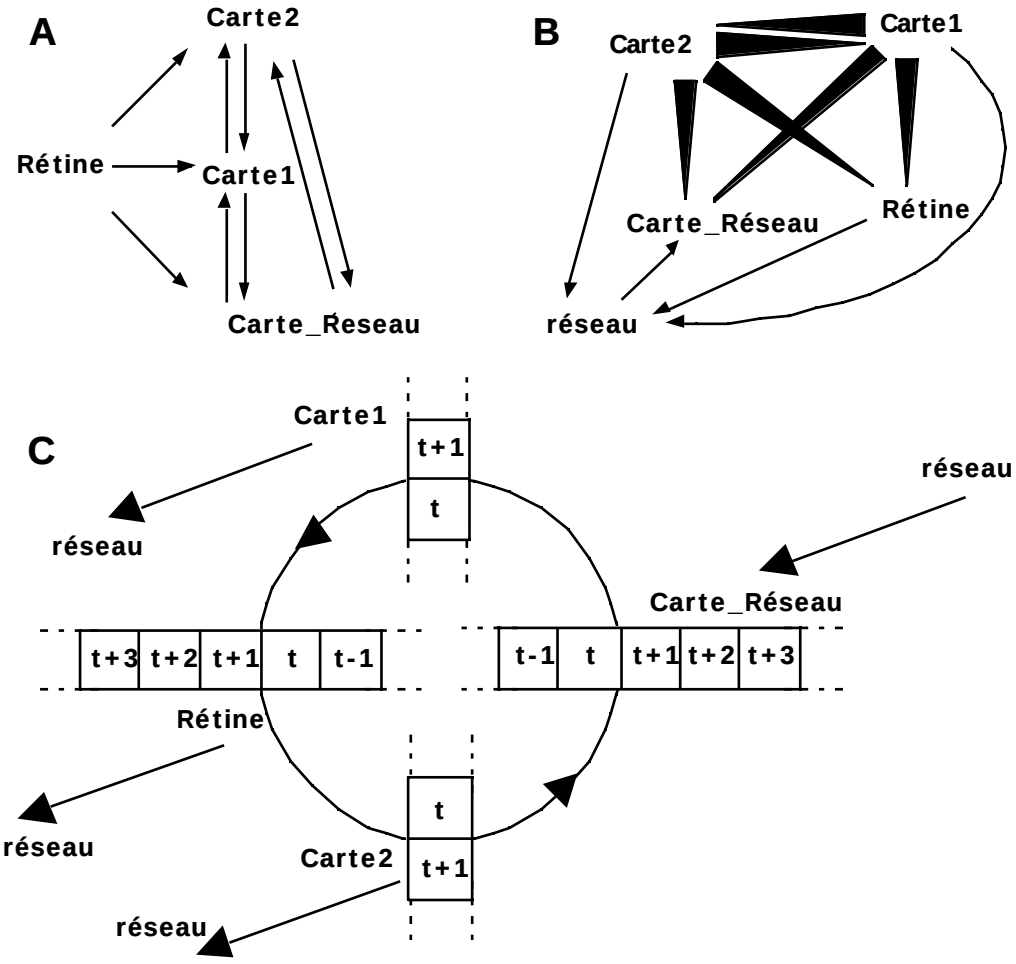


Figure 3 : Illustrations d'une propagation de décharge de neurones dans le cas d'un réseau d'ordinateurs interconnectés. Les groupes neuronaux Rétine, Carte1 et Carte2 sont locaux au processeur alors que la carte Carte_Réseau est située sur un autre processeur. A, connexions entre les différentes cartes. Les flèches indiquent la direction des projections entre les neurones. B, flot d'informations entre les cartes dans un calcul centré sur les cartes neuronales. À chaque pas de temps, seules les Carte1 et Carte2 effectuent des calculs qui se basent sur les informations disponibles dans les autres cartes. Les accès au réseau se font de façon asynchrone par rapport au déroulement du programme. C, vue détaillée de B. A chaque pas de temps t, toutes les cartes sont organisées au sein d'une boucle et les cartes sont supprimées au fur et à mesure qu'elles ont terminé de traiter leurs décharges afférentes. Les cartes Rétine et Carte_Réseau ont été placées dans la boucle pour l'homogénéité de l'ensemble, mais elles ne traitent aucune donnée.

informations en provenance du réseau arrivent de façon asynchrone, c'est-à-dire que quand un message arrive, il suspend le cours normal du programme, effectue quelques opérations puis rend la main au programme. Cela signifie que, au prochain tour dans la boucle, il est possible que la liste de décharge auparavant manquante soit disponible.

Ce type de propagation centré sur les cartes neuronales permet de s'affranchir de la majorité des problèmes de synchronisation inhérents aux programmes communiquant à travers un réseau. La solution alternative implémentée dans SpikeLab (Graßmann et Anlauf, 1998) est de propager les décharges des neurones de manière réellement événementielle : dès qu'un

neurone décharge, ou qu'une liste de décharge est générée, celle-ci se propage à travers le réseau pour aller toucher les neurones cibles. Le calcul est alors centré sur les listes de décharge des neurones. Celles-ci sont organisées en piles initialement remplies par les décharges des neurones des cartes encodant le stimulus. A chaque fois qu'une carte génère une liste de décharge, la liste est placée en haut de la pile pour être traitée et elle est dépilée quand tous les neurones cibles ont été touchés. Ce type de calcul réellement événementiel offre l'avantage de la simplicité et de ne dépendre *a priori* d'aucune contrainte temporelle (si le temps de conduction entre deux neurones est constant, la résolution temporelle est pratiquement infinie). Cependant, cela pose le problème de l'ordre de traitement des cartes : quand une carte neuronale est connectée à beaucoup d'autres, sur laquelle de ces cartes cibles les décharges des neurones doivent-elles se propager en premier ? Le second problème est celui de l'excitation récurrente : deux cartes peuvent s'exciter indéfiniment si les neurones n'ont pas de période réfractaire[4].

Le calcul centré sur les cartes neuronales présente d'autres avantages par rapport à un calcul centré sur les listes de décharge des neurones. Tout d'abord, la structure même de l'algorithme fait que les cartes vont simplement chercher les informations dans leurs cartes afférentes. Du fait de l'organisation orientée objet du programme, une carte en cours de traitement ne peut pas déterminer si une carte afférente est une carte de réseau, une carte d'entrée ou un autre type de carte : le type de la carte neuronale afférente est totalement transparent. Cela rend le calcul relativement indépendant de l'implémentation. De plus, du fait de la bande passante limitée dans un petit réseau d'ordinateurs, la distribution des décharges des neurones doit être répartie uniformément dans le temps. C'est en effet le cas pour un calcul centré sur les cartes neuronales car dès qu'une carte a terminé ses calculs, elle envoie à travers le réseau une liste de neurones ayant déchargé. Étant donné que chaque carte est traitée séquentiellement, des listes de décharges sont envoyées continuellement sur le réseau. Dans le cas d'un calcul centré sur les listes de décharge, toutes les cartes intégrant les mêmes informations (par exemple des cartes d'orientation intégrant les informations en provenance de la rétine) doivent attendre que la dernière liste de décharge soit propagée : toutes les cartes émettent des décharges pratiquement au même moment, ce qui a tendance à saturer le réseau.

[4] Période suivant immédiatement une décharge pendant laquelle un neurone ne peut pas émettre d'autres décharges.

Enfin un dernier avantage de centrer le calcul sur les cartes neuronales est que cela permet de conserver facilement les listes de décharges des pas de temps précédents, qui peuvent être gérées de façon naturelle par les cartes elles-mêmes. Si les cartes neuronales n'étaient pas explicitement présentes, il serait nécessaire de construire des tableaux complexes de correspondance pour accéder à ces listes de décharges. Comme nous allons le voir, conserver les listes de décharges se révèle très utile pour effectuer des calculs plus complexes.

2 - SpikeNET et les neurones biologiques

2.1 - Neurones "intègre et décharge" et neurones réels

Les modèles neuronaux basés sur les propriétés des canaux ioniques sont couramment utilisés dans les simulations de type biologique HH (Hodgkin et Huxley, 1952). Les modèles intégrateurs à fuite (Lapicque, 1907) - neurone "intègre et décharge" du type de ceux utilisés dans SpikeNET - sont en général considérés comme des modélisations relativement abstraites du comportement des neurones et ont du mal à s'imposer dans le milieu de la simulation biologique. De façon contradictoire, ces neurones sont trop proches de la biologie - synonyme, à tort, de lenteur de calcul - pour trouver leur place parmi les modèles de neurones formels. Une autre raison du manque d'intérêt pour les neurones IFs est que les calculs qu'ils effectuent sont encore mal compris.

On a cependant observé récemment un regain d'intérêt pour les neurones IFs en ce qui concerne l'analyse du comportement de très gros réseaux de neurones (Gerstner, 2000) du fait de leur simplicité comparée aux neurones HH (Hansel et al, 1998). De plus des travaux ont montré que ces neurones étaient capables de simuler le comportement de neurones réels dans certaines parties du système visuel (Reich et al, 1998). Des simulations ont également montré que le comportement des neurones IFs était très proche de celui des neurones HH à un seul compartiment, sous réserve d'effectuer des modifications mineures comme on le verra par la suite (Destexhe, 1997). Enfin concernant la simulation de neurones à plusieurs compartiments, Jaffe et Carnevale (Jaffe et Carnevale, 1999) ont montré à l'aide d'un modèle que, dans certaines conditions, l'arbre dendritique des neurones tendait à normaliser les entrées de façon à ce que l'effet au niveau du soma ne dépende pas de la position de l'entrée sur cet arbre dendritique. Cela justifie donc dans une certaine mesure l'utilisation de neurones à

compartiment unique et donc de neurones IF. Toutes ces études montrent que l'approximation que l'on fait en considérant que les neurones réels se comportent comme des intégrateurs à fuite n'est pas déraisonnable.

2.2 - Définitions

Nous allons maintenant tenter de donner une définition plus rigoureuse du comportement des neurones IF et montrer comment nous avons implémenté ce type de neurones dans SpikeNET. Le potentiel membranaire obéit, quand il est situé sous le seuil de décharge du neurone θ, à l'équation suivante :

$$C\frac{dV}{dt} = -g_l(V - V_l) + I_{syn}(t) \quad (1)$$

où les paramètres g_l et V_l sont respectivement la conductance et le potentiel d'inversion du courant de fuite, I_{syn} le courant synaptique résultant de l'activation des neurones voisins et C la capacitance totale de la membrane du neurone. Si le potentiel membranaire V dépasse le seuil θ alors le neurone émet une décharge et le potentiel du neurone est instantanément réinitialisé au potentiel de repos.

En l'absence d'excitation présynaptique, le courant de fuite fait converger le potentiel du neurone vers le potentiel de repos. La mise à jour du potentiel membranaire de tous les neurones à chaque pas de temps serait dramatique du point de vue du temps de calcul, notamment pour les très gros réseaux de neurones. Dans SpikeNET, du fait de la propagation événementielle des décharges, il est possible d'actualiser le potentiel du neurone par rapport au courant de fuite à chaque fois que ce neurone est stimulé. Le courant de fuite est sans conséquence dans le cas où le neurone n'est pas stimulé car il ne peut pas induire de décharge au sein du neurone. Connaissant la constante de temps du processus de fuite $\tau = C/g_l$, le temps séparant l'excitation courante (temps t) et l'excitation précédente d'un même neurone (temps

t_{last}), il est possible de recalculer le potentiel du neurone de façon exacte à l'aide de l'équation suivante[5] :

$$V(t) = V(t_{last})e^{\frac{-(t-t_{last})}{\tau}} \quad (2)$$

Concernant l'intégration des décharges afférentes, le courant synaptique est donné par I_{syn}

$$I_{syn}(t) = \sum_{i \ neurons} W_i f(t - t_{spike}(i)) \quad (3)$$

où W_i et $t_{spike}(i)$ sont respectivement le poids synaptique entre le neurone traité et le neurone i et la date de décharge du neurone i. L'implémentation de la fonction f, dans sa forme la plus simple, est une fonction en escalier prenant la valeur 1 quand t-$t_{spike}(i)$=Δt (Δt étant le pas de temps de la simulation). Nous verrons que f peut prendre des formes plus complexes dans le paragraphe suivant. Pour l'électrophysiologiste, ce type d'intégration synaptique peut sembler être une simplification abusive. Une description plus réaliste des courants synaptiques voudrait que l'on prenne en compte également les potentiels d'inversion des différentes synapses qui dépendent de la différence de concentration de certains ions entre la fente synaptique et le milieu postsynaptique :

$$I_{syn}(t) = -\sum_{i \ neurons} \bar{g}_i (V - V_i) f(t - t_{spike}(i)) \quad (4)$$

où V_l et g_l sont le potentiel d'inversion et la conductance de la synapse entre le neurone courant et le neurone i (cette formule est valide à la fois pour les synapses excitatrices et inhibitrices, la valeur de V_l déterminant le type de synapse). Bien que l'implémentation de ce type de dynamique (équation 4) ne pose aucune difficulté, je n'ai pas eu l'occasion de l'utiliser.

De tels neurones IF peuvent rendre compte de la majorité des comportements de neurones corticaux observés à ce jour. Dans le LGN, le comportement de ce type de neurone

[5] Il est possible de construire une table de correspondance à partir de l'équation pour accélérer les calculs.

est en adéquation presque totale avec celui des neurones enregistrés (Reich et al, 1997; Reich et al., 1998). Les auteurs de ces travaux montrent également que des simulations basées sur des neurones plus simples ou des fonctions probabilistes ne permettent pas d'obtenir une simulation de la même qualité. Des oscillations comme les rythmes alpha, similaires aux oscillations enregistrées dans le cerveau, ont également été observées au sein de réseaux de neurones IF (Liley et al, 1999). Golomb et Ermentrout (Golomb et Ermentrout, 1999) ont par ailleurs montré que, pour une connectivité donnée, les neurones IF se comportent comme des unités bistables, comportement que l'on a observé pour les neurones du cortex frontal. Enfin, des modèles de neurones IF, simulant l'activité des neurones dans l'hippocampe, se comportent comme des mémoires associatives (Samsonovich et McNaughton, 1997) et convergent vers des attracteurs (Hopfield et Herz, 1995; Maass et Natschläger, 1997) qui sont le fondement de l'efficacité des neurones formels. Tant au niveau de la biologie que du traitement de l'information, les neurones IFs présentent donc des propriétés remarquables.

2.3 - *Shunting Inhibition* et apprentissage

Comme cela est expliqué dans le chapitre III, il est possible d'étendre le modèle "intègre et décharge" pour y inclure un paramètre de désensibilisation en fonction du nombre de décharges reçu par le neurone. Nous avons utilisé cette propriété pour implémenter le modèle de codage par rang imaginé par Simon Thorpe (Thorpe, 1990). Dans ce cadre conceptuel, avant de propager une image, le paramètre de sensibilité de chaque neurone est fixé à 1 et décroît vers 0 au fur et à mesure qu'un neurone est stimulé. Ce paramètre de sensibilité correspondrait à l'inhibition des neurones voisins, un type d'inhibition très rapide qualifiée de *shunting inhibition* car elle divise les courants entrants (voir le chapitre III.1). Le résultat de ce mécanisme est que l'activation du neurone est maximale uniquement quand les décharges des neurones arrivent dans l'ordre des poids synaptiques, les poids les plus élevés devant être activés les premiers. Le neurone est alors sensible à l'ordre d'activation de ces afférents.

Une autre propriété que j'ai introduit dans SpikeNET est la capacité pour les connexions synaptiques de se modifier en fonction des stimulations que reçoit un neurone en provenance de ses afférents. Des travaux récents ont par exemple montré que la plasticité synaptique dépend du temps relatif entre la décharge des neurones afférents et celle du neurone efférent

(Markram et al, 1997). En conservant les listes de décharge afférentes, comme je l'ai indiqué précédemment, il est ainsi possible de reconstruire l'historique de l'excitation d'un neurone émettant une décharge et d'adapter les poids synaptiques en conséquence.

Dans les paragraphes qui suivent je décris d'autres dynamiques qui permettent de rendre compte du comportement des neurones réels. La plupart de ces mécanismes ne sont pas encore totalement implémentés dans SpikeNET et, dans tous les cas, je n'ai pas eu l'occasion de les tester dans un modèle. Cependant il m'a semblé intéressant de les décrire pour illustrer la capacité des réseaux événementiels à simuler la plupart des dynamiques des neurones réels.

2.4 - Dynamique synaptique

Dans les réseaux de neurones réels, la décharge d'un neurone peut stimuler des neurones efférents avec différents délais qui dépendent de la distance séparant les deux neurones : certaines synapses sont activées juste après la décharge neuronale et d'autres plus tardivement. Pour implémenter ce type de synapses efficacement, on peut simplement retarder la propagation des synapses activées tardivement par rapport aux autres synapses. Dans SpikeNET, les délais doivent nécessairement être des multiples du pas de temps de la simulation, mais cette contrainte n'est pas restrictive car, comme on l'a montré précédemment, augmenter le nombre de pas de temps n'a pratiquement aucun effet sur le temps de calcul. Comme cela est illustré dans la figure 4, les synapses activées tardivement seront associées aux listes de décharges neuronales des pas de temps suivants. Chaque carte contient les listes de décharges de ses neurones au pas de temps courant et aux pas de temps précédent et ce type de mécanisme est donc très facile à implémenter.

Les modèles synaptiques peuvent également avoir des dynamiques très complexes (Senn et al, 1997; Tsodycks et al, 1998) dépassant le cadre de canaux ioniques ouverts ou fermés. Par exemple, les synapses alpha, dont l'ensemble des canaux s'ouvrent instantanément mais se ferment en suivant une courbe exponentiellement décroissante, sont couramment utilisées dans les simulations de type biologique. De même les synapses dynamiques, dont la variation des paramètres obéit à un équilibre entre plusieurs états, sont relativement réalistes du point de vue de la simulation du comportement des synapses. Pour implémenter ces mécanismes, il est nécessaire de les discrétiser dans le temps et de pouvoir calculer les équations différentielles

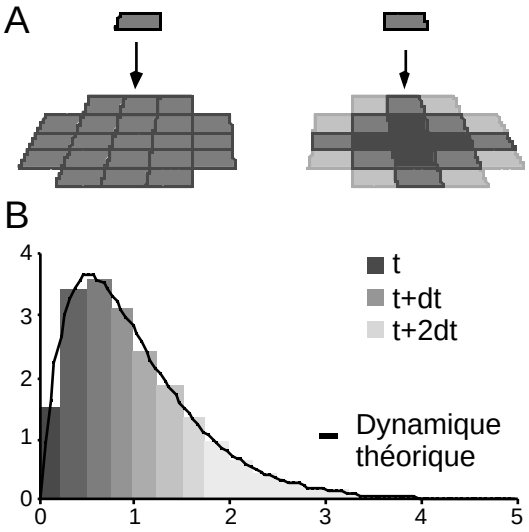


Figure 4 : A, différence entre une propagation sans délai et une propagation avec différents délais. Le carré supérieur représente un neurone afférent projetant vers un ensemble de neurones, chaque carré inférieur représentant un neurone cible. Sur la gauche, quand le neurone afférent décharge, toutes les connexions synaptiques sont traitées au même pas de temps. Sur la droite cependant, les connexions synaptiques sont traitées à des pas de temps différents introduisant un délai dans la propagation de la décharge pour certaines synapses. B, simulation de la dynamique d'une synapse. La courbe en noir, qui représente la dynamique théorique de la synapse, peut être approchée en introduisant différents poids synaptiques, à différents délais (les barres gris foncé étant traité au moment de la décharge du neurone afférent et les barres en gris clair étant traitées plus tardivement). Cela revient en fait à propager la décharge plusieurs fois et le mécanisme est donc similaire à celui utilisé pour la propagation de synapses avec des délais (A).

auxquels ils obéissent pour des variations arbitraires de pas de temps. Comme cela est indiqué dans la figure 4, propager un EPSP complexe revient à propager plusieurs fois la décharge du neurone afférent en modulant le poids synaptique par la dynamique de la synapse. On peut augmenter ou diminuer la précision temporelle pour obtenir un degré de précision acceptable. Cependant dans ce cas précis, le nombre de décharges propagées dépend du nombre d'événements utilisés pour simuler la dynamique de la synapse : trop de précision à ce niveau risque d'affecter dramatiquement le temps de calcul.

2.5 - Vers des neurones artificiels plus proches de la biologie

Concernant les simulations de type biologique, le comportement des neurones IF devient pratiquement indiscernable de celui des neurones HH en incluant la dynamique des canaux voltage-dépendants. Destexhe (1997) a montré qu'en incluant un modèle bimodal des canaux voltage-dépendants (canaux Na+ et K+) il est possible de simuler le modèle HH avec une très grande précision. Le canal passe à un état actif dès que le potentiel dépasse un certain seuil et sa dynamique est alors prise en compte pendant un court laps de temps afin de simuler le potentiel d'action (on simule alors les changements continus des paramètres de ces canaux comme on peut le faire avec des neurones HH), puis ce canal repasse à l'état inactif jusqu'à la prochaine décharge. En comparaison, les modèles simples IF sont moins précis car ils ne simulent pas la durée du potentiel d'action. Ajouter ces propriétés à SpikeNET est en fait beaucoup plus simple qu'on pourrait le croire. Il est possible, par exemple, de calculer de

manière analytique ou d'approcher à l'aide d'une table de correspondance, le potentiel à l'issue de la décharge du neurone en fonction de ses paramètres courants. Ces changements peuvent être directement appliqués au moment de la décharge du neurone, le neurone ne pouvant pas être stimulé pendant un certain laps de temps du fait de la période réfractaire

En utilisant le même type de technique, beaucoup d'autres paramètres peuvent être simulés sans perte de temps de calcul dramatique. En effet, comme on l'a vu, ces calculs peuvent être effectués à chaque fois qu'un neurone est touché, ils ne nécessitent donc pas que tous les neurones soient mis à jour à chaque pas de temps. Par exemple les neurones à compartiment peuvent également être simulés et leur état mis à jour à chaque fois que le neurone reçoit une décharge afférente. On a en effet montré que des neurones avec un nombre très réduit de paramètres et possédant deux compartiments, l'un pour le soma et l'autre pour l'arbre dendritique, rendent relativement bien compte du comportement des modèles de neurones pyramidaux très complexes (Pinsky et Rinzel, 1994).

La principale limitation de SpikeNET reste que tous les processus intervenant au sein des neurones doivent être synchronisés avec les PSP ce qui signifie que dans certains cas, il n'est pas possible de rendre compte des dynamiques du potentiel de membrane sous le seuil du neurone. Si la solution analytique d'un groupe d'équations différentielles gouvernant le comportement du neurone entre deux stimulations existe alors il est possible d'implémenter directement cette solution dans SpikeNET. Si la solution exacte n'est pas calculable, il est toujours possible de tenter d'approcher cette fonction en utilisant des tableaux numériques de correspondance ou en utilisant une fonction polynomiale pour approcher la solution. Dans le cas où la dynamique interne du neurone peut provoquer une décharge peu après la stimulation et que cette fonction n'est pas calculable *a priori,* il est toujours possible de propager des poids synaptiques nuls aux pas de temps suivant la stimulation sur ce neurone. Pour déterminer ces pas temps et la dynamique du neurone, on peut alors utiliser des méthodes standard comme un algorithme d'intégration de 1er ordre (Euler), de 2nd ordre (Runge Kutta) ou même un algorithme utilisant des pas de temps variables comme CVODE[6].

[6] "C" Variable-Coefficient Ordinary Differential Equation solver.

Une dernière extension du modèle IF concerne la probabilité de décharge de ces neurones. Dans certaines conditions, en particulier *in vitro*, les dates de décharge des neurones sont très reproductibles (Mainen et Sejnowski, 1995). Cependant dans la majorité des cas, la variation des dates de décharge est très importante. L'une des principales raisons réside dans le fait que les synapses ne propagent pas systématiquement les décharges des neurones afférents - même si un neurone afférent connecte son efférent en à peu près 10 points. Certains auteurs utilisent un seuil variable bruité pour rendre compte de la variabilité des décharges des neurones (Reich et al, 1997; Reich et al, 1998). Induire des variations du seuil à tous les pas de temps pour tous les neurones n'est pas envisageable dans SpikeNET car la perte de performance serait dramatique. Il semble cependant que si les seuils des neurones sont constants mais bruités initialement, le comportement des neurones est indiscernable de la condition où les seuils varient de façon permanente (Reich et al, 1997). Il est également possible d'utiliser d'autres méthodes pour rendre compte de la variabilité des décharges des neurones comme une réinitialisation du potentiel du neurone à un certain pourcentage du seuil (Bugmann, 1996), ou une probabilité d'échec au niveau des synapses.

2.6 - Analyse de la précision

Concernant la précision temporelle de l'intégration des décharges et du potentiel membranaire, l'erreur commise par SpikeNET est de l'ordre du pas de temps (Δt). Quand un neurone décharge entre le pas de temps t et $t+\Delta t$, une erreur locale d'ordre Δt est générée car la décharge est propagée systématiquement au temps $t+\Delta t$. Cela induit donc une erreur sur le potentiel de membrane au temps $t+\Delta t$ et une autre erreur concernant la date d'arrivée de l'EPSP. Hansel et al (1998) ont proposé d'approcher linéairement la date exacte de décharge du neurone entre le temps t et $t+\Delta t$ afin de réduire cette erreur. L'erreur commise est alors d'ordre Δt^2. Bien que le temps de calcul dépende relativement peu du pas de temps utilisé dans SpikeNET - on peut facilement diviser par 10 le pas de temps sans perte significative de performance - le potentiel de réinitialisation peut aisément être modifié pour intégrer ce type de dynamique. Pour réduire l'erreur commise sur la date d'arrivé des EPSP dans SpikeNET, les listes de décharge des neurones devraient contenir, en plus de la référence du neurone afférent, la date exacte à laquelle il a déchargé.

Dans cette courte revue des mécanismes que l'on peut implémenter à l'aide d'un modèle événementiel, je n'ai pas la prétention d'avoir réalisé une analyse exhaustive. Nous avons vu cependant que la grande majorité des modèles peuvent s'adapter à ce type de propagation. L'implémentation de tous ces mécanismes peut sembler inutile pour certains, mais je suis persuadé qu'elle ne l'est pas car, étant donné la fréquence moyenne de décharge des neurones (environ une décharge par seconde), une propagation événementielle sera toujours plus efficace que la mise à jour de l'ensemble des paramètres des neurones à chaque pas de temps.

3 - Les champs de projection dans SpikeNET

Les projections entre groupes de neurones organisés rétinotopiquement constituent l'organisation sous-jacente du système visuel. Dans le cerveau humain, le système est de plus organisé de façon hiérarchique : les neurones de V1 ont des champs récepteurs (CR) très petits alors que, dans les couches supérieures du système visuel, la taille des CRs augmente avec le niveau hiérarchique. Dans V4, les CRs s'étendent sur à peu près un quart du champ visuel et dans IT certains CRs englobent tout le champ visuel (figure 5). Dans SpikeNET, il est donc nécessaire de pouvoir faire se projeter des cartes neuronales de différentes tailles afin de tenter de comprendre les calculs effectués dans le système visuel humain.

Chaque neurone n'intègre pas directement l'ensemble des décharges de ses afférents. Pour les raisons algorithmiques que j'ai déjà expliquées, les neurones déchargent et effectuent la mise à jour des neurones cibles auxquels ils sont connectés. Il nous faut donc tout d'abord transformer les CRs des neurones en champs de projection (CP). Pour une projection entre des cartes neuronales de même taille, dont le CR de l'ensemble des neurones cibles est

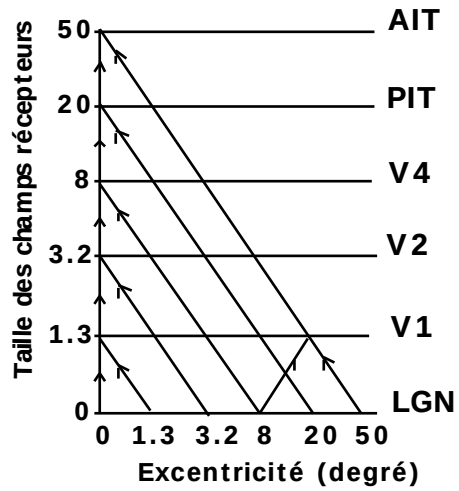


Figure 5 : évolution de la taille du champ récepteur des neurones en fonction du niveau hiérarchique du système visuel. Dans le LGN, à l'entrée du cortex visuel, la taille des champs récepteurs est de 1,3° d'angle visuel. Dans AIT cependant, au niveau hiérarchique le plus élevé dans le système visuel, le champ récepteur des neurones est en moyenne de 100° et peut inclure la totalité du champ visuel. Pour le détail des aires neuronales, se reporter au chapitre I.2. Adapté de Rolls (1992).

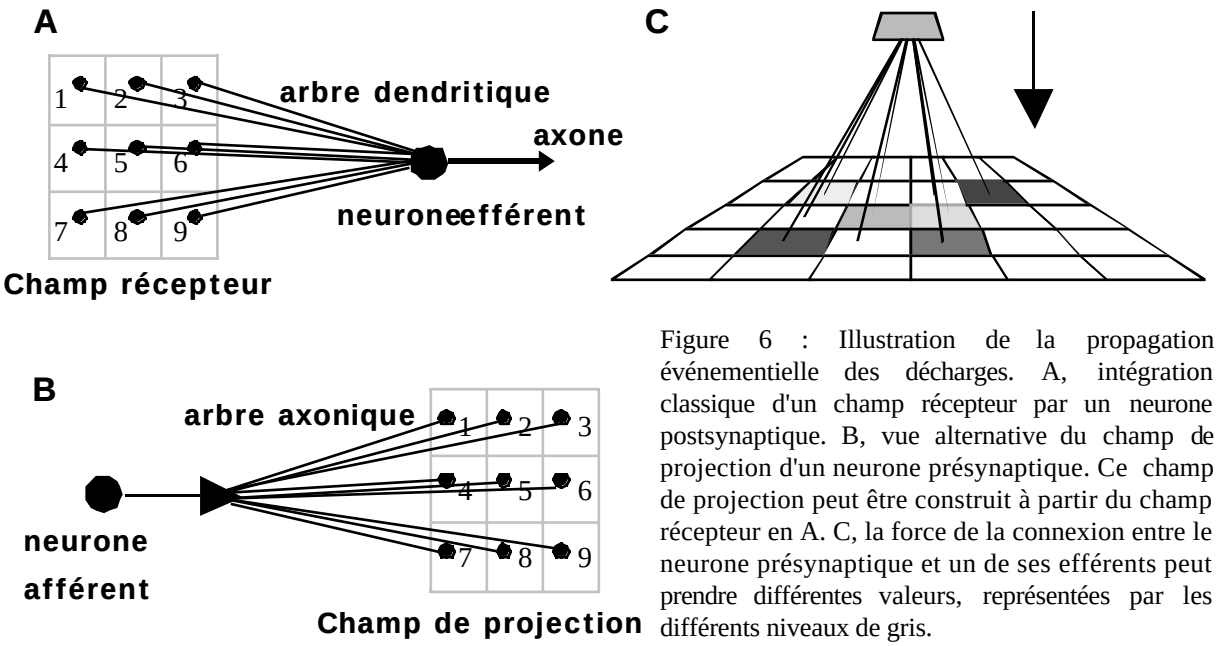


Figure 6 : Illustration de la propagation événementielle des décharges. A, intégration classique d'un champ récepteur par un neurone postsynaptique. B, vue alternative du champ de projection d'un neurone présynaptique. Ce champ de projection peut être construit à partir du champ récepteur en A. C, la force de la connexion entre le neurone présynaptique et un de ses efférents peut prendre différentes valeurs, représentées par les différents niveaux de gris.

uniforme, on obtient naturellement ce résultat en prenant la matrice miroir (horizontalement et verticalement) de la convolution des poids (figure 6). Si chaque neurone postsynaptique possède son propre CR - c'est-à-dire dans le cas de CRs non homogènes - il est alors nécessaire d'utiliser l'ensemble des CRs des neurones efférents à un neurone présynaptique pour reconstruire le CP de celui-ci. Malgré son apparente complexité, ce processus est en fait trivial. Pour des raisons de clarté nous nous focaliserons cependant sur le cas des champs récepteurs homogènes. Une autre raison pour s'intéresser tout particulièrement aux CRs homogènes est qu'ils prennent peu de place en mémoire et peuvent donc être manipulés de manière très rapide par le processeur de l'ordinateur.

Considérant deux cartes neuronales de tailles différentes projetant l'une vers l'autre, le processus permettant de convertir les CRs des neurones postsynaptiques en CPs des neurones présynaptiques est loin d'être trivial. La position d'un neurone cible dans l'espace d'entrée est en général une position non entière (qui n'est en face d'aucun neurone présynaptique en particulier) et il est donc nécessaire de déformer et de recalculer le CR des neurones cibles pour déterminer leurs connexions avec les neurones d'entrée. Nous allons maintenant analyser trois possibilités pour effectuer ce type de transformation, chaque approche donnant des résultats différents complémentaires les uns des autres.

3.1 - Zoom basé sur les champs récepteurs

Le CR des neurones postsynaptiques est toujours défini dans l'espace de coordonnées d'entrée. Pour simuler une architecture neuronale biologiquement plausible du système visuel, il est pas possible d'utiliser le pattern de connectivité d'un neurone tel qu'il peut être défini par des techniques électrophysiologiques. Il en résulte en général un CR qui peut éventuellement être approché par une fonction mathématique. Il est ensuite nécessaire d'implémenter ces résultats dans un simulateur de réseaux de neurones (en l'occurrence SpikeNET).

Dans le cas d'un CR homogène pour les neurones cibles, la solution intuitive serait d'appliquer une transformation en miroir sur la matrice des poids synaptiques comme on l'a fait dans le cas de deux cartes neuronales de même taille. Par la suite, on ajusterait la taille de cette matrice en fonction de la différence de taille des cartes d'entrée et de sortie. Cette solution n'est pas la bonne et peut mener à des erreurs très importantes : dans le cas particulier d'un zoom où les dimensions de la carte neuronale de sortie sont inférieures de moitié à celles de la carte d'entrée (en largeur et en hauteur), chaque neurone dans la carte de sortie correspond à 4 neurones dans la carte d'entrée. En utilisant le processus naïf décrit précédemment, quel que soit le neurone qui décharge en entrée (parmi les 4 neurones correspondant à une même position), l'application du champ de projection se fera exactement à la même position dans la carte de sortie et l'activation résultante des neurones postsynaptiques ne dépendra pas de la

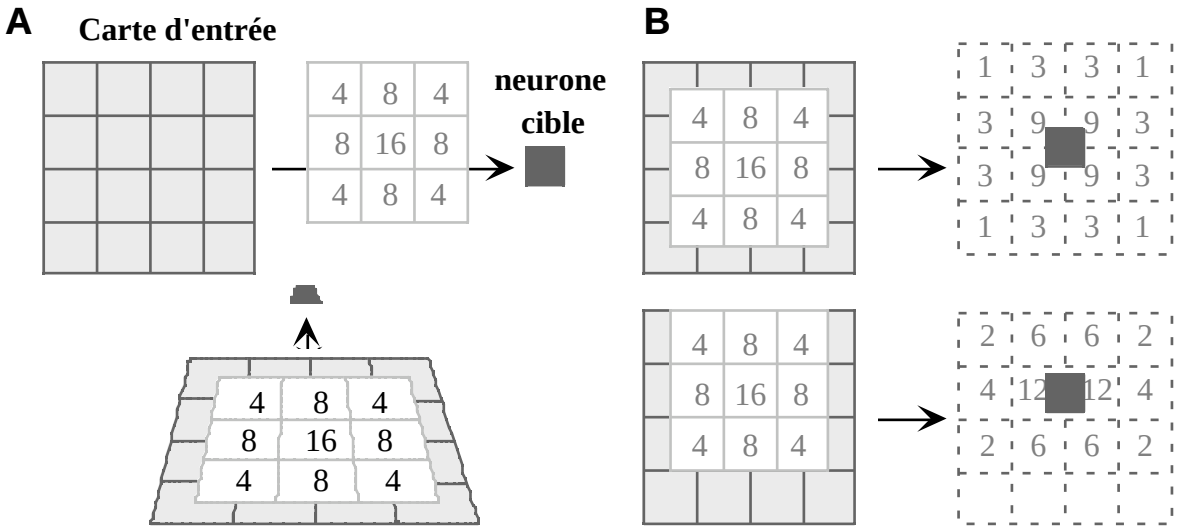


Figure 7 : illustration en deux et en trois dimensions du champ récepteur d'un neurone cible projeté dans l'espace d'entrée. B, champ récepteur effectif du neurone cible pour deux positions différentes de celui-ci. La matrice en pointillé indique la position du neurone cible (en gris foncé) dans l'espace de coordonnés défini par la carte d'entr ée. Les nombres dans la matrice en pointillés représentent le poids synaptique effectif entre les neurones d'entrée et le neurone cible.

Paramètres d'entrée de l'algorithme — *Résultat renvoyé par l'algorithme*

M, **la taille de la carte d'entrée**
N, **la taille de la carte de sortie**
BaseConvolution, **le CR ou CP commun à tous les neurones**

multiConvo, **un groupe de champs de projection effectifs pour les neurones d'entrée**

```
Procedure resample  qui renvoie output, PF de taille SC x SC
    Les paramètres d'entrée de la procédure sont deux entiers indiquant le recadrage à effectuer: shiftx and shifty

    Pour x_rf allant de -SC/2 à SC/2
        Pour y_rf allant de -SC/2 à SC/2
            x_in = (x_rf+x_shiftx)/N*M          coordonnées dans l'espace d'entrée  (on considère des CRs)
            y_in = (y_rf+y_shifty)/N*M
            output[ SC/2 - x_rf ][ SC/2 - y_rf ] = interpolation(baseConvolution, x_in, y_in)
                    ou
            x_in = x_rf/M*N - shiftx            coordonnées dans l'espace de sortie  (on considère des CPs)
            y_in = y_rf/M*N - shifty
            output[ x_rf ][ y_rf ] = interpolatation(baseConvolution, x_in, y_in)
    fin_pours
fin_procedure

La boucle principale de l'algorithme parcourt une partie de l'espace d'entrée
    multiConvo = nouvelle convolution matrice de taille M/pgcd(N,M) * M/pgcd(N,M)
    Pour x allant de 0 à M/pgcd(N,M)
        Pour y allant de 0 à M/pgcd(N,M)
            x_shift = x / M * N - Floor( x / M * N )
            y_shift = y / M * N - Floor( y / M * N )
            multiConvo[ x ][ y ] = resample(x_shift , y_shift )
        fin_pour
    fin_pour
```

Figure 8 : Description de l'algorithme utilisé pour calculer les projections des neurones d'entrée vers les neurone-cibles en fonction de leurs positions. L'algorithme engendre M/pgcd(N,M) champs de projection pour les neurones d'entrée. Pour chaque position dans la matrice des neurones d'entrée, la procédure *resample* permet de recalculer le poids synaptique exact vers le neurone de sortie considéré en effectuant une interpolation à partir de la convolution de base *baseConvolution*. La fonction d'interpolation (non représentée) renvoie une approximation du poids synaptique (en utilisant un modèle linéaire ou gaussien) à une position non entière de la convolution de base. Pour des champs récepteurs non homogènes entre les neurones d'une même carte, le même algorithme peut être utilisé mais la convolution de base change en fonction de la position considérée.

position du neurone présynaptique qui a déchargé. On perd alors les fréquences spatiales les plus hautes. Une alternative évidente, sur laquelle porte le reste de ce paragraphe, consiste à appliquer un champ de projection différent dans la carte de sortie en fonction du neurone présynaptique qui a déchargé.

Considérant la position (non entière) du CR d'un neurone cible dans l'espace d'entrée, il est nécessaire de recalculer ses poids synaptiques par rapport au pattern de connectivité global. En fonction des coordonnés du neurone de sortie dans la carte d'entrée, le CR du neurone va se superposer d'une certaine façon avec la position des neurones présynaptiques. Ainsi, dans la figure 7, deux champs récepteurs de neurones cibles se superposent différemment avec les positions des neurones d'entrée ce qui produit différents CRs effectifs pour ces deux neurones cibles. Il existe différentes techniques pour recalculer le CR du neurone cible. J'ai personnellement choisi une interpolation simple et une interpolation gaussienne. Dans une interpolation simple, pour chaque paire d'un neurone d'entrée et d'un neurone de sortie, la position du neurone d'entrée est projetée à une position non entière dans le champs

récepteur et le poids synaptique entre les deux neurones est recalculé à partir des valeurs avoisinant la position du neurone d'entrée dans le champ récepteur (à l'aide d'une interpolation linéaire ou gaussienne). Dans le cas où les CRs sont définis par des fonctions mathématiques et non par des scalaires, il est uniquement nécessaire de recalculer cette fonction pour la position du neurone d'entrée par rapport au neurone de sortie. L'algorithme permettant d'effectuer ces transformations est présenté en figure 8. Dans le cas de CRs homogènes, il n'est pas nécessaire de générer un CP par neurone d'entrée : pour une carte d'entrée de taille m^2 et une carte de sortie de taille n^2, seules $(n/s)^2$ champs de projection sont nécessaire avec $s = PGCD(m,n)$[7]. Le point central à comprendre ici est que, pour un même CR dans la carte neuronale de sortie, il est nécessaire de générer plusieurs CPs dans l'espace d'entrée.

3.2 - Zoom basé sur les champs de projection

Nous pouvons considérer les champs récepteurs des neurones de sortie, mais nous pouvons également considérer les champs de projection (CP) des neurones d'entrée. Dans ce cas, les champs de projection sont directement définis dans l'espace de sortie. Pour le champ de projection d'un neurone d'entrée dans la carte de sortie, je tente tout d'abord de déterminer les coordonnées exactes de ce neurone (souvent non entières) dans l'espace de sortie (figure 9). Cette approche est similaire à celle que nous avons utilisée dans le paragraphe précédent mais il s'agit ici de champs de projection définis dans l'espace de sortie et non plus de champs récepteurs définis dans l'espace d'entrée. Par exemple dans la figure 9, deux champs de projection de deux neurones d'entrée se superposent différemment avec les positions des neurones de sortie, donnant lieu à deux champs de projection effectifs différents pour ces neurones d'entrée. Comme je l'ai déjà mentionné précédemment il est possible d'utiliser différentes techniques pour interpoler le champ de projection initial. De plus si les champs de projection sont définis par des fonctions mathématiques alors le poids synaptique correspondant à la connexion entre un neurone d'entrée et un neurone de sortie peut être directement calculé. L'algorithme utilisé pour effectuer toutes ces opérations est pratiquement identique à celui utilisé dans le paragraphe précédent (figure 8). Comme dans le cas d'un zoom

[7] Le PGCD est le Plus Grand Commun Diviseur de deux nombres.

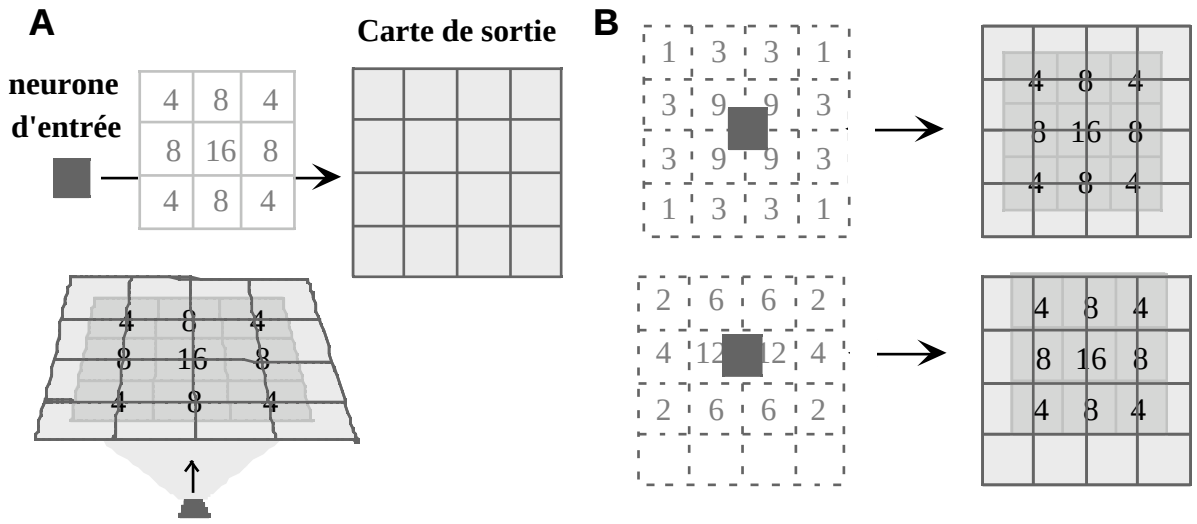


Figure 9 : illustration en deux et en trois dimensions du champ de projection d'un neurone d'entrée vers les neurones cibles. B, champ de projection effectif du neurone d'entrée pour deux positions différentes de celui-ci. La matrice en pointillés indique la position du neurone d'entrée (en gris foncé) dans l'espace de coordonnées défini par la carte de sortie. Les nombres dans la matrice en pointillés représentent le poids synaptique effectif entre le neurone d'entrée et les neurones cibles.

de champs récepteurs, pour une carte d'entrée de taille m^2 et une carte de sortie de taille n^2, seuls $(n/s)^2$ champs de projection sont nécessaires, avec $s = PGCD(m,n)$.

3.3 - Zoom et apprentissage

À l'aide de l'algorithme défini dans la figure 8, il est possible d'implémenter des projections convergentes ou divergentes. La convergence entre groupes de neurones rétinotopiquement organisés permet de simuler ce qui peut se passer entre différentes aires corticales dans le système visuel, de V1 vers V2, de V2 vers V4... De façon similaire, les projections divergentes dans SpikeNET permettent de simuler le *feedback* des aires de haut niveau vers les aires de plus bas niveau (V4 vers V2 et V2 vers V1 par exemple). Au sein de cette architecture, il serait intéressant de pouvoir implémenter des mécanismes d'apprentissage. Dans le cas de CRs non homogènes, chaque connexion synaptique peut être modifiée indépendamment des autres, et le problème consiste simplement à passer d'un CP à un CR et réciproquement. Cependant dans le cas de CRs homogènes, il est nécessaire de reporter chaque modification synaptique au niveau d'un neurone sur l'ensemble des neurones de la carte neuronale.

Pour effectuer de telles transformations, il faut conserver le lien entre chaque poids synaptique et la matrice de poids initiale. Ainsi, dès qu'un poids synaptique est modifié, il est

possible de reporter cette modification dans la matrice initiale (en extrapolant la valeur si nécessaire) et de reporter ces modifications dans l'ensemble des CPs des neurones. Cela vaut aussi bien dans le cas de zooms de CRs que de zooms de CPs.

3.4 - Faire varier l'échelle

Dans certains cas, il peut être intéressant d'effectuer l'apprentissage à une certaine échelle et d'effectuer ensuite la simulation à d'autres échelles (pour obtenir le meilleur rapport vitesse/performance par exemple). Je considère, pour simplifier, que les CRs ou CPs sont définis tous deux entre deux cartes neuronales à la même échelle. Il nous faut donc adapter cette matrice de poids en fonction du zoom que l'on veut obtenir entre la carte d'entrée et la carte de sortie. C'est une transformation en deux temps : après avoir zoomé le CR ou le CP, il

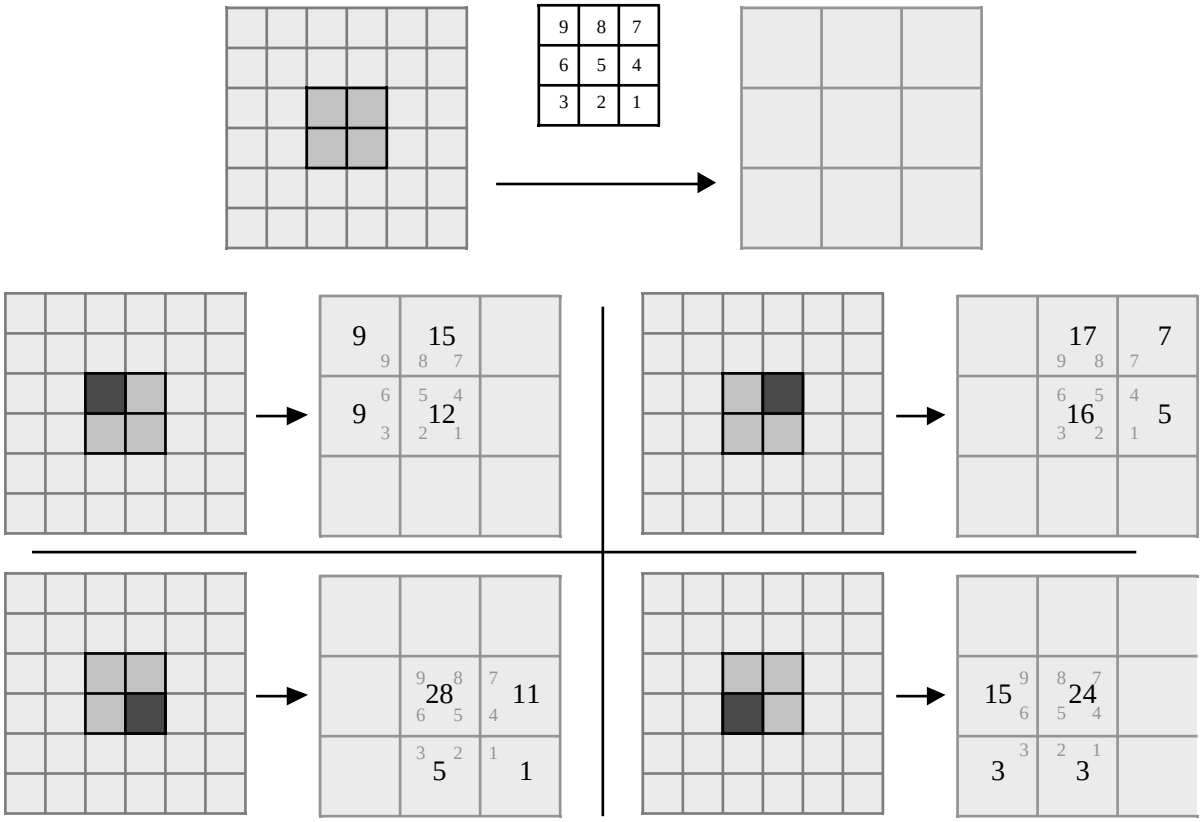


Figure 10 : en haut, projection d'une carte neuronale vers une autre carte possédant quatre fois moins de neurones. La matrice représente le champ de projection des neurones d'entrée vers la carte de sortie. En bas, préservation de l'activation de neurones hypothétiques d'une carte de sortie non réduite. Le carré gris foncé indique le neurone activé en entrée et les chiffres grisés dans la carte de sortie indiquent la propagation du champ de projection pour ce neurone. Les nombres en noir, calculés à partir des gris, indiquent la somme des activations reçues par les neurones postsynaptiques. Ces nombres constituent en fait le champ de projection du neurone présynaptique. Suivant la position du neurone qui décharge dans la carte d'entrée, quatre champs de projection sont générés. Ce type de propagation permet, au niveau de la carte de sortie, de conserver, en quelque sorte, l'activité d'une carte de sortie qui aurait gardé la même résolution que la carte d'entrée.

est nécessaire de transformer la nouvelle matrice en poids synaptiques effectifs. La première opération peut être effectuée en utilisant des techniques de traitement d'image et la seconde se base sur l'algorithme décrit dans les paragraphes précédents. Concernant la première opération, on effectuera par exemple une réduction ou une augmentation de la taille de la matrice en fonction des tailles des cartes d'entrée et de sortie. Bien évidemment, si l'on possède une fonction mathématique représentant la matrice, il est possible de modifier cette fonction en la convoluant par exemple avec une fonction gaussienne dont la largeur σ dépendra des tailles des cartes d'entrée et de sortie.

Une solution alternative à la réduction ou l'agrandissement des CRs et CPs est de tenter de préserver dans la carte neuronale de sortie zoomée l'activité des neurones de la carte de sortie non zoomée. Pour des zooms négatifs ou positifs, on propage les poids synaptiques comme si les deux cartes étaient de même taille (définie par la carte de plus grande taille). Puis, si la carte neuronale de sortie est par exemple de taille inférieure à la carte d'entrée - zooms positifs - les poids synaptiques entre neurones sont calculés afin que l'activité des neurones de la carte de sortie "résume" celle des neurones dans la carte non zoomée. Comme le montre la figure 10, en utilisant cette technique on tente de préserver l'activité des neurones d'une carte : le CP des neurones est calculé à partir du CR ou du CP défini entre les cartes non zoomées.

3.5 - Perte de sélectivité liée au zoom

Pour les zooms basés sur l'activité que nous venons de voir et ceux basés sur une conservation du CP des neurones, les CRs théoriques des neurones de sortie diffèrent en fonction de leurs positions. Cependant les CPs de tous les neurones d'entrée, définis soit dans l'espace de sortie (zoom des CPs) soit dans l'espace d'entrée (zoom de l'activité) implémentent la même fonction. À l'opposé, pour les zooms de CRs, les CRs des neurones de sortie sont préservés : les poids synaptiques effectifs peuvent être différents mais le CR théorique est le même.

Des zooms positifs induiront des pertes d'informations dans le cas d'un zoom basé sur les CRs ou les CPs. Par rapport à deux cartes neuronales de même taille, dans un processus de zoom de CR où l'on utilise la même matrice de poids, le CR des neurones dans la carte neuronale de sortie sera sous-échantillonné par rapport à une version non zoomée. Cela

signifie que même si les neurones conservent la même sélectivité en termes de reconnaissance de forme, l'ensemble des neurones n'est plus capable de détecter cet objet à toutes les positions (les neurones détectent l'objet à des positions bien précises). En appliquant un filtre gaussien sur le CR, on atténue cet effet mais les neurones ne sont plus aussi sélectifs aux hautes fréquences spatiales des stimuli. Il faut donc trouver le bon équilibre en termes de diminution de sélectivité et de reconnaissance globale pour l'ensemble des positions d'un stimulus dans l'espace d'entrée.

Dans le cas de zoom de CPs, les neurones de la carte neuronale de sortie ne présenteront pas la même sélectivité que des neurones d'une carte de sortie non zoomée. Pour une carte de sortie avec un zoom positif, les CRs des neurones de sortie sont beaucoup plus grands que ceux de la version non zoomée de cette même carte de sortie et les neurones seront donc sélectifs à des formes zoomées de stimulus dans la carte d'entrée. Dans ce type de transformation, il n'y a pas de sous-échantillonnage mais, comme dans le cas précédent, les neurones de sortie ne sont plus sélectifs aux hautes fréquences spatiales. Pour tous ces types de zoom, il faut trouver un compromis entre la sélectivité des neurones et la vitesse de calcul que l'on désire atteindre.

4 - Performances de SpikeNET

SpikeNET a été conçu pour effectuer des millions de propagations synaptiques par seconde. Du fait du nombre réduit de paramètres par neurone (16 bytes en fonction de la précision requise), SpikeNET peut utiliser efficacement la mémoire cache du processeur[8] pour atteindre des performances encore inégalées, à ma connaissance, du point de vue du nombre de connexions synaptiques traitées par seconde. De plus, l'utilisation de poids partagés, commun à plusieurs neurones au sein d'une carte neuronale, permet une utilisation optimale de cette mémoire cache du processeur. Pour de simples neurones IFs, SpikeNET est capable de mettre à jour environ 20 millions de connexions synaptiques par seconde et cela même quand les neurones sont sensibles à l'ordre de décharge de leurs afférents (Macintosh PowerPC 750 processeur à 266 Mhz). Cela est suffisant pour simuler un réseau de 400 000 neurones ayant

[8] Mémoire privée du processeur, en quantité très faible mais extrêmement rapide.

chacun 50 connexions afférentes en temps réel avec des pas de temps de 1 ms (on suppose une fréquence de décharge d'environ 1Hz ce qui correspond approximativement à celle des neurones corticaux). Notons que dans les réseaux de neurones conventionnels, il est nécessaire de recalculer l'activité de chaque neurone à chaque pas de temps et que, à puissance de calcul équivalente, il est seulement possible de simuler 20 000 connexions par seconde ce qui correspondrait à environ 400 neurones.

La performance est clairement impressionnante pour des poids partagés mais même dans le cas de poids non partagés, où chaque neurone détient ses propres poids synaptiques efférents et afférents, la vitesse de calcul est seulement divisée de moitié. De plus, comme je l'ai déjà mentionné, augmenter la résolution temporelle de 1 ms à 0,1 ms n'a pratiquement aucun effet sur le temps de calcul et ajouter une fuite aux neurones augmente le temps de calcul d'environ 30 à 40 %.

Bien qu'utiliser SpikeNET sur des ordinateurs de bureau ou de simples stations de travail soit déjà suffisamment rapide, la structure même du logiciel en fait un excellent candidat pour des implémentations parallèles. Dans ce type d'implémentation, le facteur limitant est en général la quantité de données à transférer entre les différents processeurs : l'augmentation de performance entre 4 et 8 processeurs est donc en général relativement faible. Dans SpikeNET, la seule information à transférer entre les différents processeurs est constituée des listes de neurones ayant déchargé, chaque neurone étant représenté par 1-2 bytes dans ces listes. Un réseau de 10 millions de neurones déchargeant en moyenne 1 fois par seconde peut donc être simulé avec des technologies comme fast-Ethernet déjà présentes sur le marché.

Il n'est pas réellement possible d'effectuer une comparaison entre SpikeNET et les autres simulateurs de réseaux de neurones. Soit ces réseaux modélisent les mécanismes détaillés en utilisant des techniques standard (pour une revue détaillée cf. De Schutter, 1992), soit ils traitent de neurones formels dont le comportement n'a plus rien à voir avec celui des neurones réels (ces neurones par exemple n'émettent pas de décharges). SpikeNET se trouve, de façon très surprenante, seul au milieu de ces deux extrêmes. Le seul logiciel effectuant également des calculs basés sur des événements se révèle très peu efficace pour simuler des réseaux de neurones de plus de 100 neurones (Graßmann et Anlauf, 1998) : les neurones y étant

considérés individuellement, ce système rencontre de gros problèmes de synchronisation (Graßmann, communication personnelle).

Malgré le contenu très technique de cette partie, elle m'a semblé primordiale tant du point de vue pratique que théorique. Quand on tente, par exemple, de comprendre comment les champs récepteurs des neurones doivent être transformés lors de zooms, on est au cœur des processus qui opèrent dans le système visuel. Le type de simulateur comme SpikeNET est voué à un avenir très prometteur car il permet de relier la cognition (par exemple la reconnaissance d'objet) à l'activité des neurones.

Annexe 3

Articles publiés présentés dans cette annexe

Delorme, A., Gautrais, J., VanRullen, R., et Thorpe, S. J. (1999). SpikeNET: A simulator for modeling large networks of integrate and fire neurons. Neurocomputing, 26-27, 989-996.

Delorme, A., Richard, G., et Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. Vision Research, 40(16), 2187-2200.

Fabre-Thorpe, M., **Delorme, A**., Marlot, C., et Thorpe, S. J. (2000). A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes. Journal of Cognitive Neuroscience, in press.

Autres articles publiés

Van Rullen, R., Gautrais, J., **Delorme, A**., & Thorpe, S. (1998). Face processing using one spike per neurone. Biosystems, 48(1-3), 229-239.

Delorme, A., Richard, G., et Fabre-Thorpe, M. (1999). Rapid processing of complex natural scenes: a role for the magnocellular pathway. Neurocomputing, 26-27, 663-670.

Fabre-Thorpe, M., **Delorme, A.,** et Richard, G. (1999). Singes et Hommes face au monde visuel : la Catégorisation. Primatologie, 2, 111-139.

Thorpe, S. J., **Delorme, A.**, VanRullen, R., et Paquier, W. (2000). Reverse engineering of the visual system using networks of spiking neurons. Proceedings of the IEEE 2000 International Symposium on Circuits and Systems, IEEE press. IV: 405-408.

Quelques articles en préparation ou soumis

Delorme, A., Thorpe, S.J. Face identification using one spike per neuron: resistance to image degradations. Neural Networks, submitted.

Delorme, A., Rousselet, G., & Fabre-Thorpe, M. Fast categorisation of natural scenes: top-down neuronal mechanism underlying decision processes. In preparation for Nature Neuroscience.

Delorme, A., & Thorpe, S. Modeling Early Cortical Orientation Selectivity using Fast Shunting Inhibition and Rank Order Coding. In preparation for Neural Computation.

SPIKENET : A SIMULATOR FOR MODELING LARGE NETWORKS OF INTEGRATE AND FIRE NEURONS

Arnaud Delorme, Jacques Gautrais, Rufin van Rullen & Simon Thorpe
Centre de Recherche Cerveau & Cognition
133, route de Narbonne, 31062, Toulouse, France
arno@cerco.ups-tlse.fr, gautrais@cerco.ups-tlse.fr, rufin@cerco.ups-tlse.fr, thorpe@cerco.ups-tlse.fr

Abstract

SpikeNET is a simulator for modeling large networks of asynchronously spiking neurons. It uses simple integrate-and-fire neurons which undergo step-like changes in membrane potential when synaptic inputs arrive. If a threshold is exceeded, the potential is reset and the neuron added to a list to be propagated on the next time step. Using such spike lists greatly reduces the computations associated with large networks, and simplifies implementations using parallel hardware since inter-processor communication can be limited to sending lists of the neurons which just fired. We have used it to model complex multi-layer architectures based on the primate visual system that involve millions of neurons and billions of synaptic connections. Such models are not only biological but also efficient, robust and very fast, qualities which they share with the human visual system.

Keywords : Modeling software, Natural scenes, categorization, biological visual systems.

1. Introduction

There are currently a large number of different systems that can be used for simulating neural networks. Many have been designed for simulating networks of artificial neurons and make no attempt to model the detailed biophysics of neurons. The underlying units have no structure, and their outputs typically consist of a single continuous value (often in the range 0 to 1 or from -1 to +1). While such systems have been widely used, and have had applications in a wide range of engineering and financial areas, few would regard them as being useful as tools for the computational neuroscientist.

At the other end of the spectrum there are sophisticated programs such as GENESIS and NEURON which are good for performing detailed biophysical simulations that take into account factors like the dendritic structure and complex channel kinetics, but where the level of detail makes it difficult to simulate very large networks efficiently [2, 3].

In this paper we describe SpikeNET, a neural network simulation package written in highly portable C++ code which lies between these two extremes. It is sufficiently biologically realistic to make it possible to examine the role of temporal properties such as synchronous or asynchronous spiking in neurons, and yet sufficiently simple to allow real-time simulation of large scale networks of neurons.

2. Basic Organization

The basic objects in SpikeNET are two dimensional arrays of relatively simple leaky integrate-and-fire neurons. Each unit is characterized by a small number of parameters : a membrane potential, a threshold, and (in some cases) a membrane time constant. When an afferent neuron fires, the weight of the synapse between the two neurons is added to the target neuron's potential, and we test to see whether the neuron's potential has exceeded the threshold. If so, the neuron is reset (by subtracting the threshold) and the neuron is added to the list of neurons that have fired in the current time

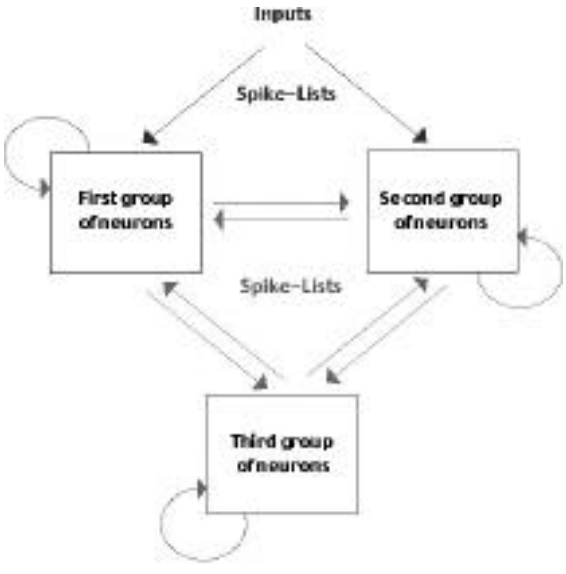


Figure 1: Basic organization of SpikeNET. SpikeNET redirects lists of spikes between different groups of neurons organized in two-dimensional arrays. Since only a small percentage of cells fire in each time-step, communication overheads are kept to a minimum.

step. Propagation of activity within SpikeNET involves sending lists of spikes between neuronal arrays as illustrated in figure 1. The event-driven nature of spike propagation is one of the reasons for the efficiency of SpikeNET as a modeling system.

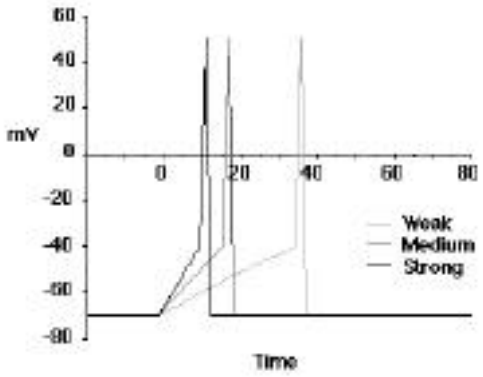


Figure 2: Basic behavior of an integrate and fire neuron. The latency of its discharge depends on the strength of the stimulation. With strong stimulation, the neuron will reach threshold quickly whereas with weak stimulation the latency will increase.

The basic cellular model can be made more complex by including a sensitivity parameter which modulates the effect of incoming action potentials. We have used this feature to implement a rank-order coding scheme which we have developed [6]. According to this scheme, the sensitivity parameter is initially fixed at 1.0, but decreases by a fixed percentage with each incoming impulse, resulting in a progressive desensitization of the post-synaptic neuron which can be thought of in terms of fast shunting inhibition [1]. The net result of this mechanism is that activation is maximal only when the spikes arrive in the order of the weights - with the highest weight synapses being activated first. If desired, this desensitization process can be made specific to particular sets of inputs such that, for example, inputs from the thalamic could mutually desensitize each other without affecting the efficacy of intra-cortical inputs to pyramidal cells. These more complex models for individual neurons are designed to mimic some of the effects of the dendritic structure of neurons while at the same time avoiding the computationally expensive detailed modeling that is normally required.

Most neurons are only affected by spikes in their afferent neurons. However, for certain "input" cells, corresponding for example to cells in the retina, we determine spike timing by a direct calculation that depends on the stimulus. Thus for retinal ganglion cells, we can perform a local "Mexican-hat" convolution on the image, and this value is used to calculate the latency of the unit's spike - the earliest latencies correspond to those cells for which the value of the convolution is highest, whereas lower activation levels result in progressively longer latencies (figure 2).

3. SpikeNET in Action

To illustrate how SpikeNET can be used, we will describe a multiscale face recognition network which extends the face-localization model described by Van Rullen et al [7], and uses an architecture loosely based on the organization of the primate visual system. Input images are first analyzed by arrays of ON-center and OFF-center cells in the "retina" at three different spatial scales. These cells send spikes to neurons in the next layer which contains neurons tuned for 8 different orientations at each spatial scale. Lateral interactions between cells in this layer were used to improve selectivity, and are similar to those described by Zhaoping Li [4]. A weak shunting inhibition was also included to make the neurons sensitive to the order of activation of their inputs. A third layer in the network contains neurons selective for faces at the three spatial scales. The connections between the level 2 orientation maps and these face-selective units were trained using a set of 200 photographs of faces and a supervised learning procedure which attaches high weights to inputs which are systematically among the first to fire, and progressively smaller weights to later firing inputs. Finally, a fourth layer of neurons contains neurons which integrate the information at the three different spatial scales in the previous layer.

As can be seen from Figure 3, the simulation is successful in that in the final map, the neurons fire if a face, at any scale, is present in the input image.

The model is clearly not very realistic. For example, no attempt was made to model change in resolution with retinal eccentricity, but the architecture illustrated here demonstrates how SpikeNET can be used to create quite complex multilayer architectures involving large numbers of units, and it shows how different hypotheses could be tested and integrated easily in a biologically plausible neural network.

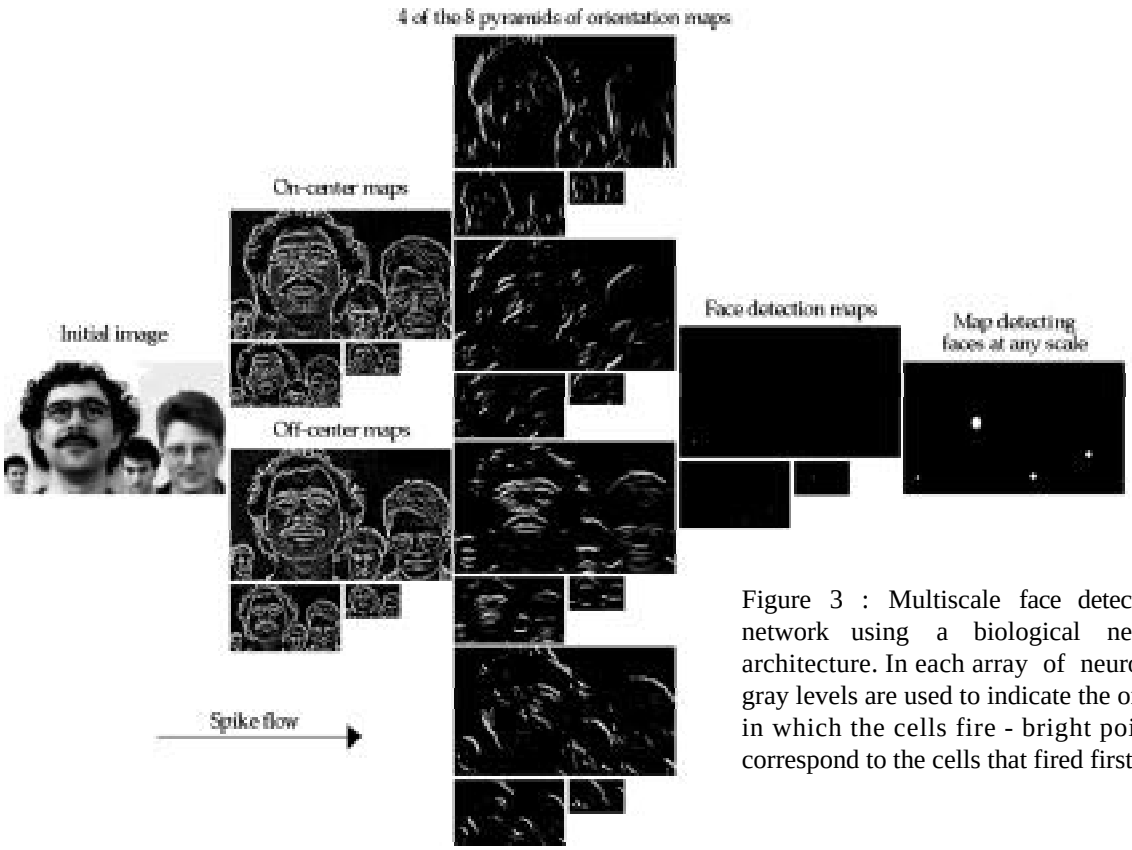


Figure 3 : Multiscale face detection network using a biological neural architecture. In each array of neurons, gray levels are used to indicate the order in which the cells fire - bright points correspond to the cells that fired first.

4. Performance of SpikeNET

SpikeNET has been designed to be computationally efficient. One of its advantages comes from the efficient use of RAM. Since the number of parameters per neuron is kept low, each neuron can require as little as 16 bytes of memory, depending on the type of precision required. More importantly, the use of shared weights means that one set of weights can be used for all the neurons in an array. As a result it is perfectly reasonable to simulate networks with tens of millions of neurons and billions of synapses on standard desktop computers.

The second advantage of SpikeNET is speed. Using a standard G3 Macintosh (PowerPC 750 processor at 266 Mhz), SpikeNET can update roughly 20 million connections per second, even when using the sensitivity parameter to modulate the effect of each synaptic input. This is sufficient to model a network of 400 000 neurons in real time, using a time step of 1 ms (assuming 49 connections per neuron, and an average firing rate of 1 spike per second, a value which is a reasonable estimate for the average firing rate of cortical neurons). Note that with a more conventional neural network simulation approach one has to recalculate every unit at every time step, and so the same computational power would only allow 20 000 connections to be calculated per millisecond, which with 49 connections per neuron would limit real-time simulation to around 400 neurons.

Performance is clearly optimal with shared weights, but even when each neuron has its own set of weights (which obviously increases RAM usage very considerably), speed only drops by a factor of around 2. Adding a decay to neurons to simulate the leaky nature of the synaptic integration process adds roughly 30-40% to the computation time (the exact value depends on the number of time steps in the simulation). Finally, note that increasing the time resolution from 1 ms to 0.1 ms has virtually no effect on computation time, since the number of spikes that are propagated does not change.

5. Parallel SpikeNET

Although running SpikeNET on a standard desktop machine is already reasonably quick, the very nature of SpikeNET makes it an ideal candidate for implementation on parallel hardware. The factor which usually prevents large scale use of parallel hardware in computing is the amount of communications needed between processors. For many problems, one sees little speed up once the computation has been split between more than 4 or 8 processors. However, with SpikeNET, the only information that needs to be transferred between processors are the Spike Lists. The format used by SpikeNET means that the identity of each neuron

which fired can be transmitted using only around 1-2 bytes, and so even a network with 10 million neurons firing at an average of one spike per second could be simulated in real time without saturating the bandwidth of a cluster of processors linked by conventional fast Ethernet technology. We are currently developing multiprocessor PCI boards which will allow real time simulation of even larger networks of neurons.

6. Final Comments

Although primarily designed as a tool for modeling biological neural networks, the level of performance obtained with SpikeNET is such that in a variety of tasks, processing architectures developed using SpikeNET can perform at least as well and in many cases substantially better than more conventional image processing techniques. To the biologist, this may not be so surprising. We know that the processing strategies and architectures used in the human visual system (for example) are the end-product of hundreds of millions of years of intense natural selection. The levels of performance achieved by the human visual system are orders of magnitude better than even the most sophisticated artificial vision systems [5]. By elucidating the computational principles which make this level of performance possible, it may well be possible not only to demonstrate the power of computational neuroscience as a paradigm for understanding biology, but may reveal the potential of the discipline in areas as diverse as machine vision and artificial intelligence.

References

[1] L. J. Borg-Graham, C. F. Monier, Y. Frégnac, Visual input evokes transient and strong shunting inhibition in visual cortical neurons, Nature 393, (1998) 369-73.

[2] J. M. Bower, D. Beeman, The book of GENESIS: Exploring realistic neural models with the GEneral SImulation System. Second Edition Springer-Verlag, New York, (1998).

[3] M. L. Hines, N. T. Carnevale, The NEURON simulation environment, Neural Computation 9, (1997) 1179-1209.

[4] Z. Li, A neural model of contour integration in the primary visual cortex, Neural Computation 10, (1998) 903-40.

[5] S. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381, (1996) 520-522.

[6] S. J. Thorpe, J. Gautrais, Rank Order Coding: A new coding scheme for rapid processing in neural networks, in: J. Bower, Ed, Computational Neuroscience: Trends in Research, (Plenum Press, New York, 1998) 113-118.

[7] R. Van Rullen, J. Gautrais, A. Delorme, S. J. Thorpe, Face detection using one spike per neurone, Biosystems (In press), (1998).

Ultra-Rapid Categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans

Delorme, A., Richard, G. & Fabre-Thorpe, M.

Centre de Recherche Cerveau et Cognition (UMR 5549) Faculté de Médecine de Rangueil. 133, route de Narbonne, 31062 – Toulouse - FRANCE

In a rapid categorisation task, monkeys and humans had to detect a target (animal or food) in briefly flashed, previously unseen, natural images. Removing colour cues had little effect on average performance. In both tasks, accuracy and latency of the fastest behavioural responses were unaffected. Impairments were restricted to a mild accuracy drop (in some of the human subjects) and a small mean RT increase (10-15 ms) observed both in monkeys and humans but only in the detection of food targets. We suggest that rapid categorisation might depend on the feed-forward processing of the early coarse achromatic magnocellular information.

Key words*:* colour, natural scenes, categorisation, primate, visual processing

INTRODUCTION

The recognition of objects and scenes appears effortless and almost instantaneous. The complex processing needed to perform an object categorisation task can be extremely fast in humans (Thorpe, Fize, & Marlot, 1996). Subjects could detect an animal in briefly flashed, previously unseen natural photographs with a high rate of success (94% of correct responses) and very short reaction times (median RT: 445 ms). Frontal ERPs differed sharply on target and non-target trials from 150 ms after stimulus onset. As the underlying visual processing probably involves all the stages along the ventral visual pathway - from retina to the highly integrative infero-temporal cortex - the authors argued that the underlying processing should be essentially feed-forward to be compatible with such time constraint. This constraint on processing speed could be even more severe in monkeys; using the same fast visual categorisation task, rhesus macaques can detect food or animal targets in natural scenes with a correct response rate close to 90% but with much faster behavioural RTs (median RT: 290 ms) than humans (Fabre-Thorpe, Richard, & Thorpe, 1998).

Which image features could be used in generating such fast responses? Although this high level of accuracy cannot be explained by the use of one single image feature, the monkeys could use a combination of different low level cues, and among them colour is an obvious candidate. There is evidence for such a view, for example, to determine whether a photograph contained a human being, Capuchin monkeys may have used the presence of a red patch (D'Amato & Van Sant, 1988). But on the other hand, colour cues were shown not to account for the monkeys' performance in a recent study using fish vs. non fish and tree vs. non-tree categorisations (Vogels, 1999a). The importance of colour could depend on whether or not this cue is a diagnostic feature of the target category (Oliva & Schyns, 2000; Tanaka & Presnell, 1999). For instance, in monkeys and humans that had to select photos of kingfishers – a very brightly coloured bird - among photos of other birds, a drop of performance was observed when the pictures were presented in black and white (BW) (Roberts & Mazmanian, 1988).

Colour differences and colour contours may also be used in image segmentation to provide information about object shape and region of interest within individual objects. However, in human object recognition, the role of colour - particularly in the early visual processes leading to fast identification - is still very controversial. Colour appears to interact with object recognition processing when object naming is required but not in verification tasks in which the object name is presented prior to the object (Ostergaard & Davidoff, 1985). Nevertheless, when performing both a verification task and a naming task with either colour photographs or BW drawings, human performance was unaffected by the absence of colour (Biederman & Ju, 1988). The authors argued for a fast access to a coarse structural mental representation of objects; colour would only be used in the recognition of blurred objects, when the shape does not provide enough information for accurate categorisation or in the case of low level vision subjects. When target and

non-target items are very similar in shape (i.e. subordinate classification), colour has indeed been shown to be relevant even when naming is not required (Price & Humphreys, 1989). Alternatively, an advantage was observed with coloured images -over BW ones- in a food object naming task but was not enhanced when the task was done with blurred images or when testing low vision patients (Wurm, Legge, Isenberg, & Luebker, 1993). These studies point towards a role of colour in late stages of processing to facilitate object recognition or naming. However, colour has recently been shown to play a role in very early visual recognition processes, in a delayed match to sample task using natural images (Gegenfurtner, 1997).

The aim of the present study is to test whether colour is an important feature in the rapid visual go/no-go categorisation task that we have used both in monkeys and humans. Processing speed is so fast that the system might have reached its limits (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2000). Thus if colour is one of the relevant features used in the early phases of visual processing, the absence of colour cues should induce an impairment in either accuracy or speed of performance. On the other hand, if the fast responses observed in our task can be produced using feed-forward processing of the fastest visual inputs to cortex, they should be independent of colour features, as the earliest responses in the visual cortex originate from the (achromatic) magnocellular stream, whereas the arrival of parvocellular chromatic information is delayed by roughly 20 ms (Nowak, Munk, Girard, & Bullier, 1995).

In the present study, the role played by colour in fast object categorisation was addressed in humans and monkeys using two different categories of target-objects: "food", and "animal". For both categories one might predict that colour could be important for segmenting images, since both food objects (such as ripe fruit) and animals (such as birds or fish) are often brightly coloured. Coloured and BW natural photographs were mixed at random to prevent subjects (humans or monkeys) from adopting different strategies when using chromatic and achromatic stimuli and to allow more direct comparison of performance in the two conditions. Monkeys and humans were tested on the same task with the same set-up and the same stimuli for a comparative study on the relative role played by colour cues in their strategies. Moreover, as monkeys performed hundreds of trials per day, the role of colour in the categorisation task could be analysed on both new and highly familiar natural scenes.

GENERAL METHODS

Subjects

Three rhesus monkeys were trained either on a Food/non Food (Rh1, male aged 6) or on an Animal/non Animal (Rh2 and Rh3, male and female both aged 4) categorisation task.

Two groups of 10 human subjects were also tested, one on each categorisation task, (mean age 37 in the food task and 35 in the animal task, with seven subjects in common). Subjects were mainly students or members of the laboratory staff; they all gave written consent to do the task and reported normal colour vision.

Task and set-up

The subjects (monkeys or humans) sat about 30-35 cm away from a tactile screen in the centre of which pictures were flashed for only 32 ms on a black background with a 1.5-3 s random intertrial between successive images. The subjects placed one hand on a capacitive tactile key located below the screen at waist level to start stimulus presentation. They had 1s to quickly release the button and touch the screen when they detected a target (animal or food) in the flashed image, otherwise they had to keep their hand on the button. This is a very demanding task : humans and monkeys have to make a succession of rapid decisions on the basis of brief stimulus presentations that prevent any foveating eye movements. Correct - go or no-go – decisions were rewarded by a beep noise. In addition, monkeys were given a drop of fruit juice. Incorrect decisions were followed by a 3-4 s display of the incorrectly classified stimuli delaying the next trial and the next possibility of reward and allowing time for ocular exploration.

The monkeys worked daily for as long as they wanted (1-3 hours), five days a week. At the end of each testing session and during week-ends ***ad libitum*** water was provided. Adequate measures were taken to minimise any discomfort to the animals. They were restrained in a primate chair (Crist Instruments, GA USA) during testing and lived in a cage (European normalisation) in between the sessions.

Stimuli

All the pictures were natural scenes taken from a large commercial CD database (Corel). Some additional photographs (roughly 10%) were added for the food task in order to increase stimulus variety and to allow further controls to

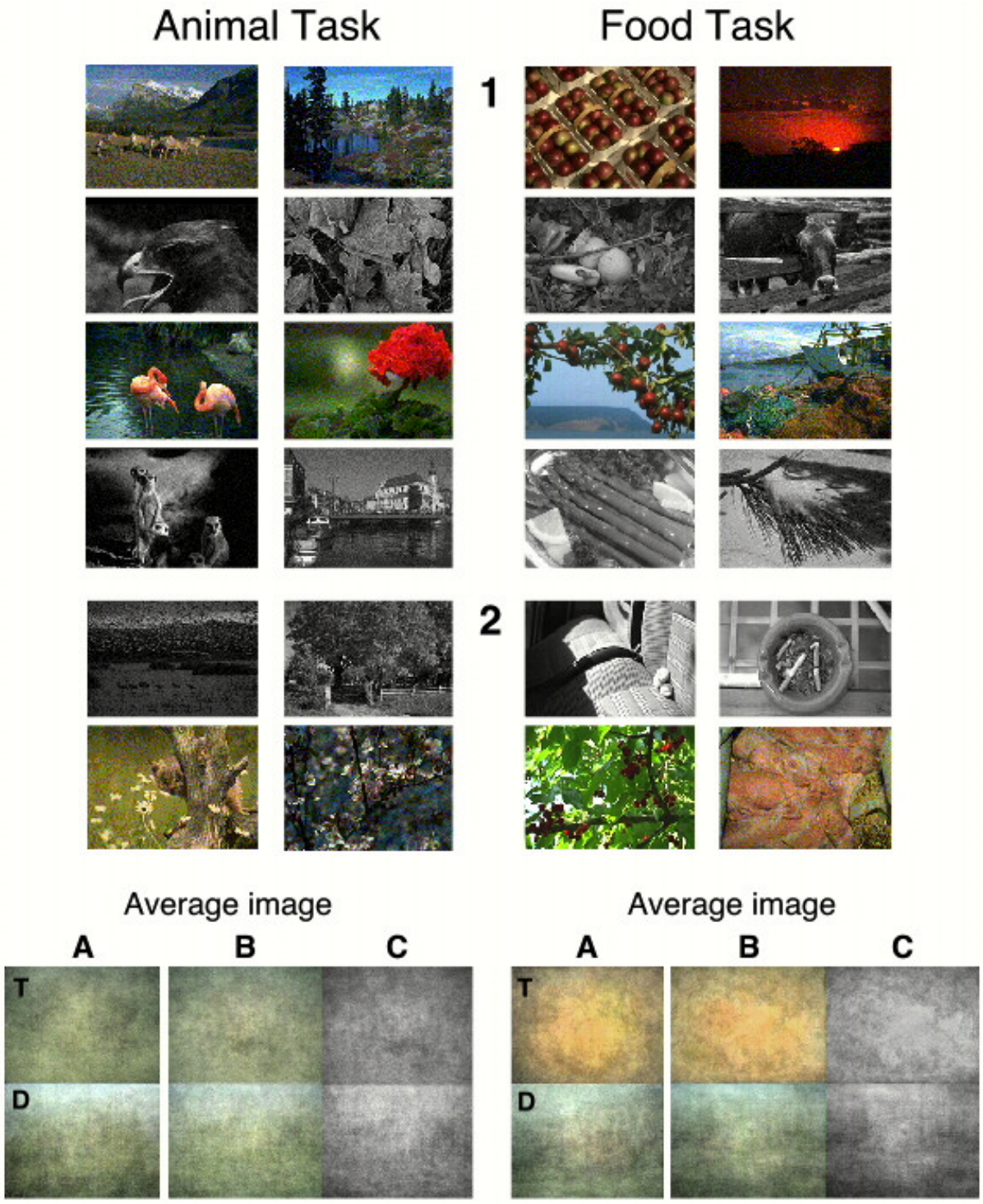


Fig. 1. Examples of the stimuli used in the animal and in the food categorisation tasks. Top : For each of the tasks, 6 targets are presented in the left column and six distractors in the right column. They are illustrated in colour or in BW as they were presented in the tasks. In all 4 columns, the first 4 photographs (1) were accurately classified by both the monkeys and the 10 human subjects, the bottom two (2) induced incorrect responses both in monkeys and in some of the human subjects.Bottom : Average images for targets (T, top row) and distractors (D, bottom row) have been calculated for each task : animal task (on the left) and food task (on the right) and for all horizontal images of each group of 200 photos. These averages allowed to search for a colour bias between sets of images and, within a given set, between targets and distractors. (A) average colour images for the set of photos presented in colour; (B) average colour pictures and (C) average BW images of the set of pictures that was then randomly chosen to be converted in BW images. Average images have been renormalised to reinforce differences that otherwise would remain undetected. For a given task all averaged images were considered together in the renormalisation process for rigorous comparison. Note that a color bias can clearly be seen in between targets and distractors especially in the food task so that the question of the role of colour cues addressed in the present study is definitely pertinent. Note also that there is no obvious colour bias between the set of images that were presented for categorisation either in colour (A) or in BW (B, C).

be performed. Targets and distractors were equiprobable and included both close-ups and general views. Food targets included photographs of fruit, vegetables, salads, cakes, biscuits, sweets... presented against natural backgrounds. Animal targets included fish, birds, mammals and reptiles also presented in their natural environments. Distractors included some of the target category of the other task, landscapes, trees, flowers, objects, monuments, cars... On target trials, the subjects had no a priori knowledge concerning the position, the size or the number of targets in the picture. Moreover, both tasks included targets that were only partly visible, or partially masked in the scene. The photographs were so varied (Fig. 1) that contextual help can effectively be ruled out.
Images (192 x 128 pixels, corresponding to an angular size of about 25°/15°) were mostly horizontal photographs (73%). They were flashed for 2 frames at a refresh rate of 62 Hz (non-interlaced), corresponding to a presentation time of 32 ms, using a programmable graphics board (VSG 2, Cambridge Research Systems) mounted in a PC-compatible computer. Colour and BW images were both converted from 24-bit colour photographs to 8-bit indexed pixels.

Evaluation of performance and Data analysis
Performance was evaluated both in terms of accuracy and speed. A go response was scored (whether correct on target trials or incorrect on distractor trials) when the subject released the key and touched the screen with the same hand in less than 1 s. The reaction time (RT) - delay between stimuli onset and button release - was recorded for all go responses. A no-go response was scored when the subject had kept pressing the key for over 1 s. Any other response (i.e. releasing the key without touching the screen) was considered as an incorrect response both on target and distractor trials. Accuracy differences were assessed using a standard two-tailed Chi^2 test on correct and incorrect responses; RT distributions were compared using two-tailed Mann-Whitney U test.

Procedure before testing monkeys with new achromatic stimuli
The training steps and first results have been reported previously (Fabre-Thorpe et al., 1998). At the end of the training period (4-6 months), monkeys were able to categorise accurately new stimuli (90,5% in the food task and 87% in the animal task) with very short RTs (mean values: 356 ms and 251 ms respectively for the food and the animal task).

To study the effects induced by the removal of colour cues, monkeys went through a number of new training and control steps. They were first trained to perform the categorisation task on a set of 200 images (taken from the set of 480-650 of images that they had already experienced) of which half were presented in BW (50 targets and 50 distractors). To prevent monkeys (and humans) from developing a new categorisation strategy in which colour cues would be ignored because stimuli were always -thus predictably- presented in BW, colour and BW images were mixed at random. Under these conditions, the monkeys scored as well as before on colour images and took 1-2 weeks to stabilise accuracy and speed with BW ones. Then, the role of colour cues on the processing of familiar images was studied, during 2-5 successive days, with a second set of 200 familiar images with all conditions equally balanced: (1) half were targets, half distractors, (2) all were seen in colour and in BW and finally (3) half were first presented in BW then in colour the second half being presented first in colour. Only the responses given by the monkey to the first BW and the first coloured presentation of each image were taken into consideration and compared. Results were replicated in the final experiment that is described in detail in the present study. They showed that monkeys can categorise BW images with very little impairment.

Testing performance with new achromatic stimuli
In the final testing phase, monkeys and humans were tested with 400 images that they had never seen before. For each task (Food vs. non-Food and Animal vs. non-Animal), 400 novel images (200 targets and 200 distractors) were chosen from the large database used previously (Fig. 1). For each task, the 400 images chosen were randomly divided into 2 sets of 200 images (100 targets, 100 distractors). For each set, the average horizontal colour image was computed separately for targets and distractors (Fig. 1). These average images illustrate the colour bias that exists between target and distractor images –especially in the food task- as well as the similarity of the two image sets. One set was then randomly chosen to be converted in BW.

Testing monkeys
Monkeys had to categorise daily 20 new images (10 in colour and 10 in BW of which half were distractors and half were targets) mixed with familiar colour and BW images that were randomly selected every day. With familiar

images monkeys are rewarded in over 90% of the trials. Mixing new images with familiar ones (i) allowed performance to be compared for new and familiar stimuli, (ii) avoided the impact of response errors on the processing of subsequent images and thus (iii) ensured the stability of the emotional state of the monkey when faced with a new stimulus. As we only consider the response given by the monkey to the first presentation of a new image, this stability is a necessary condition to obtain consistent results. For the same reason, monkeys were tested with 40-100 familiar images before being presented with new images and 2 new images were never shown in immediate succession.

Testing humans

Subjects were all familiar with the task since they already volunteered for other studies using this rapid categorisation task (Fabre-Thorpe et al., 1998; Thorpe et al., 1996). They were tested using the same set-up and the same 400 stimuli that were presented in 4 blocks of 100 new stimuli. As with the monkeys, they were given feed-back about the accuracy of their responses: a beep noise for a correct response and a 3 s re-appearance of incorrectly classified images.

RESULTS

Only mild impairments were observed for both monkeys and humans when colour cues were removed, although colour appeared as a more important feature in the food task. For humans, the accuracy impairment varied from one subject to another indicating that individual subjects may rely differently on colour cues to reach their decision.

Accuracy in monkeys

New images: For the 3 monkeys the average accuracy with previously unseen photographs was 87.2% in colour and 87.3% in BW. Whatever the task (food task: 85.5% correct in colour vs. 87.5% in BW, two-tailed Chi^2=0.343, d.f.=1, p=0.66; animal task: 88% correct in colour vs. 87.2% in BW, two-tailed Chi^2=0.104, d.f.=1, p=0.83) there was no significant difference (Fig. 2). The same result was observed for each of the 2 monkeys (Rh2 and Rh3) tested in the animal task (Table I).

Familiar images: The 3 monkeys had performed a large number of trials with familiar images (1500-1750 trials for each condition). They averaged 95.2% correct in colour and 92.8% in BW (Fig. 2). This mild 2–3% accuracy impairment for BW images is statistically significant (table I). When accuracy was

Accuracy

		Colour			BW	
Food Task		**T / D**	**Total**	**Chi2**	**Total**	**T / D**
	Familiar					
Rh1	N trials	844 / 851	1 695		1 506	735 / 771
	N Cor trials	81 6 / 800	1 61 6		1 386	676 / 71 0
	%Cor	96.7 / 94	95.3	=.0004	92	92 / 92.1
	New					
Rh1	N trials	1 00 / 1 00	200		200	1 00 / 1 00
	%Cor	83 / 88	85.5	ns	87.5	91 / 84
Animal Task						
	Familiar					
Rh2	N trials	822 / 820	1 642		1 646	822 / 824
	N Cor trials	81 2 / 763	1 575		1 561	800 / 761
	%Cor	98.8 / 93	95.9	ns	94.8	97.3 / 92.4
Rh3	N trials	91 5 / 856	1 771		1 747	883 / 864
	N Cor trials	893 / 780	1 673		1 604	862 /742
	%Cor	97.6 / 91 ,1	94.5	<.005	91 .8	97.6 / 85.9
	New					
Rh2	N trials	1 00 / 1 00	200		200	1 00 / 1 00
	%Cor	94 / 87	90.5	ns	91 .5	91 / 92
Rh3	N trials	1 00 / 1 00	200		200	1 00 / 1 00
	%Cor	83 / 88	85.5	ns	83	86 / 80

Reaction Time (ms)

		Colour			BW	
Food Task		**mean**	**med**	**U test**	**med**	**mean**
	Familiar					
Rh1	Cor Go-trials	31 0	297	<.0001	309	322
	InC Go-trials	342	320		31 8	351
	New					
Rh1	Cor Go-trials	31 2	297	ns	306	324
	InC Go-trials	346	31 6		325	356
Animal Task						
	Familiar					
Rh2	Cor Go-trials	262	258	ns	258	264
	InC Go-trials	285	276		278	283
Rh3	Cor Go-trials	247	240	ns	241	248
	InC Go-trials	271	256		266	266
	New					
Rh2	Cor Go-trials	270	266	ns	269	281
	InC Go-trials	31 4	292		300	350
Rh3	Cor Go-trials	268	252	ns	242	262
	InC Go-trials	329	293		308	356

Table I. Monkeys' performance with new and familiar images in both colour and black and white (BW) conditions. On the left part of the table, the accuracy performance is shown for each of the two tasks (Food Task and Animal Task) and each of the three monkeys (Rh1, Rh2 and Rh3). For New and Familiar targets and distractors (T/D) the total number of trials (N trials), and the percentage of correct trials (% Cor) are indicated. For familiar images the total number of correct trials (N Cor trials) is also given. The same indication are given for targets and distractors pooled together (Total) with, in the central column the statistical result of a Chi^2 test between the colour and the BW conditions. On the right part of the table, the reaction time (expressed in ms) is also given for each of the two tasks (Food Task and Animal Task) and for each of the three rhesus monkeys (Rh1, Rh2 and Rh3). For new and familiar images, the mean and median (med) reaction times are indicated for all correct go-responses (Cor Go-trials) and all incorrect go-responses (InC Go-trials). The RT distributions of correct go-responses obtained in the two conditions (Colour and BW) have been compared using a two-tailed Mann-Whitney U test; results are shown in the central column (U test).

considered separately for targets and distractors, all monkeys showed a significant bias towards go responses with colour stimuli (two-tailed Chi^2, d.f.=1, p<.01 for Rh1, p<.0001 for Rh2 and Rh3). This bias persisted with BW photographs in the animal task (Rh2, Rh3) but disappeared for Rh1 that was tested on the food categorisation. The main accuracy effect associated with the absence of colour cues is seen in the detection of targets in the food task.

New vs. familiar images : Training improves accuracy in both tasks by about 5-10% in the chromatic condition and 3-8% in the achromatic condition.

Monkeys are extremely good and equally efficient at categorising new stimuli whether in colour or not. Colour is clearly not needed to explain the monkeys performance in this rapid categorisation task. Accuracy being similar when both BW and colour images are seen for the first time, the slight advantage for coloured stimuli that appears with familiar images might be due to the monkeys larger experience with chromatic images than with their BW counterparts.

Accuracy in human subjects

For the two groups of humans, the global average categorisation accuracy (Fig. 2) with previously unseen photographs was about 2% higher with colour photographs (93.2% in colour and 91% in BW). This tendency was not significant in the food task (90.7% correct in colour vs. 89% in BW, two-tailed Chi^2=2.8, d.f.=1, p=0.0942) but reached significance in the animal task (95.7% correct for colour images and 93% correct for BW image, two-tailed Chi^2=13.6, d.f.=1, p=0.0002). When the rate of correct responses was analysed separately for targets and distractors a difference appeared between the two tasks. In the animal task, the global 2.5-3% accuracy decrease observed with BW stimuli was seen for both targets and distractors. On the other hand, in the food task, the ability to detect targets dropped by 5.9% with BW stimuli (two-

FOOD TASK

	Colour			BW	
	Accuracy				
	T/D	Total	Chi 2	Total	T / D
Fl *	93 / 90	91 ,5	ns	90	90 / 90
F2*	93 / 82	87,5	ns	89,5	89 / 90
F3*	90 / 92	91	ns	86	76 / 96
F4*	91 / 89	90	ns	89,5	88 / 91
F5	93 / 84	88,5	ns	87,5	86 / 89
Ml *	97 / 83	90	ns	88	94 / 82
M2*	95 / 89	92	ns	88,5	89 / 88
M3*	96 / 92	94	ns	91 ,5	89 / 94
M4	93 / 93	93	ns	93	92 / 94
M5	93 / 85	89	ns	86,5	82 / 91
Total	93,4 / 87,9	90,7	ns	89	87,5 / 90,5
Chi 2	<.0001				<.04
	RT on correct go-responses				
	mean	med	U test	med	mean
Fl *	461	446	<.05	462	478
F2*	376	360	=.0004	396	41 2
F3*	506	493	<.05	508	521
F4*	390	368	<.06	384	403
F5	406	394	<.008	41 2	427
Ml *	41 0	397	=.09	401	428
M2*	449	438	ns	438	453
M3*	467	449	ns	474	481
M4	468	457	ns	465	479
M5	438	421	<.06	451	463
Total	437	427	<.0001	439	453

ANIMAL TASK

	Colour			BW	
	Accuracy				
	T/D	Total	Chi 2	Total	T / D
Fl *	98 / 96	97	ns	94,5	96 / 93
F2*	95 / 96	95,5	ns	94	94 / 94
F3*	94 / 94	94	<.06	88,5	90 / 87
F4*	96 / 94	95	ns	92	91 / 93
F6	98 / 99	98,5	<.02	94	94 / 94
Ml *	99 / 93	96	ns	96	97 / 95
M2*	90 / 97	93,5	ns	90,5	89 / 92
M3*	98 / 96	97	ns	94,5	94 / 95
M6	97 / 96	96,5	ns	93	92 / 94
M7	97 / 91	94	ns	92,5	95 / 90
Total	96,2 / 95,2	95,7	=.0002	93	93,2 / 92,7
Chi 2	ns				ns
	RT on correct go-responses				
	mean	med	U test	med	mean
Fl *	41 9	411	ns	408	41 9
F2*	345	330	ns	333	361
F3*	524	520	ns	503	51 3
F4*	385	376	ns	374	388
F6	469	455	ns	470	477
Ml *	387	381	ns	380	387
M2*	445	444	ns	440	445
M3*	442	429	ns	443	457
M6	427	41 7	ns	411	430
M7	363	346	ns	348	368
Total	420	41 2	ns	41 5	424

Table II. Individual human performance in the food task (left) and in the animal task (right). Female (F) and male (M) subjects that participated in each task are listed on the left; the asterisk indicates that the same subject participated in both experiments. For each of the subjects and for both colour and black and white (BW) images, the accuracy is given as the percentage of correct responses; the mean and the median (med) RT are given in milliseconds. In both tasks, human subjects had a tendency to categorise coloured images better than BW ones. Statistical comparison using a Chi^2 is shown in the central columns. However it only reached significance in the animal task (two-tailed Chi^2=13.6, d.f.=1, p=0.0002) and mainly as a result of two of the subjects (F3 and F6). The bias towards go responses is not significant in both colour and BW conditions for the animal task. It is highly significant in the food task when stimuli are in colour (p<.0001) and tends to disappear in the BW condition (p<.04). For RTs, in the animal task, the difference between conditions never reached significance at the global or at the individual level. In the food task, the average RT difference between the two conditions was highly significant (two-tailed Mann Whitney U=359702, p<0.0001) and reached significance for most of the subjects.

tailed Chi^2=20.1, d.f.=1, p<.0001) whereas the rate of correct no-go responses increased by 2.6% with BW distractors (almost reaching significance two-tailed Chi^2=3.5, d.f.=1, p<.07) which partly compensate for the accuracy drop with targets. This detection impairment for food targets ranged from 1% to 14% (Table II, 1% for subject M4, 14% for subject F3) whereas the impairment range was much smaller with animal targets (1-5%). Note that the same impairment in detecting food targets was seen for familiar images in monkeys.

As a global result, this study shows that, in rapid categorisation tasks, removing colour cues from the stimuli has on average, very little effect on human accuracy. It also shows that human subjects rely more heavily on colour to detect food targets relatively to animal targets. However, a large amount of variability between subjects was seen when considering individual performances. Whereas most subjects showed an impaired accuracy when categorising BW photographs, for some of them the global accuracy was identical in both conditions (Table II, subjects F2 and M4 in the food task; M1 in the animal task). For the subjects tested, the strength of the accuracy impairment induced by the removing of colour cues appeared correlated with RTs. The fastest subjects categorised equally well chromatic and achromatic stimuli whereas the accuracy advantage observed with chromatic stimuli increased progressively for subjects with longer RTs (correlation : R=0.81 in the animal task, R=0.72 in the food task). Colour could be a relevant feature in the visual processing leading to decision only for subjects responding with relatively long reaction times.

Speed of performance in monkeys

New images: In both tasks (Table I for individual results), RTs for correct go responses did not depend on whether the images were shown in colour or in BW (Fig. 2). In the animal task, mean RT was 269 ms (median: 259 ms) in colour and 271 ms (median: 261 ms) in BW (two-tailed Mann Whitney U=15269, p=0.68); in the food task mean RT was 312 ms (median: 297 ms) in colour and 324 ms (median: 306 ms) in BW (two-tailed Mann Whitney U=3651, p=0.31). The RT distributions for new BW and colour photos are illustrated for both tasks (Fig. 3). In the animal task they are identical for both BW and colour conditions, and the absence of colour cues had no consequence on the earliest correct go responses that are seen with latencies as short as 200 ms in both cases. Note that these responses cannot be simply considered as just random anticipations because, as targets and distractors are equiprobable, correct and incorrect anticipated go-responses should be equally distributed. From 200 ms on, correct go-

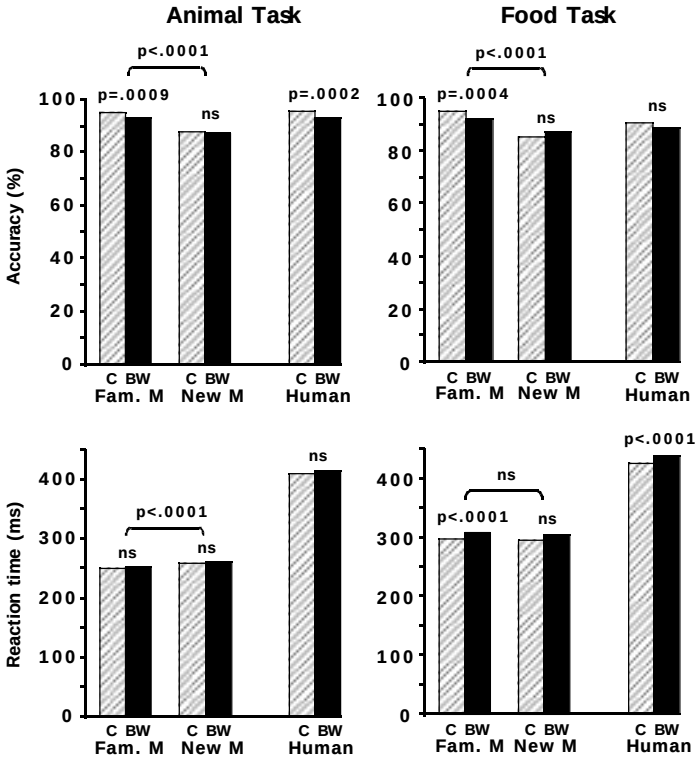


Fig. 2. Monkeys' and humans' global categorisation performance. Familiar images for monkeys (Fam. M); new images for monkeys (New M), new images for the group of 10 Humans (Human), images presented in colour (C; black bars) and in grey levels (BW; hatched bars). Top : accuracy expressed as the percentage of correct responses, for the Animal Task (left) and the Food Task (right). Note that monkeys categorise equally well new colour and BW images and that a very mild advantage is observed with colour pictures for monkeys tested with familiar images and for humans (the statistical significance is given by a Chi^2). Training leads to improved accuracy as shown by the higher scores observed for monkeys tested with familiar images; this is true in both colour and BW and with the same statistical significance. Bottom : The speed of response is illustrated for all conditions by the median reaction time (expressed in ms) for correct go-responses. Note the absence of an effect in the animal task and the 10-15 ms RT increase when food objects have to be detected in BW. This increase is significant (see text). Note that the effect of training on performance speed is only visible in the animal task.

responses significantly outnumbered incorrect go responses both in colour (12 correct vs. 1 incorrect go responses in the bin width 200-220; $p<0.02$) and in BW (12 correct vs. no incorrect go responses in the bin width 200-220; $p<0.005$). Thus, visual processing must have been completed and cannot exceed 100-120 ms as the 200 ms behavioural RT also includes time for decision and a large motor component. In the food task, the small delay seen for the earliest latencies is not very conclusive since only one monkey was tested (a total of 88 correct go responses in colour and 91 in BW) and the effect was not replicated with the larger sample of familiar photographs.

Familiar images: Concerning the performance speed, results with chromatic and achromatic familiar photographs confirmed the observations made with new images (Fig 2). No difference was seen in the animal task whereas, in the food task, RTs for correct go responses towards BW images were about 10 ms longer (median: 297 vs. 309 ms, mean: 310 vs. 322 ms) with the large sample of trials with familiar images (over 1500 trials in either condition), this shift towards longer latencies is significant (two-tailed Mann Whitney $U=242504$, $p<0.0001$). The RT distributions obtained in the food task (Fig. 3) shows that: (1) the earliest responses statistically biased towards correct go responses are produced at the same latency (220-230 ms) in both conditions, (2) the peak of the RT distribution for correct go responses is sharper for chromatic stimuli (variance 3321 in colour and 4326 in BW) with a clear mode in the bin width 270-280 ms. The effect associated with colour cues is mainly observed in the range 250-300 ms within which colour appears to facilitate the detection of food objects as more colour targets (an additional 9% relatively to the BW condition) are detected. On the other hand, in the achromatic condition, the RT distribution for correct go responses has no clear mode and extends towards long latencies responses with a greater percentage of go responses triggered after 400 ms in BW (6.5% in colour vs. 11% in BW).

New vs. familiar images : Whereas training induced a clear improvement in performance accuracy, there is little effect (if any) on the speed at which such a task is performed. In the food task, familiar and previously unseen photographs are categorised at exactly the same speed. The only effect was seen in the animal task in which a small RT decrease (10 ms) was observed with familiar images and in both monkeys. The fact that extensive training with photographs fails to speed up the neural processes underlying performance suggests that the speed at which monkeys categorise new images is already near to optimal.

In summary, the results obtained clearly demonstrate that for previously unseen images, the monkeys' abilities to categorise colour and BW images are almost identical for both accuracy and speed of response. A mild facilitation in the detection of food targets can be seen when colour cues are available, but this facilitation appears from 250 ms on. The experiment also shows that familiar images are categorised with higher accuracy than new ones but tend to be processed at about the same speed.

Speed of performance in human subjects

Speed of response: Data obtained with human subjects also showed that animal detection was not speeded up when colour cues were available whereas a mild effect could be seen in the detection of food targets. In the animal task (Fig. 2) the comparison of the overall RT distributions for correct go responses with either coloured or BW targets showed no statistical difference (RT in colour: mean=420 ms, median=412 ms; RT in BW: mean=424 ms, median=415 ms; Mann Whitney $U=441635$, $p=0.58$). The same result was obtained when comparing RT distributions individually for each subject (table II). In the food task, the results were clearly different. A statistically significant RT increase was seen with BW images (RT in colour: mean=437 ms, median=427 ms; RT in BW: mean=453 ms, median=439 ms; Mann Whitney $U=359702$, $p<0.0001$) although it was – as for the monkey - very small (about 15 ms). This result was confirmed at the individual level: the RT increase was seen in all subjects and reached significance for seven of them.

Humans are extremely efficient and can reach high accuracy scores with short RTs in both tasks. Compared to the animal task, the food task appears more difficult: the global accuracy is lower and the RTs are longer. Moreover, whereas the earliest responses start around 280 ms in the animal categorisation, they are delayed by at least 30 to 40 ms in the food categorisation (Fig. 3). Removing colour did not affect these earliest behavioural latencies. On the other hand, as it was already the case for monkeys in the food task, a greater percentage of go responses are triggered late (after 500 ms) with achromatic stimuli (17.5% in colour vs. 23,3% in BW) which is not the case in the animal task. In fact, the effect of removing

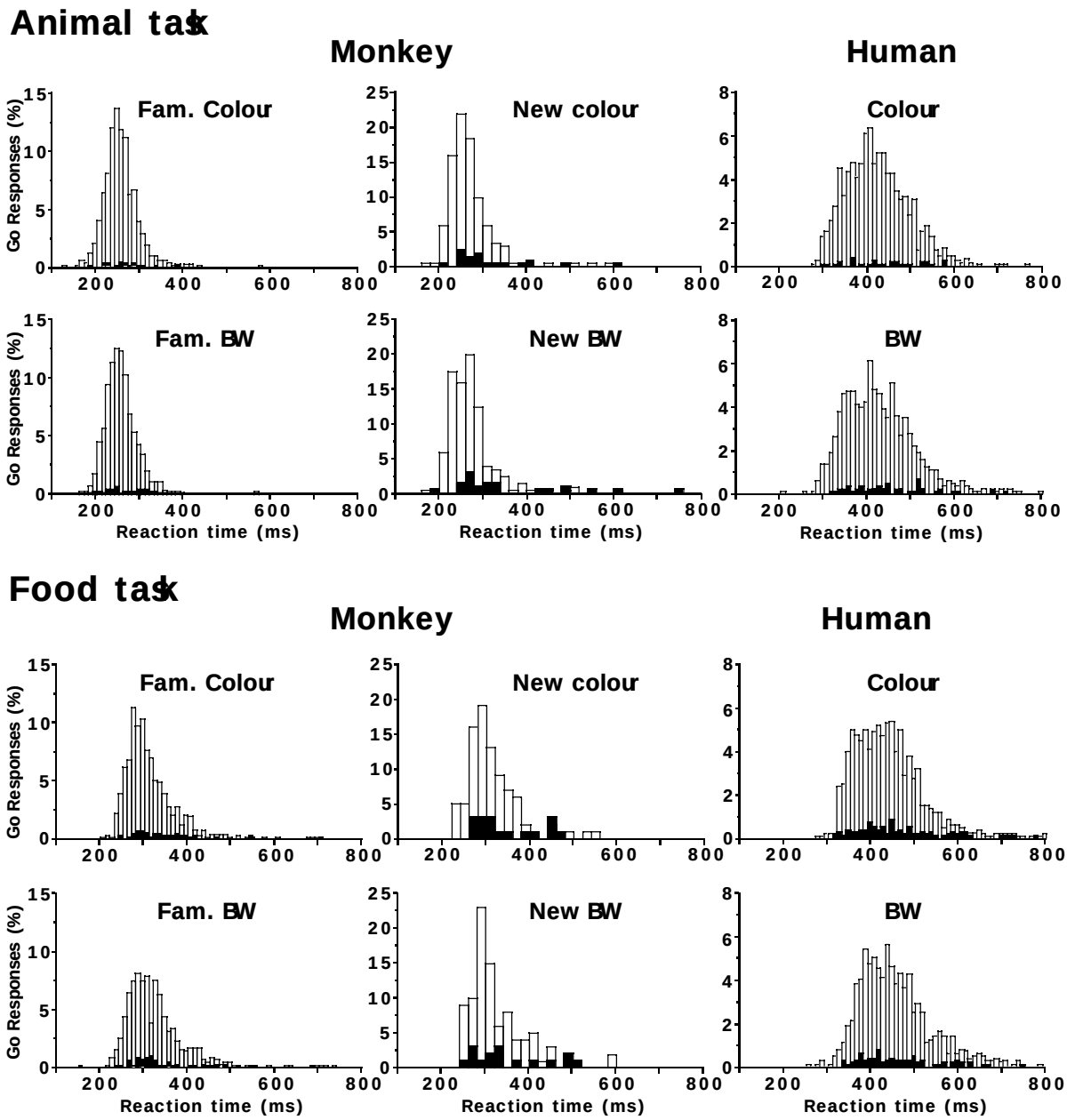


Fig. 3. Reaction time distribution of correct go responses and false positives for humans and monkeys. For each task : animal task (top) and food task (bottom), the RT distributions are shown for chromatic (top row) and achromatic (bottom row) image presentations. In each graph, the empty histogram corresponds to the correct go-responses whereas incorrect go-responses (false positives) are shown in black. Left column: monkey RT distributions (10 ms bin width) for go responses towards familiar (Fam.) stimuli. Central column : monkey RT distributions (20 ms bin width) for go responses with new photographs. Right column, RT distributions (10 ms bin width) for the group of 10 human subjects tested with the same previously unseen images. Reaction time is expressed in ms and in each bin width, go-responses are expressed as a percentage of all (correct and incorrect) go-responses.

colour cues was very mild, most stimuli were accurately and quickly identified in BW with, at the most, a 2% decrease in global accuracy and a 15 ms increase in RT. With the very varied total of 800 natural scenes proposed in these tasks, colour does not seem essential to allow fast and accurate categorisation of natural images. When available, colour cannot speed up the visual processes underlying the earliest, ultra-rapid, behavioural responses.

Comparison between humans and monkeys

When comparing the speed at which humans and monkeys perform the task, monkeys are

much faster than humans; on average this advantage reaches roughly 130 ms in both tasks. Monkeys appear to behave like the fastest humans, combining short RTs and nearly no advantage for coloured stimuli. Note that (1) like humans, monkeys are faster to perform the animal categorisation task, (2) like humans, monkeys' go responses are delayed only in the food categorisation of BW stimuli, (3) this delay reaches about 10 ms in monkeys, a value compatible with the 15 ms delay observed in humans, (4) as for humans, colour appears more important as a feature for detecting food than animals, (5) monkeys and humans tend to make incorrect decisions in response to the same stimuli both in chromatic and achromatic conditions (Fig. 1).

GENERAL DISCUSSION

The first results replicate the data previously reported (Fabre-Thorpe et al., 1998; Thorpe et al., 1996): both monkeys and humans are fast and accurate at categorising natural images that they have never seen before, even without contextual help and without using eye movements. However studying the processing of 800 previously unseen natural scenes by humans and monkeys, and an additionnal 800 familiar natural scenes by monkeys, the main finding is that this sort of rapid visual processing of natural images is only very mildly affected by the removal of colour information. Humans' impairment without colour cues varies from one individual to another (some of them bein unaffected) and monkeys are fast and rely very little on colour.

The use of colour information in between tasks and species

Colour appears as a more relevant feature in the categorisation of food objects. The use of colour features to determine whether an item belongs to a category might depend on the pertinence of colour in identifying objects from that category. Although very few objects could be recognised on the b asis of their colour only, colour could be more or less "diagnostic" in the recognition of certain categories of object (Biederman & Ju, 1988; Oliva & Schyns, 2000; Tanaka & Presnell, 1999). Monkeys were shown to encode pictures of fruits mainly along two dimensions : the type of fruit (apples or grapes) and their colour (red or yellow), ignoring other dimensions like their size or number (Sands, Lincoln, & Wright, 1982). For food, the colour is less arbitrary – i.e. varying in a smaller range - than in other categories like animals for example. This is shown by the yellow to orange predominance in the average food images that were computed from our sets of photographs (see Fig. 1). Colour could also help decision making in the case of ambiguous photographs such as close-ups of a rose vs. a green salad. Using top-down influences, the visual system could be "pre-set" to detect an object with special colour attributes; this could explain why RT are globally 10-15 ms shorter with chromatic stimuli. On the other hand, for animal categorisation, colour is clearly not essential, perhaps because it has no "diagnostic" value. Although it can be argued that cage-reared monkeys may not rely on animal colour to generalise their training, the fact that the same results were found in humans shows that fast detection of animal does not depend on colour cues.

Colour could also be used to help segmentation of the target-objects from the background. In that case, the contribution of colour may depend on how well objects can be segregated from their background and it could be argued that fast categorisation responses may only be seen when colour is not necessary for target segregation. In fact, the natural scenes used in our studies are very complex and we recently showed that ultra-rapid visual processing is not restricted to "easy to process" animal-targets (Fabre-Thorpe et al., 2000). It therefore appears that, at least for the animal task, short latency behavioural responses can be observed with the vast majority of targets. In the case of the food task, the effect of colour is seen around the mean RT when more chromatic targets are detected. Thus colour may not be used in the earliest processing stages but may be involved in later steps that can help improve object detection. In primates, trichromacy is thought to have evolved for segregating flowers and fruit from background foliage (Mollon, 1989; Regan et al., 1998) and probably plays a vital role in every day tasks such as choosing ripe fruit. In our data, the small 10-15 ms RT increase observed when colour is removed is mainly due to a small subset of targets that take abnormally long to detect in greyscale. In absence of colour, it is possible that more detailed analysis would be necessary for target detection. Thus the additional delay might be explained if, as suggested by (Smid, Jakob, & Heinze, 1997), it is faster to combine the coarse information about an object shape with its colour than to analyse its detailed local shape features. However, even if colour is used in the detection of some targets, and if it can improve pattern recognition in some cases (Syrkin & Gur, 1997), it is clearly not the

most crucial aspect of the object used by monkeys and humans to perform the rapid categorisations studied here.

Neural correlates in IT

It seems likely that the processing leading to food or animal categorisation involves all the processing stages along the ventral visual pathway known to play a crucial role in object recognition (Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). Thus, the data obtained in the present study support the idea, previously developed (Fabre-Thorpe et al., 1998; Thorpe et al., 1996; Thorpe & Imbert, 1989), that visual information processing in this task is mainly feed-forward. For the monkeys' earliest behavioural responses, visual processing must be restricted to roughly 100 ms. This finding is in agreement with neuronal responses in IT that typically have onset latencies of 80-100 ms (Oram & Perrett, 1992; Perrett, Rolls, & Caan, 1982) and with the short latency (100 ms) of the differential IT responses between target and non-target stimuli reported in categorisation tasks (Vogels, 1999b). The feed-forward aspect of processing is also supported by the similar latencies of the responses triggered towards new and familiar stimuli. Extensive training could have been expected to allow the bypass of some processing loops resulting in a RT decrease. This floor effect on the processing speed of natural scenes is also seen in humans (Fabre-Thorpe et al., 2000).

The minor role played by colour cues reported in the present study is in agreement with the characteristics of IT neuronal responses. IT neurones respond to faces, and specific objects (Booth & Rolls, 1998; Logothetis, Pauls, & Poggio, 1995; Perrett et al., 1982). Some can selectively respond to several objects that belong to the same category (Vogels, 1999b). Indeed, in monkeys performing a categorisation task with trees as targets, a quarter of IT cells responded differentially to tree vs. non tree stimuli. These responses are sensitive to image scrambling, but largely invariant to stimulus transformations such as changes in position or size. They are observed whether the objects are presented as coloured pictures, line-drawings, silhouettes or illusory contours (Chadaide, Kovàcs, Köteles, & Benedek, 1999), with simplified versions of the objects or with a combination of their features (Kobatake & Tanaka, 1994; Tanaka, 1997). IT cells are generally reported to be selective to shape irrespectively of colour, and only a very small subset of neurones needs a combination of shape and colour information to respond with maximal amplitude (Chadaide et al., 1999; Tanaka, Saito, Fukada, & Moriya, 1991). This view is also supported by two recent studies. In the first one (Booth & Rolls, 1998), macaques had been given a number of new objects to manipulate in their cages. Cellular recording in IT showed a small subset of totally "view-invariant" neurones, suggesting that objects rather than the visual features were coded. Moreover, most of them exhibited similar responses for colour or greyscale object images. The second study (Vogels, 1999b) used a tree vs. non tree categorisation in macaques and although the colour content of the image could affect the average response of IT cells, chromatic and achromatic stimulus presentations often elicited similar neuronal responses. Thus processing in IT cells is consistent with the mild effect observed in our task when colour cues are removed.

Fast processing using the magnocellular pathway ?

The hypothesis that the use of colour features is time consuming is supported by at least three different results reported here. (1) humans with the largest accuracy advantage for colour stimuli were those which had the longest mean reaction times, (2) both monkeys and humans are slower in performing the food task in which colour is a more relevant feature, (3) the earliest behavioural responses do not depend on colour cues. Visual information can reach cortical area V1 using either the magnocellular (M) system or the parvocellular (P) system. Traditionally, the M system has been associated with the extraction of structure from motion whereas the P system is thought to be concerned with the fine analysis of static images. A possible explanation for the late use of colour is based on the FACADE model (Bradski & Grossberg, 1995; Grossberg, 1994) and the sequential use of parvocellular information : boundaries would first be formed and then control the surface filling-in of brightness and colour information. Boundaries could directly activate object recognition processes but the use of colour would take longer. Alternatively, we would like to suggest that in our task, the absence of a clear accuracy advantage when colour cues are available could be related to differences in the temporal dynamics of processing in the M and P pathways of the visual system. Chromatic information in the parvocellular stream reaches visual cortex roughly 20 ms after the magnocellular inputs that mainly transmit motion and luminance based information (Nowak & Bullier, 1997; Nowak et al., 1995). This delay

could be even longer if, as suggested recently, colour vision depends upon the koniocellular pathway (Calkins & Sterling, 1999). As has been argued elsewhere, the sort of rapid visual categorisation performed in our tasks could depend on the unidirectional processing of the first 10-20 ms of activity in each cortical area, the analysis would be based on coarse and near colour-blind magnocellular information. It follows that colour would only be important for images that could not be accurately categorised on the basis of such early information.

Coarse to fine processing has already been proposed by a number of authors (Carpenter & Grossberg, 1987; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994), and in a recent study (Sugase, Yamane, Ueno, & Kawano, 1999) the discharge of IT neurones in response to faces has been shown to have a fast phasic component related to the presentation of a face regardless of its other characteristics and a second tonic component developing with a 40-50 ms delay which is linked to finer information concerning for example, the owner of the face (human or non human) or the facial expression. The coarse magnocellular information might thus be sufficient to get access to a global shape representation that might be adequate in most cases in our task for the ultra-rapid categorisation of natural scenes – i.e. the fast behavioural responses produced in our task both by monkeys and humans. Such fast process could be used as a header to improve further processing of colour and fine details.

Acknowledgements :
This work was supported by the CNRS, by the Cognisciences Program, and by the Midi-Pyrénées Region. Authorization for experiments with humans (CCPPRB N° 9614003).

References

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. ***Cognitive psychology, 20***, 38-64.

Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. ***Cerebral Cortex, 8***, 510-523.

Bradski, G., & Grossberg, S. (1995). Fast-Learning VIEWNET Architectures for Recognizing Three-dimentional Objects from Multiple Two-dimensional Views. ***Neural Networks, 8***, 1053-1080.

Calkins, D. J., & Sterling, P. (1999). Evidence that circuits for spatial and color vision segregate at the first retinal synapse. ***Neuron, 24***, 313-321.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. ***Computer Vision, Graphic and Image Processing, 37***, 54-115.

Chadaide, Z., Kovàcs, G., Köteles, K., & Benedek, G. (1999). Selectivity of macaque inferior temporal neurons for shapes with different surface attributes. ***Perception, 28 (suppl.)***, 97.

D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (***Cebus apella)***. ***Journal of Experimental Psychology: Animal Behavior Processes, 14***(1), 43-55.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2000). A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes. ***Journal of Cognitive Neuroscience***, in press.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. ***Neuroreport, 9***(2), 303-308.

Gegenfurtner, K. R. (1997). Sensory and cognitive contributions of color to the recognition of natural scenes. ***Investigative Ophthalmology and Visual Science, 39***, S156.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. ***Perception & Psychophysics, 55***(1), 48-121.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. ***Journal of Neurophysiology, 71***(3), 856-867.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. ***Curr. Biol., 5***(5), 552-563.

Milner, A. D., & Goodale, M. A. (1993). Visual pathways to perception and action. ***Progress in Brain Research, 95***(317), 317-337.

Mollon, J. D. (1989). "Tho she kneeled in that place were she grew...". The use and origins of primate colour vision. ***J. Exp. Biol., 146***, 21-38.

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rocklund, & A. Peters (Eds.), ***Extrastriate cortex in primates*** (pp. 205-241). New-York: Plenum Press.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. ***Visual Neuroscience, 12***, 371-384.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. ***Cognitive Psychology, in press***.

Oram, M. W., & Perrett, D. I. (1992). Time Course of Neural Responses Discriminating Different Views of the Face and Head. ***Journal of Neurophysiology, 68***(1), 70-84.

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. ***Journal of Experimental Psychology : Learning, Memory, Cognition, 11***(3), 579-587.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. ***Perception, 21***, 147-160.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. ***Experimental Brain Research, 47***(3), 329-342.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. ***The Quaterly Journal of Experimental Psychology A, 41***(4), 797-827.

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (1998). Frugivory and colour vision in Alouatta seniculus, a trichromatic platyrrhine monkey. ***Vision Research, 38***, 3321-3327.

Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. ***Journal of Experimental Psychology: Animal Behavior Processes, 14***(3), 247-260.

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. ***Journal of Experimental Psychology: General, 111***(4), 369-389.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges : Evidence for time and scale dependent scene recognition. ***Psychological Science, 5***, 195-200.

Smid, H. G. O. M., Jakob, A., & Heinze, H. J. (1997). The organisation of multidimensional selection on the basis of color and shape : An event-related brain potential study. ***Perception and Psychophysics, 59***(5), 693-713.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. ***Nature, 400***(6747), 869-873.

Syrkin, G., & Gur, M. (1997). Colour and luminance interact to improve pattern recognition. ***Perception, 26***, 127-140.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. ***Percept. Psychophys., 61***(6), 1140-1153.

Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. ***Current Opinion in Neurobiology, 7***(4), 523-529.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. ***Journal of Neurophysiology, 66***, 170-189.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. ***Nature, 381***(6582), 520-522.

Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), ***Connectionism in Perspective.*** (pp. 63-92). Amsterdam: Elsevier.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), ***Analysis of visual behaviour*** (pp. 549-585). Cambridge, MA-USA: MIT Press.

Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 1 : behavioural study. ***European Journal of Neuroscience, 11***, 1223-1238.

Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys. Part 2 : single-cell study. ***European Journal of Neuroscience, 11***, 1239-1255.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improve object recognition in normal and low vision. ***Journal of Experimental Psychology : Human perception and performance, 19***(4), 899-911.

A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes

Michèle Fabre-Thorpe, Arnaud Delorme, Catherine Marlot & Simon Thorpe

Centre de Recherche Cerveau & Cognition (UMR 5549, CNRS-UPS)
Faculté de Médecine de Rangueil, 133, Route de Narbonne, 31062, Toulouse, France
mft@cerco.ups-tlse.fr, arno@cerco.ups-tlse.fr, marlot@cerco.ups-tlse.fr, thorpe@cerco.ups-tlse.fr

The processing required to decide whether a briefly flashed natural scene contains an animal can be achieved in 150 ms (Thorpe, Fize, & Marlot, 1996). Here we report that extensive training with a subset of photographs over a three-week period failed to increase the speed of the processing underlying such Rapid Visual Categorizations: completely novel scenes could be categorized just as fast as highly familiar ones. Such data imply that the visual system processes new stimuli at a speed and with a number of stages that cannot be compressed. This rapid processing mode was seen with a wide range of visual complex images challenging the idea that short reaction times can only be seen with simple visual stimuli and implying that highly automatic feed-forward mechanisms underlie a far greater proportion of the sophisticated image analysis needed for everyday vision than is generally assumed.

Both humans and monkeys are able to categorize natural images accurately and very rapidly (Fabre-Thorpe, Richard, & Thorpe, 1998; Thorpe et al., 1996). The nature of the underlying mechanisms is currently the subject of intense debate (Biederman & Kalocsai, 1997; Logothetis & Sheinberg, 1996; Tanaka, 1997; Treisman & Kanwisher, 1998; Wallis & Bülthoff, 1999), but it is generally believed that this sort of object recognition must involve bi-directional information processing, in which bottom-up information originating in the retina interacts with feed-back mechanisms in a system preset by top-down knowledge (Frith & Dolan, 1997; Grossberg, Mingolla, & Ross, 1997; Humphreys, Riddoch, & Price, 1997; Kersten, 1997; Kosslyn, 1994; Mesulam, 1998; Przybyszewski, 1998; Rao & Ballard, 1999; Sastry, Shah, Singh, & Unnikrishnan, 1999; Tallon-Baudry, Bertrand, Delpuech, & Permier, 1997; Ullman, 1996; Vecera & O'Reilly, 1998). Indeed, recent neurophysiological studies have demonstrated the importance of lateral and feedback interactions in fundamental visual processes that include texture segmentation and figure-ground effects (Hupé et al., 1998; Lamme, Super, & Spekreijse, 1998) as well as visual attention (Desimone, 1998; Duncan, Humphreys, & Ward, 1997; Hillyard, Teder-Salejarvi, & Munte, 1998; Luck, Chelazzi, Hillyard, & Desimone, 1997; Roelfsema, Lamme, & Spekreijse, 1998; Vidyasagar, 1998).

On the other hand, it is likely that at least some forms of visual processing can be achieved on the basis of purely feed-forward mechanisms (Thorpe & Imbert, 1989), although it is a difficult experimental issue to determine the relative importance of bottom-up and top-down effects. One potential hallmark of relatively automatic feed-forward processing mechanisms is that their time course should be relatively fixed. The argument supposes that when processing involves extensive use of bi-directional interactions between bottom-up and top-down mechanisms one should predict that processing time will depend heavily on experience, and indeed, this is very often the case. Experiments on visual priming have demonstrated that, in many tasks, behavioral reaction times are reduced by experience with particular stimuli (Schacter & Buckner, 1998; Warren & Morton, 1982; Wiggs & Martin, 1998), and there is clear evidence that training can influence visual processing, even at relatively early stages in the visual system (Ahissar & Hochstein, 1997; Dolan et al., 1997; Karni & Bertini, 1997; Tovee, Rolls, & Ramachandran, 1996). Furthermore, there have been a number of recent reports showing that stimulus repetition can have pronounced effects on the Event-Related Potentials (ERPs) to a variety of visual stimuli, and that in some cases these effects appear to occur at remarkably short latencies (Begleiter, Porjesz, & Wang, 1995; Debruille, Guillem, & Renault, 1998; George, Jemel, Fiori, & Renault, 1997; Seeck et al., 1997). Note that the existence of such familiarity effects is not enough to distinguish feed-forward and feed-back mechanisms, because effects of familiarity can arise even in feed-forward models (Mozer, 1991). On the other hand, if familiarity has no effect on processing speed, only preoptimized and largely hard-wired feedback mechanisms would be compatible.

In this study, we have looked at the effects of familiarity on processing speed and accuracy in a go/no-go visual categorization task. When natural images that they have never seen before are flashed for only 20 ms, humans can detect the presence of an animal with high accuracy (94% correct) and with a median reaction time of 445 ms (Thorpe et al., 1996). Moreover, the associated ERPs show differential cerebral activity between

target and non-target trials that develops from 150 ms after stimulus onset suggesting that the visual processing in such a complex categorization could be mainly based on feed-forward mechanisms. In the present study, we investigated whether repetitive presentations of the stimuli during 14 days spread over three weeks, and thus history and knowledge, would speed-up the information processing in this fast go/no-go visual categorization task.

Results

The task required the subjects to respond by releasing a button whenever a briefly flashed image contained a living animal. A set of 2000 color photographs was selected from a commercial photo library of which 50% were targets showing one or more animals in their natural environments. They were extremely varied and included mammals, birds, fish, reptiles, insects, arthropods and crustacean. The remaining pictures were distractors and were also extremely varied, including not just pictures of natural scenes (landscapes, fruits, flowers…), but also a range of man-made environments. As a test of the sophistication of the decision strategies used by the subjects, human beings together with a number of "trick" images of animal-like objects (paintings, embroideries, statues, balloons...) were used as distractors. During the first part of the experiment, each subject was trained on 14 separate days spread over three weeks with a subset of 200 randomly selected images. Performance improved significantly during this period, both in terms of accuracy (94% on day 1, 97% on day 14) and reaction time which decreased from an initial mean value of 480 ms (median RT = 467 ms) to 436 ms (median RT = 420 ms) on day 14. On the two subsequent days, the familiar stimuli were randomly mixed with 1200 completely novel ones and testing was combined with Event-Related Potential recordings. Behavioral and electrophysiological responses with novel and familiar stimuli were directly compared.

The behavioral results showed that accuracy with familiar stimuli was higher than for the novel ones (96.9 % correct vs. 94.7 %) This improvement was mainly the result of a higher success rate with some of the more difficult targets (1.2 % of the familiar targets were missed compared with 6.3 % of the novel targets), and mean reaction time was also significantly shorter for familiar than for novel targets (424 ms for familiar vs. 444 ms for novel, $p<0.0001$ paired t-test, DF=13). However, analysis of the reaction time distributions plotted in Figure 1 reveals that this apparent increase in processing speed is virtually entirely due to the elimination of long reaction time responses to some of the familiar targets that were initially difficult to detect. In contrast, the initial parts of the two distributions that correspond to the earliest reaction times can be superimposed, showing that processing was no faster for the highly familiar images. In fact, the 10th percentile point of the reaction time distribution is identical for novel and familiar natural scenes;

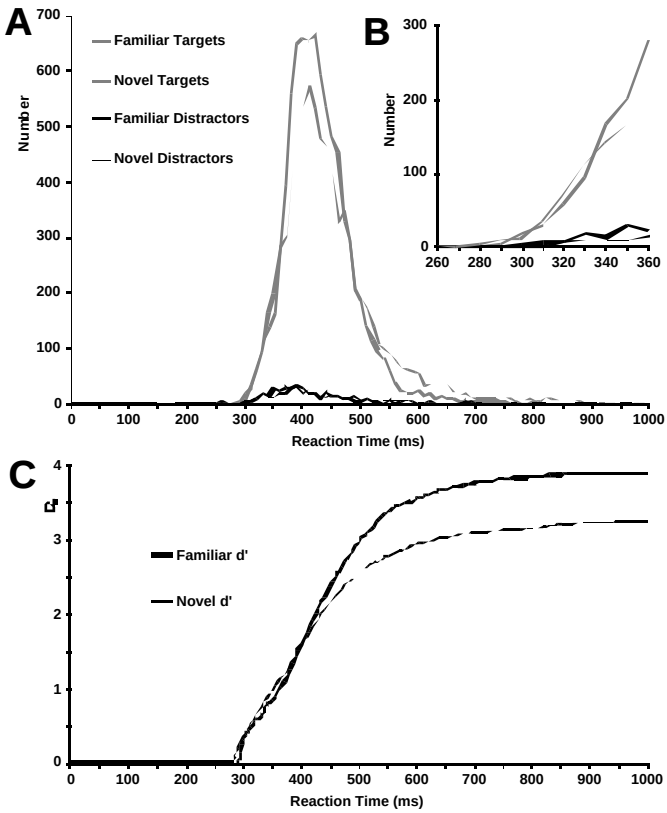
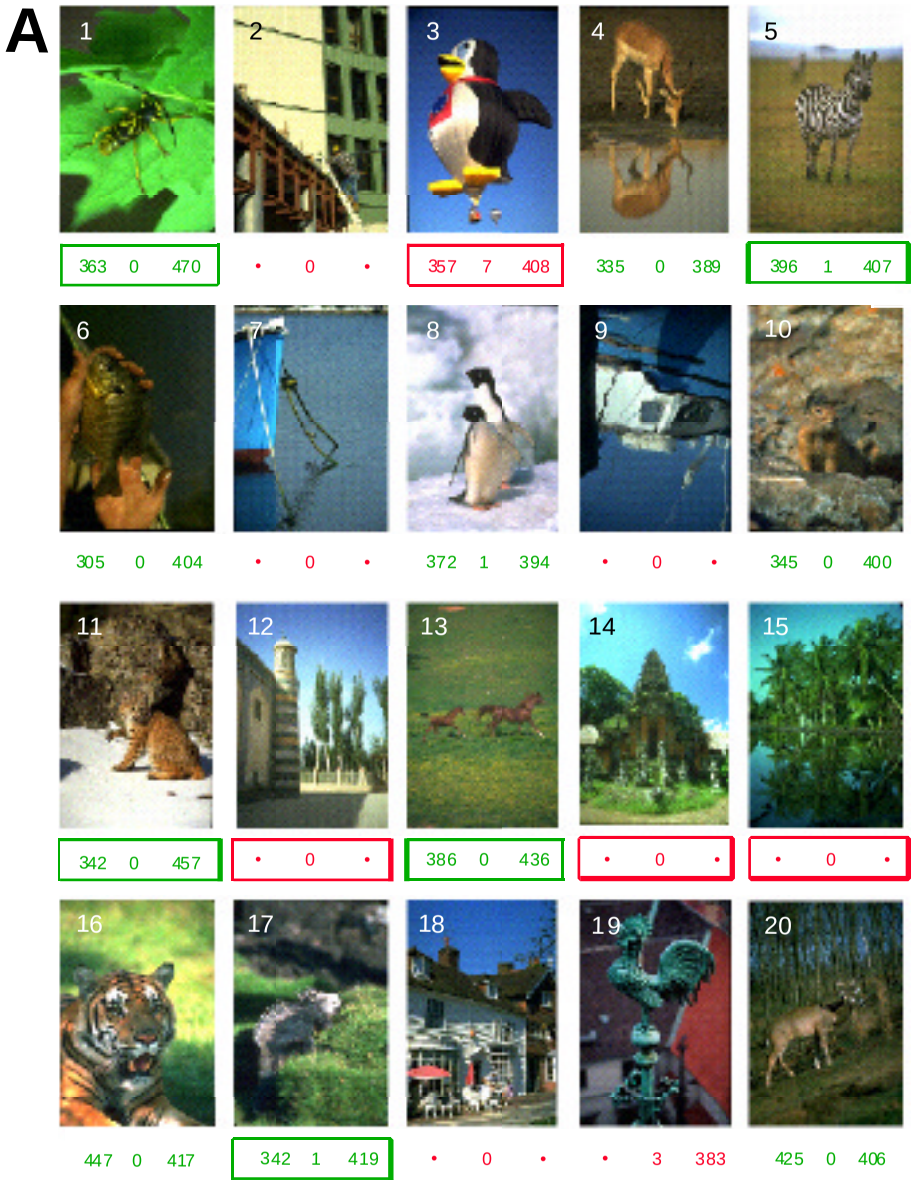
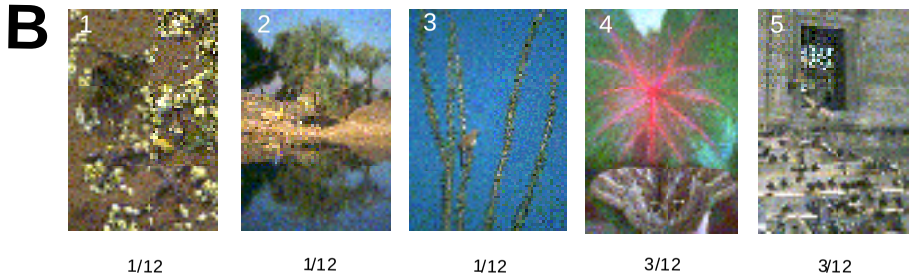
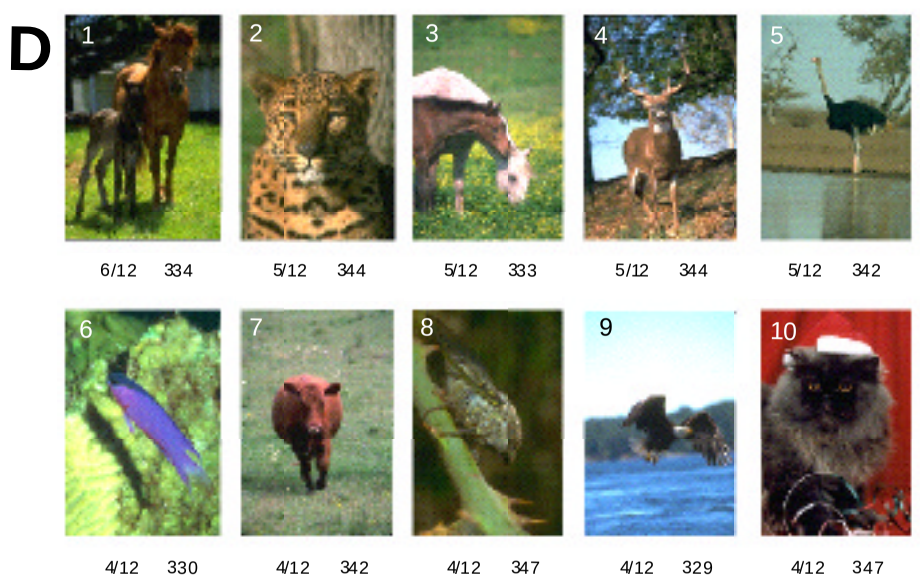


Figure 1. A: Reaction time distributions for Novel (thin lines) and Familiar (thick lines) stimuli. The values give the number of trials in a 10 ms bin. Gray curves show the data for correct go responses to target stimuli, whereas the black lines provide data for incorrect go responses to distractors. B: The inset shows a higher magnification view of the reaction time distributions for the critical region between 260 and 360 ms. Note that (i) the curves are effectively identical for novel and familiar targets, and (ii) that there is a strong bias in favor of responding to targets from very early on, thus ruling out the possibility that the subjects were making anticipatory responses. C: A plot of how the sensitivity index d' improves as a function of time for familiar and novel stimuli. The d' was calculated from the formula $d'=z_n - z_s$, where z_n is chosen such that the area of the normal distribution above that value is equal to the false-alarm rate, and where z_s is chosen to match the hit rate. While it is clear that d' is higher for familiar stimuli if we take into account all the behavioral responses, no advantage is seen before 400 ms (36.1% of the responses).

10.5% of responses to novel targets had reaction times of 360 ms or less, whereas the corresponding figure for familiar targets was 10.4%. Furthermore, by using a measure of d' to assess sensitivity, we found that only after 400 ms was there any evidence that performance was actually better for familiar stimuli (Figure 2C). The reason is that while the probability of a hit (correct detection of a target) earlier than 400 ms was higher for familiar targets than novel ones (0.372 vs. 0.312), so was the false alarm rate (0.027 for familiar distractors, and 0.019 for novel ones). Thus, over a third of the images used in the present study was processed so quickly that an improvement in performance with familiarity could not be seen.







Journal of Cognitive Neuroscience, ***in press***

Figure 2. A. A representative sequence of 20 images taken from one subject (RvR). Targets are shown with green text, distractors in red, and a box surrounds the text for images that were familiar to RvR. The left number gives the reaction time for RvR on trials where he responded. The center number gives the number of errors made by the 12 subjects for whom the stimulus was novel (for the other two subjects, the image was familiar), and the right hand number gives the mean reaction time for those 12 subjects. Note that all 12 subjects to whom they were presented as novel successfully detected the targets in images 1, 4, 6, 10, 11, 13, 16 and 20. The only targets that posed a problem were images 5, 8 and 17 that were each missed by one of the 12 subjects. Similarly, the distractor stimuli in images 2, 4, 7, 9, 12, 14, 15 and 18 were successfully ignored by all the subjects. Only images 3 and 19, which were "trick" distractors provoked errors – 7 out of 12 in the case of image 3, and 3 out of 12 for image 19. B. The 5 images that resulted in the highest number of errors. Below each image is indicated the number of subjects (out of 12) that correctly detected the target. C. The 5 images that were correctly detected by all the subjects but that resulted in the highest number of particularly long reactions times (over 505 ms). Below each image is indicated : on the left, the number of subjects (out of 12) with reaction times above 505 ms, and on the right, the mean reaction time for these subjects. D. A set of images which produced the highest proportion of very short (<360 ms) reaction times in the 12 subjects for whom they were completely novel. Below each image is indicated : on the left, the number of subjects (out of 12) with reaction times below 360 ms, and on the right the mean reaction time for these subjects. Images 1-5 shows the subset of images for which 5 or 6 subjects (out of 12) had very shorts RTs; images 6-10 are a representative subset of the 18 images for which 4 subjects (out of 12) had very short RTs.

Note that the absence of such short reaction times cannot be the result of subjects making random anticipations. If subjects were just guessing, they could not do better than 50% correct, since the probability of target and distractor trials was equal. However, the comparison of the reaction time distributions for correct “go” responses on targets and incorrect "go" responses on distractor trials clearly shows that even the shortest reaction times (those between 280 and 300 ms) are very strongly and statistically significantly biased in favor of targets. Thus, even such fast responses allow enough time for both visual processing and response execution.

The sequence shown in figure 2A illustrates the wide range of images used in these experiments, and demonstrates the remarkable sophistication of the decision strategies required to perform the task. Despite this great variety, when presented as novel, 511 of the 700 targets were correctly detected by all 12 subjects. Similarly, 533 of the 700 distractors seen as novel were correctly ignored by all 12 subjects. Errors, whether for novel targets or distractors, tended to be concentrated on a relatively small subset of the images in that over half of them (53.7%) were produced by only 5.5% of the images. Indeed, behavioral responses to this particular subset of targets were also characterized by particularly long reaction times (mean RT, 557 ms). One may wonder what factors make certain target images difficult to detect. To investigate this issue we examined all the stimuli that produced either a high proportion of misses, or an abnormally large number of “go” responses with long reaction times. Figure 2B shows five of the most difficult images as defined by the error rates, and figure 2C shows the five images that, although correctly categorized by all 12 subjects, had the highest proportion of long reaction times. Note that with three weeks of training these same targets could often be accurately detected (89% correct), with a reduced mean reaction time (470 ms) and that the elimination of these long latency responses accounts for much of the training-related changes in the RT distributions seen in Figure 1. Various reasons could account for the difficulty in spotting the animals in such images. They include small target size in the scene (B2, B3, C1), the presence of several small targets (B5), the unusual or ambiguous aspect of some animals (B1, C3), targets for which contours are difficult to extract (B3, C1, C2), and images in which the animal is not the main object (B2, B4, B5). Further analysis would be required to pin-point all the factors that make certain images difficult to analyze.

Given that a small subset of the images was particularly difficult to categorize, it is natural to suppose that there might be another subset that was particularly easy to process. In such a case, the lack of improvement seen for the most rapidly categorized images following training could be explained by the existence of a floor effect: no further improvement was possible because the processing required was already so trivial. To test this possibility, we looked at the images for which the proportion of subjects that responded with particularly short reaction times was unusually high. Of the 700 novel targets, 23 images induced a RT < 360 ms for 4 or more subjects (all were correctly detected by all 12 subjects). Ten of them are illustrated in figure 2D. Intriguingly, unlike the difficult images, there seems to be little to distinguish them from the rest of the set. They are highly varied, including not only mammals, but also birds, fish and insects. Furthermore, they include examples that, by almost any criterion, would be considered very challenging for current models of visual processing (non-canonical views, overlapping objects, low contrast images etc.). This makes it very difficult to argue that the lack of an improvement in rapid processing could be explained by supposing that these images were abnormally easy to process.

This clear contrast between “easy” and “hard” targets is made even more obvious in table 1 that gives the numbers of images for which 4 or more subjects had abnormally long or short reaction times. While the distribution for long RTs clearly points to a subset of particularly difficult images, the distribution for short RTs is exactly what would be expected on the basis of a purely random distribution.

Further evidence for the lack of an early effect of familiarity comes from the analysis of the differential Event-Related Potentials between target and distractor trials that were recorded in parallel during the final two days of testing. As in the previous study (Thorpe et al., 1996), we found a very strong

Number of Subjects out of 12	Image distribution		
	Observed RT < 361 ms	Expected with p= 0.118	Observed RT > 505 ms
0-3	488	488	463
4	18	18	27
5	4	4	10
6	1	1	9
7	0	0	1
8-10	0	0	0
11	0	0	1
Total	511	511	511

Table 1. Distribution of slow and fast behavioral responses to novel targets. The analysis was restricted to the 511 novel targets that were correctly identified by all 12 subjects so that each target had induced the same number of behavioral go responses (a total of 12*511 = 6132 behavioral responses). Of these, 724 were triggered with RTs of 360 ms or less (a limit below which familiarity effects could not be seen); they were compared to the similar number of responses (722) that were triggered with abnormally long RTs (505 ms or more). The table shows how many of the 511 targets were detected by k subjects (from 0 to 11) with either a short RT (left hand column) or a long RT (right hand column). For comparison, the central column gives the expected number of images for which k of the 12 subjects would have had short RTs if the reaction times were totally random (binomial distribution with p(RT≤360ms) equal to 724/6132 = 0.118). The distribution seen for short RTs (left hand column) fits exactly the prediction of a random distribution, which means that there was no evidence to support the view that certain images were associated with particularly fast responses. In contrast, the distribution of images with long RTs is heavily skewed (right hand column). For example, one image (see Figure 2C, image 1) produced RTs of over 505 ms in 11 of the 12 subjects. This is strong evidence that even when we only take into consideration images that were perfectly identified, a particular subset of these targets was especially hard to detect. Note that this analysis does not include even more difficult targets that were missed by one or more subjects (see figure 2B).

and robust difference in the frontally recorded evoked potentials (averaging together all the waveforms obtained either on correct target and distractor trials) that developed from 150 ms after stimulus onset. Indeed, the strong differential reaction at 150 ms is actually preceded by a deflection in the opposite direction that becomes statistically significant at around 120 ms. Differential event-related electrical and magnetic responses of this type have now been seen by a number of authors in relation to higher order visual processing (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Eimer, 1998; Linkenkaer-Hansen et al., 1998; Schendan, Ganis, & Kutas, 1998). Surprisingly, there was no difference in either the onset latency of the differential activity or its slope for familiar and for novel images (see Figure 3), although it is clear that (as in a number of other studies, (Buckner et al., 1998; Hertz, Porjesz, Begleiter, & Chorlian,

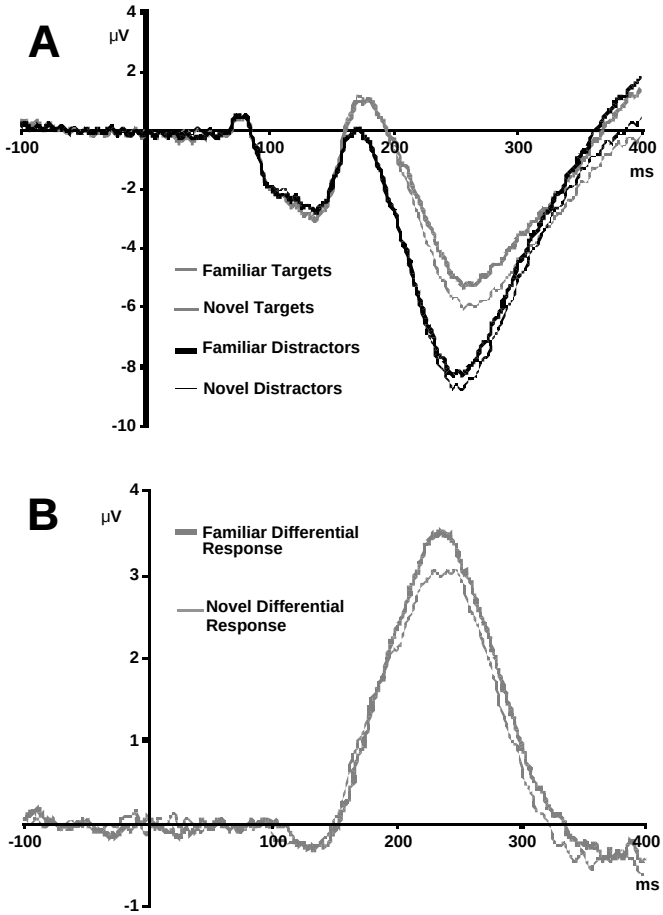


Figure 3. A. Grand average evoked potentials obtained from electrode Fz using 14 subjects. Separate curves are shown for Familiar Targets and Distractors, and for Novel Targets and Distractors. Note that familiarity has essentially no effect on either set of curves until at least 200 ms post stimulus. B. Plots of the evolution of the difference between the response to Targets and Distractors separately for Familiar and Novel stimuli. Note that the onset and slope of the differential responses are essentially identical for novel and familiar stimuli. In addition, there is a small differential response of opposite polarity that precedes the 150 ms deflection. This difference is statistically significant for the time steps 113 to 136 ms (two tailed t-test, t>2.65, df=13, p < .02).

1994)) the responses to familiar stimuli were in general somewhat smaller than with novel ones. On the other hand, there was evidence that the amplitude of the differential response between target and distractor trials was slightly larger in amplitude with familiar stimuli, but this effect was not evident until 30-50 ms after the start of the differential response.

Discussion

The present results argue for the existence of a specific mode of visual processing that cannot be speeded up by extensive training; a mode which we term Ultra-Rapid Visual Categorization. It may well be that this very rapid processing mode is used in many everyday visual tasks, but it is only when

subjects are forced to respond very quickly (as in the current task) that such processing can be revealed at the behavioral level. Both the reaction time data and the differential ERP effects indicate that the underlying visual processing can be just as fast with completely novel images as it can be for images with which the subjects are highly familiar. The existence of such a floor effect is certainly a surprising finding, given that familiarity normally facilitates processing, as shown by the extensive literature on the effects of priming. Although we are not claiming that priming did not occur, since we did indeed see a clear 20 ms decrease in mean reaction time for familiar images that was associated with a 2.2% increase in accuracy, we would claim that priming has no effect on either the shortest behavioral reaction times (i.e. those under 360 ms) or the onset latency of the differential category-specific ERP signal.

It appears, therefore, that even extensive training with particular stimuli cannot result in significantly faster processing than that seen with totally novel images. This appears to be in contradiction with the generally accepted view that visual processing involves extensive bi-directional mechanisms based on contextual knowledge. Rather, it implies that even a challenging visual task of the sort used here can be performed using highly automatic and data-driven routines, probably involving extensive use of feed-forward mechanisms. The processing complexity needed for the classification of some of the stimuli (humans from other animals, or living animals from artificial ones) also implies that these feed-forward mechanisms can be much more sophisticated than previously thought.

There is, in fact, a range of arguments in favor of the view that some types of processing can indeed be achieved using feed-forward mechanisms (Thorpe et al., 1989). One argument comes from studies that have demonstrated that many neurons are already highly selective even at the very beginning of the visual response, a result which suggests that iterative feed-back mechanisms are not required to build up the response selectivity (Celebrini, Thorpe, Trotter, & Imbert, 1993; Oram & Perrett, 1992; Rolls & Tovee, 1994). For example, this has been shown for face-selective neurons in the monkey temporal lobe that start firing 100 ms or less after the onset of a visual stimulus. Such data need to be considered in relation to the remarkably short behavioral reaction times that we have seen in the present task and which cannot be explained by random anticipations. In humans, such early responses start at around 280 ms, but in monkeys, the input-output sequence linking the retina to the hand muscles can take as little as 180 ms (Fabre-Thorpe et al., 1998). This is only 80 or so milliseconds after the onset of firing of neurons in inferotemporal cortex (Perrett, Rolls, & Caan, 1982), a structure that is almost certainly needed to perform this sort of task. However, IT cortex has no direct outputs to the motor system, and so information probably needs to be sent via relays in prefrontal, premotor and motor cortex before being sent down the pyramidal tract to the motoneurons in the spinal cord. This leaves very little time to do anything other than a feed-forward pass.

A second argument comes from the finding that, in this same rapid visual categorization task, the removal of chromatic information from such images has essentially no effect on performance, either in monkeys, or in humans with fast reaction times (Delorme, Richard, & Fabre-Thorpe, 1999). One interpretation of this result depends on the fact that luminance-based information in the magnocellular pathways reaches visual cortex slightly before the chromatic input from the parvocellular pathways (Nowak & Bullier, 1997). In this case, any processing or behavior relying on just the initial part of the response in each cortical area and using feed-forward processing would not be influenced by chromatic information. Rather it would be dominated by relatively coarse luminance based magnocellular information as has been suggested by experiments on coarse to fine visual processing (Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994).

A third argument comes from the existence of neural network models using only feed-forward mechanisms that are able to perform non-trivial visual processing tasks (Mel, 1997; Rolls & Treves, 1998; Wallis & Rolls, 1997). For example, Van Rullen et al. recently demonstrated that a simple multi-layer feed-forward network of asynchronously spiking neurons can be trained to locate faces in natural images (Van Rullen, Gautrais, Delorme, & Thorpe, 1998).

Finally, recent fMRI and ERP data has demonstrated that even visual stimuli that have been masked so that the subject is unaware of their presence are capable of significantly modulating activity in the motor system (Dehaene et al., 1998). There is therefore reason to believe that sophisticated visual processing is indeed possible using rapid feed-forward mechanisms, and it makes sense from an evolutionary point of view to use feed-forward mechanisms whenever possible.

While few would argue with the suggestion that at least some information can be extracted from an image on the basis of the first wave of information passing through the visual system, the present data suggests that the complexity of the analysis which can be achieved using such mechanisms is substantially greater than has been suspected in the past. One particularly important issue concerns the question of image segmentation, since it is generally assumed that cluttered natural images of the type used in the present experiments must first be correctly segmented before object recognition is possible. Recent neurophysiological experiments on texture segmentation and figure-ground effects imply that this is a time-consuming process that involves feed-back circuits (Lamme et al., 1998), so it is difficult to see how such a complex process could be achieved in a feed-forward network. On the other hand, the present results suggest that the visual processing needed to detect the presence of an animal in a natural scene (and trigger a behavioral response) may be possible without having first to complete these relatively time-consuming low level processes. It could be that detecting an animal in a natural scene can be done

with a massively parallel search for a large number of component features, some of which may be sufficient on their own to indicate - with a high probability - the presence of an animal. For example, the presence of an eye, a fin, a claw or teeth would all be sufficient to trigger a behavioral response in an animal detection task, even under conditions where the image has not yet been completely segmented. In this respect it may be worth noting that top-down modulation of visual processing may well play a vital role in pre-setting the visual system in such a way that neuronal mechanisms sensitive to components of animals are primed. Note, however, that this must be done before the start of a trial.

Whereas all these arguments support the view that feed-forward processing could play a critical role in the present task, it is certainly not possible to rule out the use of feedback loops. Certain critical steps in visual processing may involve such mechanisms, but if this is the case, the current evidence provides a number of serious constraints. First, their number must be very restricted in order to be compatible with the short behavioral reaction times observed both in the present study and in previous ones (Fabre-Thorpe et al., 1998; Thorpe et al., 1996). Second, the use of feedback must presumably be restricted to relatively hard-wired and preoptimized processing that does not depend on specific knowledge about particular images.

The present data support the idea that categorization can involve two different types of mechanism that have different time courses. Ultra-Rapid Visual Categorization would be the result of an initial wave of fast feed-forward visual processing that can quickly trigger behavioral responses. In our task, many targets seem to be detected that way, but to deal with particularly difficult stimuli, additional time consuming processing would be necessary. In everyday life, when constraints in response speed are not so high and accuracy is at a premium, most responses will only be triggered once analysis is sufficiently detailed. This additional processing would almost certainly involve extensive top-down and interactive mechanisms, and would be expected to be highly subject to the effects of training. Indeed, this is precisely what was observed here. Extensive training certainly did improve performance, but only on the (relatively limited) subset of targets that were initially difficult to detect. It seems likely that such images may be impossible to categorize on the basis of the first wave of visual information and require a longer processing phase with more detailed or more specific analysis. Such two-stage visual processing has recently been shown for the analysis of faces (Sugase, Yamane, Ueno, & Kawano, 1999) and is thought to require a number of iterative processes (Oram & Richmond, 1999).

It should be stressed that even if very rapid and automatic uni-directional mechanisms are often sufficient to trigger a response in our task, there is certainly no reason to believe that this would signal the end of visual processing. Indeed, there is actually no good reason to suppose that the ability of subjects to perform this challenging task involves conscious visual perception at all. When presented with a flashed photograph of the type used in the current experiments, subjects report that the first thing they perceive is a fully segmented scene in which the relations of all the objects in the image are clear. But this end-result may involve a much more complex process than that required to trigger (or not) the behavioral response in the task used here. Furthermore, the rapid detection of animal features using fast feed-forward mechanisms could be used to help the segregation process occurring at lower levels in the visual system via feedback connections. The precise limits of what can be achieved by the first feed-forward pass through the visual system is clearly something that will need a great deal of further research, but it is already clear that the power of this type of processing has been seriously underestimated in the past.

Methods

Subjects : Fourteen subjects participated in the study (7 males, 7 females, mean age 30.5 years, range 21-46).They were divided into 7 pairs (1 male, 1 female). 1400 images were used in this experiment, and each pair of subjects was assigned a different subset of 200 images on which performance was tested Monday to Friday for three weeks. During the testing session, the subjects were seated 1m from a color monitor in a dimly lit room. They were required to start each trial by pressing a button. After a short delay, an 8-bit color photograph (256 pixels wide by 384 pixels high) was flashed for 20 ms using a programmable graphics board (VSG 2.1, Cambridge Research Systems) and the subject was required to respond by releasing the button if the image contained an animal. On distractor trials, the subjects had to keep pressing the button for at least 1 second. Trials were grouped in blocks of 100, with 50 targets and 50 distractors shuffled at random before the start of each block. The 20 ms presentation prevented any exploratory eye movement.

Behavioral procedure and ERP recordings : On the fourteenth and fifteenth days of testing (Thursday and Friday of week 3) testing was associated with EEG recording. The training session with the 200 familiar images was immediately followed by a testing session in which the familiar images were mixed 3 times with 600 of the remaining 1200 novel images. Each day's testing involved 12 blocks of 100 trials, each of which contained 50 of the familiar stimuli together with 50 of the novel images, chosen at random. Over the course of the final two days, each subject was tested on a total of 2400 trials, 1200 novel stimuli, and 6 repetitions of each of the 200 familiar ones. By assigning different sets of 200 stimuli for training with each of the 7 pairs of subjects, we ensured that all 1400 images were used equally often as familiar and novel stimuli. The brain electrical activity was recorded using a 32-channel SynAmps amplification system (Neuroscan Inc.) using an sampling rate of 1 kHz and linked

ears as the reference. Potentials on each trial were baseline corrected using the signal during the 100 ms that preceded the onset of the stimulus. Trials with artifacts related to ocular movements and blinks were excluded from the analysis, as were trials contaminated by strong alpha frequency activity. As in the previous study by Thorpe et al, comparisons between responses on target and distractor trials were made using paired t-test.

Acknowledgements :
This work was supported by the CNRS, by the Cognisciences Program, and by the Midi-Pyrénées Region. The experimental procedures used were authorized by the local ethical committee (CCPPRB N° 9614003).

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. Nature, 387, 401-406.
Begleiter, H., Porjesz, B., & Wang, W. (1995). Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. Electroencephalogr Clin Neurophysiol, 94, 41-49.
Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. J Cognitive Neurosci, 8, 551-565.
Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philos Trans R Soc Lond B Biol Sci, 352, 1203-1219.
Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., & Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron, 20, 285-296.
Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake monkey. Visual Neuroscience, 10, 811-825.
Debruille, J. B., Guillem, F., & Renault, B. (1998). ERPs and chronometry of face recognition: following-up Seeck et al. and George et al. Neuroreport, 9, 3349-3353.
Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. Nature, 395, 597-600.
Delorme, A., Richard, G., & Fabre-Thorpe, M. (1999). Rapid processing of complex natural scenes : A role for the magnocellular visual pathways? Neurocomputing, 26-27, 663-670.
Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. Philos Trans R Soc Lond B Biol Sci, 353, 1245-1255.
Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S. J., & Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. Nature, 389, 596-599.
Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. Curr Opin Neurobiol, 7, 255-261.
Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? Neuroreport, 9, 2945-2948.
Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. NeuroReport, 9, 303-308.
Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. Philos Trans R Soc Lond [Biol], 352, 1221- 1230.
George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: A spatio- temporal ERP study. Neuroreport, 8, 1417-1423.
Grossberg, S., Mingolla, E., & Ross, W. D. (1997). Visual brain and visual perception: how does the cortex do perceptual grouping? Trends Neurosci, 20, 106-111.
Hertz, S., Porjesz, B., Begleiter, H., & Chorlian, D. (1994). Event-related potentials to faces: the effects of priming and recognition. Electroencephalogr Clin Neurophysiol, 92, 342-351.
Hillyard, S. A., Teder-Salejarvi, W. A., & Munte, T. F. (1998). Temporal dynamics of early perceptual processing. Curr Opin Neurobiol, 8, 202-210.
Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. Philos Trans R Soc Lond B Biol Sci, 352, 1275-1282.
Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature, 394, 784-787.
Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. Curr Opin Neurobiol, 7, 530-535.
Kersten, D. (1997). Perceptual categories for spatial layout. Philos Trans R Soc Lond [Biol], 352, 1155- 1163.
Kosslyn, S. M. (1994). Image and brain : the resolution of the imagery debate. Cambridge, Mass.: MIT Press.
Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. Curr Opin Neurobiol, 8, 529-535.
Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. Neurosci Lett, 253, 147-150.
Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. Annu Rev Neurosci, 19, 577-621.
Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol, 77, 24-42.
Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. Neural Comput, 9, 777-804.
Mesulam, M. M. (1998). From sensation to cognition. Brain, 121, 1013-1052.
Mozer, M. C. (1991). The perception of multiple objects : a connectionist approach. Cambridge, Mass.: MIT Press.
Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rockland, & A. Peters (Eds.), Extrastriate cortex in primates (Vol. 12, pp. 205-241). New York: Plenum.
Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. J Neurophysiol, 68, 70-84.
Oram, M. W., & Richmond, B. J. (1999). I see a face - a happy face. Nature Neuroscience, 2, 856-858.
Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. Perception, 21, 147-160.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res, 47, 329-342.
Przybyszewski, A. W. (1998). Vision: Does top-down processing help us to see? Curr Biol, 8, R135-R139.
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci, 2, 79-87.
Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. Nature, 395, 376-381.
Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc R Soc Lond B Biol Sci, 257, 9-15.
Rolls, E. T., & Treves, A. (1998). Neural Networks and Brain Function. New-York: Oxford University Press.
Sastry, P. S., Shah, S., Singh, S., & Unnikrishnan, K. P. (1999). Role of feedback in mammalian vision: a new hypothesis and a computational model. Vision Research, 39, 131-148.
Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. Neuron, 20, 185-195.
Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. Psychophysiology, 35, 240-251.
Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges : Evidence for time and scale dependent scene recognition. Psychological Science, 5, 195-200.
Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., & Schomer, D. L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. Neuroreport, 8, 2749-2754.
Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. Nature, 400, 869-873.
Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Permier, J. (1997). Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. J Neurosci, 17, 722-734.
Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. Curr Opin Neurobiol, 7, 523-529.
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520-522.
Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), Connectionism in Perspective. (pp. 63-92). Amsterdam: Elsevier.
Tovee, M. J., Rolls, E. T., & Ramachandran, V. S. (1996). Rapid visual learning in neurones of the primate temporal visual cortex. Neuroreport, 7, 2757-2760.
Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. Curr Opin Neurobiol, 8, 218-226.
Ullman, S. (1996). High-level vision : object recognition and visual cognition. Cambridge, Mass.: MIT Press.
Van Rullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. Biosystems, 48, 229-239.
Vecera, S. P., & O'Reilly, R. C. (1998). Figure-ground organization and object recognition processes: an interactive account. J Exp Psychol Hum Percept Perform, 24, 441-462.
Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. Neuroreport, 9, 1947-1952.
Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. Trends in Cognitive Sciences, 3, 22-31.
Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. Prog Neurobiol, 51, 167-194.
Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. Br J Psychol, 73, 117-129.
Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. Curr Opin Neurobiol, 8, 227-233.

Références

Abbott, L. F., et Song, S. (1999). Temporally Asymmetric Hebbian Learning, Spike Timing and Neuronal Response Variability. In M. S. Kearns, Solla, S.A. and Cohn, D.A. (Ed.), Advances in Neural Information Processing Systems vol. 11. Cambridge MA: MIT Press.

Adorjan, P., Levitt, J. B., Lund, J. S., et Obermayer, K. (1999). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. Vis Neurosci, 16, 303-318.

Aiple, F., et Krüger, J. (1988). Neuronal synchrony in monkey striate cortex: interocular signal flow and dependency on spike rates. Exp Brain Res, 72, 141-149.

Albrecht, D. G. (1995). Visual cortex neurons in monkey and cat: effect of contrast on the spatial and temporal phase transfer functions. Vis Neurosci, 12(6), 1191-1210.

Alonso, J. M., Usrey, W. M., et Reid, R. C. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. Nature, 383, 815-819.

Apicella, P., Scarnati, E., Ljungberg, T., et Schultz, W. (1992). Neuronal activity in monkey striatum related to the expectation of predictable environmental events. J Neurophysiol, 68(3), 945-960.

Arnett, D. W. (1975). Correlation analysis of units recorded in the cat dorsal lateral geniculate nucleus. Exp Brain Res, 24, 111-130.

Aubertin, A., Fabre-Thorpe, M., Fabre, N., et Géraud, G. (1999). Fast visual categorization and speed of processing in migraine. C R Acad Sci III, 322(8), 695-704.

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., et Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proc R Soc Lond B Biol Sci, 264(1389), 1775-1783.

Baizer, J. S., Ungerleider, L. G., et Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J Neurosci, 11(1), 168-190.

Bar, M., et Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. Proc Natl Acad Sci U S A, 96(4), 1790-1793.

Barcelo, F., Suwazono, S., et Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. Nat Neurosci, 3(4), 399-403.

Baylis, G. C., Rolls, E. T., et Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res, 342(1), 91-102.

Berman, N. J., Douglas, R. J., Martin, K. A., et Whitteridge, D. (1991). Mechanisms of inhibition in cat visual cortex. J Physiol, 440, 697-722.

Berry, M. J., Warland, D. K., et Meister, M. (1997). The structure and precision of retinal spike trains. Proc Natl Acad Sci U S A, 94, 5411-5416.

Bi, G. Q., et Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci, 18(24), 10464-10472.

Biederman, I. (1987). Recognition by components: a theory of visual pattern recognition. Psychol Rev, 94(2), 115-147.

Biederman, I., et Ju, G. (1988). Surface versus edge-based determinants of visual recognition. Cogn Psychol, 20, 38-64.

Booth, M. C., et Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex, 8(6), 510-523.

Borg-Graham, L. J., Monier, C., et Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. Nature, 393, 369-373.

Boucart, M., et Humphreys, G. W. (1992). The computation of perceptual structure from collinearity and closure: normality and pathology. Neuropsychologia, 30(6), 527-546.

Bovet, D., et Vauclair, J. (2000). Picture recognition in animals and humans. Behav Brain Res, 109(2), 143-165.

Bower, J. M., et Beeman, D. (1998). The book of GENESIS: Exploring realistic neural models with the GEneral SImulation System (Second Edition). New York: Springer-Verlag.

Boyce, S. J., et Pollatsek, A. (1992). Identification of objects in scenes: the role of scene background in object naming. J Exp Psychol: Learn Mem Cogn, 18, 531-543.

Bradski, G., et Grossberg, S. (1995). Fast-Learning VIEWNET Architectures for Recognizing Three-dimensional Objects from Multiple Two-dimensional Views. Neural Networks, 8, 1053-1080.

Bringuier, V., Chavane, F., Glaeser, L., et Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science, 283, 695-699.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., et Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron, 20, 285-296.

Bugmann, G. (1996). Biologically plausible neural computation. Biosystems, 40, 11-19.

Bullier, J., Hupé, J. M., James, A., et Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. J Physiol Paris, 90(3-4), 217-220.
Burnod, Y. (1991). Organizational levels of the cerebral cortex: an integrated model. Acta Biotheor, 39(3-4), 351-361.
Burnod, Y., Baraduc, P., Battaglia-Mayer, A., Guigon, E., Koechlin, E., Ferraina, S., Lacquaniti, F., et Caminiti, R. (1999). Parieto-frontal coding of reaching: an integrated framework. Exp Brain Res, 129(3), 325-346.
Cabeza, R., et Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. J Cogn Neurosci, 12(1), 1-47.
Carandini, M., et Ferster, D. (1997a). A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. Science, 276, 949-952.
Carandini, M., Heeger, D. J., et Movshon, J. A. (1997b). Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci, 17, 8621-8644.
Cavanagh, P., Saida, S., et Rivest, J. (1995). The contribution of color to depth perceived from motion parallax. Vision Res, 35, 1871-1878.
Cave, C. B., et Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. Perception, 22, 229-248.
Celebrini, S., et Newsome, W. T. (1995). Microstimulation of extrastriate area MST influences performance on a direction discrimination task. J Neurophysiol, 73(2), 437-448.
Celebrini, S., Thorpe, S. J., Trotter, Y., et Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. Vis Neurosci, 10(5), 811-825.
Chadaide, Z., Kovacs, G., Sary, G., Köteles, K., et Benedek, G. (1999). Selectivity of macaque inferior temporal neurons for shapes with different surface attributes. Perception, 28, 97.
Chagnac-Amitai, Y., et Connors, B. W. (1989). Synchronized excitation and inhibition driven by intrinsically bursting neurons in neocortex. J Neurophysiol, 62, 1149-1162.
Chao, L. L., Martin, A., et Haxby, J. V. (1999). Are face-responsive regions selective only for faces? Neuroreport, 10(14), 2945-2950.
Chelazzi, L., Duncan, J., Miller, E. K., et Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol, 80(6), 2918-2940.
Conan Doyle, A. (1917). The Adventure of Wisteria Lodge. Available at http://www.tirkzilla.com/holmes/.
Cook, M., et Mineka, S. (1990). Selective associations in the observational conditioning of fear in rhesus monkeys. J Exp Psychol: Anim Behav Process, 16, 372-389.
Cooper, E. E., et Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. J Exp Psychol: Learn Mem Cogn, 26(2), 470-488.
Coull, J. T., Frith, C. D., Buchel, C., et Nobre, A. C. (2000). Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. Neuropsychologia, 38(6), 808-819.
Crick, F., et Koch, C. (1998). Consciousness and neuroscience. Cereb Cortex, 8(2), 97-107.
Crook, J. M., Kisvarday, Z. F., et Eysel, U. T. (1997). GABA-induced inactivation of functionally characterized sites in cat striate cortex: effects on orientation tuning and direction selectivity. Vis Neurosci, 14, 141-158.
Damasio, A. R. (1999). Le sentiment même de soi. Paris: Odile Jacob.
Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., et Damasio, A. R. (1996). A neural basis for lexical retrieval. Nature, 380, 499-505.
Dawkins, R. (1990). Le gène égoïste. Paris: Armand Colin.
De Graef, P., De Troy, A., et D'Ydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scenes. Can J Psychol, 46, 489-508.
De Schutter, E. (1992). A consumer guide to neuronal modeling software. Trends Neurosci, 15, 462-464.
Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., et Le Bihan, D. (1998). Imaging unconscious semantic priming. Nature, 395, 597-600.
Delorme, A., Gautrais, J., VanRullen, R., et Thorpe, S. J. (1999). SpikeNET: A simulator for modeling large networks of integrate and fire neurons. Neurocomputing, 26-27, 989-996.
Delorme, A., Richard, G., et Fabre-Thorpe, M. (1999). Rapid processing of complex natural scenes: a role for the magnocellular pathway. Neurocomputing, 26-27, 663-670.
Delorme, A., Richard, G., et Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. Vision Res, 40(16), 2187-2200.
Desimone, R. (2000). Neuronal Mechanisms for Selective Attention. Paper presented at the Fourth International Conference on Cognitive and Neural Systems, Boston, MA.
Desimone, R., et Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J Neurophysiol, 57(3), 835-868.
Destexhe, A. (1997). Conductance-based integrate-and-fire models. Neural Comput, 9(3), 503-514.
DiCarlo, J. J., et Maunsell, J. H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. Nat Neurosci, 3(8), 814-821.

Dupuy, J. P. (1999). Aux origines des sciences cognitives. Paris: La découverte.
Edelman, S., et Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Res, 32(12), 2385-2400.
Egeth, H. E., et Yantis, S. (1997). Visual attention: control, representation, and time course. Annu Rev Psychol, 48, 269-297.
Engel, A. K., Konig, P., Kreiter, A. K., et Singer, W. (1991). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. Science, 252, 1177-1179.
Eriksen, C. W., et Collins, J. F. (1969). Temporal course of selective attention. J Exp Psychol, 80(2), 254-261.
Fabre-Thorpe, M., Delorme, A., et Richard, G. (1999). Singes et Hommes face au monde visuel: la Catégorisation. Primatologie, 2, 111-139.
Fabre-Thorpe, M., Delorme, A., Marlot, C., et Thorpe, S. J. (2000). A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes. J Cogn Neurosci, in press.
Fabre-Thorpe, M., Richard, G., et Thorpe, S. (1998). Rapid categorization of natural images by rhesus monkeys. Neuroreport, 9, 303-308.
Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. Behav Brain Res, 76(1-2), 181-189.
Felleman, D. J., et Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex, 1(1), 1-47.
Ferrera, V. P., Nealey, T. A., et Maunsell, J. H. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. J Neurosci, 14(4), 2080-2088.
Ferster, D., et Jagadeesh, B. (1992). EPSP-IPSP interactions in cat visual cortex studied with in vivo whole- cell patch recording. J Neurosci, 12, 1262-1274.
Ferster, D., et Spruston, N. (1995). Cracking the neuronal code. Science, 270, 756-757.
Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am, 4, 2379-2394.
Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol, 67, 103-112.
Fize, D. (2000). Bases Cérébrales de la Catégorisation Visuelle Rapide. Thèse, EHESS, Toulouse.
Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.
Frank, K., Poo, M., et Sejnowski, T. (2000). An M cell model of calcium dynamics and calmodulin activation in dendritic spines. Paper presented at the Ninth Annual Computational Neuroscience Meeting, Brugge, Belgium.
Freund, T. F., Martin, K. A., Soltesz, I., Somogyi, P., et Whitteridge, D. (1989). Arborisation pattern and postsynaptic targets of physiologically identified thalamocortical afferents in striate cortex of the macaque monkey. J Comp Neurol, 289, 315-336.
Friedman, D. (1990). Cognitive event-related potential components during continuous recognition memory for pictures. Psychophysiol, 27, 136-148.
Fujita, I., Tanaka, K., Ito, M., et Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature, 360, 343-346.
Fukai, T. (1995). Oscillations for rapid selection of neural activities based on spike timing. Neuroreport, 7, 273-277.
Gabbott, P. L., Martin, K. A., et Whitteridge, D. (1988). Evidence for the connections between a clutch cell and a corticotectal neuron in area 17 of the cat visual cortex. Proc R Soc Lond B Biol Sci, 233, 385-391.
Gallant, J. L., Connor, C. E., et Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. Neuroreport, 9(7), 1673-1678.
Gauthier, I., Skudlarski, P., Gore, J. C., et Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci, 3(2), 191-197.
Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., et Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci, 2(6), 568-573.
Gautrais, J., et Thorpe, S. (1998). Rate coding versus temporal order coding: a theoretical approach. Biosystems, 48(1-3), 57-65.
Gawne, T. J., Kjaer, T. W., et Richmond, B. J. (1996). Latency: another potential code for feature binding in striate cortex. J Neurophysiol, 76(2), 1356-1360.
Gazeres, N., Borg-Graham, L. J., et Frégnac, Y. (1998). A phenomenological model of visually evoked spike trains in cat geniculate nonlagged X-cells. Vis Neurosci, 15, 1157-1174.
Gegenfurtner, K. R., et Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. Curr Biol, 10(13), 805-808.
Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. Neural Comput, 12, 43-89.

Ghose, G. M., et Freeman, R. D. (1992). Oscillatory discharge in the visual system: does it have a functional role? J Neurophysiol, 68, 1558-1574.
Gibbon, J., et Church, R. M. (1992). Comparison of variance and covariance patterns in parallel and serial theories of timing. J Exp Anim Behav, 57(3), 393-406.
Glaser, W. R., et Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. J Exp Psychol Gen, 118, 13-42.
Golomb, D., et Ermentrout, G. B. (1999). Continuous and lurching traveling pulses in neuronal networks with delay and spatially decaying connectivity. Proc Natl Acad Sci U S A, 96, 13480-13485.
Goodale, M. A., Milner, A. D., Jakobson, L. S., et Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. Nature, 349, 154-156.
Gouras, P., et Kruger, J. (1979). Responses of cells in foveal visual cortex of the monkey to pure color contrast. J Neurophysiol, 42(3), 850-860.
Graßmann, C., et Anlauf, J. K. (1998). Distributed, Event Driven Simulation of Spiking Neural Networks. Paper presented at the Proceedings of the International ICSC/IFAC Symposium on Neural Computation.
Gray, C. M., Konig, P., Engel, A. K., et Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature, 338, 334-337.
Gray, E. M. (1987). Visual recognition of landmarks by the homing pigeon. Thèse, Harvard University, Boston, MA.
Gregory. (1972). Eye and Brain. Oxford: Oxford University Press.
Gross, C. G., Rocha-Miranda, C. E., et Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol, 35(1), 96-111.
Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. Percept Psychophys, 55, 48-121.
Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. Spatial Vision, 12, 163-186.
Hackley, S. A., et Valle-Inclan, F. (1999). Accessory stimulus effects on response selection: does arousal speed decision making? J Cogn Neurosci, 11(3), 321-329.
Hanna, A., et Remington, R. (1996). The representation of color and form in long-term memory. Mem Cognit, 24, 322-330.
Hansel, D., Mato, G., Meunier, C., et Neltner, L. (1998). On numerical simulations of integrate-and-fire neural networks. Neural Comput, 10, 467-483.
Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., et Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci, 14(11 Pt 1), 6336-6353.
Haxby, J. V., Ungerleider, L. G., Horwitz, B., Maisog, J. M., Rapoport, S. I., et Grady, C. L. (1996). Face encoding and recognition in the human brain. Proc Natl Acad Sci U S A, 93(2), 922-927.
Heller, J., Hertz, J. A., Kjaer, T. W., et Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision. J Comput Neurosci, 2(3), 175-193.
Hernstein, R. J., et Loveland, D. H. (1964). Complex Visual Concept in Pigeon. Science, 146, 549-551.
Hernstein, R. J., Loveland, D. H., et Cable, C. (1976). Natural concepts in pigeons. J Exp Psychol: Anim Behav Process, 2(4), 285-302.
Herrnstein, R. J. (1990). Levels of Categorization. In G. M. Edelman, W. E. Gall, et C. W. M. (Eds.), Signal and sense: Local and global order in perceptual maps (pp. 385-414). New York: Wiley-Liss.
Hertz, S., Porjesz, B., Begleiter, H., et Chorlian, D. (1994). Event-related potentials to faces: the effects of priming and recognition. Electroencephalogr Clin Neurophysiol, 92, 342-351.
Hines, M. (1989). A program for simulation of nerve equations with branching geometries. Int J Biomed Comput, 24, 55-68.
Hines, M. L., et Carnevale, N. T. (1997). The NEURON simulation environment. Neural Comput, 9, 1179-1209.
Hirsch, J. A., Alonso, J. M., Reid, R. C., et Martinez, L. M. (1998). Synaptic integration in striate cortical simple cells. J Neurosci, 18, 9517-9528.
Hodgkin, A. L., et Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation nerve. J Physiol (London), 117, 500-544.
Holmes, G. (1918). Disturbances of vision by cerebral lesions. Brit J Ophthalmol, 2, 353-384.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A, 79(8), 2554-2558.
Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. Nature, 376, 33-36.
Hopfield, J. J., et Herz, A. V. (1995). Rapid local synchronization of action potentials: toward computation with coupled integrate-and-fire neurons. Proc Natl Acad Sci U S A, 92, 6655-6662.
Horton, J. C., et Stryker, M. P. (1993). Amblyopia induced by anisometropia without shrinkage of ocular dominance columns in human striate cortex. Proc Natl Acad Sci U S A, 90(12), 5494-5498.

Hubel, D. H., et Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. J Physiol (Lond), 195(1), 215-243.
Hubel, D. H., et Wiesel, T. N. (1972). Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. J Comp Neurol, 146(4), 421-450.
Hummel, J. E., et Biederman, I. (1992). Dynamic Binding in a Neural Network for Shape Recognition. Psychol Rev, 99(3), 480-507.
Hummel, J. E., et Stankiewicz, B. J. (1996). An Architecture for Rapid, Hierarchical Structural Description. In T. Invi et J. McClelland (Eds.), Attention and Performance XVI: Information Integration in Perception and Communication (pp. 93-121). Cambridge, MA: MIT Press.
Humphreys, G. W., et Boucard, M. (1997). Selection by color and form in vision. J Exp Psychol : Hum Percept Perform, 23, 136-153.
Humphreys, G. W., et Forde, E. M. E. (2000). Category-specific deficits: A review and presentation of the Hierarchical Interactive Theory (HIT). Available at http://psgsuni.bham.ac.uk/staff/humphreysg/ghmanusp.htm, submitted.
Humphreys, G. W., Riddoch, M. J., et Quinlan, P. T. (1988). Cascade processes in picture identification. Cog Neuropsychol, 5, 67-103.
Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., et Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature, 394, 784-787.
Hurlbert, A., et Poggio, T. (1988). Making Machines (and AI) See, Daedalus: Proc Am Acad Arts Sci (Vol. 117, pp. 213-239).
Imbert, M. (1999). Etude du cortex cérébral des primates : comparaison des aires visuelles chez le macaque et chez l'homme. Primatologie, 2, 1-28.
Ito, M., Tamura, H., Fujita, I., et Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol, 73(1), 218-226.
Itti, L., et Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res, 40(10-12), 1489-1506.
Jaffe, D. B., et Carnevale, N. T. (1999). Passive normalization of synaptic integration influenced by dendritic architecture. J Neurophysiol, 82, 3268-3285.
Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. Exp Brain Res, 78(1), 193-202.
Jemel, B., George, N., Chaby, L., Fiori, N., et Renault, B. (1999). Differential processing of part-to-whole and part-to-part face priming: an ERP study. Neuroreport, 10(5), 1069-1075.
Jordan, J. R., Geisler, W. S., et Bovik, A. C. (1990). Color as a source of information in the stereo correspondence process. Vision Res, 30, 1955-1970.
Joseph, J. E., et Proffitt, D. R. (1996). Semantic Versus Perceptual Influence of Color in Object Recognition. J Exp Psychol : Learning, Mem and Cogn, 22, 407-429.
Kandel, E., Schwartz, J., et Jessel, T. (Eds.). (1991). Principle of Neural Science (third edition). New York, Amsterdam, London, Tokyo: Elsevier.
Kanwisher, N., McDermott, J., et Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci, 17(11), 4302-4311.
Kastner, S., De Weerd, P., Desimone, R., et Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science, 282, 108-111.
Kelly, J. P. (1985). Anatomy of the central visual pathways. In E. Kandel et J. Schwartz (Eds.), Principles of Neural Science (second edition). New York and London: Elsevier.
Kendrick, K. M., et Baldwin, B. A. (1987). Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. Science, 236, 448-450.
Kessels, R. P., Postma, A., et de Haan, E. H. (1999). P and M channel-specific interference in the what and where pathway. Neuroreport, 10(18), 3765-3767.
Keysers, C., Xiao, D., Foldiak, P., et Perrett, D. I. (2000). The Speed of Sight. J Cog Neurosci, in press.
Kinchla, R. A. (1974). Detecting target elements in multi-element arrays: a confusability model. Percept Psychophys, 15, 149-158.
Kinchla, R. A. (1992). Attention. Annu Rev Psychol, 43, 711-742.
Kistler, W. M., et van Hemmen, J. L. (2000). Modeling synaptic plasticity in conjuction with the timing of pre- and postsynaptic action potentials. Neural Comput, 12(2), 385-405.
Koch, C., et Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol, 4(4), 219-227.
Koechlin, E., Anton, J. L., et Burnod, Y. (1996). Dynamical computational properties of local cortical networks for visual and motor processing: a bayesian framework. J Physiol Paris, 90(3-4), 257-262.
Koechlin, E., Anton, J. L., et Burnod, Y. (1999). Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. Biol Cybern, 80(1), 25-44.

Kohonen, T. (1982). Self Organized Formation of Topologically Correct Feature Maps. Biol Cybern, 43, 59-69.
Kolb, F. C., et Braun, J. (1995). Blindsight in normal observers. Nature, 377, 336-338.
Kolers, P. A., et von Grunau, M. (1975). Visual construction of color is digital. Science, 187, 757-759.
Komatsu, H., et Ideura, Y. (1993). Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. J Neurophysiol, 70(2), 677-694.
Körner, E., Gewaltig, M. O., Körner, U., Richter, A., et Rodemann, T. (1999). A model of computation in neocortical architecture. Neural Networks, 12, 989-1005.
Kotchoubey, B., Grozinger, B., Kornhuber, A. W., et Kornhuber, H. H. (1997). Electrophysiological analysis of expectancy: P3 in informed guessing. Int J Neurosci, 91(1-2), 105-122.
Kovacs, I., Papathomas, T. V., Yang, M., et Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. Proc Natl Acad Sci U S A, 93(26), 15508-15511.
Kufler, S. W. (1953). Discharge patterns and functional organization of the mammalian retina. J Neurophysiol, 16, 37-68.
Lapicque, L. (1907). Recherches quantitatives sur l'exitation électrique des nerfs traité comme une polarisation. J Physiol Pathol Gen, 9, 620-635.
Lee, D. D., et Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788-791.
Legge, G. E., Parish, D. H., Luebker, A., et Wurm, L. H. (1990). Psychophysics of reading : Comparing color contrast and luminance contrast. J Opt Soc Am, 7, 2002-2010.
Lestienne, R., et Strehler, B. L. (1988). Differences between monkey visual cortex cells in triplet and ghost doublet informational symbols relationships. Biol Cybern, 59, 337-352.
Leuthold, H., et Sommer, W. (1993). Stimulus presentation rate dissociates sequential effects in event-related potentials and reaction times. Psychophysiol, 30(5), 510-517.
Li, L., Miller, E. K., et Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol, 69, 1918-1929.
Libet, B. (1981). Timing of cerebral processes relative to concomitant conscious experience in man. In G. Adam, I. Meszaros, et E. I. Nanyai (Eds.), Advances in Physiological Sciences. Elmsford NY: Pergamon Press.
Liley, D. T., Alexander, D. M., Wright, J. J., et Aldous, M. D. (1999). Alpha rhythm emerges from large-scale networks of realistically coupled multicompartmental model cortical neurons. Network, 10, 79-92.
Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. Trends Neurosci, 20, 38-43.
Liu, Z. (1996). Viewpoint dependency in object representation and recognition. Spat Vis, 9, 491-521.
Livingstone, M. S., et Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J Neurosci, 7, 3416-3468.
Livingstone, M. S., Freeman, D. C., et Hubel, D. H. (1996). Visual responses in V1 of freely viewing monkeys. Cold Spring Harb Symp Quant Biol, 61, 27-37.
Livingstone, M. S., et Hubel, D. H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. Science, 240, 740-749.
Logothetis, N. K. (1998). Object vision and visual awareness. Curr Opin Neurobiol, 8(4), 536-544.
Logothetis, N. K., et Sheinberg, D. L. (1996). Visual object recognition. Annu Rev Neurosci, 19, 577-621.
Logothetis, N. K., Pauls, J., et Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Curr Biol, 5(5), 552-563.
Lorenceau, J. (1987). Recovery from contrast adaptation: effects of spatial and temporal frequency. Vision Res, 27(12), 2185-2191.
Lorenceau, J. (1996). Motion integration with dot patterns: effects of motion noise and structural information. Vision Res, 36(21), 3415-3427.
Lorenceau, J., Chavane, F., Georges, S., et Frégnac, Y. (1998). Spatiotemporal dynamics of contour integration: psychophysics and physiological correlates. Paper presented at the 21th European Conference on Visual Perception, Oxford (UK).
Maass, W., et Natschläger, T. (1997). Networks of spiking neurons can emulate arbitrary Hopfield nets in temporal coding. Network: Computation in Neural Systems, 8, 355-372.
MacEvoy, S. P., Kim, W., et Paradiso, M. A. (1998). Integration of surface information in primary visual cortex. Nat Neurosci, 1(7), 616-620.
MacLeod, K., et Laurent, G. (1996). Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. Science, 274, 976-979.
Madey, S. F., et Gilovich, T. (1993). Effect of temporal focus on the recall of expectancy-consistent and expectancy-inconsistent information. J Pers Soc Psychol, 65(3), 458-468.
Maex, R., et Orban, G. A. (1996). Model circuit of spiking neurons generating directional selectivity in simple cells. J Neurophysiol, 75, 1515-1545.

Mainen, Z. F., et Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. Science, 268, 1503-1506.
Markram, H., et Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature, 382, 807-810.
Markram, H., Lubke, J., Frotscher, M., et Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science, 275, 213-215.
Marr, D. (1982). Vision. San Francisco: Freeman, Cooper et Co.
Martin, A., Wiggs, C. L., Ungerleider, L. G., et Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature, 379, 649-652.
Matt, J., Leuthold, H., et Sommer, W. (1992). Differential effects of voluntary expectancies on reaction times and event-related potentials: evidence for automatic and controlled expectancies. J Exp Psychol: Learn Mem Cogn, 18(4), 810-822.
Maunsell, J. H., et Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. J Neurophysiol, 68(4), 1332-1344.
Maunsell, J. H., Ghose, G. M., Assad, J. A., McAdams, C. J., Boudreau, C. E., et Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. Vis Neurosci, 16(1), 1-14.
McCormick, D. A., Connors, B. W., Lighthall, J. W., et Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J Neurophysiol, 54, 782-806.
McKeefry, D. J., et Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. Brain, 120, 2229-2242.
McKeefry, D., et Zeki, S. (1998). Activation of human area V4 in a delayed match-to-sample colour task. Paper presented at the ECVP, Oxford, UK.
McKoon, G. (1981). The representation of pictures in memory. J Exp Psychol: Hum Learn, 7, 216-221.
Mechler, F., Victor, J. D., Purpura, K. P., et Shapley, R. (1998). Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli. J Neurosci, 18(16), 6583-6598.
Mel, B. W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. Neural Comput, 9(4), 777-804.
Mel, B. W., Ruderman, D. L., et Archie, K. A. (1998). Translation-invariant orientation tuning in visual "complex" cells could derive from intradendritic computations. J Neurosci, 18(11), 4325-4334.
Menon, R. S., Ogawa, S., Strupp, J. P., et Ugurbil, K. (1997). Ocular dominance in human V1 demonstrated by functional magnetic resonance imaging. J Neurophysiol, 77(5), 2780-2787.
Miltner, W., Braun, C., Johnson, R., Jr., Simpson, G. V., et Ruchkin, D. S. (1994). A test of brain electrical source analysis (BESA): a simulation study. Electroencephalogr Clin Neurophysiol, 91(4), 295-310.
Missal, M., Vogels, R., et Orban, G. A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. Cereb Cortex, 7(8), 758-767.
Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision. J Exp Biol, 146, 21-38.
Monroy, M. (1996). Scènes, mythes et logiques. Paris: ESF - sciences - psychologie - psychanalyse.
Montaron, M. F., et Fabre-Thorpe, M. (1996). Effect of lesioning the nucleus accumbens on attentive preparation and performance of a reaching movement in the cat. Behav Brain Res, 79(1-2), 31-40.
Moore, C. J., et Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. Brain, 122, 943-962.
Mozer, M. C., Colagrosso, M., et Huber, D. E. (2000). Temporal Dynamics of Information Transmission in Neural System. Paper presented at the Conference on Neural Systems, Boston, MA.
Nagy, A. L., et Kamholz, D. W. (1995). Luminance discrimination, color contrast, and multiple mechanisms. Vision Res, 35, 2147-2155.
Nakamura, K. (1998). Neural processing in the subsecond time range in the temporal cortex [letter]. Neural Comput, 10, 567-595.
Nelson, S., Toth, L., Sheth, B., et Sur, M. (1994). Orientation selectivity of cortical neurons during intracellular blockade of inhibition. Science, 265, 774-777.
Neuenschwander, S., Martinerie, J., Renault, B., et Varela, F. J. (1993). A dynamical analysis of oscillatory responses in the optic tectum. Brain Res Cogn Brain Res, 1, 175-181.
Nobre, A. C., Allison, T., et McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. Nature, 372, 260-263.
Nowak, L. G., et Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rocklund, et A. Peters (Eds.), Extrastriate cortex in primates (pp. 205-241). New-York: Plenum Press.
Nowak, L. G., Munk, M. H. J., Girard, P., et Bullier, J. (1995). Visual lantencies in areas V1 and V2 of the macaque monkey. Vis Neurosci, 12, 371-384.

O'Brien, G., et Opie, J. (1999). A Connectionist Theory of Phenomenal Experience. Behav Brain Sci, 22, 127-148.
O'Craven, K. M., Downing, P. E., et Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. Nature, 401, 584-587.
O'Regan, J. K., Rensink, R. A., et Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. Nature, 398, 34.
O'Regan, J. K., et Noë, A. (2000). A sensorimotor account of vision and visual consciousness. Behav Brain Sci, submitted.
Ochsner, K. N., Chiu, C. Y., et Schacter, D. L. (1994). Varieties of priming. Curr Opin Neurobiol, 4, 189-194.
Okatan, M., et Grossberg, S. (2000). Frequency-Dependent Synaptic Potentiation, Depression, and Spike Timing Induced by Hebbian Pairing in Cortical Pyramidal Neurons. Neural Networks, submitted.
Olshausen, B. A., Anderson, C. H., et Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J Neurosci, 13(11), 4700-4719.
Olshausen, B. A., Anderson, C. H., et Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size- and position- invariant object representations. J Comput Neurosci, 2(1), 45-62.
Opara, R., et Worgotter, F. (1996). Using visual latencies to improve image segmentation. Neural Comput, 8(7), 1493-1520.
Oram, M. W., et Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. J Neurophysiol, 68(1), 70-84.
Oram, M. W., et Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J Neurophysiol, 76(1), 109-129.
Oram, M. W., Wiener, M. C., Lestienne, R., et Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. J Neurophysiol, 81, 3021-3033.
Orlov, T., Yakovlev, V., Hochstein, S., et Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. Nature, 404, 77-80.
Ostergaard, A. L., et Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. J Exp Psychol: Hum Percept Perform, 11, 579-587.
Paquier, W., Delorme, A., et Thorpe, S. (2000). Motion processing using one spike per neuron. Paper presented at the Ninth Annual Computational Neuroscience Meeting, Brugge, Belgium.
Palmer, S. E. (1999). Vision Science. Cambridge, MA: MIT Press.
Palmer, S. E., Rosch, E., et Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long et A. Baddeley (Eds.), Attention and Performances IX. Hillsdale, NJ: Lawrence Erlbaum.
Parker, D. M., Lishman, J. R., et Hughes, J. (1992). Temporal integration of spatially filtered visual images. Perception, 21(2), 147-160.
Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., et Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp Brain Res, 86(1), 159-173.
Perrett, D. I., Rolls, E. T., et Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res, 47(3), 329-342.
Perrinet, L. (1999). Apprentissage non supervisé d'un réseau asynchrone de neurones. Mémoire de DEA, Cerco, Toulouse.
Pinsky, P. F., et Rinzel, J. (1994). Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons. J Comput Neurosci, 1(1-2), 39-60.
Polich, J., et Bondurant, T. (1997). P300 sequence effects, probability, and interstimulus interval. Physiol Behav, 61(6), 843-849.
Posner, M. I., et Keele, S. W. (1970). Retention of abstract idea. J Exp Psychol, 83, 304-308.
Potter, M. C. (1999). Understanding Sentences and Scenes: The Role of Conceptual Short-Term Memory. In V. Coltheart (Ed.), Fleeting Memories. Cambridge, MA: MIT Press.
Pratarelli, M. E. (1994). Semantic processing of pictures and spoken words: evidence from event-related brain potentials. Brain Cogn, 24(1), 137-157.
Price, C. J., et Humphreys, G. W. (1989). The effect of surface detail on object categorization and naming. Quarterly J Exp Psychol, 41A, 797-828.
Propatias, A. D., Vanier, M., et Bower, J. M. (1998). Simulating Large Networks of Neurons. In C. Koch et I. Segev (Eds.), Methods in Neuronal Modeling. Cambridge, MA: MIT Press.
Puce, A., Allison, T., et McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. Cereb Cortex, 9(5), 445-458.
Reich, D. S., Victor, J. D., et Knight, B. W. (1998). The power ratio and the interval map: spiking models and extracellular recordings. J Neurosci, 18(23), 10090-10104.
Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., et Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. J Neurophysiol, 77, 2836-2841.

Reinagel, P., et Zador, A. M. (1999). Natural scene statistics at the centre of gaze. Network, 10(4), 341-350.
Reinagel, P., Godwin, D., Sherman, S. M., et Koch, C. (1999). Encoding of visual information by LGN bursts. J Neurophysiol, 81, 2558-2569.
Rensink, R. A., O'Regan, J. K., et Clarck, J. J. (1997). To see or not to see: The need of attention to perceive changes in scenes. Physiol sci, 8, 368-373.
Reynolds, J. H., Chelazzi, L., et Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci, 19(5), 1736-1753.
Reynolds, J. H., Pasternak, T., et Desimone, R. (2000). Attention increases sensitivity of V4 neurons. Neuron, 26(3), 703-714.
Richmond, B. J., et Optican, L. M. (1990). Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. J Neurophysiol, 64(2), 370-380.
Riehle, A., Grun, S., Diesmann, M., et Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. Science, 278, 1950-1953.
Riesenhuber, M., et Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat Neurosci, 2(11), 1019-1025.
Roberts, W. A., et Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys and people. J Exp Psychol, 14(3), 247-260.
Rodemann, T., et Körner, E. (2000). Two separate processing streams in a cortical-type architecture. Paper presented at the Ninth Annual Computational Neuroscience Meeting, Brugge, Belgium.
Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., et Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. Nature, 397, 430-433.
Roland, P. E., et Gulyas, B. (1995). Visual memory, visual imagery, and visual recognition of large field patterns by the human brain: functional anatomy by positron emission tomography. Cereb Cortex, 5(1), 79-93.
Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond B Biol Sci, 335(1273), 11-20.
Rosenblatt, F. (1961). Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Washington, DC: Spartan Books.
Rousselet, G. (2000). De la reconnaissance à la catégorisation visuelles: contraintes temporelles sur les décisions perceptives de l'humain. Mémoire de DEA, CERCO.
Ruderman, D. L. (1997). Origins of scaling in natural images. Vision Res, 37, 3385-3398.
Ruf, B., et Schmitt, M. (1997). Hebbian learning in networks of spiking neurons using temporal coding. In J. Mira, R. Moreno-Diaz, et J. Cabestany (Eds.), Biological and artificial computation: From neuroscience to technology (Vol. 1240 of Lecture Notes in Computer Science, pp. 380-389). Berlin: Springer.
Rugg, M. D., Cox, C. J., Doyle, M. C., et Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. Neuropsychologia, 33(4), 471-484.
Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C., et Brammer, M. J. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. Proc Natl Acad Sci U S A, 94(17), 9406-9411.
Salzman, C. D., Britten, K. H., et Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. Nature, 346, 174-177.
Samaria, F. S., et Harter, A. C. (1994). Parametrisation of a Stochastic Model for Human Face Identification. Paper presented at the Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision, Sarasota (Florida).
Samsonovich, A., et McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. J Neurosci, 17, 5900-5920.
Sands, S. F., Lincoln, C. E., et Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. J Exp Psychol Gen, 111, 369-389.
Schiller, P. H., Finlay, B. L., et Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. V. Multivariate statistical analyses and models. J Neurophysiol, 39(6), 1362-1374.
Schyns, P. G. (1999). Diagnostic recognition: task constraints, object information and their interactions. In M. J. Tarr et H. H. Bülthoff (Eds.), Object Recognition in Man, Monkey and Machine. Cambridge MA: MIT Press.
Schyns, P. G., et Oliva, A. (1994). From blops to boundary edges: evidence for time and spatial scale dependent scene recognition. Psychol Sci, 5, 195-200.
Sekuler, A. B., Lee, J. A., et Shettleworth, S. J. (1996). Pigeons do not complete partly occluded figures. Perception, 25, 1109-1120.
Seltzer, B., et Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res, 149(1), 1-24.
Senn, W., Tsodyks, M., et Markramn, H. (1997). An algorithm for synaptic modification based on precise timing of pre and postsynaptic action potentials. Lecture Notes in Computational Science, 97, 121-126.

Seriès, P., Lorenceau, J., Georges, S., Alais, D., et Frégnac, Y. (1999). Spatiotemporal dynamics of long range interaction in a model of primary visual cortex. Paper presented at the 22th European Conference on Visual Perception, Oxford (UK).
Seriès, P., Lorenceau, J., Georges, S., et Frégnac, Y. (2000). Spatial and temporal contextual modulation in a detailed model of V1. Paper presented at the 9th Annual Computational Neuroscience Meeting, Brugge (Belgium).
Shadlen, M. N., et Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci, 18, 3870-3896.
Sheinberg, D. L., et Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci U S A, 94(7), 3408-3413.
Sillito, A. M. (1975). The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. J Physiol (Lond), 250, 305-329.
Silva, L. R., Amitai, Y., et Connors, B. W. (1991). Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science, 251, 432-435.
Simondon, G. (1969). Du mode d'existence des objets techniques, (pp. 103). Paris: Aubier Montaigne.
Sommer, W., Leuthold, H., et Soetens, E. (1999). Covert signs of expectancy in serial reaction time tasks revealed by event-related potentials. Percept Psychophys, 61(2), 342-353.
Sompolinsky, H., et Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. Curr Opin Neurobiol, 7, 514-522.
Sperling, G., et Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. Attention Perf, 8, 347-360.
Stemmler, M., Usher, M., et Niebur, E. (1995). Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. Science, 269, 1877-1880.
Storck, J., Jaekel, F., et Deco, G. (2000). Temporal clustering with networks of spiking neurons and dynamical synapses. Paper presented at the Ninth Annual Computational Neuroscience Meeting.
Suarez, H., Koch, C., et Douglas, R. (1995). Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. J Neurosci, 15, 6700-6719.
Subramaniam, S., Biederman, I., et Madigan, S. A. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. Visual Cognition, 7, 511-535.
Sugase, Y., Yamane, S., Ueno, S., et Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. Nature, 400, 869-873.
Syrkin, G., et Gur, M. (1997). Colour and luminance interact to improve pattern recognition. Perception, 26(2), 127-140.
Talbot, S. A., et Marshall, W. H. (1941). Physiological studies on neurophysiological mechanisms of visual localization and discrimination. Am J Opthalmol, 24, 1225-1264.
Tallon-Baudry, C., et Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci, 3(4), 151-162.
Tanaka, K. (1996). Inferotemporal Cortex and Object Vision. Annu Rev Neurosci, 19, 109-139.
Tanaka, M. (1995). Object sorting in chimpanzees (Pan troglodytes): classification based on physical identity, complementarity, and familiarity. J Comp Psychol, 109(2), 151-161.
Theunissen, F., et Miller, J. P. (1995). Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci, 2, 149-162.
Thorpe, S. J. (1990). Spike arrival times : a highly efficient coding scheme for neural networks. In R. Eckmiller, G. Hartman, et G. Hauske (Eds.), Parallel processing in neural systems. Amsterdam: Elsevier.
Thorpe, S. J. (2000). Pattern recognition in the visual system and the nature of neural coding. In A. Carbone, M. Gromov, et P. Prusinkiewicz (Eds.), Pattern Formation in Biology, Vision and Dynamics (pp. 382-391). Singapour: World Scientific.
Thorpe, S. J., et Gautrais, J. (1998). A new coding scheme for rapid processing in neural networks. In J. Bower (Ed.), Computational Neuroscience: Trends in Research (pp. 113-118). New York: Plenum Press.
Thorpe, S. J., et Imbert, M. (1989). Biological constraints on connectionist models. In R. R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, et L. Steels (Eds.), Connectionism in Perspective (pp. 63-92). Amsterdam: Elsevier.
Thorpe, S. J., Fize, D., et Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520-522.
Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., et Büllthoff, H. H. (2000). Spotting animals out of the corner of the eye. Vision Res, submitted.
Tipper, S. P., et Behrmann, M. (1996). Object-centered not scene-based visual neglect. J Exp Psychol: Hum Percept Perform, 22, 1261-1278.
Tootell, R. B., Dale, A. M., Sereno, M. I., et Malach, R. (1996). New images from human visual cortex. Trends Neurosci, 19(11), 481-489.

Tovee, M. J., et Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. Visual Cognition, 2, 35-59.
Tovee, M. J., Rolls, E. T., et Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J Neurophysiol, 72(3), 1049-1060.
Troscianko, T. (1994). Contribution of colour to the motion aftereffect and motion perception. Perception, 23, 1221-1231.
Trotter, Y., et Celebrini, S. (1999). Gaze direction controls response gain in primary visual-cortex neurons. Nature, 398, 239-242.
Tsodyks, M., Pawelzik, K., et Markram, H. (1998). Neural networks with dynamic synapses. Neural Comput, 10(4), 821-835.
Tulving, E., et Schacter, D. L. (1990). Priming and human memory systems. Science, 247, 301-306.
Tulving, E., Markowitsch, H. J., Craik, F. E., Habib, R., et Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cereb Cortex, 6, 71-79.
Uexküll. (1984). Mondes animaux et mondes humains. Paris: Denoel.
Ullman, S. (1995). Sequence Seeking and Counter Streams: A Computational Model for Biderectional Information Flow in the Visual Cortex. Cereb Cortex, 5(1), 1-11.
Ungerleider, L. G., Courtney, S. M., et Haxby, J. V. (1998). A neural system for human visual working memory. Proc Natl Acad Sci U S A, 95(3), 883-890.
Van Rullen, R., Gautrais, J., Delorme, A., et Thorpe, S. (1998). Face processing using one spike per neurone. Biosystems, 48(1-3), 229-239.
Van Vreeswijk, C., Abbott, L. F., et Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. J Comput Neurosci, 1, 313-321.
VanRullen, R., Delorme, A., et Thorpe, S. (2000c). Object recognition using spiking neurons II: Spatial attention explained by temporal precedence of information. Paper presented at the Neural binding of space and time symposium. Leipzig, Germany.
VanRullen, R., et Thorpe, S. J. (2000a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual categories. Perception, submitted.
VanRullen, R., et Thorpe, S. J. (2000b). The time course of visual processing: from early perception to decision-making. J Cog Neurosci, in press.
Varela, F. J. (1989). Autonomie et connaissances. Paris: Le Seuil.
Varela, F. J. (1996). Neurophenomenology : A Methodological Remedy for the Hard Problem. In J. Shear (Ed.), Journal of Consciousness Studies (Vol. Special Issues on the Hard Problems).
Veneau, E. (1996). Codage impulsionnel par rang et apprentissage. Mémoire de DEA, Cerco, Toulouse.
Verghese, P., et Nakayama, K. (1994). Stimulus discriminability in visual search. Vision Res, 34, 2453-2467.
Vetter, T., Hurlbert, A., et Poggio, T. (1995). View-based models of 3D object recognition: invariance to imaging transformations. Cereb Cortex, 5(3), 261-269.
Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: parietal guiding the temporal. Brain Res Rev, 30(1), 66-76.
Vidyasagar, T. R., Pei, X., et Volgushev, M. (1996). Multiple mechanisms underlying the orientation selectivity of visual cortical neurones. Trends Neurosci, 19, 272-277.
Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur J Neurosci, 11(4), 1223-1238.
Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci, 11(4), 1239-1255.
Vogels, R., et Orban, G. A. (1996). Coding of stimulus invariances by inferior temporal neurons. Prog Brain Res, 112, 195-211.
Volgushev, M., Pei, X., Vidyasagar, T. R., et Creutzfeldt, O. D. (1993). Excitation and inhibition in orientation selectivity of cat visual cortex neurons revealed by whole-cell recordings in vivo. Vis Neurosci, 10, 1151-1155.
Volgushev, M., Vidyasagar, T. R., et Pei, X. (1995). Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. Vis Neurosci, 12, 621-628.
von der Heydt, R., et Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. J Neurosci, 9(5), 1731-1748.
von der Malsburg, C. (1995). Binding in models of perception and brain function. Curr Opin Neurobiol, 5(4), 520-526.
Wachsmuth, E., Oram, M. W., et Perrett, D. I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cereb Cortex, 4(5), 509-522.
Wallis, G. (1994). Neural Mechanisms Underlying Processing in the Visual Areas of the Occipital and Temporal Lobes. Thèse, Oxford University, Oxford.

Wallis, G., et Rolls, E. T. (1997). Invariant face and object recognition in the visual system. Prog Neurobiol, 51(2), 167-194.
Wang, G., Tanaka, K., et Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. Science, 272, 1665-1668.
Watanabe, S. (1997). Visual discrimination of real objects and pictures in pigeons. Anim Learn Behav, 25, 185-192.
Watzlawick, P., Beavin, J. H., et Jackson, D. D. (1972). Une logique de la communication. Paris: Seuil.
Weber, P. B., et Ojemann, G. A. (1995). Neuronal recordings in human lateral temporal lobe during verbal paired associate learning. Neuroreport, 6(4), 685-689.
Wiener, N. (1961). Cybernetics or Control and Communication in the Animal and the Machine. Cambridge, MA: MIT Press.
Wiesel, T. N., et Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. J Neurophysiol, 29, 1115-1156.
Wiskott, L., et von der Maalsburg, C. (1995). Recognizing faces by dynamic link matching. Paper presented at the ICANN '95, Paris, France.
Wurm, L. H., Legge, G. E., Isenberg, L. M., et Luebker, A. (1993). Color Improves Object Recognition in Normal and Low Vision. J Exp Psychol, 19, 899-911.
Würtz, R. P. (1997). Neuronal theories and technical systems for face recognition. Paper presented at the ESANN, Brussels.
Yoshioka, T., Dow, B. M., et Vautin, R. G. (1996). Neuronal mechanisms of color categorization in areas V1, V2 and V4 of macaque monkey visual cortex. Behav Brain Res, 76, 51-70.
Zeki, S., et Shipp, S. (1988). The functional logic of cortical connections. Nature, 335, 311-317.