Une première vague de potentiels d'action, une première vague idée de la scène visuelle - rôle de l'asynchronie dans le traitement rapide de l'information visuelle.

Rufin Vanrullen

Université Paul Sabatier

Thèse

présentée pour l'obtention du doctorat de

Sciences Cognitives

UNE PREMIERE VAGUE DE POTENTIELS D'ACTION,

UNE PREMIERE VAGUE IDEE DE LA SCENE VISUELLE.

Rôle de l'asynchronie dans le traitement rapide de l'information visuelle.

Par

Rufin VANRULLEN

M.	Michel Imbert	Président
M.	Jeanny Herault	Rapporteur
M.	Francisco Varela	Rapporteur
M.	Bernard Doyon	
M.	Andrew James	
M.	Simon Thorpe	Directeur de Thèse

Centre de Recherche Cerveau et Cognition UMR 5549 CNRS-UPS

Faculté de Médecine de Rangueil, 133 Route de Narbonne, 31062 Toulouse

TABLE DES MATIERES

I. INTRODUCTION *1*

1. Anatomie du système visuel **3**
1.1 Architecture hiérarchique 3
1.2 Différentes voies, différentes propriétés 8
1.2.1 Voie ventrale: traitement de la forme et identification des objets 9
1.2.2 Voie dorsale: traitement de l'information spatiale et de mouvement 10
1.3 Connexions cortico-corticales, feed-back 11

2. Les premiers modèles neuronaux du traitement visuel **13**
2.1 Description structurelle ou représentation par vues ? 14
2.2 Implémentation neuronale 15

3. Décours temporel d'activation des aires visuelles, catégorisation visuelle rapide (150 ms). **18**
3.1 Electrophysiologie chez le singe 18
3.2 Psychophysique chez l'homme 20
3.3 Implications théoriques 22

4. Les modèles de "Spiking Neurons" **24**

5. Catégorisation visuelle en 50 ms? Implications pour les modèles. **28**
5.1 Résultats expérimentaux 29
5.2 Implications théoriques fondamentales 31
5.3 Catégorisation visuelle: 50 ou 150 ms? 32
5.3.1 Seeck et al.: erreur conceptuelle de 1er ordre... 32
5.3.2 Mêmes paradigmes, résultats contradictoires: erreur conceptuelle de 2ème ordre? 33

6. Résumé, formulation des objectifs **33**

II. LA DUREE DU TRAITEMENT VISUEL *34*

1. Catégorie "animal": un cas particulier? **34**
1.1 Le problème 34
1.2 Article n° 1 : VanRullen & Thorpe, 2000. Is it a bird, is it a plane? Ultra-rapid visual categorization of natural and artifactual categories. *Perception.* 34
1.3 Résumé 52

2. Séparation des mécanismes visuels de bas- et haut-niveau **52**
2.1 Un nouveau paradigme 53
2.2 Article n° 2 : VanRullen & Thorpe, 2000. The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience.* 53
2.3 Activité précoce: interprétation 62
2.3.1 La métaphore de l'ordinateur 62
2.3.2. La sortie de la rétine diffère selon la catégorie, mais la catégorisation ne s'effectue pas dans la rétine 63
2.3.3 Le stimulus diffère selon la catégorie, mais la catégorisation ne s'effectue pas avant la rétine! 65
2.4 Activité de haut-niveau: interprétation 67
2.4.1 Relation entre l'activité électrique et la réaction motrice 67
2.4.2 Décision ou attention? 71
2.4.3 Conscience visuelle 72
2.5 Deux mécanismes distincts : bases neuronales 73
2.5.1 Méthode 74
2.5.2 Activité précoce : bases neuronales 74
2.5.3 Activité à 150 ms : bases neuronales 77

3 Synthèse: la durée du traitement visuel **79**

III. THEORIE: CONTRAINTES, MODELES, SIMULATIONS *82*

1. Contraintes biologiques: **82**
1.1 Traitement feed-forward 82
1.2 Un spike par neurone 83

2. Coder l'information avec un spike par neurone **84**
2.1 Article 3 : VanRullen & Thorpe, 2000. Rate coding vs. Temporal Order Coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*. 85
2.2 Complements 112
2.3 Le décodage de l'ordre 114
2.4 Synchronie et oscillations: un épiphénomène? 115
2.4.1 Synchronie 117
2.4.2 Oscillations 119
2.4.3 Codage d'information par synchronie 120
2.4.4 La synchronie seule ne peut encoder l'information spatiale dans la rétine 122
2.4.5 Une succession de vagues de potentiels d'action explique les observations expérimentales de synchronie et oscillations 124
2.4.6 Codage par la synchronie, ou codage par l'asynchronie? 127
2.5 Synthèse 129

3 Transmettre l'information en feed-forward **129**
3.1 Que peut-on faire de façon “feed-forward”? 129
3.2 Article 4 (en préparation) : VanRullen, Delorme & Thorpe. Feed-forward contour integration based on asynchronous spike propagation. *NeuroComputing* 130
3.3 Attention: précédence temporelle pour les régions d'intérêt 136
3.3.1 L'hypothèse: Article 5. VanRullen & Thorpe, 1999. Spatial attention in asynchronous neural networks. *NeuroComputing* 136
3.3.2 Les arguments 145
3.3.3 Résumé 164

4. Une solution pour le traitement visuel rapide? **164**
4.1 Article 6 : VanRullen, Gautrais, Delorme & Thorpe, 1998. Face processing using one spike per neurone. BioSystems. 164
4.2 Détection de visages à plusieurs tailles 176
4.3 Une représentation de haut-niveau de l'entrée visuelle 181

IV. SYNTHESE ET PERSPECTIVES *183*

1. Une théorie du traitement visuel rapide **185**

2. Questions en suspens **188**

3. Conclusion **191**

I. INTRODUCTION

Le monde visuel est probablement la source d'informations la plus riche de notre expérience sensorielle. Une large majorité des tâches que nous réalisons quotidiennement se trouvent grandement simplifiées par notre capacité à percevoir des objets distants, et c'est certainement pourquoi l'évolution a favorisé le développement chez bon nombre d'espèces animales d'un système neuronal de traitement d'information entièrement dédié à cette finalité.

Alors que des disciplines comme la philosophie ou la psychologie se sont intéressées à cette capacité de voir depuis plusieurs millénaires, ce n'est qu'au début du XXème siècle, avec la découverte des neurones par Santiago Ramon y Cajal (1852-1934) et Camillo Golgi (1843-1926), couronnée par un Prix Nobel en 1906, que la discipline des Neurosciences a pu voir le jour. Cette discipline à part entière constitue aujourd'hui une source majeure de nos connaissances sur le fonctionnement du système visuel.

Peu avant la seconde Guerre Mondiale se sont développées les techniques d'enregistrement sur les fibres du nerf optique, puis dans les années cinquante au niveau cellulaire. Ceci a permis de comprendre le mécanisme de traitement neuronal de l'information visuelle, et plus tard d'introduire la notion de champ récepteur, tout d'abord au niveau de la rétine (Hartline, 1938; 1940a; 1940b; Kuffler 1953; Lettvin, et al. 1959; Rodieck 1965; Barlow, Hill et Levick, 1964; Barlow et Levick, 1965), puis dans le cortex visuel primaire (Hubel et Wiesel, 1959; 1962; 1968). Ce n'est sans doute pas un hasard si ces découvertes, qui furent permises par des avancées techniques importantes, telles les micro-électrodes, se sont produites à un moment de l'Histoire scientifique qui voyait également l'avènement d'une nouvelle ère technologique: la révolution informatique.

Des premiers "super-calculateurs" à l'invention des transistors, qui permit une miniaturisation progressive des ordinateurs, le concept d'Intelligence Artificielle a servi de moteur sous-jacent au développement de l'Informatique. La puissance computationnelle des ordinateurs laissait envisager la possibilité de reproduire, égaler ou dépasser le cerveau humain. Si de nombreux scientifiques aux débuts de l'ère cybernétique se sont tournés vers une approche symbolique de l'Intelligence Artificielle, une autre approche tout aussi importante, guidée par les travaux de Warren McCulloch et Walter Pitts (1943) et dénommée "connexionniste", a cherché à s'inspirer des principales propriétés connues du cerveau : une capacité de calcul globale distribuée sur un large ensemble d'unités ou de processeurs à capacité relativement limitée : les neurones.

Ainsi l'avancée scientifique dans le domaine des neurosciences s'est elle toujours accompagnée d'une composante théorique, computationnelle, que l'on désigne parfois sous le terme "modélisation" lorsqu'elle permet de reproduire ou d'interpréter des résultats expérimentaux, aussi bien électrophysiologiques que psychologiques ou cliniques. Ce qui est vrai dans un sens l'est aussi dans l'autre : les progrès des Neurosciences ont alimenté les recherches sur les modèles connexionnistes, mais les résultats théoriques obtenus en modélisation ont bien souvent influencé ou guidé les études neurophysiologiques, et modelé la conception aujourd'hui acceptée du traitement neuronal de l'information. L'état de l'art en Neurosciences de la Vision est donc le résultat d'une avancée parallèle, voire conjointe, entre recherche expérimentale et théorique.

Nous ne remonterons pas dans cette introduction aux tout-premiers modèles neuronaux, décrits par McCulloch et Pitts (1943), Rosenblatt (1958) ou McClelland et Rumelhart (1986), et basés sur une conception relativement abstraite des neurones, de leur fonctionnement et de l'architecture des systèmes neuronaux. Nous nous intéresserons plutôt à tracer, en parallèle, d'un côté les découvertes majeures et l'état de l'art en Neurosciences de la Vision, et de l'autre, les différents modèles qui en ont découlé, en gardant toujours à l'esprit que l'influence de la recherche expérimentale sur la théorie a également une réciproque, et que certains de ces modèles ont en quelque sorte "façonné" une grande partie de l'approche expérimentale.

Dans un premier temps nous décrirons l'anatomie fonctionnelle du système visuel telle qu'elle a été dessinée des années 50 à nos jours, avec une architecture hiérarchique comprenant des étapes successives où les traitements effectués sont de plus en plus complexes; une séparation de l'information en 2 voies aux propriétés complémentaires, l'une portant l'information sur la forme et l'identité des objets, l'autre sur les propriétés spatiales et de mouvement de la scène visuelle; une connectivité complexe, bien souvent réciproque, entre ces différentes aires visuelles. Nous évoquerons les modèles classiques de la vision, largement inspirés de cette anatomie fonctionnelle, et tâcherons de définir les hypothèses communes de ces modèles quant au traitement neuronal de l'information et en particulier quant au type de **codage neuronal** utilisé.

Nous présenterons ensuite un ensemble de résultats plus récents, aussi bien électrophysiologiques que psychophysiques, qui mettent en évidence le décours temporel d'activation des aires visuelles. Ces résultats ont remis en question les théories classiques citées plus haut, et les modélisateurs ont dû se pencher au cours de la dernière décennie sur des modèles prenant en compte une structure temporelle plus fine du traitement et transfert de l'information visuelle, afin de respecter ces nouvelles contraintes biologiques. En particulier, le fait que les neurones réels génèrent une réponse sous la forme de potentiels d'action (spikes) et non sous forme analogique, est un élément fondamental du traitement

d'information et du codage neuronal, qui ne peut être "capturé" par un ensemble d'équations différentielles utilisant le taux de décharge neuronal comme variable principale. Cette propriété bien établie des neurones réels reste pourtant largement ignorée, et n'a été explorée de façon explicite que par de rares auteurs (e.g. Perkel, Gerstein et Moore, 1967a,b; Perkel et Bullock, 1968), et très épisodiquement, durant l'histoire des neurosciences. L'établissement des contraintes temporelles qui pèsent sur le système visuel des primates a engendré au cours de la dernière décennie un regain d'intérêt pour les potentiels d'action et leur pouvoir computationnel, que nous essaireons de décrire dans cette introduction.

Nous verrons enfin que ces nouveaux modèles de la vision et des mécanismes neuronaux en général se heurtent à leur tour à une série d'expériences publiées ces dernières années, dont les résultats tendraient à renforcer encore les contraintes temporelles mentionnées plus haut. Ainsi, avant de pouvoir s'orienter vers une théorie biologiquement plausible du traitement de l'information visuelle, il faudra déterminer avec précision quelles sont les contraintes, notamment temporelles, qui pèsent sur le système visuel des mammifères.

1. Anatomie du système visuel

1.1 Architecture hiérarchique

La rétine

Les premiers enregistrements électriques dans le système visuel ont été effectués dans la rétine, ou plus précisément dans le nerf optique, qui transmet la réponse des cellules ganglionnaires de la rétine vers le cerveau, via le corps genouillé latéral. Les neurones de la rétine ont tout d'abord été définis comme répondant à différents niveaux d'illumination (Hartline, 1938). Ensuite a été introduite la notion de champ récepteur d'un neurone (Hartline, 1940), correspondant à la zone du champ visuel à l'intérieur de laquelle une stimulation visuelle engendre une réponse du neurone. Notre connaissance des propriétés encodées dans la rétine s'est rapidement accrue, grâce à de nombreuses études chez la grenouille (e.g. Lettvin et al, 1959) et les mammifères (Kuffler, 1953; Barlow, Hill et Levick, 1964; Barlow et Levick, 1965; Rodieck, 1965).

Bien que cela ne suffise pas pour capturer toute la complexité des traitements rétiniens, les résultats d'enregistrements électrophysiologiques et d'études anatomiques dans la rétine peuvent être résumés comme suit:

- les cellules ganglionnaires sont divisées en 2 types selon qu'elles répondent préférentiellement à un incrément de lumière au centre de leur champ récepteur (ON-center cells) ou à un décrément (OFF-center cells). Dans les deux cas le pourtour de la cellule (appelé "surround") montre une sélectivité opposée ou "antagoniste". Ces neurones sont donc en fait sensibles à des changements de contraste.

- ces neurones sont de plus séparés en fonction du caractère soutenu (tonique) ou transient (phasique) de leur réponse: les premiers (les cellules beta) répondent de façon continue à la stimulation visuelle, les derniers (les cellules alpha) uniquement lors d'un brusque changement ("transient") de cette stimulation. Ces deux catégories de neurones sont aussi appelées respectivement X et Y cells chez le chat, ou P ("parvocellulaire") et M ("magnocellulaire") chez le primate. Les neurones M ont, en général, des latences de réponse plus courtes que les neurones P. Il existe une troisième catégorie, plus rare et encore mal connue (voir cependant Hendry et Reid, 2000), que nous laisserons ici de côté: les cellules W (ou gamma), correspondant à la voie "koniocellulaire".

- chacune de ces catégories de neurones peut avoir des champs récepteurs plus ou moins grands, c'est-à-dire qu'elle traite l'information de contraste (ou de contraste spatio-temporel) à plusieurs échelles spatiales. Cependant les cellules M ont, en moyenne, des champs récepteurs plus larges que les cellules P.

- enfin les neurones rétiniens peuvent être sélectifs à la couleur, c'est à dire encoder des contrastes de longueur d'onde entre leur centre et leur pourtour (par exemple "rouge" au centre et "vert" sur le pourtour: "red-green color-opponent" cell). La grande majorité des neurones sélectifs à la couleur est composée de cellules P.

Le noyau géniculé latéral

Au niveau du corps genouillé latéral (LGN), relais thalamique de l'information visuelle entre la rétine et le cortex, les différentes caractéristiques de réponses neuronales présentes dans la rétine (ON ou OFF-center, M ou P) sont conservées de manière séparée grâce à une organisation en différentes couches. Nous ne décrirons pas plus avant les propriétés fonctionnelles des neurones du LGN, qui restent encore mal connues. Notons simplement qu'une majorité des travaux en neurosciences visuelles considèrent le LGN comme un simple relais d'information, dans lequel pas ou peu de traitements sont effectués. De rares auteurs comme Atick et Dan (Atick, 1992; Dan et al. 1996) ont émis l'hypothèse que le rôle

du LGN (tout comme la rétine) était de recoder l'entrée visuelle afin de décorréler au maximum les réponses qui sont nécessairement corrélées (a priori), compte-tenu de la statistique des images naturelles (cette hypothèse est discutée par Field, 1994, et Olshausen et Field, 1996).

Le manque de données sur les propriétés des neurones du LGN est d'autant plus regrettable que l'on sait que près de 80% des entrées de ce noyau thalamique sont constituées par des connexions en retour (feed-back) du cortex visuel. Il est peu vraisemblable que le système visuel s'offre le "luxe" d'envoyer autant d'informations à une structure qui ne constitue qu'un "relais" du traitement neuronal. Une possibilité, avancée par Crick (1984) et Koch et Ullman (1985), est que ces projections feed-back extensives puissent servir à implémenter un mécanisme d'attention spatiale, en renforçant spécifiquement l'activité des neurones dont le champ récepteur se situe dans la "fenêtre attentionnelle". Cette hypothèse est soutenue par des observations expérimentales d'une suppression d'activité neuronale dans le LGN, autour de la région d'intérêt (Vanduffel, Tootell et Orban, 2000). Une autre possibilité, non exclusive, est que les connexions corticothalamiques pourraient modifier activement la transmission de l'information visuelle en fonction de la correspondance entre l'activité corticale courante, i.e. l'interprétation courante de la scène visuelle, et le signal rétinien afférent (Varela et Singer, 1987; Gove et al, 1995).

Le cortex visuel primaire (V1)

L'aire qui reçoit la plus grande partie des informations du LGN est dénommée le cortex visuel primaire ou V1. Grâce à un nombre toujours croissant de publications, et en particulier aux travaux pionniers de Hubel et Wiesel (1959, 1962, 1968), les propriétés et l'architecture des neurones de V1 sont aujourd'hui relativement bien connues.

Ces neurones corticaux présentent une sélectivité plus complexe que les neurones du LGN, qui sont activés par des contrastes de luminance. On distingue les cellules simples (qui effectuent une sommation linéaire de l'activité dans leur centre et leur pourtour), et les cellules complexes, (pour lesquelles la sommation est non-linéaire). La principale caractéristique des neurones de V1 est leur sélectivité à l'orientation, c'est-à-dire que ces neurones répondront préférentiellement à une barre ou un bord d'une certaine orientation (par exemple, horizontale) à l'intérieur de leur champ récepteur. Certains répondent uniquement à l'information en provenance de l'un des deux yeux, en particulier dans la couche IVc qui reçoit les afférences thalamiques, mais la plupart des neurones sont binoculaires, avec toutefois une dominance pour l'un ou l'autre des yeux. Ces neurones peuvent donc également être sensibles à une "disparité binoculaire", indice de profondeur calculé à partir de différences entre les informations provenant de chacun des yeux. Il a été récemment démontré que la réponse des neurones de V1 pouvait aussi être influencée par

la direction du regard (Trotter et Celebrini, 1999). Enfin, la majorité des cellules simples sont sélectives à la couleur.

L'organisation des neurones sur la surface du cortex est dite modulaire. Les neurones sont regroupés en fonction de leur orientation préférée (on parle de colonnes d'orientation), et également en fonction de la dominance oculaire (colonnes de dominance oculaire). On trouve entre ces colonnes des zones appelées "blobs", portant l'information de couleur et de surface.

Le traitement d'information effectué dans V1 est extrêmement complexe, et ne peut simplement se résumer en quelques lignes (pour une synthèse plus complète, voir par exemple Bruce et Green, 1990, Chapitre III). En effet, la représentation de l'information visuelle à la sortie de V1 servira de base à la plupart des mécanismes mis en jeu dans les autres aires du cortex visuel. Cette représentation se doit donc d'être complète, tout en minimisant les éventuelles redondances, qui constitueraient un coût non négligeable pour le cerveau. Notons cependant que le système visuel ne s'arrête pas à V1, même si de nombreux auteurs utilisent à tort le terme "cortex visuel" pour simplement désigner le cortex visuel "primaire".

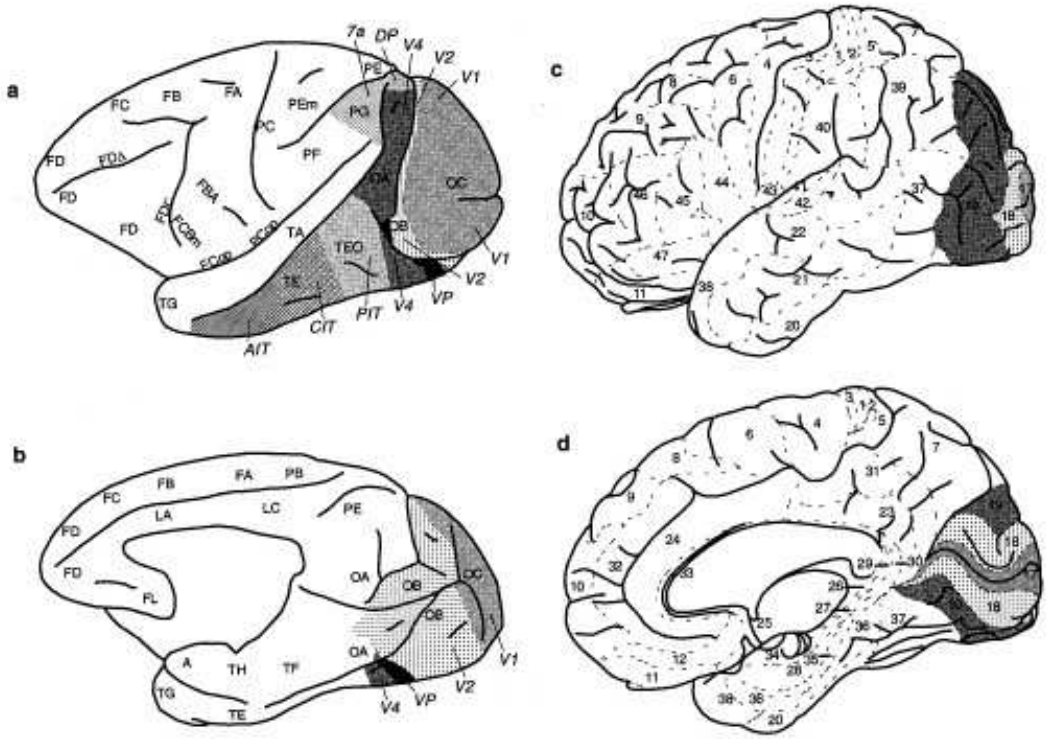


Figure 1. Les différentes aires visuelles chez le singe (a-b) et l'homme (c-d, classification de Brodmann). Vues de côté (a,c) et médiales (b,d). Reproduit d'après Logothetis et Sheinberg (1996).

Les aires extra-striées: V2 et au-delà

Les aires situées au-delà de V1 sont appelées aires visuelles extrastriées. Chez le singe, plus de 30 aires visuelles ont été mises en évidence (Zeki, 1969 ; Van Essen, 1979), et bientôt la moitié auront été identifiées chez l'homme, par des techniques d'imagerie comme le PET (Haxby et al, 1991) ou l'IRMf (Ungerleider et Haxby, 1994 ; Sereno et al. 1995 ; Tootel et al. 1996). Les aires visuelles du singe et de l'homme sont représentées schématiquement sur la figure 1.

Les neurones de l'aire V1 projettent majoritairement vers l'aire corticale V2, organisée en bandes (fines, épaisses et interbandes). Les neurones de V2 ont des champs récepteurs plus larges que ceux de V1, mais la plupart montrent une sélectivité à l'orientation similaire à celle observée dans V1 (Zeki, 1978). De là, l'information est distribuée dans les différentes aires, comme le montre la figure 2. Globalement, on observe une hiérarchie anatomique et fonctionnelle, avec des traitements de plus en plus complexes effectués dans les différentes aires au-delà de V1 (théorie de "feature detection" de Barlow, 1972; Barlow et al, 1972). Ces aires extrastriées s'organisent selon 2 voies principales, que nous allons explorer plus en détail.

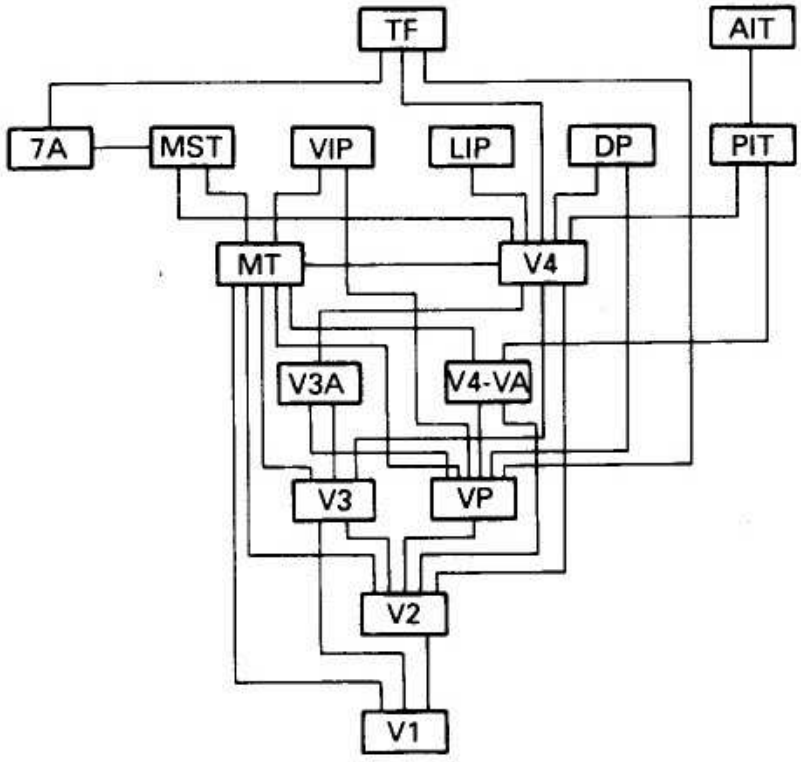


Figure 2. Distribution hiérarchique de l'information visuelle dans les différentes aires corticales. Reproduit d'après Maunsell et Newsome (1987).

1.2 Différentes voies, différentes propriétés

Outre les nombreuses études électrophysiologiques des cellules de V1, on a commencé à mieux connaître vers la fin des années 70 les autres aires du cortex visuel, et les propriétés des neurones de ces aires (Zeki, 1978; Van Essen, 1979). Un certain nombre d'observations faites notamment à la suite de lésions du cortex visuel ont mené Leslie Ungerleider et Mortimer Mishkin (1982; Mishkin et al, 1983) à la conclusion que ces aires extrastriées pouvaient être divisées en 2 voies principales, relativement séparées, l'une portant l'information concernant la forme et l'identité des objets (la voie du "what", ventrale ou encore temporale), et l'autre portant l'information sur la structure spatiale et de mouvement de la scène visuelle (la voie du "where", dorsale ou encore pariétale). Cette séparation est également présente chez l'homme (Haxby et al. 1991; pour revue, voir Ungerleider et Haxby, 1994). La figure 3 illustre très schématiquement cette organisation.

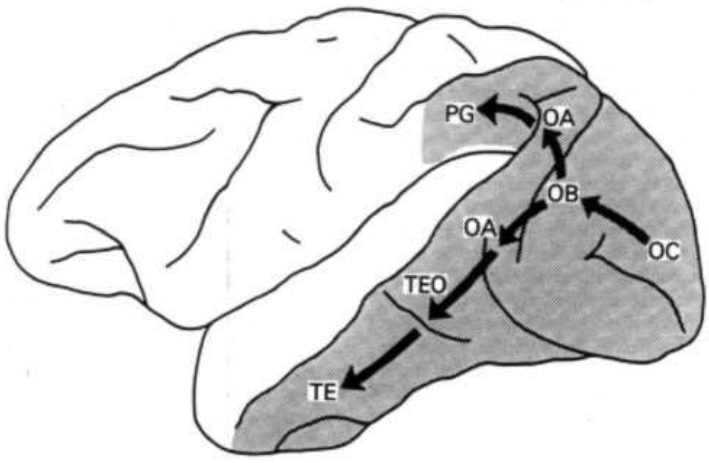


Figure 3. La voie ventrale, reliant le cortex visuel occipital au cortex inferotemporal antérieur, est supposée porter l'information de forme et d'identité des objets. La voie dorsale, en direction du cortex pariétal, extrait les propriétés spatiales et de mouvement. Reproduit d'après Mishkin et al. (1983).

D'un point de vue physiologique, la voie dorsale reçoit principalement ses informations des neurones magnocellulaires. La voie ventrale, elle, utilise aussi bien les informations parvocellulaires, magnocellulaires et koniocellulaires. La séparation en 2 voies se fait à la sortie de V2, mais l'organisation en bandes de V2 reflète déjà cette structure.

1.2.1 Voie ventrale: traitement de la forme et identification des objets

Le long de la voie ventrale, l'information traverse les aires V1, V2, V4, avant d'atteindre le cortex inferotemporal, composé des aires PIT (postérieur inférotemporal), CIT (central inféro-temporal) et enfin AIT (antérieur inférotemporal). L'organisation rétinotopique, évidente dans V1 et V2, s'estompe progressivement dans V4 et le cortex inferotemporal. En revanche, les champs récepteurs des neurones s'agrandissent au fur et à mesure qu'on avance dans cette voie. Typiquement, les neurones de V4 ont des champs récepteurs 30 fois plus larges que ceux de V1, et jusqu'à 100 fois cette taille dans AIT.

Les représentations de l'information visuelle deviennent de plus en plus complexes et abstraites. Certains neurones de V4 montrent une sélectivité à la couleur, et leurs réponses peuvent être corrélées à la couleur perçue plutôt qu'à la longueur d'onde du stimulus (Zeki 1980, 1983). Cependant, V4 n'est pas "l'aire de la couleur", et les sélectivités des neurones de V1 et V2 sont suffisantes pour bien des tâches impliquant une discrimination de la couleur (Kulikowski et al, 1994). De plus, une lésion de V4 entraîne une dégradation des performances dans des tâches de discrimination de formes (Walsh et al, 1992; Schiller, 1995; Merigan 1996; Merigan et al, 1998). De fait, les enregistrements cellulaires dans V4 soulignent une sélectivité importante à la forme, que ce soit pour des barres ou gratings orientés (Desimone et Schein, 1987) ou pour des stimuli plus complexes, "non-cartésiens" comme des cercles concentriques ou des spirales (Gallant et al, 1993, 1996). Enfin, V4 montre une organisation modulaire, avec un regroupement en colonnes des neurones codant pour des propriétés similaires (Ghose et T'so, 1997).

V4 sert d'entrée corticale principale (mais pas unique) au cortex inferotemporal, où les neurones peuvent répondre sélectivement à des parties d'objets ou même à des objets entiers, parfois de façon invariante à la position spatiale, la taille, ou l'orientation. Les enregistrements pionniers de Gross et al (1972), puis Perrett et al (1982) et Desimone et al (1984) ont démontré que les neurones du cortex inferotemporal pouvaient répondre de façon sélective à des stimuli complexes comme des visages de singe ou d'homme (Perrett et al, 1991). Depuis, le travail extensif de Keiji Tanaka (1993, 1996, 1997) a montré que ces neurones peuvent répondre à différents types d'objets, qu'il existe dans le lobe temporal une organisation modulaire similaire à celle observée dans le cortex visuel primaire et les premières aires extrastriées (Fujita et al, 1992; Wang et al, 1996), et qu'à l'intérieur même de IT, les représentations des objets ont une complexité croissante (Kobatake et Tanaka, 1994), allant d'éléments ou parties simples dans PIT à une représentation de l'objet invariante à certaines transformations dans AIT. Les expériences de Logothetis et al. (1994, 1995) et Bülthoff et al. (1995) sur l'apprentissage de nouveaux objets suggèrent que cette représentation est plastique dans une certaine mesure, et que les neurones de IT peuvent

"apprendre" à répondre pour des classes de stimuli que les singes ne connaissent pas a priori. Cependant, certaines études électrophysiologiques suggèrent que les modules de IT ne constitueraient pas un "alphabet" de formes de bases, mais plutôt que la représentation d'un objet dans IT serait distribuée sur les réponses de plusieurs neurones de la population (Gochin et al, 1994; Rolls et al, 1997a,b).

Dans tous les cas, il semblerait donc que la voie ventrale soit spécialisée dans l'extraction des propriétés du stimulus visuel nécessaires à son identification ou à sa catégorisation. Parce que les réponses des neurones de IT sont modulées par l'interprétation "consciente" de la stimulation rétinienne, le cortex inférotemporal pourrait correspondre à une étape de traitement "au-delà de la résolution des ambiguités, où l'activité reflète la représentation interne des objets, plutôt que les effets de la stimulation rétinienne sur des cellules encodant des caractéristiques simples ou des formes primaires" (Sheinberg et Logothetis, 1997).

L'absence de rétinotopie dans AIT et l'invariance des réponses à la position suggère que la localisation du stimulus doit être traitée dans d'autres aires corticales.

1.2.2 Voie dorsale: traitement de l'information spatiale et de mouvement

C'est dans la voie dorsale (ou pariétale) que se fait l'intégration des propriétés spatiales et dynamiques de la scène visuelle, notamment la position des objets et leurs relations spatiales ainsi que leur mouvement. Les informations cheminent de V1 et V2 à travers V3, V3A, MT qu'on dénomme également V5, MST avant d'atteindre les aires postérieures pariétales comme l'aire 7A et IP (intrapariétale).

Les neurones des aires V3 et V3A répondent fortement au contraste, et ont une certaine sélectivité à la direction de mouvement (il existe en fait une différence qualitative entre ces aires chez l'homme et le singe ; voir Tootell et al, 1997). Les cellules de MT ont une forte sélectivité à la direction de mouvement, et certaines sont corrélées au mouvement apparent (perçu) plutôt qu'à la direction locale (Movshon, et al. 1985; Britten et al, 1996), ce qui n'est pas le cas pour les neurones de V1 par exemple.

La taille des champs récepteurs augmente dans la voie dorsale, de la même façon que pour la voie ventrale (Raiguel et al, 1997). De même, les neurones sont sensibles à des propriétés du stimulus de plus en plus complexes. Ainsi, dans MT comme MST, les neurones peuvent être impliqués dans les mouvements de poursuite oculaire (Komatsu et Wurtz, 1988), mais l'invariance à la position du stimulus est plus prononcée dans MST que dans MT (Lagae et al, 1994). Les neurones de MST peuvent également être sélectifs au flux optique (Duffy et Wurtz, 1991a, 1991b), à des translations (Duffy, 1998), et en déduire un mouvement de l'animal lui-même dans son environnement (Britten et van Wezel, 1998).

Enfin, les aires postérieures pariétales comme l'aire 7A sont supposées participer au contrôle et à la direction de l'attention spatiale (Lawler et Cowey, 1987; Steinmetz et Constandinis, 1995) grâce à une représentation de la saillance (Gottlieb et al, 1998; Colby et Goldberg, 1999), et dans l'encodage de l'intention de mouvement (par exemple pour les saccades oculaires; Andersen et al 1997; Snyder et al, 1997, 2000). Il a été proposé que la voie dorsale ne serait donc pas uniquement la voie du "where" mais aussi (ou plutôt) celle du "how", c'est-à-dire impliquée dans les mouvements et le comportement de l'animal envers les objets de la scène visuelle (Milner et Goodale, 1995). Signalons également qu'une certaine sélectivité à la forme peut être présente dans les régions pariétales, comme il l'a été récemment démontré par Sereno et Maunsell (1998) pour l'aire LIP.

Il apparaît donc que les aires visuelles peuvent être séparées en 2 voies fonctionnellement distinctes, la voie temporale traitant de la forme, de la couleur et de l'identité des objets, la voie dorsale s'intéressant à leur position, leurs relations spatiales et leur mouvement. Ces voies projettent toutes deux vers le cortex frontal, où ces 2 types d'informations sur le stimulus visuel pourraient être liées pour donner lieu à une perception consciente de la scène visuelle (Koch et Braun, 1996).

Mais à l'intérieur même du système visuel, la séparation anatomique entre les 2 voies n'est pas totale. De plus, l'organisation hiérarchique dans chacune des voies n'implique pas une transmission d'information séquentielle, de V1 vers les aires temporales ou pariétales.

1.3 Connexions cortico-corticales, feed-back

Les études anatomiques, utilisant principalement des techniques de marquage ou traçage, ont mis en évidence un pattern de connexions dense et complexe entre les aires visuelles corticales (Morel et Bullier, 1990, Baizer et al, 1991; Salin et Bullier, 1995; Bullier et al, 1996). Ce pattern tend à s'opposer à une vue hierarchique simple où l'information circulerait séquentiellement de la rétine à travers les 2 voies ventrale et dorsale vers les cortex temporal et pariétal. De plus, une large majorité des connexions sont réciproques: 2 aires visuelles reliées par une connexion "feed-forward" ont de grandes chances d'être également reliées par une connexion "feed-back". Ceci est exprimé de manière simple dans la figure 4.

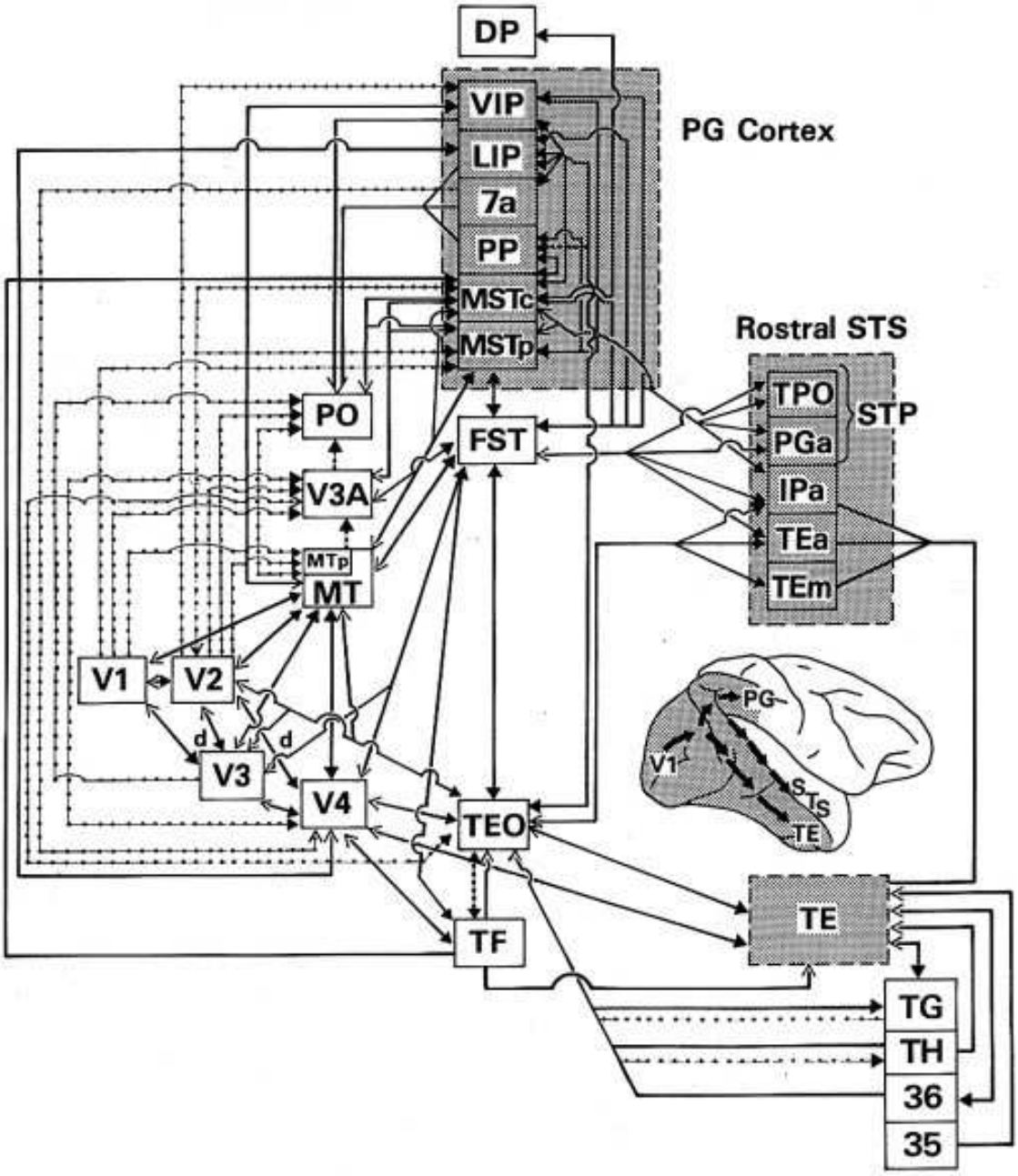


Figure 4. Organisation schématique de la hiérarchie des aires visuelles et des connexions cortico-corticales. Les lignes solides représentent des connexions provenant de représentations centrales et périphériques du champ visuel, alors que les pointillés correspondent à des représentations périphériques uniquement. Les flèches solides sont des connexions feed-forward, les flèches ouvertes des connexions feed-back, et les flèches solides réciproques des connexions "intermédiaires". Le diagramme montre la divergence des voies temporale et pariétale, et les sites potentiels d'interaction entre ces voies par le sulcus temporal supérieur (STS). Reproduit d'après Distler et al (1993).

Il apparaît que la plupart des aires sont interconnectées, et que les voies ventrale et dorsale sont reliées à chaque niveau ou presque, notamment par le STS rostral entre le lobe temporal et le cortex pariétal (Baizer et al, 1991). Enfin, notons qu'à l'intérieur même de chaque aire, les neurones voisins sont connectés entre eux, et qu'il existe également des connexions de plus longue portée, entre neurones distants.

Fonctionnellement, le rôle de ces connexions feed-back reste encore aujourd'hui largement méconnu, même si de nombreuses hypothèses existent (Mumford, 1991, 1992; Lamme, Super et Spekreijse, 1998; Lee et al, 1998). Il semblerait, d'après des études d'inactivation réversible de certaines aires visuelles (e.g. Hupé et al, 1998), que ce type de connexion permettre d'influencer, c'est-à-dire d'augmenter ou de réduire sélectivement la réponse des neurones dans les premières aires visuelles, et pourrait ainsi faciliter des mécanismes tels que la segmentation de scène, la ségrégation figure-fond, ou l'attention sélective (Roelfsema, Lamme et Spekreijse, 1998).

Comme on le verra par la suite, cette propriété d'interconnexion des aires corticales a façonné l'interprétation de l'architecture fonctionnelle du système visuel, et les premiers modèles et théories de la vision computationnelle.

2. Les premiers modèles neuronaux du traitement visuel

La perception visuelle a été depuis les débuts des neurosciences l'un des champs d'investigation les plus privilégiés, aussi bien par l'expérimentation que par la théorie ou la modélisation connexionniste. Nous ne pourrons pas passer en revue ici tous les modèles proposés, car il existe au moins autant de modèles que de fonctions du système visuel auxquelles la science s'intéresse, et le terme "modèle" lui-même prend différentes significations selon le niveau d'analyse auquel on se place. Nous nous pencherons donc en particulier sur le problème de la reconnaissance d'objets, c'est-à-dire comment les différents objets composant la scène visuelle peuvent être séparés, puis identifiés ou catégorisés par le sujet humain ou animal.

Selon la très célèbre "métathéorie" de David Marr (1982), un problème théorique comme celui de la vision peut être approché à 3 niveaux différents:

- au niveau computationnel (i.e. conceptuel), le système est traduit en termes d'entrées et de sorties désirées.
- au niveau algorithmique, des processus ou mécanismes sont définis, qui permettent de résoudre le problème posé, c'est-à-dire de transformer les entrées du système en sorties désirées.

- au niveau implémentationnel, ces mécanismes sont expliqués en termes de la structure physique du système dans lequel la fonction désirée doit être réalisée ("implémentée"). Pour le système visuel des primates, il s'agit d'expliquer comment un ensemble de neurones interconnectés transforment l'information visuelle en représentations suffisament abstraites et complexes pour permettre à l'animal d'interagir avec l'environnement perçu.

Pour être un "modèle", une théorie de la vision doit donc pouvoir reproduire ou tout du moins expliquer le "comportement" de l'homme (ou de l'animal), ou de son système visuel. Le manque de données exhaustives sur les réactions de l'homme aux différents stimuli visuels (qui est la raison pour laquelle nous aurons toujours besoin de l'approche psychophysique ou de psychologie expérimentale) ouvre la brèche à des théories de la vision qui se situent au niveau conceptuel (le niveau computationnel de Marr): ne sachant pas exactement quelle est la "sortie désirée" du système pour une entrée visuelle précise, il est toujours possible de construire des théories conceptuelles de la vision, qui peuvent donner lieu à des prédictions, démontrées ou réfutées par l'expérimentation.

2.1 Description structurelle ou représentation par vues ?

Un grand nombre de modèles de la reconnaissance d'objets se situent aujourd'hui entre le niveau algorithmique et conceptuel. En particulier, la dichotomie entre modèles structuraux ou basés sur la vue (view-based) relève de ce niveau intermédiaire. Les modèles structuraux suggèrent que l'image rétinienne est transformée, après un certain nombre d'étapes intermédiaires comme l'esquisse en 2 dimensions et demi ($2^{1/2}$ D sketch) de Marr et Nishihara (1978), en une représentation centrée sur l'objet lui-même, et qui contient une description structurale de l'objet en termes de ses composantes et de leurs relations (par exemple les géons de Biederman, 1987; Hummel et Biederman, 1992). La théorie de reconnaissance "view-based" (Bülthoff et al, 1995; Tarr et Bülthoff 1995, 1998; Tarr et al, 1998) suppose que les objets sont reconnus par rapport à des vues prédéfinies ou "templates", appris par le système (Fukushima et Miyake, 1982; Riesenhuber et Poggio, 1999). L'invariance à l'angle de vue peut néammoins être obtenue en ayant plusieurs "templates" par objet, un pour chaque angle de vue possible, avec un mécanisme d'interpolation entre vues proches (Poggio et Edelman, 1990; Edelman et Poggio, 1991). De même l'invariance à la position ou à la taille peut être obtenue en ayant un template pour chaque position ou taille possible de l'objet (ce qui nécessiterait des ressources probablement incompatibles avec le cerveau humain), ou grâce à un ensemble de transformations comme des translations, rotations, agrandissements ou rétrécissements

(Pitts et McCulloch, 1947; Anderson et Van Essen, 1987; Van Essen et al, 1992; Olshausen et al, 1993).

Le débat entre représentations structurales ou view-based n'est toujours pas résolu de façon définitive, et nous ne tenterons pas ici d'y apporter une solution. Notons simplement qu'il ne s'agit aucunement d'un problème indécidable, puisqu'il s'agit de déterminer, du point de vue conceptuel ou algorithmique, quelles sont les représentations utilisées par les systèmes visuels biologiques pour parvenir à reconnaître des objets de façon invariante à différentes transformations : l'accumulation de données expérimentales sur le système visuel lui-même et sur le fonctionnement des neurones le composant permettra à court ou moyen terme de décider laquelle de ces 2 approches (ou éventuellement une troisième alternative) est utilisée par notre cerveau. Nous nous attarderons plutôt ici sur les modèles de la vision possédant une composante au niveau implémentationnel, c'est-à-dire ceux qui proposent explicitement un moyen de construire des systèmes visuels implémentés dans des réseaux de neurones (indépendamment de l'approche choisie, i.e. structurelle ou view-based).

2.2 Implémentation neuronale

Neurone formel, transmission de valeurs analogiques

Les premiers enregistrements électrophysiologiques de Adrian (1926) et Hartline (1938, 1940) ont défini la réponse des cellules neuronales à une stimulation comme une série de potentiels d'actions (ou "spikes"), éléments unitaires (indivisibles), dont la fréquence augmente avec l'intensité de la stimulation. Cette simple observation est aujourd'hui encore la base commune de la plupart des théories du traitement neuronal. Si le premier neurone formel de McCulloch et Pitts (1943) avait une réponse binaire (0 ou 1), qui pouvait être interprétée comme le reflet d'un potentiel d'action, on s'est très vite tourné vers des modèles de neurones formels où la réponse était graduelle, analogique, reflétant une fréquence d'émission de ces potentiels d'action (Rosenblatt, 1958). Aux débuts de la cybernétique, Von Neumann (1958) écrit:

> *"Consequently, intensities of quantitative stimuli are rendered by periodic or nearly periodic pulse trains, the frequency always being a monotone function of the intensity of the stimulus. This is a sort of frequency-modulated system of signalling; intensities are translated into frequencies."*

Une version classique du neurone formel est représentée figure 5.

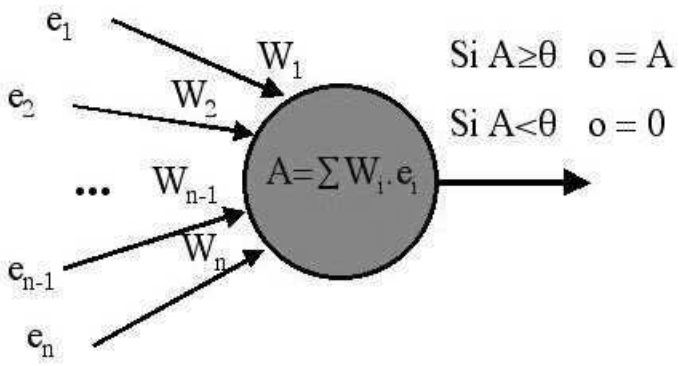


Figure 5. Le neurone formel classique calcule son activité comme la somme de ses entrées pondérées par les poids des connexions correspondantes. Sa réponse, si le seuil Θ est dépassé, est une fonction (ici l'identité) de son activation.

Du point de vue théorique, le neurone est donc considéré comme un convertisseur d'intensité en fréquence de décharge. Notons que cette interprétation a également dominé les travaux experimentaux, et que les études en électrophysiologie caractérisent la plupart du temps la réponse du neurone ou sa sélectivité comme une fréquence moyenne de décharge en fonction de la stimulation.

Tous les premiers modèles connexionnistes de la vision, et en particulier ceux que nous allons maintenant présenter, se conforment implicitement ou explicitement à cette hypothèse de codage et transmission de l'information analogique correspondant à une fréquence moyenne de décharge. Ils diffèrent donc essentiellement par l'architecture choisie et l'algorithme qu'ils implémentent.

Exemples de modèles

Les tout-premiers modèles connexionnistes des années 40 et 50 ne s'appliquent pas seulement à la vision, mais se rejoignent néammoins dans le fait qu'ils implémentent tous une reconnaissance de forme ("forme" étant défini ici par rapport à un signal d'entrée, qui n'est pas forcément visuel). Ceci s'applique cependant très bien au problème de la vision, et c'est certainement la raison pour laquelle la première couche d'un perceptron (Rosenblatt, 1958) porte le nom de "rétine". Ces modèles ont la propriété de pouvoir classifier des entrées, de façon linéaire ou non (on parle alors de perceptron multi-couche). Un autre caractère important des réseaux de neurones issus du perceptron est leur capacité à "apprendre", c'est-à-dire à organiser les connexions entre neurones pour répondre aux entrées selon les sorties désirées. Nous ne traiterons pas ici du problème de l'apprentissage

mais plutôt du fonctionnement de ces réseaux: l'architecture et la connectivité étant données, comment ces modèles réalisent-ils la fonction désirée?

Un des auteurs les plus prolifiques en modèles de la vision par réseaux de neurones depuis les années 60 est certainement Stephen Grossberg. Il serait impossible de résumer ici tout le travail fourni par Grossberg et son équipe. Nous nous contenterons de décrire les quelques propriétés de ces modèles qui nous intéresserons pour la suite. L'une des premières hypothèses spécifiques de Grossberg (1976) portait sur l'utilisation d'interactions latérales récurrentes entre neurones dans le but d'adapter automatiquement le gain des réponses à l'activité "ambiante": par le biais d'un feedback non linéaire, le réseau peut ainsi normaliser ses entrées. La théorie de Grossberg s'est ensuite raffinée pour se transformer en ART (Adaptive Resonance Theory): les entrées du système visuel sont constamment confrontées à ses "attentes" (top-down expectations). Lorsqu'il y a incohérence, un traitement spécial est effectué (Grossberg et Mingolla, 1985). Cette théorie est assez proche de celle de Shimon Ullman (1989, 1990), reprise en 1995. Ullman suppose que chaque neurone d'une voie "ascendante" possède un double dans une voie "descendante", et que la correspondance entre la voie montante et la voie descendante permet de "router" l'information au fur et à mesure qu'elle circule dans le système, c'est-à-dire qu'elle régit les "transformations" effectuées sur le pattern d'entrée. L'appariement (identification de l'objet) peut donc se réaliser à différents niveaux du système. Ces deux modèles de traitement de l'information attribuent donc une grande importance aux connexions en retour, des aires placées à un haut niveau de la hiérarchie visuelle vers les aires inférieures.

D'autres modèles considèrent que la reconnaissance des objets peut se faire de façon feed-forward, c'est-à-dire en utilisant un transfert unidirectionnel de l'information à travers la hiérarchie des aires visuelles. Citons par exemple le Neocognitron de Fukushima (1982), dans lequel l'information circule successivement à travers des couches de cellules simples, sélectives à des formes de plus en plus abstraites, et des cellules complexes, intégrant l'information sur une certaine étendue spatiale, et permettant ainsi d'obtenir une invariance à la position et à la taille dans les derniers niveaux du système. Ce type de modèle avec agrandissement des tailles des champs récepteurs le long de la hiérarchie visuelle a servi d'inspiration à de nombreux autres modèles comme celui de Wallis et Rolls (1997). De même le modèle de Riesenhuber et Poggio (1999), une amélioration du modèle de Poggio et Edelman (1990) et Vetter et al. (1995), fait appel à une séquence de cellules simples et complexes, les cellules complexes effectuant une transformation non linéaire (tout simplement l'opération MAX) sur leurs entrées. Notons également que, même si ce dernier modèle est construit avec une architecture strictement feed-forward, les auteurs proposent l'utilisation de connexions récurrentes pour implémenter la fonction MAX (qui n'est pas implémentée explicitement dans cette version du modèle).

Il apparaît donc que toutes les implémentations neuronales des modèles de reconnaissance d'objets utilisent des réseaux où l'information transmise est analogique, correspondant à la fréquence d'émission de potentiels d'action des neurones biologiques. De plus, la plupart de ces modèles font appel à des connexions feed-back, c'est-à-dire à un transfert d'information récurrent.

3. Décours temporel d'activation des aires visuelles, catégorisation visuelle rapide (150 ms).

Les modèles présentés à la section précédente correspondent en fait à des tentatives de reproduire, par la simulation ou la théorie, le comportement des animaux face au monde visuel, en respectant les contraintes définies par nos connaissances des systèmes visuels biologiques: architecture, connectivité, fonctionnalité des neurones réels. Il existe pourtant un autre ensemble de contraintes pesant sur les théories du traitement visuel neuronal: les contraintes temporelles. Pour les systèmes biologiques, ces contraintes sont fondamentales: un animal qui mettrait plusieurs minutes à détecter la présence d'un prédateur dans son champ visuel aurait peu d'occasions d'améliorer son temps de réaction la fois suivante! Ces données se doivent d'être prises en compte dans les théories et modèles de la vision. Dans cette section, nous définirons donc les contraintes temporelles qui régissent le comportement des primates et le fonctionnement de leur système visuel.

3.1 Electrophysiologie chez le singe

Les premiers enregistrements unitaires de cellules du cortex inférotemporal (IT) n'ont pas seulement permis de mettre en évidence la sélectivité de ces neurones à des stimuli complexes comme des visages, et ainsi d'identifier cette aire corticale comme l'aire de la "reconnaissance d'objets". Ils ont également montré que la latence de réponse de ces cellules pouvait être extrêmement courte: de l'ordre de 100 ms après l'apparition du stimulus (Perrett et al. 1982). De plus, la sélectivité aux visages de ces neurones est présente dès les premières millisecondes de la réponse (Oram et Perrett, 1992). Il semblerait donc que, au niveau neuronal, l'identification du stimulus visuel puisse se faire en seulement 100 ms chez le singe. Cette contrainte est extrêmement forte car, comme nous le verrons plus tard, elle limite grandement les architectures et algorithmes possibles des modèles de reconnaissance d'objets.

On retrouve le même type de contrainte temporelle tout au long de la voie ventrale du système visuel (voir figure 6). Les premières décharges des cellules (magnocellulaires) du LGN sont observées après seulement 25 ms, et 50% des cellules ont déjà déchargé après

30-35 ms (Schmolesky et al. 1998), ce qui limite fortement le temps d'intégration de l'information visuelle par les cellules rétiniennes. De même, les premières cellules de V1 déchargent après seulement 30 ms, et en moyenne aux alentours de 50 ms. La latence médiane des neurones de V2 est de 80 ms, 100 ms pour V4 et 110 ms pour IT. Ces latences ne sont bien sûr que des estimations. Elles varient selon le type de stimulus utilisé (donc selon l'étude de laquelle elles sont issues), et selon le type de cellules (par exemple magno- ou parvocellulaire) auquel on s'intéresse. Notons par exemple que les cellules des bandes épaisses et des bandes pâles de V2 (dominées par la voie magnocellulaire) s'activent en général avant les cellules de la couche 4Cβ de V1 (dominées par la voie parvocellulaire), qui elles-mêmes s'activent avant les neurones des bandes fines de V2 (Nowak et Bullier, 1995). Ceci a conduit Bullier et Nowak (1995) à remettre en question la stricte hiérarchie des aires visuelles définie par Van Essen et al. (1990). Notons pour finir qu'il est tout à fait regrettable qu'aucune étude n'ait pu à ce jour mettre en évidence de façon systématique les latences de décharge des neurones des différentes aires visuelles composant la voie ventrale, dans les mêmes conditions expérimentales (i.e. avec la même stimulation visuelle et la même tâche réalisée par le même animal).

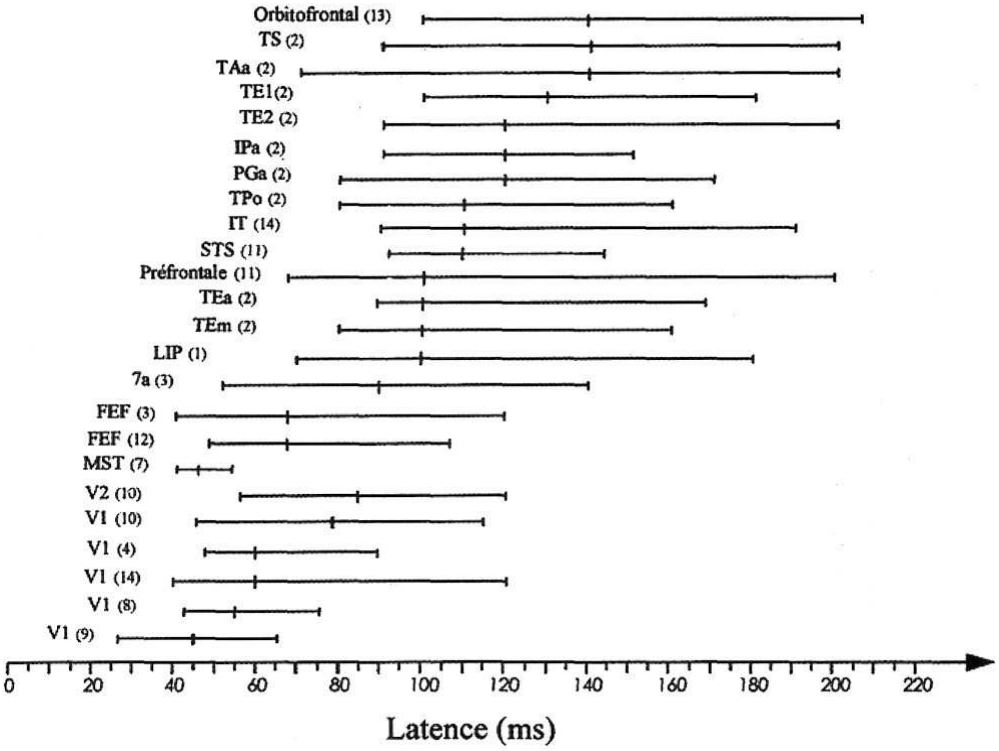


Figure 6. Latences d'activation des neurones dans les différentes aires visuelles. Pour chaque aire, les latences vont du 10ème au 90ème centile, la médiane est représentée par le trait vertical. Les numéros entre parenthèses correspondent à différentes études. Reproduit d'après Nowak et Bullier (1998).

3.2 Psychophysique chez l'homme

Chez l'homme, la durée du traitement visuel est plus difficile à mesurer que chez le singe: éthiquement, on ne peut pas envisager d'implanter des électrodes dans le cortex humain (cela se fait néammoins chez le sujet épileptique). La première mesure de la durée du traitement visuel vient du temps de réaction motrice (reaction time: RT). Pour détecter la présence d'un stimulus visuel (quel qu'il soit, par exemple un flash lumineux), le temps de réaction (qu'on appelle alors temps de réaction simple) est de l'ordre de 200 ms (Luce, 1986). Pour des tâches plus complexes (reconnaissance ou catégorisation d'objets), le temps de réaction se situe typiquement entre 400 et 1200 ms, et varie surtout selon la réponse demandée: prononcer le nom de l'objet requiert une opération motrice plus complexe, donc plus longue, que simplement appuyer sur une touche. La mesure du temps de réaction est donc fortement influencée par la composante motrice de la tâche à réaliser, et ne donnera en général qu'une faible indication sur la durée réelle du traitement visuel.

Les études RSVP (Rapid Serial Visual Presentation; Potter et Levy, 1969) de Potter (1976; 1999) et Intraub (1999) apportent une précision supplémentaire sur la nature des traitements que l'homme peut effectuer rapidement. Lorsqu'on présente à un sujet une série d'images différentes en succession rapide (par exemple 10 images par seconde), le sujet ne peut reporter ce qu'il a vu. Par contre, si on lui demande avant la présentation de la série de répondre dès qu'il verra une "cible" (définie comme "a boat", "a picnic on the beach", "two people drinking", etc...; Potter, 1975; Intraub, 1980), le sujet est capable de détecter la cible au-dessus du niveau de la chance. Il semble donc qu'une certaine information sémantique soit disponible très rapidement (100 à 300 ms) après la présentation de l'image (dans ce que Potter nomme un buffer de "very short-term conceptual memory").

Ces études de psychologie expérimentale ont été plus récemment complétées par des enregistrements d'activité électrique cérébrale. Les potentiels évoqués visuels (Event-related potentials: ERPs) montrent une négativité avec un pic à 170 ms (appelée N170) qui peut être spécifique à la présence d'un visage (Bötzel et al. 1995; Jeffreys, 1996 ; Schendan et al, 1998) ou d'autres objets (Rossion et al, 2000). Ces résultats sont confirmés par des études plus récentes en magnéto-encéphalographie (MEG ; Halgren et al. 2000) qui permettent de localiser (de façon plus précise que l'EEG) cette activité dans la partie ventrale du lobe temporal humain (l'équivalent du cortex inférotemporal du singe).

Une des études les plus intéressantes à cet égard est sans doute celle de Thorpe et al (1996) qui a montré que l'activité électrique à partir de 150 ms pouvait refléter une catégorisation du stimulus visuel. Les sujets devaient relâcher un bouton le plus vite possible si l'image, présentée pendant seulement 20 ms, contenait un animal. L'étude des différences

entre potentiels évoqués pour les stimuli cibles ("animal") et distracteurs ("non-animal") montre qu'une population de neurones doit être différemment activée dans les 2 conditions, après seulement 150 ms (voir figure 7). Les propriétés de ce type de catégorisation visuelle ultra-rapide (ultra-rapid visual categorization: URVC) ont depuis fait l'objet de nombreuses études, et sont ainsi mieux connues (voir par exemple la thèse de Denis Fize, 2000). Tout d'abord, cette capacité ne dépend pas directement d'une allocation de l'attention spatiale: la performance n'est pas affectée lorsque le stimulus peut apparaître aléatoirement à plusieurs positions, ce qui requiert pourtant une répartition des ressources attentionnelles (Fabre-Thorpe et al, 1998; Fize et al, en préparation). L'information de couleur dans les stimuli ne semble pas non plus être un élément déterminant de la performance: en l'absence de couleurs, la performance des sujets rapides n'est pas affectée (Delorme et al, 2000), et la différence de potentiels évoqués entre cibles et distracteurs survient à la même latence (i.e. 150 ms; Fize 2000). Ce type de tâche de catégorisation ultra-rapide peut également être réalisé en vision périphérique extrême, avec toutefois une diminution de la performance, mais dans des conditions où les sujets sont souvent incapables de reporter verbalement le stimulus présenté (Thorpe et al, 1999). Enfin, cette catégorisation ultra-rapide ne peut être encore accélérée par une connaissance préalable des images présentées, même après 3 semaines de pratique intensive (Fabre-Thorpe et al, 2000), ce qui suggère que cette durée de traitement minimale de 150 ms serait en quelque sorte "incompressible".

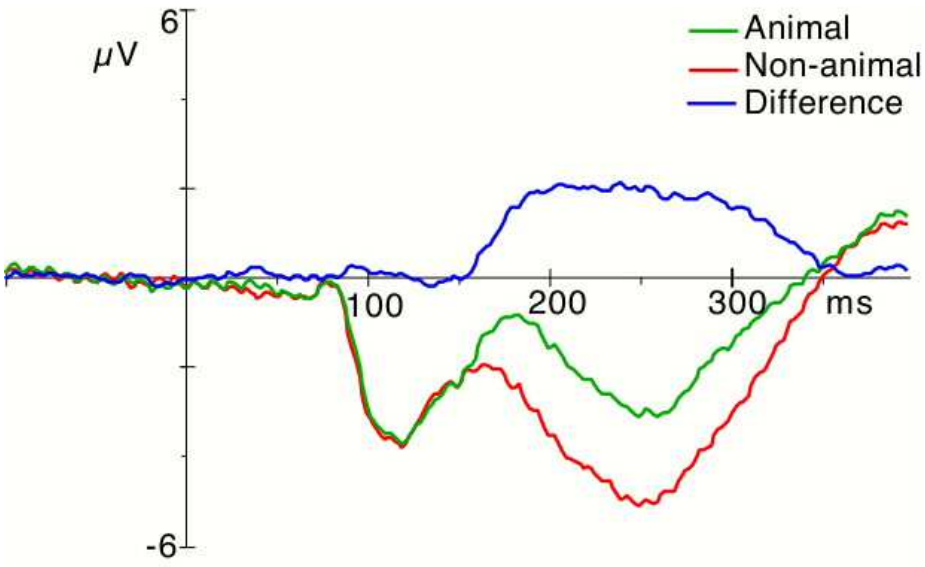


Figure 7. Les potentiels évoqués par les cibles ("animal") d'une tâche de catégorisation et par les distracteurs ("non-animal") divergent à partir de 150 ms. Certains neurones répondent différemment à ces 2 types d'images. Au niveau neuronal, la catégorisation peut donc être réalisée en 150 ms. Reproduit d'après Thorpe et al. (1996).

Cette durée de 150 ms correspond en fait certainement à la latence d'activation des neurones du cortex inférotemporal du singe macaque, enregistrée par Perrett et al (1982) et d'autres auteurs, neurones qui s'activent spécifiquement pour une catégorie donnée de stimulus visuel, 100 ms environ après la stimulation. Cette différence d'environ 50 ms entre singe et homme est corroborrée par le fait que les singes macaques peuvent réaliser le même type de tâche de catégorisation, mais avec des temps de réaction plus courts (de 50 à 100 ms) que les sujets humains les plus rapides (Fabre-Thorpe et al, 1998). De fait, de récentes études où des électrodes sont implantées sous le crâne chez des sujets humains épileptiques, montrent que les régions occipito-temporales ventrales (correspondant au cortex inféro-temporal du singe) s'activent spécifiquement pour des visages humains ou d'autres objets, avec des latences de réponse entre 150 et 200 ms (Allison et al. 1999).

3.3 Implications théoriques

Les études électrophysiologiques chez le singe et chez l'homme démontrent que certains neurones du lobe temporal répondent de façon différente (ou spécifique) à différentes catégories d'objets, après seulement 100 ms chez le singe ou 150 ms chez l'homme. Ces observations ont des implications fondamentales pour les théories du traitement neuronal de l'information visuelle (Thorpe et Imbert, 1989).

Compte-tenu du nombre d'aires visuelles à travers lesquelles l'information doit circuler, du nombre d'étapes synaptiques dans chacune de ces aires (une dizaine en tout), de la fréquence maximale de décharge des neurones corticaux (entre 100 et 200 Hz), et de la vitesse de conduction des fibres axonales (environ 1 m/s), il semble que:

- chaque neurone à chaque étape ne peut décharger qu'une ou 2 fois au maximum
- l'information doit se transmettre de manière essentiellement feed-forward

Les délais d'activation entre différentes aires corticales qui ont été présentés plus haut (par exemple, 30 ms pour le LGN et 50 ms pour V1) amènent à des conclusions identiques. Ceci est résumé dans la figure 8.

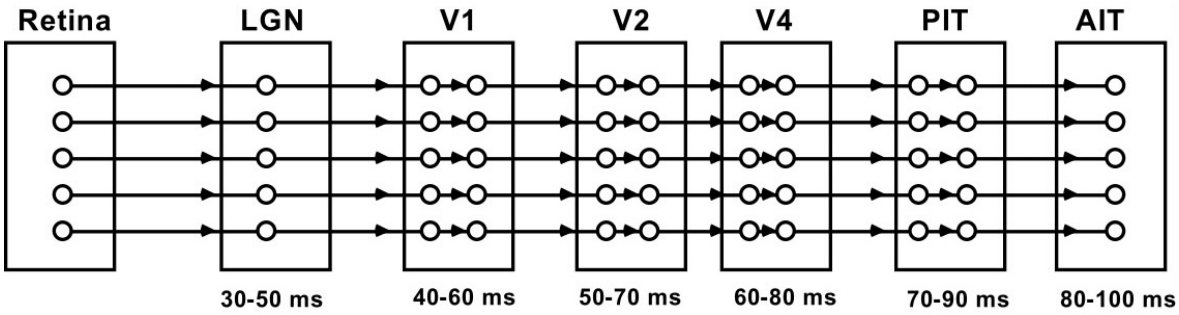


Figure 8. Les latences d'activation des différentes aires corticales, le nombre d'étapes synaptiques, la vitesse de transmission des fibres axonales, et le taux de décharge des neurones corticaux (typiquement inférieur à 100 Hz) impliquent que l'information visuelle doit circuler en feed-forward, et que seulement les tout-premiers spikes de chaque neurone pourront être pris en compte.

Cette idée d'un traitement visuel basé sur une propagation feed-forward de l'information, et l'utilisation d'un seul spike par neurone, a récemment trouvé un support expérimental direct. Keysers et al. (sous presse) ont enregistré chez le singe macaque les réponses de neurones du STS, sélectifs à des patterns complexes comme des visages, lors d'une stimulation visuelle avec un paradigme RSVP (Rapid Serial Visual Presentation). Ces neurones répondent à leur stimulus préféré à l'intérieur de la séquence de stimulation, même lorsqu'il est présenté pour seulement 14 ms. Dans ces conditions, environ 7 images ont déjà été présentées à la rétine et sont simultanément représentées dans les différentes étapes de traitement, lorsque les neurones du STS commencent à répondre, i.e. 90-100 ms après l'apparition de leur stimulus préféré. L'observation de la figure 8 suggère sans équivoque que le traitement visuel sous-jacent à une telle capacité doit être basé sur une architecture en "pipeline", où les représentations des stimuli visuels traversent les différentes étapes neuronales en une succession très rapide (i.e. feed-forward), et où à chaque étape, le temps disponible pour intégrer l'information et la distribuer à l'étape suivante est extrêmement court (environ 10 ms), ne laissant place qu'à la génération d'un seul potentiel d'action par neurone au maximum.

Aucun des modèles du traitement neuronal de l'information visuelle présentés dans la section 2.2 ne peut respecter les 2 types de contraintes mentionnées. Même si certains de ces modèles font appel à une transmission feed-forward de l'information, ils se basent tous sur la fréquence d'émission des spikes comme codage de l'information transmise entre 2 neurones. Or, avec seulement 1 spike pris en compte pour chaque neurone à chaque étape,

le codage par fréquence n'a aucun sens: il se ramène à un codage binaire, distinguant les neurones qui ont déchargé de ceux qui sont restés silencieux.

La dernière décennie a donc connu une sorte de "révolution" dans la façon de concevoir le traitement de l'information visuelle. Aussi bien au niveau expérimental que théorique, de nombreux auteurs ont dû prendre en compte la structure temporelle précise des réponses neuronales ("spike trains"), afin de pouvoir respecter les contraintes temporelles imposées par la biologie.

4. Les modèles de "Spiking Neurons"

Les contraintes temporelles mentionnées à la section précédente ont donc été le principal vecteur du regain d'intérêt qui s'est observé ces dernières années pour les potentiels d'action et leur puissance computationnelle (voir par exemple le livre "Spikes : exploring the neural code" ; Rieke et al, 1997; Miller, 1994; Koch, 1997; Gerstner et al, 1997; Ferster et Spruston, 1995; Koch et Laurent, 1999). Ce regain s'est manifesté également de la part des électrophysiologistes, par une étude plus systématique et plus détaillée de la structure temporelle des réponses neuronales (les spikes trains). On s'est vite rendu compte que cette structure pouvait contenir bien plus d'informations que le taux de décharge seul (Richmond et al, 1987, 1990; Richmond et Optican, 1987, 1990; Optican et Richmond, 1987; McClurkin et al, 1991a,b; Geisler et al, 1991; Bialek et al, 1991; Bialek et Rieke, 1992; Eskandar et al, 1992; McClurkin et al, 1996; Borst et Theunissen, 1999), et que différentes composantes pouvaient porter différentes informations, indépendamment l'une de l'autre. Par exemple, Cattaneo et al (1981) observent dans le cortex visuel primaire du chat des spikes "isolés" portant l'information de contraste, et des regroupements de spikes portant l'information de fréquence spatiale, ou de vitesse. Gawne et al (1996) démontrent que la latence du premier spike émis par des neurones de V1 encode efficacement le contraste du stimulus, alors que l'amplitude de réponse permet d'encoder son orientation. Des résultats récents suggèrent que les premières réponses de neurones du cortex inféro-temporal peuvent encoder spécifiquement la catégorie d'un stimulus (visage humain, visage de singe), alors qu'après 50 ms ces réponses pourront décrire des aspects plus détaillés du visage comme son identité ou son expression (Sugase et al, 1999).

Les modèles récents de traitement neuronal de l'information que nous allons présenter dans cette section portent le nom de modèles de "Spiking Neurons" : à l'encontre des modèles "classiques" que nous avons décrits auparavant, ceux-ci prennent en compte le fait que la sortie des neurones réels n'est pas caractérisée par une fonction continue, mais

par une séquence discrète d'éléments indivisibles : les spikes. Ces modèles peuvent être séparés grossièrement en 2 classes : ceux qui font appel à la **synchronie** des décharges neuronales pour réaliser la fonction désirée, et ceux qui utilisent les **temps** (ou intervalles de temps) précis d'occurrence des spikes.

La synchronisation des décharges.

Cette catégorie de modèles trouve ses racines dans différentes études expérimentales qui ont mis en évidence une synchronisation des réponses neuronales entre neurones voisins (Gray et al. 1989), ou appartenant à différentes aires (Engel et al, 1991), voire même à différents hémisphères cérébraux (Engel et al, 1991). Ces observations sont très bien résumées par Singer (1993), Singer et Gray (1995) ou encore König et Engel (1995). Nous ne considérons pour l'instant ces résultats que dans une perspective historique, mais nous y reviendrons plus en détail au chapitre III section 2.4.

La synchronisation des décharges neuronales a pour propriété évidente de créer un lien entre 2 ou plusieurs populations de neurones : ceux qui déchargent simultanément sont "liés". C'est pourquoi ce type de codage de l'information a été principalement appliqué à la résolution du "binding problem" (Treisman, 1996). Dans la plupart des modèles, les différentes propriétés ou attributs du stimulus sont encodés dans différentes "cartes" ou populations neuronales (typiquement, l'identité des objets dans IT, leur position dans le cortex pariétal ; leur couleur dans V4, leur direction de mouvement dans MT). Il se pose alors le problème de lier les activités neuronales de chacune de ces populations correspondant au même objet : le binding problem.

Cette stratégie de "binding par synchronie" a été employée par Hummel et Biederman (1992) pour améliorer le modèle de Biederman (1987) et implémenter dans un réseau de neurones le liage des différentes propriétés extraites par ce modèle. D'autres implémentations similaires existent, et sont compilées dans von der Malsburg (1995).

L'approche choisie par Abeles (1991), et reprise entre autres par Neven et Aertsen (1992) et Diesmann et al. (1999), présente un certain nombre de particularités conceptuelles qui méritent d'être décrites. Cette approche porte le nom de "synfire chains" : l'information est représentée par une chaîne (une séquence) de neurones, organisés en classes selon le moment auquel ils déchargent ("fire"). Les neurones qui déchargent de manière synchrone sont regroupés, et la séquence d'activations synchrones représente la transformation de l'entrée visuelle en un événement neuronal qui porte une signification comportementale (par exemple l'identification d'un objet de la scène visuelle). Les neurones qui sont regroupés par la synchronisation n'appartiennent pas forcément à la même population ou aire visuelle. Le transfert d'informations ne peut donc pas être défini comme feed-forward ou récurrent : chaque neurone qui participe à une classe, de par sa synchronisation avec les autres

neurones de la classe, peut éventuellement participer à une classe ultérieure dans la séquence, voire plusieurs. Dans ce type de réseau, chaque neurone impliqué dans un pattern d'activité synchrone (i.e. une représentation) peut rapidement se désynchroniser puis resynchroniser avec une autre population (donc une autre représentation) : c'est une des propriétés majeures des synfire chains.

Le spike timing

Le fait que la durée du traitement visuel soit trop courte pour pouvoir utiliser un codage de l'information par le taux de décharge n'implique pas directement que c'est la synchronie des décharges qui doive encoder cette information. En fait, de nombreux auteurs qui ont tenté de décrire des modèles respectant la contrainte de "pas plus d'un spike par neurone" (Thorpe et Imbert, 1989), se sont tournés vers un autre type de traitement de l'information utilisant les "spiking neurons" : des modèles où la latence d'émission des spikes est prise en compte.

Thorpe (1990) fut le premier à proposer un tel type de codage d'information (le "spike arrival time") et à décrire sa puissance computationnelle. En utilisant seulement un spike par neurone (ce qui s'accorde avec les contraintes temporelles du système visuel), il est toujours possible de transmettre une information analogique entre 2 populations de neurones, si les neurones efférents décodent simplement les latences respectives auxquelles ces spikes ont été reçus. En effet ces latences, tout comme la fréquence de décharge supposée par les modèles classiques, reflètent le niveau d'activation des neurones. Ceci est expliqué dans la figure 9.

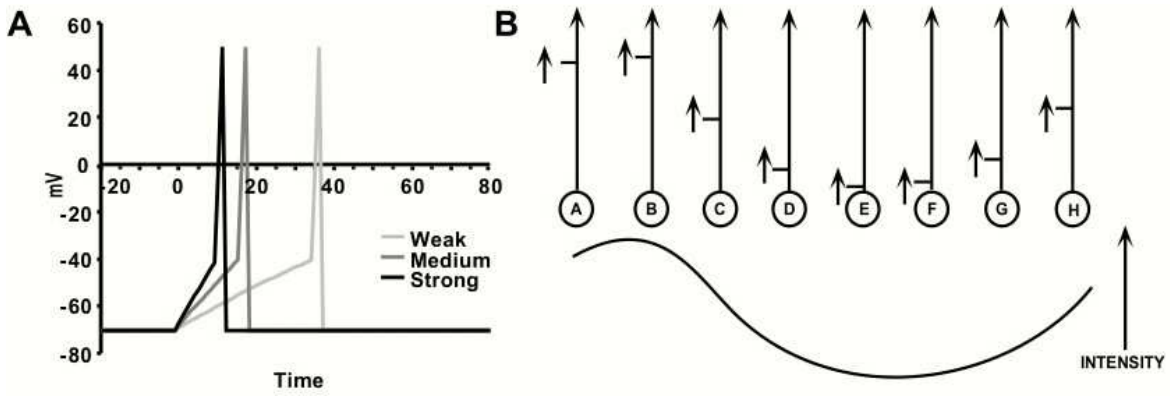


Figure 9. Les propriétés des neurones "integrate-and-fire" (version commune du spiking neuron) impliquent que plus un neurone est activé, plus il déchargera tôt (A). Un pattern d'intensité analogique comme celui en (B) peut donc être encodé par les latences respectives auxquelles les neurones déchargent.

Cette idée a ensuite été reprise par Hopfield (1995), qui a proposé une méthode pour implémenter un tel type de codage, basée sur des lignes à délai (les connexions entre neurones se font par des axones plus ou moins longs suivant la latence que l'on veut décoder) et des neurones détecteurs de coïncidence. Cette méthode, parfaitement valide du point de vue théorique, a cependant bien peu de support biologique

L'idée d'utiliser des lignes à délai a cependant été reprise par Bunomano et Merzenich (1998) qui implémentent un système de reconnaissance de formes invariant à la position en se débarrassant de l'information spatiale et ne considérant que l'histogramme des latences de réponse des neurones. Le modèle réalise une performance de 93% lors d'une tâche de reconnaissance de chiffres à différentes positions. Là encore, la plausibilité biologique du modèle est loin d'être évidente : si l'information spatiale se perd progressivement du fait de l'agrandissement des champs récepteurs le long de la voie ventrale, il paraît néammoins peu probable que toute structure spatiale soit perdue en une seule opération, et ce à l'intérieur même des champs récepteurs des neurones de IT.

Dans le même temps Thorpe et Gautrais (1997, 1998) développent un modèle de codage neuronal qui permet d'éviter le problème du décodage de la latence : le "rank order coding". Les neurones efférents ne s'intéressent plus à la date précise d'émission des spikes par leurs afférents, mais plutôt à l'ordre dans lequel ils reçoivent ces spikes. La puissance computationnelle est toujours remarquable : avec 8 neurones, 8! soit plus de 40000 signaux différents peuvent être encodés (voir figure 9B), et ce en utilisant seulement un spike pour chaque neurone. Le décodage de l'ordre fait appel à une désensibilisation du neurone à mesure qu'il reçoit des spikes. Le neurone attribue beaucoup d'importance aux premiers spikes qu'il reçoit, puis graduellement de moins en moins d'importance : il décode bien l'ordre d'arrivée. Ce code par l'ordre nécessite évidemment, tout comme le codage par latences, un mécanisme de remise à zéro. Il serait sinon sujet à des confusions entre les premières décharges correspondant à une image et les dernières de l'image précédente, confusions qui seraient dramatiques, puisque les dernières décharges ne portent que peu d'information pour ce code, alors que les premières sont interprétées comme la part la plus importante du signal. Bien que cette remise à zéro n'ait pu à ce jour être démontrée dans les systèmes visuels biologiques, il semble exister des mécanismes candidats comme les oscillations sous-liminaires de l'activité des neurones, ou encore les microsaccades visuelles (Martinez-Conde et al, 2000) .

Notons que le modèle du cortex visuel de Nakamura (1998) se rapproche implicitement du codage par rang de Thorpe et Gautrais. Ce modèle utilise la latence des premières décharges des neurones ainsi qu'un mécanisme de compétition par inhibition latérale qui a pour résultat de filtrer l'information selon son importance. Le modèle de Thorpe

et Gautrais réalise en fait la même opération grâce à son mécanisme de désensibilisation neuronale, sans invoquer d'inhibition latérale entre neurones voisins.

Le potentiel théorique des réseaux utilisant le spike timing a été mis en évidence de façon formelle par les travaux de Maass (1996, 1997), qui a entre autres démontré que ce type de modèle pouvait approximer n'importe quelle fonction continue de plusieurs variables. La validité de ces résultats a été étendue par Ruf (1997) à des neurones aux paramètres "biologiquement plausibles" (modélisés par les équations de Hodgkin-Huxley). De plus, des mécanismes d'apprentissage biologiquement plausibles (Fregnac et al, 1994; Markram et al, 1997; Fregnac et Shulz, 1999) peuvent être implémentés simplement dans de tels modèles (Gerstner et al, 1996; Ruf et Schmidt, 1997; Natschlager et Ruf, 1998)

Cependant (même si les contraintes temporelles provenant des systèmes biologiques ne laissent pas vraiment d'alternative), l'idée d'utiliser la latence de décharge comme vecteur de l'information n'est pas aujourd'hui universellement acceptée. Ceci provient en partie de la question de la plausibilité biologique du code : même si on n'invoque pas de lignes à délai, il reste toujours une notion de remise à zéro des neurones qui n'a pour l'instant que peu de support biologique. C'est pourquoi d'autres auteurs, comme Horn et Levanda (1998), restent plus timides et choisissent une position intermédiaire, en utilisant par exemple l'intervalle entre 2 spikes consécutifs (ISI) comme code d'information.

Pour résumer, qu'on utilise un code par synchronisation des décharges neuronales ou par décodage des latences (absolues ou relatives) de décharge, il semble que la prise en compte de la structure temporelle précise des spike trains soit à ce jour le seul moyen connu de modéliser les traitements de l'information mis en jeu dans la voie ventrale du système visuel, tout en respectant les contraintes temporelles imposées par la biologie.

5. Catégorisation visuelle en 50 ms? Implications pour les modèles.

Nous avons vu aux sections précédentes que le décours temporel d'activation des aires visuelles, et donc la durée du traitement visuel estimée à 150 ms chez l'homme (par exemple pour une tâche de détection d'animaux), avait remis en cause les modèles neuronaux du traitement de l'information visuelle, et donné lieu à une sorte de révolution théorique. Les seuls modèles pouvant respecter ces contraintes biologiques semblent être ceux qui transmettent l'information de façon feed-forward, et prennent en compte la structure temporelle précise des réponses neuronales.

Plus récemment, de nouvelles études expérimentales ont jeté un "pavé dans la mare" en proposant que, dans certains cas, le traitement visuel pourrait être encore plus rapide : il faudrait seulement 50 ms pour accéder à une représentation "sémantique" de l'information visuelle, comme la catégorie de l'objet, ou son genre (masculin ou féminin) dans le cas des visages. Nous présenterons ici brièvement ces résultats, puis leurs implications potentielles pour les modèles du traitement de l'information visuelle.

5.1 Résultats expérimentaux

La première expérience à reporter des activations neuronales précoces (i.e. avant 150 ms) reflétant des propriétés de "haut niveau" des objets de la scène visuelle a été réalisée par Seeck et al. (1997). Ces auteurs ont enregistré l'activité électrique de surface chez des sujets sains (potentiels évoqués) et l'activité intracraniale dans le lobe temporal (inférotemporal et hippocampe) et dans le lobe frontal chez des patients épileptiques, lors d'une tâche de "reconnaissance de visages". Il s'agit en fait de répondre dès qu'un des visages présentés a déjà été vu lors de la même série d'images (ce qui correspond plutôt à la détection d'une répétition, ou d'une familiarité). Les visages sont présentés par paires, séparés de seulement 150 ms. L'intervalle entre 2 paires est d'environ 2 secondes. La paire peut (avec une probabilité d'environ 3/11) contenir un (seul) visage répété (c'est-à-dire qui a déjà été présenté durant la série): le sujet doit alors répondre en appuyant sur un bouton; dans le cas contraire il appuie sur un autre bouton. Lorsque les auteurs comparent les potentiels évoqués par les visages répétés et les visages non répétés (nouveaux), ils s'aperçoivent que les activités pour les 2 catégories divergent avant même 50 ms, que ce soit avec les enregistrements de surface ou intracraniaux, notamment pour ces derniers au niveau des électrodes temporales, hippocampales et orbito-frontales. Les auteurs concluent donc que des aspects de "haut niveau" du traitement visuel, comme la reconnaissance d'un visage vu auparavant, surviennent seulement 50 ms après la présentation du stimulus, et que le lobe temporal doit être impliqué dans ce type de traitement.

La même année, George et al. (1997) enregistrent les potentiels évoqués visuels lors d'une tâche de détection de visages dont les stimuli sont des "Mooney faces", qui peuvent être soit à l'endroit, soit inclinés de 180° (à l'envers). Ce type de stimulus ressemblant à une photographie de très fort contraste, est composé de 2 parties (par exemple une partie blanche et l'autre noire). Lorsqu'un "Mooney face" est à l'endroit, un visage peut être détecté. Lorsqu'il est à l'envers, la détection devient très difficile, et le sujet perçoit la plupart du temps 2 formes sans aucune structure. Dans cette expérience, les sujets devaient répondre s'ils détectaient un visage, dans 8 séries de 40 images. Les 4 dernières séries étaient l'exacte répétition des 4 premières. Ainsi, les auteurs ont pu comparer les potentiels évoqués par la première et la deuxième présentation d'un même stimulus. Ces 2 potentiels

diffèrent dès la période 50-110 ms après présentation du stimulus, que ce soit pour les images perçues comme des visages (Mooney faces à l'endroit), ou pour les formes non perçues comme des visages (Mooney faces à l'envers). Cette différence semble être générée dans les régions inférotemporales. Elle est interprétée comme un effet de répétition ou de familiarité, qui n'est pas spécifique à la reconnaissance des visages.

L'année suivante, Debruille et al. (1998) mettent en place un paradigme expérimental pour tester le temps que met réellement le système visuel à reconnaître un visage familier (et non une simple répétition d'image). Les images cibles sont des photographies de personnes populaires en France (où est réalisée l'expérience), les distracteurs des photographies de personnes populaires dans d'autres pays occidentaux (qui ne sont donc pas connues à priori des sujets français). Aucune différence n'est observée entre ces 2 catégories sur la période 50-70 ms après la présentation du stimulus (à l'encontre des résultats de Seeck et al. 1997), mais une différence significative est présente entre 76 et 130 ms. Les visages seraient donc reconnus en moins de 100 ms.

Une autre étude toute récente s'est intéressée au traitement du genre sur des images de visages ou de mains humaines. Mouchetant-Rostaing et al. (2000a) ont présenté à des sujets des photographies de visages et de mains dans 3 conditions: dans la condition 1, les sujets ne doivent pas et ne peuvent pas catégoriser les images selon le genre (ils doivent répondre lorsque les visages portent des lunettes, ou lorsqu'un torse humain est présenté parmi des mains humaines, et toutes les images d'une même série représentent des personnes de même sexe); dans la condition 2, ils peuvent de façon "incidentelle" catégoriser les images selon le genre, mais sont impliqués dans une autre tâche (la même que précedemment, mais les deux genres sont mélangés dans chaque série); dans la 3ème condition, les sujets doivent explicitement ("intentionnellement") catégoriser les images selon le genre. Seuls les distracteurs de chaque condition sont comparés. Une différence d'activité électrique (EEG) entre les conditions 1 et 2 est observée entre 40 et 90 ms après la présentation du stimulus, pour les visages comme pour les mains humaines. Ceci est interprété comme un reflet d'une catégorisation rapide et automatique de l'information visuelle. Curieusement, cette différence précoce n'est pas observée entre les conditions 2 et 3. Ce type de "catégorisation précoce" a été retrouvé par les mêmes auteurs avec des stimuli "non-biologiques" (i.e. formes hachurées; Mouchetant-Rostaing et al, 2000b).

Des résultats similaires ont également été obtenus par Magnéto-Encéphalographie (MEG). Halgren et al. (2000) ont montré qu'une source d'activité (située cette fois dans le cortex occipital) pouvait différentier dès 110 ms les images de visages d'autres stimuli, ainsi qu'entre les visages d'expression neutre et les visages tristes ou joyeux.

Enfin, Johnson, Guirao-Garcia et Olshausen (1999) ont reproduit l'expérience de catégorisation "animal/non-animal" de Thorpe et al. (1996), en incluant une troisième catégorie d'images contenant des fleurs. La moitié des sujets devaient répondre sur les

images contenant des animaux, l'autre sur celles contenant des fleurs. Dans les 2 tâches, la différence d'activité électrique entre cibles et distracteurs survient à 100 ms environ après la présentation du stimulus. De plus, cette différence est aussi présente entre les 2 catégories non-cibles. Elle reflèterait donc une catégorisation implicite.

Cette série de résultats expérimentaux semble donc suggérer que des traitements de haut niveau comme la catégorisation d'objets, la reconnaissance de visages ou la discrimination de leur genre pourraient se faire en moins de 100 ms, voire avant 50 ms post-stimulus. Si tel était le cas, les modèles du traitement de l'information visuelle seraient probablement tous remis en question, y compris ceux qui font appel aux spiking neurons.

5.2 Implications théoriques fondamentales

La catégorisation ou l'identification des stimuli en 150 ms constitue déjà pour le système visuel une contrainte "maximale": elle implique que l'information ne peut circuler que vers l'avant, et que pas plus d'un spike par neurone ne peut être pris en compte. Contraindre encore cette durée semble presque impossible.

Il n'est pas raisonnablement envisageable que l'information puisse être encodée avec moins d'un spike par neurone! Ce n'est donc pas dans le code de l'information qu'il faut chercher le gain de temps nécessaire. Pratiquement, le seul moyen de raccourcir le temps de traitement serait de "sauter" des étapes neuronales: compte-tenu de la fréquence de décharge des neurones corticaux, du temps d'intégration de l'information visuelle dans la rétine, des constantes de temps synaptiques et des vitesses de conduction des fibres axonales, un objet serait identifié après seulement 1 (voire 2) étapes de traitement après la rétine! Concrètement, la reconnaissance d'objets aurait donc lieu dans V1 (l'activité électrique visuelle à des latences de 50-70 ms a en effet été identifiée avec l'activité du cortex visuel primaire; Clark et al. 1995; Clark et Hillyard 1996), ou alors la rétine serait directement connectée (avec éventuellement un relais sous-cortical) au cortex inférotemporal! Et dans ce cas, quelle serait la véritable utilité de la hiérarchie des aires visuelles extrastriées, que les chercheurs en neurosciences visuelles ont accepté comme un postulat de départ pendant plus de 30 ans?

D'un point de vue théorique, il est donc fondamental de connaître la durée exacte du traitement visuel minimal nécessaire pour accéder à une représentation de l'entrée visuelle suffisamment abstraite (par exemple, sa catégorie). S'il s'avérait que 50 ms sont suffisantes, les modèles du traitement visuel connaîtraient une nouvelle révolution. Par contre, il est possible que les résultats expérimentaux présentés à la section précédente ne soient pas valides dans l'absolu, qu'ils soient simplement le reflet d'une erreur conceptuelle dans le paradigme expérimental.

5.3 Catégorisation visuelle: 50 ou 150 ms?

5.3.1 Seeck et al.: erreur conceptuelle de 1er ordre...

Si l'on s'intéresse plus précisément au protocole expérimental utilisé par Seeck et al. (1997), il apparaît que les conclusions de cette étude ne sont pas théoriquement valides.

Les auteurs basent leurs résultats sur la comparaison des potentiels évoqués par les visages répétés (notés "cibles") et les visages nouveaux (notés "distracteurs"). Un prérequis pour pouvoir effectuer ce type de comparaison est que les 2 signaux comparés doivent être obtenus dans les mêmes conditions expérimentales. Or, dans cette étude, une image "cible" est précédée dans 50% des cas par une autre image (présentée 150 ms auparavant), qui est forcément un distracteur, de par le protocole choisi. Le signal "cible" est donc composé de:

- 50% de cibles précédées d'un distracteur
- 50% de cibles "non précédées" (i.e. premières de la paire d'images).

82 paires d'images sur 112 sont composées de 2 distracteurs; les 30 paires restantes sont composées d'une cible et d'un distracteur, soit dans cet ordre (15 paires), soit dans l'ordre inverse (15 paires). Le signal obtenu pour les distracteurs est donc composé de:

- 82 images soit 42.3% de distracteurs précédés d'un distracteur
- 82+15=97 images soit 50% de distracteurs "non précédés"
- 15 images soit 7.7% de distracteurs précédés d'une **cible**.

7.7% des images de ces 2 catégories ont donc été obtenues dans des conditions différentes: soit précédées d'un distracteur, soit d'une cible 150 ms plus tôt. Si l'on considère que le signal induit par les distracteurs diffère réellement du signal induit par les cibles entre 150 ms et 200 ms (résultats de Thorpe et al. 1996 et d'autres études citées à la section 3.2), alors les moyennes "cible" et "distracteur" effectuées par Seeck et al devraient différer... entre 0 et 50 ms après la présentation du stimulus! Puisque les 2 catégories comparées ne diffèrent dans leurs conditions que pour 7.7% des images, cette différence ne s'observe de façon statistiquement significative qu'à partir de 50 ms, c'est-à-dire 200 ms après la présentation du premier stimulus de la paire pour ces 7.7% d'images.

Seeck et al. ont donc commis une erreur conceptuelle de 1er ordre: comparer des potentiels évoqués qui ne reflètent pas uniquement les catégories d'images (visage familier/nouveau) pour lesquels ils ont été enregistrés. Leurs conclusions sont donc erronées, et ne remettent pas du tout en question l'idée d'un traitement visuel en 150 ms; au contraire, les résultats supportent même cette idée.

5.3.2 Mêmes paradigmes, résultats contradictoires: erreur conceptuelle de 2ème ordre?

Ce type d'erreur conceptuelle n'a pas (a priori) été commis dans les autres expériences présentées à la section 5.1. S'il est vrai que les 2 conditions comparées par George et al (1997), i.e. la première et la deuxième présentation d'une même image, ont été obtenues respectivement au début et à la fin de la séance pour chaque sujet (écart entre première et deuxième présentation 8 minutes et 30 secondes), s'il est vrai que les sujets étaient impliqués dans différentes tâches pour les différentes conditions comparées par Mouchetant-Rostaing et al. (2000a,b), il n'en reste pas moins vrai que ces différences précoces d'activité électrique peuvent aussi être le reflet de différences de "haut niveau" (i.e. catégorie, familiarité, genre...) entre scènes visuelles. En particulier, le protocole expérimental de Johnson et al (1999) est similaire à celui utilisé par Thorpe et al. (1996), et les résultats obtenus sont pourtant très différents: catégorisation en 150 ms ou en moins de 100 ms?

Il semble donc que pour réconcilier ces différents résultats, il soit nécessaire de pouvoir séparer les activités électriques correspondant à des différences de bas niveau ou de haut niveau entre différentes catégories d'images. Par exemple, quel type d'activité (et à quel moment?) est spécifique du caractère "cible" ou "distracteur" d'une image, indépendemment du type d'image utilisé? Quel type d'activité (de bas niveau) est caractéristique d'une catégorie d'images, et ne dépend pas de la tâche effectuée?

6. Résumé, formulation des objectifs

La durée du traitement visuel est une contrainte fondamentale pour les modèles de la vision. Or, cette durée est aujourd'hui remise en question.

Avant de pouvoir développer une théorie du traitement visuel, il sera donc nécessaire de déterminer de façon définitive sa durée. Plus précisément, nous nous attacherons à répondre à la question suivante: combien de temps faut-il au système visuel pour activer une représentation de haut-niveau de l'information visuelle, c'est-à-dire une représentation portant une signification sémantique (e.g. catégorie du stimulus) ou comportementale (e.g. cible, distracteur) pour l'animal?

II. LA DUREE DU TRAITEMENT VISUEL

"Is it a bird? Is it a plane? Oops... Too late!"

Dans ce chapitre nous tenterons de définir précisément la durée minimale requise, après la présentation d'un stimulus visuel, pour parvenir à une représentation de "haut niveau" de cette entrée. Nous partirons du postulat que cette durée est de l'ordre de 150 ms pour une tâche de catégorisation animal/non-animal (Thorpe et al. 1996), et examinerons successivement la possibilité que cette durée soit sous-estimée, du fait de la catégorie choisie, puis la possibilité que cette durée soit sur-estimée, possibilité qui serait supportée par les différents résultats expérimentaux présentés au chapitre I section 5.1.

1. Catégorie "animal": un cas particulier?

1.1 Le problème

La durée du traitement visuel obtenue par Thorpe et al. (1996) pour une tâche de catégorisation animal/non-animal pourrait être une sous-estimation de la durée de catégorisation dans le cas général. En effet, la catégorie "animal" pourrait constituer un cas particulier pour le traitement visuel, du fait de son importance biologique: les animaux ont toujours été présents aux côtés de l'homme au cours de l'évolution, et porté une signification importante pour sa survie (nourriture, prédateur □). Il est donc envisageable que l'évolution ait favorisé cette catégorie de stimulus visuel, par exemple en lui dédiant un système de traitement spécial, "précâblé", et plus rapide. Il est donc nécessaire de comparer le temps requis pour catégoriser des animaux au temps de catégorisation d'une classe de stimuli portant un minimum de signification biologique. C'est ce que nous avons choisi de faire dans l'article suivant, en utilisant une catégorie d'objets fabriqués par l'homme, et qui ne sont présents dans son environnement que depuis quelques siècles, voire quelques décennies pour certains d'entre eux : les moyens de transport.

1.2 Article n° 1 : **VanRullen & Thorpe, 2000.** Is it a bird, is it a plane? Ultra-rapid visual categorization of natural and artifactual categories. ***Perception.***

Submitted to ***Perception***

Is it a Bird? Is it a Plane?
Ultra-Rapid Visual Categorisation of Natural and Artifactual Objects.

Rufin VanRullen*, Simon J. Thorpe
Centre de Recherche Cerveau et Cognition, UMR 5549, CNRS-UPS, 133 rte de Narbonne, 31062 Toulouse Cedex, France.
email : rufin@cerco.ups-tlse.fr thorpe@cerco.ups-tlse.fr

Abstract : Visual processing is known to be very fast in ultra-rapid categorisation tasks where the subject has to decide whether a briefly flashed image belongs to a target category or not. Human subjects can respond in under 400 ms, and ERP studies have shown that the underlying processing can be done in less than 150 ms. Monkeys trained to perform the same task have proved even faster. However, most of these experiments have only been done with biologically relevant target categories such as animals or food. Here we performed the same study on human subjects, alternating between a task in which the target category was "animal", and a task in which the target category was "means of transport". These natural images of clearly artificial objects contained targets as varied as cars, trucks, trains, boats, aircraft and hot air balloons. However, the subjects performed almost identically in both tasks, with reaction times not significantly longer in the "means of transport" task. These reaction times were much shorter than in any previous study on natural image processing. We conclude that, at least for these two super-ordinate categories, the speed of ultra-rapid visual categorisation of natural scenes does not depend on the target category, and that this processing could rely primarilly on feed-forward, automatic mechanisms.

Keywords : categorisation, visual processing, natural categories, artificial categories, vision, speed of processing.

1. Introduction

The analysis of the visual scene is to human observers an effortless process that seems almost instantaneous. However, it requires a tremendous amount of processing, such that even the most sophisticated artificial vision systems are very far today from reproducing human performance. Yet a number of recent studies have underlined the fact that this kind of processing is achieved with relative ease and impressive speed by human and non-human primates even under the most constraining conditions. In a task where the subject has to decide whether a briefly flashed photograph (presentation time 20 ms) contains an animal or not, humans were able to respond in under 400 ms (Thorpe et al 1996), with a bias towards correct responses starting 280-290 ms after stimulus onset (Fabre-Thorpe et al 1998a). Monkeys have proved even faster, with mean reaction times around 250 ms, and a bias towards correct responses starting before 200 ms (Fabre-Thorpe et al 1998a). In the same study, another monkey was able to categorise images containing "food" vs "non-

* Corresponding author.

food", and in a recent experiment by Vogels (1999a), two monkeys were trained to categorise natural images of "trees" vs. "non-trees" with average reaction times inferior to 250 ms. This special kind of rapid visual processing displays a number of interesting properties. First, it does not require foveal vision, and images can be presented randomly at different eccentricities without a concurrent cost in accuracy (Fabre-Thorpe et al. 1998b; Thorpe et al, 1999). Second, the presence of colour information in the stimulus images is not necessary for this form of rapid visual categorisation (Delorme et al, in press), suggesting a major role of the magnocellular pathways in the underlying information transmission and processing mechanisms. Finally, these rapid reaction times can not be shortened by familiarity with the stimulus images, even after 3 weeks of intensive training (Fabre-Thorpe et al, in press).

These results are fundamental for current theories of visual processing because they give an upper limit to the time required by the visual system to analyse a complex scene. Because the reaction time also includes the time needed to generate the motor command, the actual duration of visual processing itself is likely to be much shorter. Indeed, Thorpe et al (1996) demonstrated that event-related potentials (ERPs) to target and distractor images diverge strongly at roughly 150 ms post-stimulus. Single unit recordings in the anterior temporal cortex of rhesus monkeys showed a selectivity to exemplars of the target category ("tree") as early as 80-90 ms after stimulus presentation (Vogels 1999b). Neurons in IT are also known to respond selectively to faces of humans or monkeys after only 100 ms (Perrett et al 1982; Oram and Perrett 1992). The implications of these findings for computational theories of visual processing are difficult to avoid: given the number of synaptic stages between the retina and higher-level visual areas such as IT, processing is very likely to be done with a single feed-forward pass through the system, and with very few spikes emitted by each neuron (Thorpe and Imbert 1989).

While the specific "ultra-rapid visual processing" described in these studies can certainly not account for the full range of tasks that can be performed by the human visual system, an ability to analyse the visual environment rapidly is undoubtedly a strong and critical component of biological visual systems. The survival of an animal often depends on the speed with which predators, prey, or food can be recognised. Therefore, it is fundamental to determine precisely the conditions under which this form of ultra-rapid visual processing can occur. In particular, the results presented above cannot be generalised yet to the categorisation of all kinds of complex visual scenes, because in the experiments mentioned the target categories used had a high "biological" relevance to the behaviour of the primate subject. Animals, food, and even trees have been present in the primates' natural environment for millions of years, and have a strong meaning in terms of the survival of the observer. Therefore it is possible that they constitute a special case for vision, and that their processing might rely on hardwired innate mechanisms, genetically inherited, or at least highly specialised. Indeed, there is some evidence for such differences in that speed of name retrieval differs when identifying pictures of natural objects vs. artifacts (Humphreys 1999).

Here we present an experiment where human subjects had to categorise complex scenes on the basis of the presence or absence of a clearly artificial category: "means of transport". The target images were natural colour photographs chosen to be as varied as

possible, including many examples of cars, trucks, trains, boats, airplanes, helicopters, hot air balloons, etc... (see examples in figures 2 and 3). To allow direct comparison between performance in this task and performance in a categorisation task involving a "natural" category, subjects had to alternate between blocks of 96 images where the target category was "means of transport" and blocks of 96 images where it was "animal". Furthermore, half of the distractor images in each task were targets for the other task, i.e. means of transport in the animal categorisation task, and animals in the means of transport task. The results presented in the next sections demonstrate that visual categorisation of artificial objects in complex scenes is neither slower nor less accurate than categorisation of biologically relevant images such as animals.

2. Materials and methods.

2.1 Procedure

The subjects were 16 volunteers, 8 males and 8 females with ages ranging from 21 to 50. They were seated in a dark room, at approximately 120 cm from a computer screen piloted from a PC computer. Two categorisation tasks with a go-no go paradigm were performed in alternation by all subjects. In each task they viewed 10 series of 96 colour images, half of which were targets and half were distractors. Whether the first block was an "animal" or a "means of transport" block was randomly decided for each subject. Then the subjects alternated between blocks of 96 images of one task and of the other. A trial was organised as follows: a white fixation point (smaller than 0.1° of visual angle) was drawn in the middle of the black screen; subjects had to press a touch-sensitive button to start the trial; the image, which subtended roughly 10° of visual angle, was flashed at the center of the screen for 20 ms; no backward masking was applied after image presentation; subjects had to release the button within 1 second if the image contained an element of the target category, and maintain pressure otherwise. The inter-trial interval was 2 seconds plus or minus a random delay of no more than 200 ms, to prevent the subjects from locking their response to the expected time of presentation. Note that the very short presentation time ensures that there was no possibility of making exploratory eye movements during image presentation. An important issue is the changes in the experimental procedure that we made compared to previous similar studies (e.g. Thorpe et al 1996). The purpose of these changes was to shorten the reaction times as much as possible. Therefore, the image size was enlarged to 10° of visual angle (instead of 5°) to increase stimulus energy, and we used a touch-sensitive plate instead of a computer mouse to record the subject's response.

2.2 Images

The pictures were complex colour scenes taken from a large commercially available CD-ROM library allowing access to several thousand stimuli. The images in each category were chosen to be as varied as possible. The animal category included pictures of mammals as well as birds, fish, insects, and reptiles. The means of transport category included images of cars, trucks, trains, civil and military airplanes, helicopters, boats, hot air balloons, and even rockets. Subjects had no a priori information on the size, position or number of the targets in a single photograph. There was also a very wide range of distractor images, which could be outdoor or indoor scenes, natural landscapes or street scenes with buildings

and roads, pictures of food, fruits, vegetables or plants, houses, man-made objects or tools. Examples of photographs are shown in figures 2, 3 and 4. In each block, subjects were presented with 96 photographs, 48 targets and 48 distractors. Half of the distractors were targets from the other task, i.e. vehicles in the animal task, and animals in the means of transport task. Furthermore, half of the vehicle images (24 per series in the means of transport task, 12 per series in the animal task) were images of cars and the other half of different means of transport, to allow further data analysis of intra-category differences. To prevent effects of image-specific learning, an image was presented once for each subject and could not appear both as a target in one task and a distractor in the other task. Image sequences were randomised so that different subjects were presented different image sequences over the same series of 96 trials, and that the presentation of a target was not predictable. Finally, all subjects alternately viewed 10 blocks of 96 trials for each task, making a total of 1920 images. The order of appearance of the series was randomly determined for each subject, as well as whether they would start with an animal or a means of transport categorisation series.

3. Results

3.1 Primary results

Reaction times and percentage correct are shown for each subject in table 1. Note that, although it was clearly not the purpose of the present study, we could not demonstrate any systematic pattern of differences between male and female subjects. Performance appears remarkable, given the complexity and the very short presentation times of the photographs. The percentage of correct responses is close to 95% in both tasks, and the median reaction time slightly above 350 ms. Some of the subjects show even better performance, with median reaction times under 300 ms and percentages correct close to 93% in both tasks. We also indicate the time at which go responses on targets significantly outnumber go responses on distractors at the $p<.001$ level (this time will now be referred to as "discrimination time"). For this we applied a χ^2 test using 10 ms bins (1 d.f.), and determined the first bin for which $p<.001$ for at least 5 consecutive bins in the case of data from a single subject, and for at least 30 consecutive bins when using the data pooled from all 16 subjects. Because of the strictness of the criteria used, no correction was made for repeated testing. The obtained discrimination time can be inferior to 250-260 ms in some of the subjects.

	Animal Task				Transport Task			
Subject	**Percentage Correct (%)**	**Mean RT (ms)**	**Median RT (ms)**	**Discrimination time (ms)**	**Percentage Correct (%)**	**Mean RT (ms)**	**Median RT (ms)**	**Discrimination time (ms)**
bat	98.12	413.6	400	335	96.35	424.1	411	345
cha	97.08	397.9	387	325	97.92	431.1	412	335
del	89.58	332.1	321	285	85.62	348.2	328	295
fab	97.92	358.1	347	305	98.44	374.9	350	305
fiz	91.77	389.3	372	315	93.23	410.5	397	325
gil	96.56	467.3	458	395	94.48	472.3	457	375
lag	96.46	440.5	429	375	95.31	446.5	432	365
mar	96.77	305.2	293	265	91.77	336.6	307	265
mas	90.21	293.3	290	265	87.19	316.4	296	275
paq	92.60	332.0	317	285	90.73	335.4	318	275
per	95.62	398.5	380	325	95.94	395.5	369	315
pet	93.44	382.1	376	335	93.23	399.1	392	305
tay	93.33	334.0	327	275	92.81	336.1	332	265
tho	93.02	327.5	316	275	95.00	328.4	306	255
trz	96.98	357.3	344	285	97.08	351.4	333	285
van	92.71	296.0	288	255	92.60	309.1	293	245
Mean	94.51	364.0	352.8	306.2	93.61	376.0	358.3	301.9
Overall	**94.51**	**363.8**	**350**	**225**	**93.61**	**375.6**	**357**	**245**

Table 1. Percentage correct, mean and median reaction times are presented for each of the 16 subjects in both tasks, as well as the time after stimulus presentation (±5 ms) when the distribution of reaction times for correct responses becomes significantly (χ^2 test over 10 ms periods, 1 degree of freedom, $p<.001$ for more than 5 consecutive periods) different from the distribution for false positives (discrimination time). Mean values over the group of subjects are also shown. The last line shows the performances obtained when the statistics are calculated using the complete data set obtained from all subjects (here the discrimination time is obtained using a χ^2 test over 10 ms periods, 1 degree of freedom, $p<.001$ for more than 30 consecutive periods).

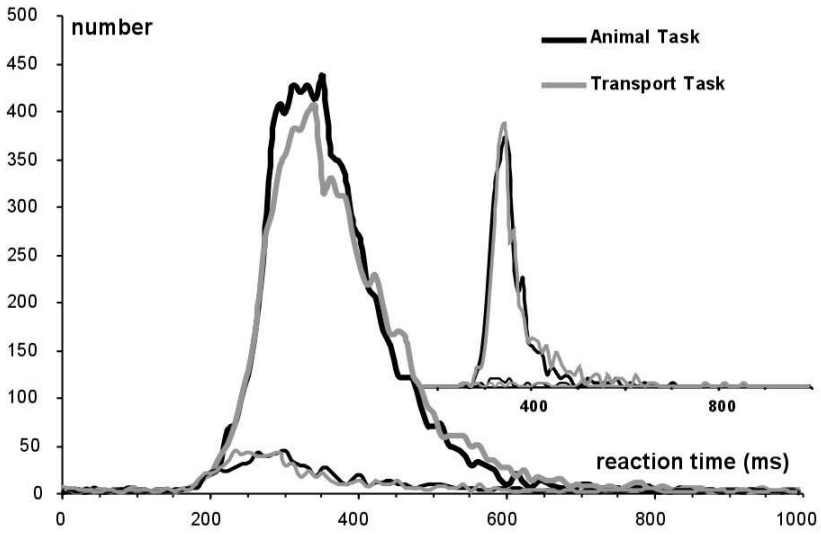


Figure 1. Histograms of reaction times for the 2 tasks. False positives reaction times are plotted as thin lines. The small inserted curve shows data from the subject fab. Note the low number of errors, the good match of the curves for the 2 tasks, and the narrowness of the histogram.

When considering the performance for all subjects, one can ask whether there is a significant difference between the animal and the transport task. Table 2 shows that there is no such difference (significance is defined as $p<.05$ for a paired t-test across subjects, 15 degrees of freedom) for either percentage correct or median reaction time. There is also no significant difference in the moment at which correct go responses significantly outnumber false positives (discrimination time). Thus subjects can perform the transport categorisation task as quickly and as efficiently as the animal task. This constitutes the primary result of this study. The statistical difference between the two tasks observed for mean reaction times only reflects the fact that there are more long-latency responses in the transport task, but the lack of a difference for the medians implies that neither task is actually faster than the other.

	Difference Animal Task *vs* Transport Task *(paired t-test, DF=15)*
% Correct	n.s. *(p=.082)*
Mean Reaction Time	p<.001
Median Reaction Time	n.s. *(p=.067)*
Discrimination time	n.s. *(p=.186)*

Table 2. Statistically significant differences between the 2 tasks (paired t-test, degrees of freedom=15). Discrimination time denotes the time (±5 ms) when correct go responses significantly (χ^2 test over 10 ms periods, 1 degree of freedom, $p<.001$ for more than 5 consecutive periods) outnumber false positives.

More detailed analysis of the reaction time histograms (figure 1) shows that the distributions for the animal and transport tasks are statistically indistinguishable (i.e. $p>.1$ for every 10 ms period, using a χ^2 test with 1 degree of freedom) before 280-290 ms. Responses with reaction times below 220 ms are probably due to anticipations, because the distributions of responses to targets and distractors are roughly equal. However, from 220-230 ms in the animal task, and 240-250 ms in the transport task (difference not significant, see table 2) the proportion of targets significantly ($p<.001$, see table 1 for a description of the statistical test) outnumbers distractors, whereas targets and distractors have the same probability of presentation (50%) : this means that the same categorisation tasks with a response time limit of 250 ms would yield a performance significantly above chance! This time window includes not only the time required for visual processing by itself, but also the decision process and the motor output. This severely limits the time available for visual processing, and therefore is a very important result in regard to current computational theories of visual processing.

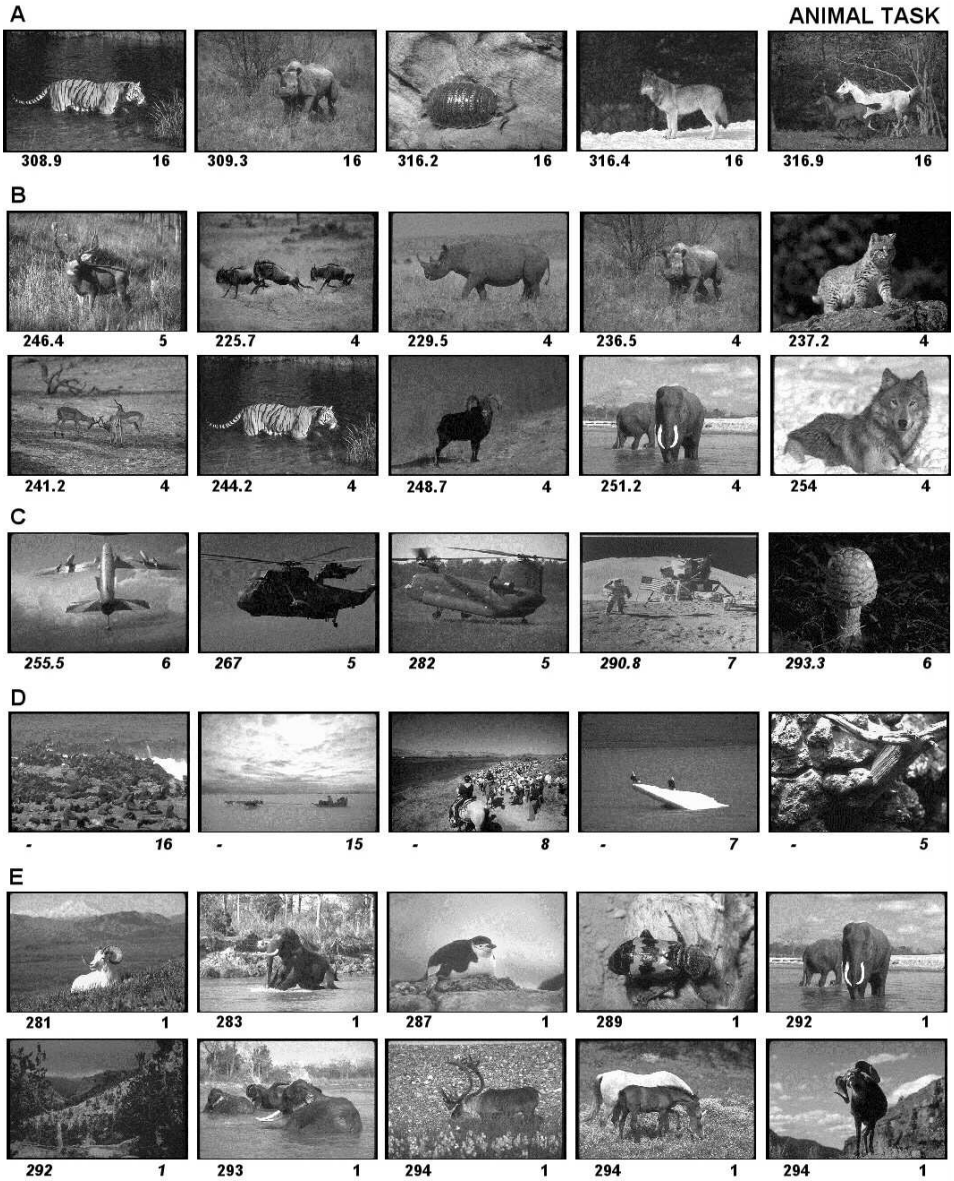


Figure 2. Examples of images presented in the "animal" categorisation task. Original images were presented in colour. Mean reaction times and number of responses (out of 16 subjects) are mentioned when available below each image, except for panel D which shows the number of missed trials (out of 16 subjects). Errors are presented in italics. A. Fastest targets. Target images having the shortest mean reaction times. Note that all of these images were correctly categorised by all subjects. B. Targets categorised by the most subjects under 270 ms. The number of responses and mean reaction times are calculated only on those trials with response times inferior to 270 ms and superior to 150 ms (to sort out obvious anticipation trials). C. False positives from 5 or more subjects with the shortest reaction times. D. Most commonly missed targets. The number of subjects that did not respond to these targets is indicated. E. Fastest trials from subject fab, including potential false positives.

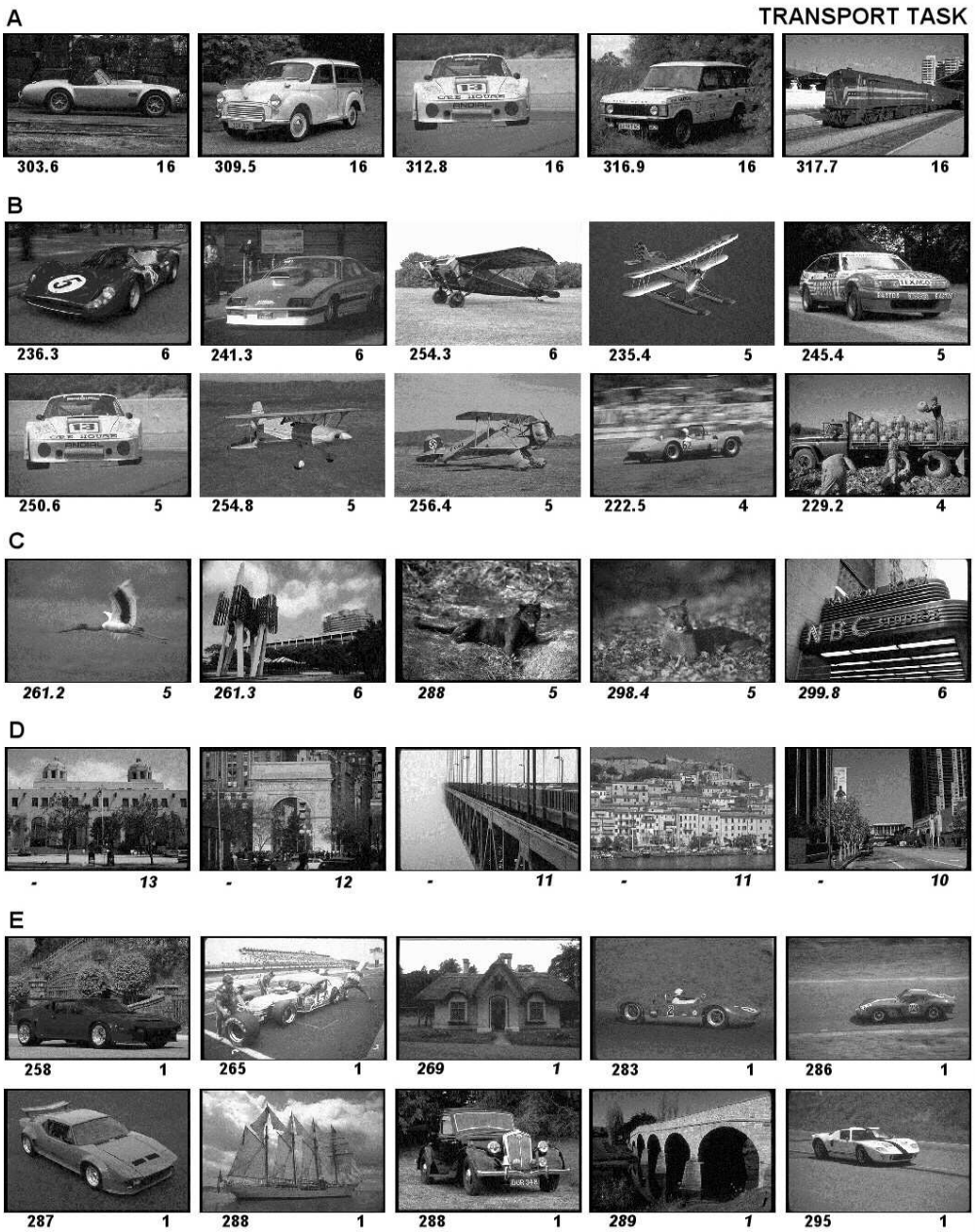


Figure 3. Examples of images presented in the "transport" categorisation task. Notations as in figure 2. See text for details.

The examples of images that were categorised the most rapidly, as well as the images that were incorrectly categorised by the most subjects (figures 2 and 3) illustrate the complexity of the two tasks. Target objects are often highly visible, but there does not seem to be any preferred size, ("canonical") orientation, nor any preferred subcategory (car images represented half of the set of vehicle images, thus a high proportion of cars would be expected). Since half of the distractor images belonged to the target category of the other task, they are as likely to appear in the false positive trials as other "regular" distractors. However, images of aircraft or helicopters on a sky background (figure 2C) seem to be frequently mistaken for animals (most probably birds or flying insects). The targets missed by the most subjects seem to have in common the small size and uncentered position of the target objects in these images. Nevertheless, these objects were perfectly visible since images subtended 10° of visual angle. This effect might thus be due to the very short presentation time (20 ms). Finally, the fastest trials from subject fab (figure 2E and 3E) show that correct responses can be obtained on varied images even with very short reaction times. Only one error is observed out of these 10 fastest trials in the animal task, and two in the transport task.

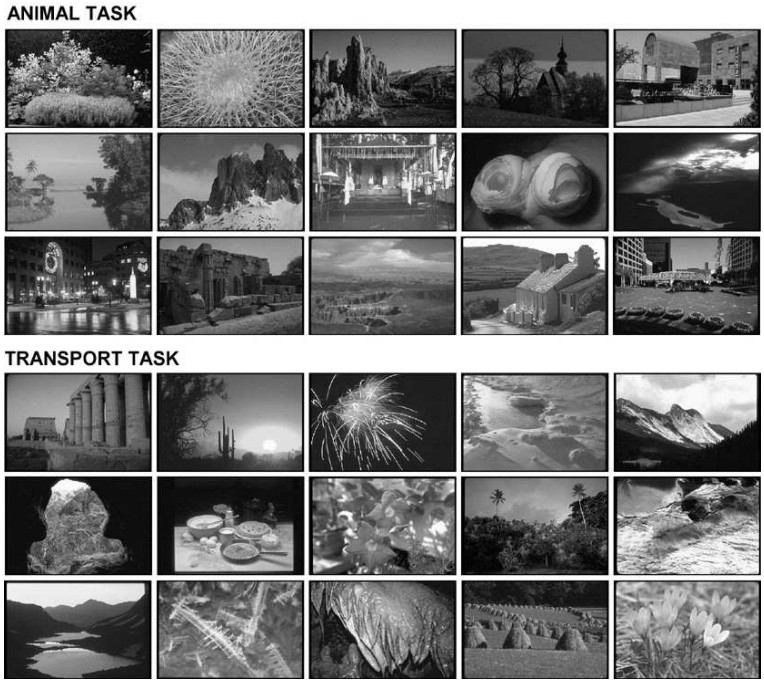


Figure 4. Examples of distractor images (non-animal, non-vehicle) correctly categorised by all 16 subjects in the 2 tasks. In the "animal" task, out of 240 distractor images, 116 were correctly categorised by all 16 subjects, and 231 by more than half the subjects. In the "transport" task, out of 240 distractor images, 108 were correctly categorised by all 16 subjects, and 236 by more than half the subjects. The great variety of distractor images and the performance obtained argue against the possibility that categorisation could be based on a limited set of simple low-level features.

Of the 480 distractor images that were neither animals nor means of transport, 224 were correctly ignored by all 16 subjects. Figure 4 provides examples of such stimuli and illustrates the very wide range of photographs presented to the subjects. Clearly, distractor images cannot be defined uniquely by a simple feature such as their configuration (for example, the presence of the sky), their overall luminance, contrast or range of spatial frequencies. The strategy used by the subjects must therefore reflect the diversity of the images to be categorised.

3.2 Influence of the other categorisation task

Our experimental procedure also allows comparison between the false positive trials on "regular" distractors and on distractors that would be targets in the other task (i.e. means of transport in the animal task, animals in the means of transport task). There were far more errors for stimuli that changed status (900 for all subjects in both tasks) than for the ones that were always to be treated as a distractor (532) despite the fact that the probabilities of presentation are equal (figure 5). Yet there is no significant difference (i.e. $p>.1$ for every 10 ms period, using a χ^2 test with 1 degree of freedom) between the distributions of responses for these 2 categories of distractors before 230-240 ms (where significance level reaches $p<.02$), strengthening the idea that RTs after this critical limit reflect the complex visual processing that is required for category discrimination.

There seems to be a strong influence of the other task. Subjects had difficulties withholding responding to images that would be targets in the other task, i.e. they cannot completely switch from a task to the other without a concurrent cost in accuracy. Although the alternation between the two tasks certainly leads to biases in the observed performances, it is important to stress that this procedure means that performance can be compared under similar conditions for the two tasks. Moreover, mixing targets from the other task among the other distractors ensures that the subjects cannot rely on systematic low-level differences between the target and distractor images.

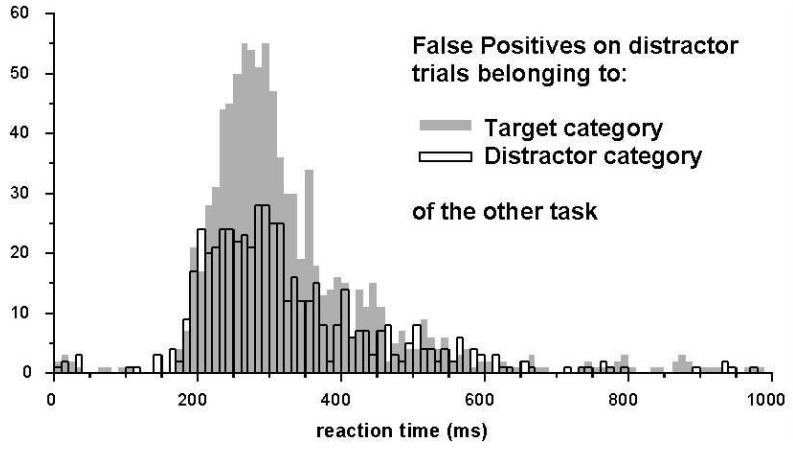


Figure 5. Histogram of reaction times for the false positive trials in both tasks. Trials are separated according to the type that the image would have in the other task : distractor or target. Note that the probability of presentation is identical for both groups.

3.3 Intra-category differences

The subdivision of the means of transport photographs into the two subcategories "cars" and "other vehicles" made before the start of the experiment allows us to investigate intra-category differences in processing. We can compare the car images and the other targets in the transport categorisation task. Cars appeared to lead to a significantly higher proportion of responses with reaction times in the period from 250-260 ms to 400 ms (i.e. a χ^2 test with 1 degree of freedom was significant at the $p<.01$ level for every 10 ms period in this interval; see figure 6). Nevertheless, the distributions for cars and other vehicles both significantly differ from the distribution of false positives as early as 240-250ms (χ^2 test with 1 degree of freedom significant at the $p\leq.001$ level for more than 30 consecutive 10 ms periods). Therefore both categories can be successfully (i.e. better than chance) discriminated from distractors in less than 250 ms. Although the categorisation of car images would give a better performance than the categorisation of "other vehicles" if the analysis was restricted to responses with reaction times below 400 ms, if we take into account all the responses made before the response time limit of 1s, there is no such advantage: the percentage correct on cars and other vehicles were respectively 96.02% and 96.95%, i.e. virtually equal. The difference observed between cars and other vehicles might result from the fact that cars were more common than other forms of transport. If subjects happened to realise this, then they might, voluntarily or not, bias or prime visual processing to favour features present in photographs of cars. Another related possible explanation of the discrepancy between these 2 sub-categories could be the level of categorisation, i.e. basic-level for cars vs. superordinate level for "other vehicles": if the subjects were actually searching for "cars" more actively than for other vehicles, performance would reflect a basic-level categorisation, even if subjects were instructed to categorise images at the super-ordinate level (with the target category "means of transport"). Indeed, numerous experiments have shown an advantage for categorisation of images at the basic vs. superordinate level (Rosch et al 1976), although other studies have reported that this advantage could decrease when objects are presented in whole scenes (Murphy and Wisniewski 1989), and ERP recordings have recently shown that this effect was only visible roughly 300 ms after stimulus presentation (Tanaka et al 1999).

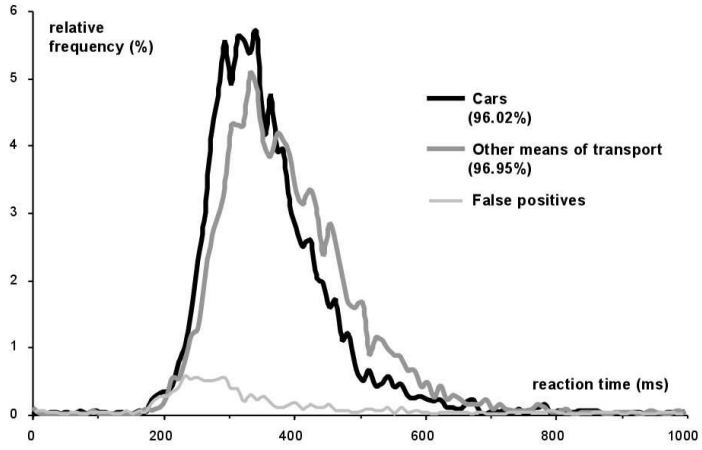


Figure 6. Histograms of reaction times in the vehicle categorisation task. Trials are separated according to the sub-category "car" or "other means of transport". Percentages correct are indicated for each sub-category. False positives are also shown (thin line).

4. Discussion

The results obtained in this experiment bring together evidence that human subjects can successfully differentiate complex visual categories in less than 250 ms. Even if median reaction times in the categorisation tasks presented in this study are around 350 ms, there is a significant bias towards correct responses from as early as 240-250 ms, for animal categorisation as well as means of transport. The difference in the distributions of false positives for "regular" distractors and the distractor images belonging to the target category of the other task also begins shortly before 240 ms. A very plausible explanation of this effect is that enough visual processing is done already to allow distractor images to be rejected (or, equally, targets to be recognised), while the other task performed in alternation biases the responses and makes it more likely for the subject to respond on distractors that would be target trials in the other task. Finally, this remarkable speed of processing is not specific to a basic-level categorisation, because both categories "animals" and "means of transport" are clearly at the super-ordinate level.

To our knowledge, no other study on the categorisation of natural scenes has ever reported reaction times as short as the ones presented here. The difference between reactions times reported here and those in previous similar experiments (e.g. Thorpe et al 1996) almost certainly arises from the improvements of the experimental procedure that we described in section 2.1. First, in the present experiment, the dimensions of the images were twice those used previously, thus increasing the amount of energy in the image. Second, we used a touch sensitive plate which means that the subjects were not required to apply pressure to keep the button pressed, as was the case previously. This certainly means that the motor reaction times can be significantly improved. Even simple reaction times to sinusoidal gratings can be longer than 350 ms depending on the contrast and spatial frequency of the stimulus, and are rarely found to be shorter than 200 ms (Breitmeyer 1975; Ejima and Ohtani 1987). With some subjects, the detection of a small flashed (50 ms) spot can take as long as 280 ms on average, even when the intensity of the stimulus is 100 times greater than the threshold (Lennie 1981). This reaction time can be reduced to 220-230 ms under gap conditions where the fixation point is turned off shortly before the target spot appears (Fischer and Rogal 1986). It is commonly said that minimal visual reaction times are not shorter than 180 ms, and an average estimated value of 250 ms fits the data from various studies fairly well (Luce 1986). These studies, however, cannot easily be compared to ours because of the simple kind of stimulation that they used. An experiment using natural images as the stimuli triggering detection would allow us to compare our reaction times with simple reaction times obtained under the same conditions.

Such an experiment had been conducted in the laboratory previously to the one reported here. Stimuli were natural images of animals or distractors that could include means of transport, and were presented at half the size used here. The device recording the subjects' response was the same type of touch-sensitive plate that we used. Five subjects of the present study (cha, lag, mar, mas, paq) were asked to perform 3 series of 100 trials where they had to respond to each image whatever the category involved. They performed these series in alternation with 3 series of 100 trials of the animal/non animal categorisation task. Their mean reaction time on the latter task was 411 ms (with 92.3% correct responses), and 227 ms on the simple detection task. However, a considerable proportion of

the trials in the simple detection series were anticipatory responses, as supported by the fact that in 4% of the trials the response occured before stimulus presentation, and in 2% it occured within 100 ms. Even more strikingly, the 2 fastest subjects in this task (both with mean reaction times under 200 ms) averaged 12% of anticipatory responses. These anticipations could not be avoided even with randomised (2200±800 ms) inter-stimulus intervals. Therefore the reported average reaction time of 227 ms is likely to be underestimated. When comparing this simple reaction time to the ones obtained in our categorisation study, it appears that the average "cost in reaction time" of processing a complex visual scene is probably around 100-120 ms, and in any case no more than 150 ms. Furthermore, we have shown that the animal and means of transport categorisation tasks could be performed accurately (i.e. above chance) with reaction times below 250 ms, a value surprisingly close to the simple reaction time of 227 ms. Thus, the complex processing required to detect objects belonging to a target visual category can be done almost as fast as a simple perceptual signal detection task. A very tempting interpretation of these results is that categorisation does not occur after perception, but these two mechanisms are performed in parallel, with some or all of the involved processing being shared or common. This theory has already been proposed by Schyns and Oliva (1999) who showed that the categorisation task being performed could strongly modify the underlying perceptual mechanisms. Here we strengthen this argument by demonstrating that reaction times in a high-level visual categorisation task can be very close to simple visual reaction times.

It is important to stress once again that all reaction times mentioned above include not just visual processing, but also the decision process and the production of the motor output. The time actually required for visual processing by itself might thus be considerably shorter. If we substract 80-100 ms, e.g. the minimum time required to generate a "reaching command" (Kalaska and Crammond 1992), it would appear that the necessary visual mechanisms involved in the categorisation tasks presented here can take no longer than 150 ms. Of course, visual processing does not abruptly stop after 150 ms, but this time of processing appears sufficient for the visual system to successfully discriminate between visual categories. The same value of 150 ms had previously been derived from ERP recordings (Thorpe et al 1996) and our results only strengthen this demonstration.

But the main finding of this study is that the previously obtained results on the speed of visual processing can be generalised from natural to artificial categories. We have shown that the categorisation of means of transport is neither slower in reaction times nor less accurate than the categorisation of animals, a natural, highly biologically relevant category. It is very unlikely that the length of the decision process or the motor command would be significantly different between these two categorisations. Therefore we can easily conclude that visual processing is as fast for man-made categories as it is for natural ones. This clearly rules out the possibilty that the Ultra-rapid Visual Categorisation of animals reported previously could be a special case. Rather, it seems likely that similarly rapid go/no-go visual categorisation could be performed for a wide range of categories. Clearly, further experiments will be needed to test the precise limits of this ability. In particular, it will be important to test whether more specific basic level categories (such as cars) are really detected earlier than more global categories (such as means of transport). The

surprisingly good performance and very short reaction times obtained here cast doubt on the intuitive idea that visual processing would require a basic-level identification of the stimulus before its potential super-ordinate level categorisation (Rosch et al 1976). Indeed, this implies that basic-level categorisation should be performed faster than the kind of categorisation that we report here. While the organisation of the semantic system can certainly explain why naming tasks are performed slower at the super-ordinate level (Humphreys et al 1999), it seems unlikely however that any visual processing task requiring a high-level analysis of the visual scene could be performed with reaction times much shorter than the ones that we obtained here for a super-ordinate categorisation.

A further point that is worth investigating is whether the observed performance can be explained by a simple low-level visual feature or property. For example, it has been shown that pigeons strongly rely on texture information to categorise natural scenes (Troje et al 1999). But that does not mean that texture information is sufficient to perform the task. If a single low-level property was sufficient to characterise distractor images in the two tasks (for example the presence of the sky), and categorisation was based on this property, then including animals in the distractors for the "means of transport" categorisation (and vice-versa) should severely impair performance (because this property would not be sufficient to discriminate between animals and vehicles). However, the performances obtained by the subjects in our experiment are if anything better than the ones reported in Thorpe et al (1996) for example. Therefore, distractor images can not be described by a simple feature. Furthermore, the very wide range of pictures used in each target category effectively rules out the possibility of a single low-level visual dimension accounting for all the variance in these images. In his study involving a "tree" vs. "non-tree" categorisation, Vogels (1999a, 1999b) systematically investigated the contribution of low-level features such as texture, color and size to the categorisation performance. None of these features was found to account for the subjects' performance. Recently, another experiment involving the same sort of rapid categorisation task (animal vs non-animal, food vs non-food) used here demonstrated that fast visual processing in man and monkeys does not rely on colour cues (Delorme et al, in press). Finally, a recent categorisation study using the same experimental procedure used here showed that the categorisation of simple forms on the basis of a single low-level feature such as shape could be done 50 ms faster (for mean reaction times) than the animal-non animal task (Aubertin et al 1999). If the latter also relied on a single low-level dimension, one would certainly not expect such a discrepancy.

Another related issue is the influence of context on categorisation performance. Although contextual effects can not be simply ruled out, and certainly take place for some images, these effects alone cannot account for the obtained performance. For example, the target images in figures 2 and 3, panels D, were missed by the most subjects, although they present prototypical backgrounds where the presence of a target would be expected (nature scenes for the "animal" task, street and city scenes for the "means of transport" task). Similarly, many of the distractor images presented in figure 4 represent typical contexts for the presence of an animal or a vehicle. Nevertheless, none of the subjects responded to any of these distractors. Moreover, even if subjects were using the scene context as a cue for categorisation, this would not make the task intrisically easier: a street scene certainly makes the presence of a vehicle more likely, but categorising a photograph as a street scene

also constitutes a rather challenging task for the visual system.

The fact that the visual mechanisms described here are too complex to rely on a simple stimulus feature estimation does not mean, of course, that it cannot be based on a combination of many such features. A recent neural network model (Campbell et al 1997) demonstrated that it is possible to classify outdoor complex scenes on the basis of a set of low-level features. However, the kind of processing required by the tasks used in our experiment should not be considered as "simple". We believe that the critical issue in defining the complexity of a visual task is not the complexity but rather the number of features that need to be used in parallel to perform the task. Whereas it is true that a few responses could be obtained, for example in the "means of transport" task, with a simple "tyre detector" (responding to a gray disc and a black surround), this single stimulus dimension could not account for the overall performance obtained in our task. Many such feature detectors (tyre, windscreen, headlights, airplane wings, sail, propeller blade, etc…), working in parallel on the input image, would be needed to reach human performance in this task. And (without entering the debate between local and distributed representation) even if no specific "car" cell was activated, the concurrent activation of the neural representations of a tyre and a windscreen and headlights (etc…) could still unequivocally define the object in the visual scene as a car. The number, rather than the complexity, of feature detectors needed to perform a given visual task can thus be considered as an indicator of the complexity of the task itself, and of the visual representations involved.

Finally, the implications of the present findings for current computational theories of visual processing need to be stressed. We have demonstrated that, even in the case of a demanding visual categorisation task requiring visual processing that would severely challenge the most sophisticated artificial vision systems, the entire input-output visuo-motor sequence can be completed in under 250 ms. Given the delays in the motor pathways, this means that the underlying visual processing has been done in not much more than 150 ms, a value that fits well with the latencies for differential ERP activity seen in the same sort of task (Thorpe et al 1996). Few if any of the currently available models of visual processing are compatible with such constraints. Given the number of stages involved in visual processing and the mean firing frequencies of neurones in the visual system, this very short time window implies that most of the processing has to be done in a feed-forward way, with probably no more than one spike emitted by each neurone (Thorpe and Imbert 1989). Although these limitations seem to strongly constrain neural network models of visual processing rather than to provide a way of improving them, recent simulations have demonstrated the power of neural coding schemes based on the temporal information transmitted by single spikes (Delorme et al 1999). For example, one such model was able to detect human faces in natural images with a much higher level of performance than classical models of face processing (VanRullen et al 1998). We believe that temporal constraints on visual processing, as highlighted by psychological studies on human subjects as well as electrophysiological recordings in animals, constitute the basis of a vivid framework in which models can be developed that could eventually approach human performance.

Acknowledgements

This work was supported by the CNRS, by the Cognisciences Program, and by the Midi-Pyrénées Region. The experimental procedures used were authorised by the local ethical committee (CCPPRB N° 9614003). The authors wish to thank M. Masset and A. Delorme for designing and carrying out the experiment on simple reaction times.

References

Aubertin A, Fabre-Thorpe M, Fabre N, Geraud G, 1999 "Fast visual categorization and speed of processing in migraine" *Compte-Rendus Académie Sciences III* **322** 695-704

Breitmeyer B G, 1975 "Simple reaction time as a measure of the temporal response properties of transient and sustained channels" *Vision Research* **15** 1411-2

Campbell N W, Thomas B T, Troscianko T, 1997 "Automatic segmentation and classification of outdoor images using neural networks" *International Journal Neural Systems* **8** 137-44

Delorme A, Gautrais J, Van Rullen R, Thorpe S J, 1999 "SpikeNET : a simulator for modeling large networks of integrate and fire neurons." *NeuroComputing* **24**

Delorme A, Richard G, Fabre-Thorpe M, 2000. "Rapid Categorisation of natural scenes is colour blind: A study in monkeys and humans." *Vision Research,* in press.

Ejima Y, Ohtani Y, 1987 "Simple reaction time to sinusoidal grating and perceptual integration time: contributions of perceptual and response processes" *Vision Research* **27** 269-76

Fabre-Thorpe M, Richard G, Thorpe S J, 1998a "Rapid categorization of natural images by rhesus monkeys" *Neuroreport* **9** 303-8

Fabre-Thorpe M, Fize D, Richard G, Thorpe S J 1998b. Rapid categorization of extrafoveal natural images: Implications for biological models. In J. Bower (Ed.), *Computational Neuroscience: Trends in Research* (pp. 7-12). New York: Plenum Press.

Fabre-Thorpe M, Delorme A, Marlot C, Thorpe S (in press). A limit to the speed of processing in Ultra-Rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience.*

Fischer B, Rogal L, 1986 "Eye-hand-coordination in man: a reaction time study" *Biological Cybernetics* **55** 253-261.

Humphreys G W, Price C J, Riddoch M J, 1999 "From objects to names: a cognitive neuroscience approach" *Psychological Research* **62** 118-30

Kalaska J F, Crammond D J, 1992 "Cerebral cortical mechanisms of reaching movements" *Science* **255** 1517-23

Lennie P, 1981 "The physiological basis of variations in visual latency." *Vision Research* **21** 815-824

Luce R D, 1986 *Response Times* (Oxford: Oxford University Press)

Murphy G L, Wisniewski E J, 1989 "Categorizing objects in isolation and in scenes: what a superordinate is good for" *Journal of Experimental Psychology : Learning Memory Cognition* **15** 572-86

Oram M W, Perrett D I, 1992 "Time course of neural responses discriminating different views of the face and head." *Journal of Neurophysiology* **68** 70-84

Perrett D I, Rolls E T, Caan W, 1982 "Visual neurons responsive to faces in the monkey temporal cortex." *Experimental Brain Research* **47** 329-342

Rosch E, Mervis C B, Gray W D, Johnson D M, Boyes-Braem P, 1976 "Basic objects in natural

categories" *Cognitive Psychology* **8** 382-**439**

Schyns P G, Oliva A, 1999 "Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations" *Cognition* **69** 243-65

Tanaka J, Luu P, Weisbrod M, Kiefer M, 1999 "Tracking the time course of object categorization using event-related potentials" *Neuroreport* **10** 829-35

Thorpe S J, Imbert M, 1989 "Biological constraints on connectionist models.", in *Connectionism in Perspective.* Eds R Pfeifer, Z Schreter, F Fogelman-Soulié and L Steels (Amsterdam: Amsterdam) pp 63-92

Thorpe S J, Fize D, Marlot C, 1996 "Speed of processing in the human visual system" *Nature* **381** 520-2

Thorpe, S.J., Gegenfurtner, K., Fabre-Thorpe M. & Bülthoff H.H. 1999 Categorisation of complex natural images in extreme peripheral vision
Perception, **28** Supplement, 61.

Troje N F, Huber L, Loidolt M, Aust U, Fieder M, 1999 "Categorical learning in pigeons: the role of texture and shape in complex static stimuli" *Vision Research* **39** 353-66

VanRullen R, Gautrais J, Delorme A, Thorpe S J, 1998 "Face processing using one spike per neuron." *Biosystems* **48** 229-239.

Vogels R, 1999a "Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study" *European Journal of Neuroscience* **11** 1223-38

Vogels R, 1999b "Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study" *European Journal of Neuroscience* **11** 1239-55

1.3 Résumé

Les temps de réaction obtenus dans l'article 1 démontrent clairement que la catégorisation d'images basée sur la présence ou l'absence de la catégorie cible "animal" n'est pas plus rapide qu'une catégorisation où la cible est définie comme "moyen de transport" : le caractère "naturel" ou "artificiel", la signification biologique des images ne semble pas déterminer la rapidité avec laquelle elles sont traitées. La durée de traitement de 150 ms mise en évidence par Thorpe et al. (1996) pour une tâche de catégorisation animal/non-animal est confirmée ici pour une tâche impliquant une catégorie cible artificielle (moyens de transport) : le temps de réaction minimal ("discrimination time") étant inférieur à 240-250 ms, la durée minimale d'initiation d'un mouvement simple étant de 80-100 ms (Kalaska et Crammond, 1992), on retrouve bien une estimation de la durée du traitement visuel de l'ordre de 150ms.

Cette vitesse de traitement ne semble pas non plus dépendre du "niveau" de catégorisation : "basic-level" ou "super-ordinate level". En effet, les temps de réaction obtenus ici pour une catégorisation au niveau "super-ordinate" sont si proches des temps de réaction simples (lorsque la tâche est une simple détection de signal visuel) qu'il semble presque impossible qu'une catégorie cible différente, même au niveau "basique", puisse raccourcir ces temps de réaction. De même, ceci jette le doute sur les résultats expérimentaux présentés au chapitre I section 5.1. Si la durée de traitement lors d'une tâche de catégorisation complexe était 100 ms plus courte que pour la tâche "animal/non-animal", on s'attendrait à des temps de réaction 100 ms plus courts également, i.e. inférieurs à des temps de réaction simples !

2. Séparation des mécanismes visuels de bas- et haut-niveau

Nous examinerons dans cette section la possibilité que la durée du traitement visuel de 150 ms mise en évidence entre autres par Thorpe et al. (1996) soit une sur-estimation du temps que met réellement le système visuel pour accéder à une représentation de haut niveau de l'entrée visuelle. Cette hypothèse est suggérée par les résultats expérimentaux récents présentés au chapitre I section 5.1. Résoudre ce problème revient en fait à répondre à la question suivante: l'activité électrique précoce (avant 150 ms) reflète-t-elle des propriétés sémantiques du stimulus visuel **indépendemment de ses propriétés physiques ou sensorielles**? Le seul moyen de répondre à cette question est de pouvoir comparer deux classes de stimuli "sémantiquement" différents (par exemple en termes de la réponse demandée au sujet), mais possédant les mêmes propriétés physiques ou sensorielles.

2.1 Un nouveau paradigme

Plutôt que d'enregistrer simplement l'activité électrique pour deux différentes catégories, et de considérer que le système visuel a effectué la catégorisation dès que ces 2 activités électriques divergent, nous proposons ici d'utiliser une seule catégorie de stimuli, et d'enregistrer l'activité évoquée dans 2 conditions différentes: lorsque cette catégorie doit engendrer une réponse, et lorsqu'elle doit être ignorée. Alors seulement l'activité différentielle entre ces 2 conditions pourra être considérée comme le reflet d'une catégorisation.

Plus précisément, un nouveau type de paradigme expérimental peut permettre de réaliser cette comparaison. Il s'agit pour les sujets de réaliser en alternance 2 tâches de catégorisation, impliquant 2 catégories cibles différentes. Nous utiliserons ici les catégories "animal" et "moyens de transport". En incluant des cibles d'une tâche parmi les distracteurs de l'autre, on obtient pour une même catégorie d'objets 2 conditions qui ne diffèrent que par leur signification comportementale: "cible" ou "non-cible". En comparant ces deux conditions, on peut donc observer directement les traitements "haut-niveau", qui reflètent le mécanisme de catégorisation lui-même, toutes propriétés sensorielles (bas-niveau) des stimuli étant égales par ailleurs. Ceci permet donc d'éviter les erreurs conceptuelles mentionnées en I.5.3.

C'est ce paradigme expérimental, que nous avons appelé "dual-task paradigm", qui a été utilisé dans l'article suivant.

2.2 Article n° 2 : **VanRullen & Thorpe, 2000.** The time course of visual processing: from early perception to decision-making. ***Journal of Cognitive Neuroscience.***

The time course of visual processing: from early perception to decision-making.

Rufin VanRullen*
Centre de Recherche Cerveau & Cognition

Simon J. Thorpe
Centre de Recherche Cerveau & Cognition

Abstract

■ Experiments investigating the mechanisms involved in visual processing often fail to separate low-level encoding mechanisms from higher-level behaviorally relevant ones. Using an alternating dual-task event-related potentials (ERP) experimental paradigm (animals or vehicles categorization) where targets of one task are intermixed among distractors of the other, we show that visual categorization of a natural scene involves different mechanisms with different time courses: a perceptual, task-independent mechanism, followed by a task-related, category-independent process. Although average ERP responses reflect the visual category of the stimulus shortly after visual processing has begun (e.g. 75-80 ms), this difference is not correlated with the subject's behavior until 150 ms post-stimulus. ■

INTRODUCTION

How do we perceive and understand a natural scene? What are the mechanisms involved in the extraction of its meaning, and how do they temporally relate to each other? Current theories of visual processing suggest a distinction between two different mechanisms: a perceptual process extracting information about different properties of the visual input, followed by a higher-level decision process evaluating the relevance of this visual information, in terms of the goals and expectations of the subject, in order to prepare and generate the appropriate behavioral response (Shadlen and Newsome, 1996; Romo and Salinas, 1999; Schall & Thompson, 1999). How these two mechanisms can be dissociated in time and space is a fundamental issue that is not easy to address (Schall and Thompson, 1999; Nichols and Newsome, 1999). In humans in particular, using reaction times as a dependent variable precludes the separation of the respective durations of perception, decision and motor response stages (Luce, 1986).

Event-related potentials (ERPs), which can be recorded even in the absence of a behavioral response, have been shown to reflect high-level properties of the visual stimulus such as its identity or category after roughly 150 ms (Thorpe et al., 1996; Bötzel et al., 1995; Schendan et al., 1998; Allison et al., 1999; Mouchetant-Rostaing et al., 2000). This holds for faces (Jeffreys, 1996) as well as other objects (Rossion et al., 2000). More recently, magneto-encephalography (MEG) has demonstrated a similar pattern of results (Halgren et al., 2000). Likewise, experiments using electrophysiological recordings in monkeys have reported neural responses highly selective to a specific visual category even before 100 ms (Perrett et al., 1982; Oram and Perrett, 1992; Vogels, 1999). A number of recent ERP studies in humans (Seeck et al., 1997; George et al., 1997; Debruille et al., 1998; Mouchetant-Rostaing et al., 2000) seem to suggest, however, that high-level properties of the visual stimulus could be extracted much faster, sometimes reporting "face recognition" (Seeck et al., 1997) or "face gender discrimination" (Mouchetant-Rostaing et al., 2000) effects as early as 50 ms post-stimulus.

* Corresponding author.

However the fact that neural activity varies with the properties of the visual input is not sufficient to conclude that the subject can actively recognize the identity of the categories involved. For example, it could simply reflect differences in the low-level properties of the visual stimuli, differences that are difficult if not impossible to investigate systematically.

Here we present an alternating dual-task paradigm ERP experiment that allowed us to compare the processing of the same visual category having different task-related status, and conversely, different visual categories having the same behavioral status. We were able to dissociate *(i)* a low-level sensory analysis activity and *(ii)* a high-level task related mechanism that was clearly independent of the sensory properties of the stimulus.

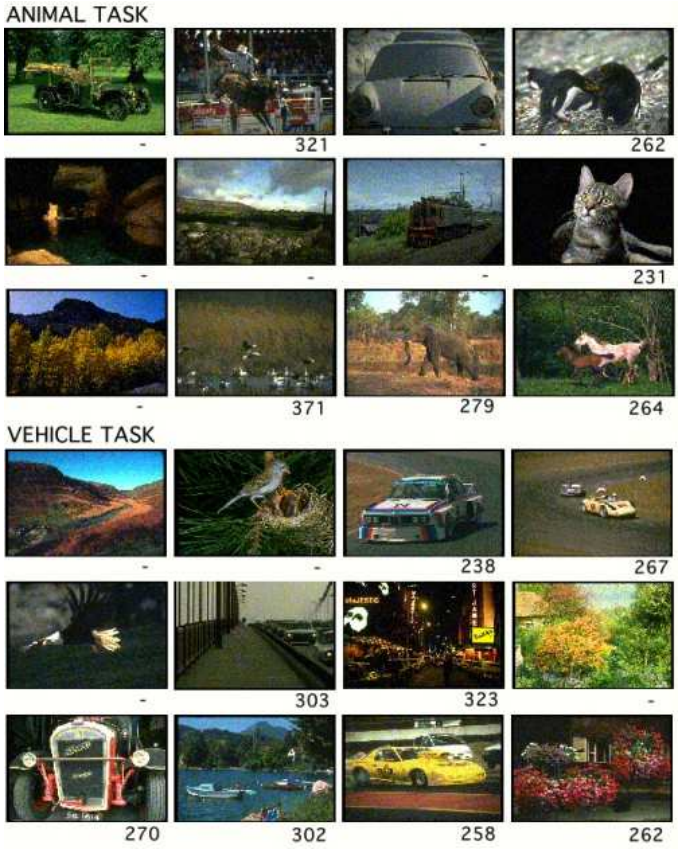


Fig.1. Two sequences of 12 images presented in the animal and the vehicle categorization tasks, with reaction times from the fastest subject (median reaction times 288 ms in the animal task, 293 ms in the vehicle task, percentages correct 92.7% and 92.6% respectively).

RESULTS

Animal and vehicle tasks

In one task, replicating the go/no go experiment from Thorpe et al. (1996), 16 subjects were seated in front of a computer screen and had to release a button if the natural photograph that was flashed for only 20 ms contained an animal. ERPs were recorded concurrently on 32 electrode sites. In the second task, the same 16 subjects were asked to respond on images belonging to the target category "means of transport". In each task, half of the non-target images belonged to the target category of the other task (i.e. vehicles in the animal task, animals in the vehicle task). The other half were distractor scenes that contained no animal or vehicle. To compare the two tasks under the same conditions, 10 series of 96 trials for each task were performed in alternation (20 in total). In different blocks, images of one particular category could be treated either as targets or as non-targets, so as to allow comparison of the processing of the same visual category in target and non-target trials. Furthermore, half of the vehicle scenes were pictures of cars, in order to perform intra-category comparisons. Images in each category were chosen to be as varied as possible. Figure 1 shows two examples of image sequences that were viewed by one subject. Animal images could contain mammals but also birds, fish or insects. Vehicle images included cars as well as trucks, trains, boats, airplanes, helicopters or hot air balloons. The size, number and position of the targets in the scenes were unpredictable. There was also a very wide range of distractor scenes that could contain landscapes, trees, roads with buildings, man-made objects, etc...

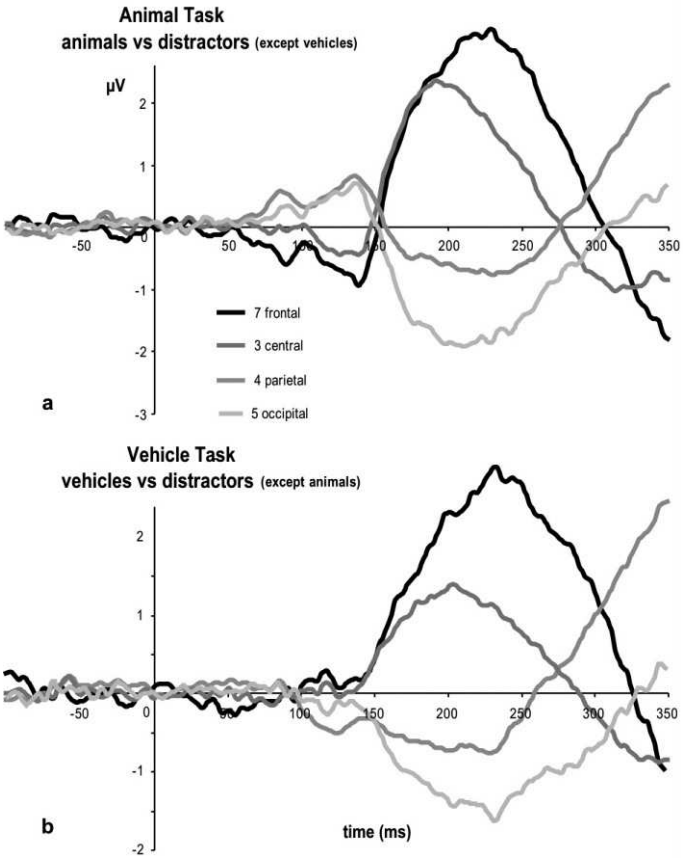


Fig. 2. Difference waveforms between target and distractor images in the two (animal and vehicle) categorization tasks (not including the non-target images that belong to the target category of the other task). Grand-average over all 16 subjects. Each panel shows the difference waveforms for the 7 frontal, 3 central, 4 parietal and 5 occipital electrodes.

Behavioral results have been presented elsewhere (VanRullen and Thorpe, submitted) and showed no difference in performance between the animal and vehicle categorization tasks. Percentages correct are around 94% and median reaction times slightly above 350 ms. We also presented evidence that categorization could be performed above chance in less than 250 ms, a surprisingly short value which gives an upper limit to the duration of perceptual processes and the beginning of the subject's decision. ERPs can strengthen these constraints and provide important insights about the time course of the underlying visual processing.

In each task, we find a strong differential activity between targets and distractors that can best be seen on frontal electrode sites after 150 ms (see black lines, Figure 2), as had been reported for an animal categorization task (Thorpe et al., 1996). At the same moment there are also clear differential responses at virtually all recording sites (central, parietal, occipital), thus stressing the magnitude of the effect. Furthermore, these previous results are reproduced here for a task involving a different target category. It is important to stress that "means of transport" being a clearly artificial category, there does not seem to be any advantage of processing natural, biologically relevant visual stimuli.

In addition to this strong differential activity between targets and distractors starting around 150 ms, Figure 2 also illustrates an earlier, weaker differential activity arising roughly 75 ms after stimulus presentation. This effect, which was not reported in the previous study by Thorpe et al. (1996), can certainly be attributed to changes made in the experimental protocol. In particular, the angular size of images was twice the size used previously, which might increase signal-to-noise ratio. Furthermore, ERPs were recorded here using an array of 32 average-referenced electrodes, instead of 20 linked-ears-referenced electrodes in the previous study. In regard to the distinction between perceptual and higher-level task-related decision mechanisms mentioned above, how should each of these two differential activities be interpreted? Is it the correlate of low-level, systematic visual differences between the target and distractor categories involved, or does it reflect the subject's decision that the target category has beendetected? The next sections will attempt to separate the causes of these two differential effects.

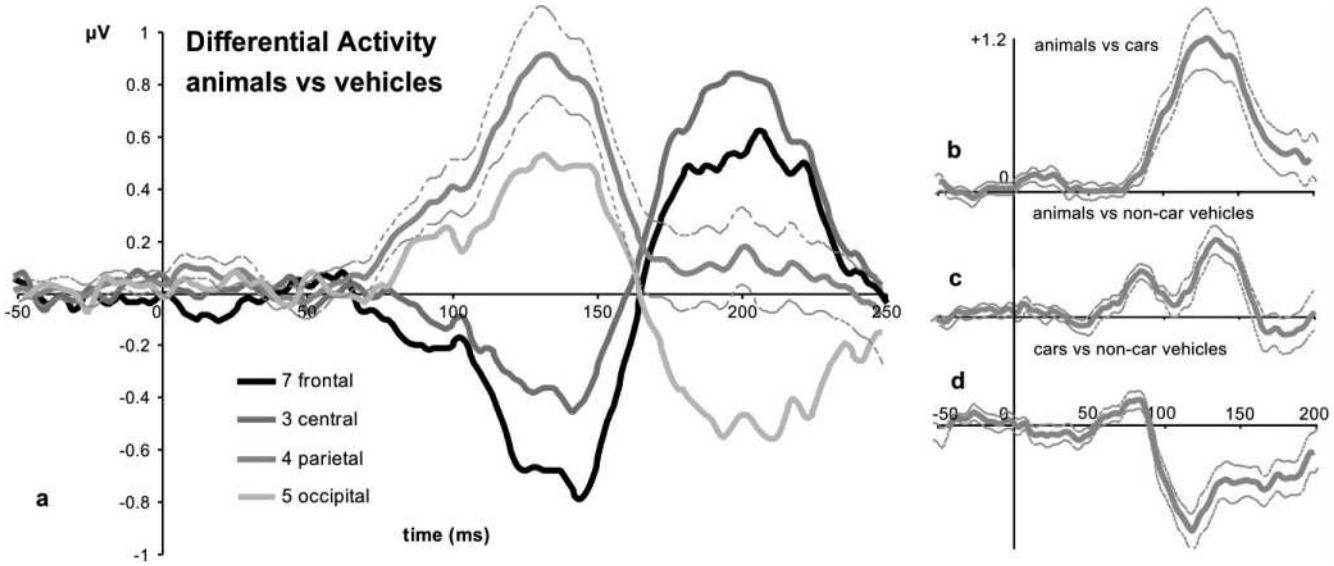


Fig. 3. Effects of the visual category. Each panel shows the differential activity between 2 visually defined categories. The task-related status (target or non-target) is equally represented in each category. For simplicity, inter-subjects standard error of the mean (SEM) is plotted only for parietal electrodes (dotted gray curves). **a.** Animals vs. vehicles on the 7 frontal, 3 central, 4 parietal and 5 occipital electrode groups. **b-d.** Difference waveforms on parietal electrodes. **b.** Animals vs. cars. **c.** Animals vs. non-car vehicles. **d.** Cars vs. non-car vehicles. Statistical t-tests (15 degrees of freedom) were performed for each of these difference waveforms between categories on parietal electrodes where the effect is best seen. At the p<.01 level, the potentials evoked by animal images significantly diverge from the vehicle-evoked potentials at 77* ms post-stimulus and for 82 consecutive samples (panel **a**). More specifically, animal and car trials diverge after 92* ms (+74 samples, panel **b**) and the statistical difference between animal and non-car vehicle trials appears at 76* ms (+25 samples, panel **c**). Finally, the differential activity between cars and non-car vehicles pictures significantly differs from zero after 81 ms and for 4 consecutive samples, after which the sign of the difference is reversed, and reaches significance again at 98* ms (+91 samples, panel **d**). The sign * indicates significance according to the criterion of Rugg et al (1995), modified for stronger precision: 15 consecutive t-test values below the p<.01 level.

Visual category effects

The different visual categories involved (animals, vehicles, further separated as cars and other vehicles) can be compared to one another in a task-independent manner. For each subject, we averaged the waveforms of images belonging to the same category independent of their target/non-target status (by weighing each task-related status according to the number of corresponding trials), then compared these category-specific waveforms to one another. The difference waveforms are presented in figure 3. The differential activities between any

two visual categories appear to diverge from zero after 75-80 ms, which is confirmed by the statistical t-tests performed on parietal electrodes (see caption on figure 3 for details). These differences are totally independent of the task, and on the status of the images in these tasks (target and non-target trials are equally represented in each category-specific waveform). In particular, objects in the car and the non-car vehicle categories have exactly the same status in both tasks: targets in the vehicle categorization task, non-targets in the animal categorization task. In this case, the difference (Figure 3d) can only be attributed to intrinsic visual properties of these categories, and not to differences in the task being performed.

Task-related effects

What is the specific activity induced by the categorization task at hand, i.e. the difference between the processing of targets vs. non-targets? To answer this question we can group together all the waveforms corresponding to different categories of stimulus when they were targets, and compare the result with the result of grouping together waveforms for the same categories of stimulus, but when they were treated as non-targets. Figure 4a shows the differential waveforms of the animals and vehicles grouped together, between the "target" and "non-target" conditions. Figures 4b-e show the same difference separately for each category, on frontal electrode sites. The earliest statistically significant effects for these differential activities are obtained between 156 and 180 ms (see caption on figure 4 for details). This difference in processing between target and non-target trials, starting at 150 ms post-stimulus and reaching significance around 160 ms, clearly constitutes the equivalent in our dual-task paradigm of the difference between target and distractor trials reported in a previous study (Thorpe et al., 1996). However the effect that we demonstrate here is totally independent of the visual category involved, because the difference is observed even when comparing the ERPs to target and non-target trials on images belonging to the same visual category. The underlying mechanism is thus only correlated with the subject's decision, rather than with any visual property of the stimulus. Note that no differential activity between targets and non-targets is observed before 150 ms, which implies that the earlier difference that we found at 75-80 ms does not carry any information about the task-related status (target or non-target) of the visual stimulus.

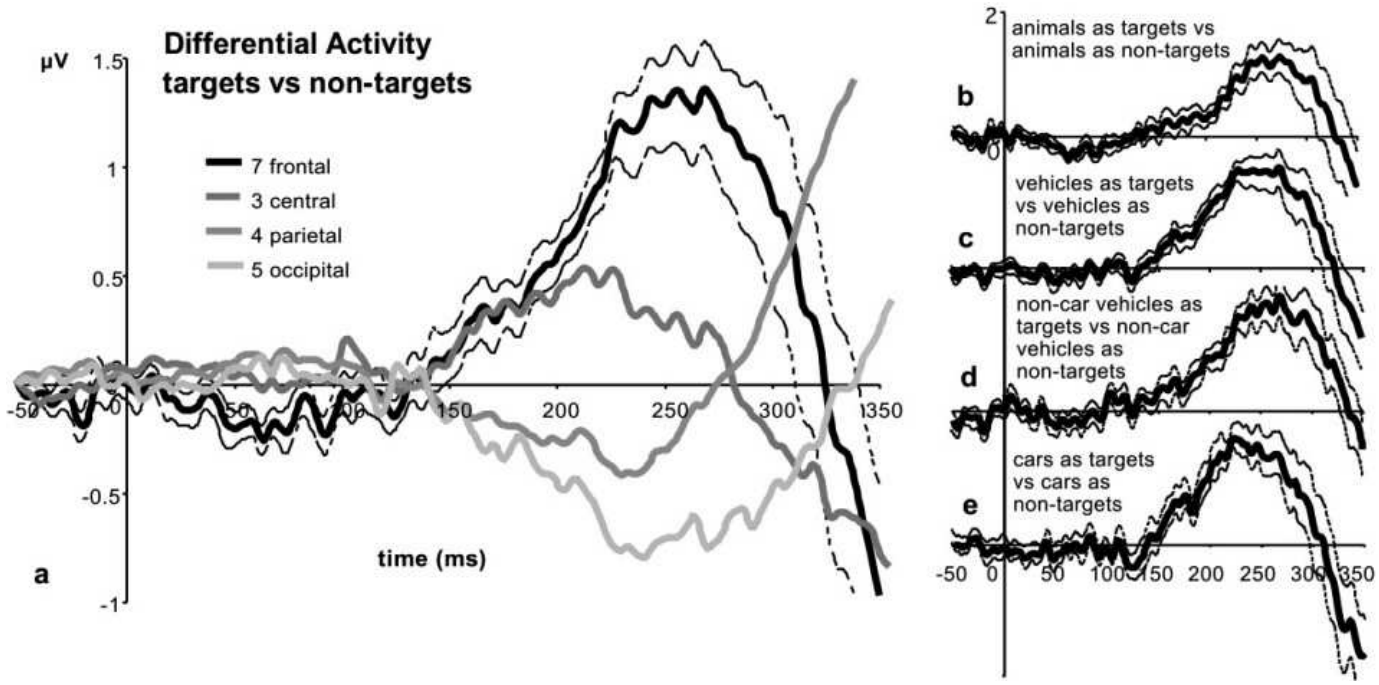


Fig. 4. Effects of the categorization task. Each panel shows the differential activity between target and non-target trials for a visually defined category. In each condition (target and non-target), the visual categories involved are represented equally. For simplicity, SEM is plotted only for frontal electrodes (dotted black curves). For readability, the y axis represents electrical activity on a different scale than figure 3. **a.** Targets vs. non-targets in both tasks (animals and vehicles). **b-e.** Differential activity on frontal electrodes. **b.** Animal category, targets vs. non-targets. **c.** Vehicle category, targets vs. non-targets. **d.** Non-car vehicle category, targets vs. non-targets. **e.** Car category, targets vs. non-targets. For the difference shown in panel **a**, significance at the p<.01 level (t-test, 15 degrees of freedom) is reached on frontal electrodes at 175* ms (+120 samples) but as early as 156* ms (+77 samples) on central electrodes. On frontal electrodes, significance at the p<.01 level is achieved at 212* ms (+72 samples) for the difference shown in panel **b**, at 164 ms (+6 samples) for vehicles as targets vs. non-targets (panel **c**), and more specifically at 189* ms (+114 samples) for other vehicles (panel **d**) and at 165 ms (+4 samples) for cars (panel **e**). However, earlier significant effects are seen on central electrodes for the animal images (156* ms, +25 samples) and on occipital electrodes for vehicles (160 ms, +4 samples) and non-car vehicle images (177* ms, +139 samples). The sign * indicates significance according to the modified Rugg et al (1995) criterion, as in figure 3.

Two separate mechanisms

We isolated two different mechanisms. The first is an early perceptual, category-dependent, task-independent process, starting at 75-80 ms after stimulus onset. The second is a task-related, category-independent mechanism clearly correlated with the subject's decision that a target was or was not present in the image. This latter effect is found to start after 150 ms. What specific neural structures are involved in these two separate processes? Given, on the one hand, the latency of the first perceptual activity (i.e. 75-80 ms, peaking at around 120 ms), and the specific requirements of the task, probably involving mechanisms such as visual features extraction (e.g. shape), this first differential activity could arise in extrastriate visual areas such as V2 or V4. Similar occipital regions have been found in a recent MEG study (Halgren et al. 2000) to respond differently to different categories of stimuli with latencies around 110 ms, i.e. of the same order as the one found here. At first sight it seems surprising that a differential activity carrying information about the visual category of the stimulus perceived could be supported by "early" extrastriate visual areas. But at virtually any stage in the visual system, the information being extracted is likely to be correlated with the visual input, and thus to differ from one visual category to the other; this does not imply that the identity of the stimulus is actively encoded in these regions, and that visual processing is already terminated.

On the other hand, when neural activity is correlated with the decision of the subject rather than with the visual input, it implies that sufficient processing has been done to allow target detection (Shadlen and Newsome, 1996; Romo and Salinas, 1999; Newsome et al., 1989). The objects in the visual scene have therefore been recognized and categorized: such a process is likely to occur at a rather late stage in the ventral occipito-temporal processing stream. Occipito-temporal activations have already been reported using event-related fMRI methods for an animal/non animal categorization task (Fize et al., 2000). Intracranial recordings in humans have also shown that ventral occipito-temporal regions can be highly responsive to faces and other objects with latencies between 150 and 200 ms (Allison et al., 1999). fMRI experiments also report regions in the ventral occipito-temporal cortex of humans that are differentially activated for different categories (e.g. faces, houses, chairs) of visual stimulus (Aguirre et al., 1998; Epstein and Kanwisher, 1998; Ishai et al., 1999), and that can be similarly activated by semantic processing of the same categories even in the absence of visual input (Chao et al., 1999). These areas probably constitute the functional equivalent of monkey infero-temporal cortex regions that have been extensively investigated, demonstrating selectivity for complex visual features and object identity (Tanaka, 1993, 1996; Rolls and Tovee, 1995; Booth and Rolls, 1998; Vogels, 1999). Altogether, these studies suggest that the sources of the decision-related activity that we isolated here could be located in occipito-temporal regions.

DISCUSSION

We have shown and separated here two distinct processing stages in high-level visual categorization tasks. The first one corresponds to a "perceptual" stage and is found to reflect the category of the picture constituting the visual input as early as 75-80 ms. Other equivalently early differences correlated with high-level properties of the visual stimuli have been reported before (Seeck et al., 1997; George et al., 1997; Mouchetant-Rostaing et al., 2000). Yet it remained questionable how much of these activities could be attributed to concurrent low-level visual differences, changes in the subject's attentional state or in the experimental procedure between the different conditions compared. Here, with equally balanced conditions in terms of the relevance of the stimuli towards the task, and a systematic alternation between the two tasks, allowing comparison between trials that are temporally interleaved, we demonstrate that this early differential activity between visually-defined categories is not correlated with high-level task-related properties. Therefore, it probably only reflects systematic visual differences in the "low-level" properties of the visual categories involved. More specifically, it is highly likely that this category-specific activity simply reflects visual encoding mechanisms and the extraction of numerous basic visual properties, the global energy of which can vary across categories. However, the difference demonstrated here should not be understood as a simple statistical difference between any two ensembles of images. For example there is no such early differential activity when comparing the processing of vehicle pictures when they are targets and when they are non-targets: these two subsets were randomly extracted from the same original database.

The second mechanism that we found is correlated with the subject's behavior; it can also be seen when comparing the same visual category on trials where images are targets and on non-target trials. Thus this differential activity can be thought of as a correlate of the decision of the subject, regardless of the visual category that is to be treated as a target (i.e. of the visual details of the task).

It is important to note that subject reaction times were considerably shorter than those reported in a previous similar study (Thorpe et al., 1996). This is probably in large part due to changes in the experimental paradigm: the angular size of the images was larger, and we used a capacitive plate to detect subject responses rather than a mouse button release. The result was that median reaction times were about 100 ms shorter (VanRullen and Thorpe, submitted). Despite this, the latency of the onset of the differential response at 150 ms was very similar to that reported previously, indicating that only the

motor component of the task has been "compressed". The time required for target detection, on the other hand, appears fairly constant across these two studies. Further evidence for the fixed latency of the differential activity at 150 ms comes from another recent study showing that even 3 weeks of intensive training failed to decrease this latency (Fabre-Thorpe et al., 2000).

A direct consequence is that experiments reporting early "high-level" differential ERP activities in humans, or equivalently early differences in electrophysiological recordings in monkeys, should be treated with caution. A neuron or neuron assembly responding differently to different "categories" of visual input is not by itself evidence that sufficient visual processing has been done to allow categorization. To put it another way, what the experimenter reads out is not necessarily what the brain is actually reading out. As an example, if one was to "record" electrical activity in a computer while it is processing an integer variable to determine whether it is a prime number, one would find that the least significant bit of the binary-encoded integer is on average more "active" on prime than on non-prime numbers. However, it takes much more from the computer to decide if the variable is a prime number than to just check whether it is an odd or even integer. Before processing itself has even started, the preliminary encoding mechanisms can sometimes reflect, on average, high-level properties of the input variables. Comparing the average signal for trials presenting that property with trials that do not is clearly not sufficient to claim that the property has been extracted. A major advantage of dual-task paradigms such as the one used here is to allow the study of one particular high-level property independent of the underlying concurrent low-level properties.

Whether the two distinct activities shown here enter visual awareness or not is clearly not a straightforward question to answer, and probably beyond the scope of the present study. Also, much more attention has been devoted to the localization of visual awareness (Crick and Koch, 1995; Koch and Braun, 1996a, 1996b; Leopold and Logothetis, 1996; Logothetis, 1998; Sahraie et al., 1997) than to its latency (Libet et al., 1983). A rather obvious conclusion is that the first perceptual activity, starting at around 75-80 ms, is not directly and immediately accessible to awareness; it would otherwise be correlated with the subject's behavior. Indeed, subjects could make use of that information to generate the motor command. In contrast, numerous theories suppose that neural activity at the final stages of the ventral "what" stream is correlated with the aware percept (Milner and Goodale, 1995; Bar and Biederman, 1999; Crick and Koch, 1995; Koch and Braun, 1996). Along with such views the second, decision-related activity arising at 150 ms could, directly or indirectly, participate in the awareness of the visual scene. The present data however is not sufficient to reject nor to confirm such an interpretation.

METHODS

Experimental procedure.

16 subjects, 8 males and 8 females with ages ranging from 21 to 50 were seated in a dark room, free of movement, at 120 cm from a computer screen piloted from a PC computer. Two categorization tasks with a go-no go paradigm were performed in alternation by all subjects. In each task they viewed 10 series of 96 images, half of which were targets and half were non-targets. A trial was organized as follows: a white fixation point (smaller than 0.1° of visual angle) was drawn in the middle of the black screen; subjects had to press a touch-sensitive button to start the trial; after a delay of 1 second the image, that subtended roughly 10° of visual angle in width and 7° in height, was flashed at the center of the screen for 20 ms, and subjects had to release the button within 1 second if the image contained an element of the target category, and maintained pressure otherwise; the stimulus onset asynchrony (SOA) was 2 seconds plus or minus a random delay of no more than 200 ms, to prevent the subjects from locking their response to the expected time of presentation. After the SOA, the subject's pressure on the touch-sensitive button was tested: the image was flashed only if the subject was ready, otherwise a delay of 1 second was added. Note that the very short presentation time ensures that there is no exploratory eye movement during image presentation.

Images.

The pictures were natural photographs of complex scenes taken from a large commercially available CD-ROM library. The images in each category were chosen to be as varied as possible. The animal category included pictures of mammals as well as birds, fish, insects, or reptiles. The means of transport category included images of cars, trucks, trains, civil and military airplanes, helicopters, boats, hot air balloons, and even rockets. There was no a priori information on the size, position or number of the targets in a single photograph. There was also a very wide range of distractor images, which could be outdoor or indoor scenes, natural landscapes or street scenes with buildings and roads, pictures of food, fruits, vegetables or plants, houses, man-made objects or tools. In each series, subjects were presented with 96 photographs, 48 targets and 48 non-targets. Half of the non-target images were targets from the other task, i.e. vehicles in the animal task, and animals in the means of transport task. Furthermore, half of the vehicle images (24 per series in the means of transport task, 12 per series in the animal task) were images of cars and the other half of different means of transport, to allow further data analysis of intra-category differences. To prevent contamination by the effects of learning, an image was presented once for each subject and could not appear both as a target in one task and a non-target in the other task. Image sequences were randomized so that different subjects were presented different image sequences over the same series of 96 trials, and that the

presentation of a target was not predictable. Finally, all subjects alternately viewed 10 series of 96 trials for each task, and the order of appearance of the series was randomly decided for each subject, as well as whether they would start with an animal or a means of transport categorization series.

ERP recordings

Electric brain potentials were recorded from 32 electrode sites equally distributed over the skull. Data acquisition was made at 1000 Hz (corresponding to a sample bin of 1ms) using a Synamps recording system coupled with a PC computer. Recorded potentials were average-referenced on electrode CZ, and low-pass filtered at 100 Hz. Trials were checked for artifacts and discarded using a [-50; +50 μV] criterion over the interval [-100; +400 ms] on frontal electrodes for eye movements and a [-30; +30 μV] criterion on the period [-100; +100 ms] on parietal electrodes for "alpha" brain waves. Electrodes were grouped into frontal (FP1, FP2, F3, F4, F7, F8, FZ in the 10-20 system nomenclature), central (C3, C4, CZ), parietal (PZ, PO3, PO4, POZ) and occipital groups (O1, O2, PO7, PO8, OZ) according to position, so as to represent the whole midline. Inter-subjects statistical t-tests (15 degrees of freedom) were performed at the $p<.01$ level on the parietal electrodes group for the differential activities shown in Figure 3 and on frontal electrodes for the differential activities shown in Figure 4. According to the criterion of Rugg et al (1995) (see also Thorpe et al, 1996), a difference potential is considered significant if 15 consecutive t-test values are below the $p<.05$ level. Here we strengthen the precision of this criterion by using $p<.01$. Times at which significant differences emerge according to this criterion are indicated in the text by the sign *.

REFERENCES

Aguirre, G.K., Zarahn, E. & D'Esposito, M. An area within human ventral cortex sensitive to 'building' stimuli: Evidence and implications, *Neuron* **21**:373-383 (1998).

Allison, T., Puce, A., Spencer, D.D. & McCarty, G. Electrophysiological studies of human face perception. I : Potentials generated in occipito-temporal cortex by face and non-face stimuli, *Cereb. Cortex* **9**:415-430 (1999).

Bar, M. & Biederman, I. Localizing the cortical region mediating visual awareness of object identity, *Proc Natl Acad Sci USA* **96**, 1790-3 (1999)

Booth, M.C.A. & Rolls, E.T. View-invariant representations of familiar objects by neurons in the Inferior Temporal Visual Cortex., *Cereb. Cortex* **8**, 510-523 (1998).

Bötzel, K., Schulze, S., & Stodieck, S. R. Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res, 104*(1), 135-143 (1995).

Chao, L., Haxby, J.V. & Martin, A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects, *Nat Neurosci* **2**, 913-919 (1999).

Crick, F. & Koch, C. Are we aware of neural activity in primary visual cortex?, *Nature* **375**, 121-3 (1995).

Debruille, J. B., Guillem, F., & Renault, B. ERPs and chronometry of face recognition: following-up Seeck et al. and George et al. *Neuroreport, 9*(15), 3349-3353 (1998).

Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment, *Nature* **392**, 598-601 (1998).

Fabre-Thorpe, M., Delorme, A., Marlot, C. & Thorpe, S. A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes, *J. Cog. Neurosci*. (in press).

Fize, D., *et al.* Brain areas involved in rapid categorization of natural images: an event-related fMRI study, *NeuroImage* (in press).

George, N., Jemel, B., Fiori, N. & Renault, B. Face and shape repetition effects in humans: a spatio-temporal ERP study, *Neuroreport* **8**, 1417-23 (1997).

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. Cognitive Response Profile of the Human Fusiform Face Area as Determined by MEG. *Cereb Cortex, 10*(1), 69-81 (2000).

Ishai, A., Ungerleider, L.G., Martin, A., Shouten, J.L. & Haxby, J.V. Distributed representation of objects in the human ventral visual pathway, *Proc Natl Acad Sci USA* **96**, 9379-9384 (1999).

Jeffreys, D. A. Evoked potential studies of face and object processing. *Visual Cognition, 3*, 1-38 (1996).

Koch, C. & Braun, J. Towards the neuronal correlate of visual awareness, *Curr Opin Neurobiol* **6**, 158-64 (1996a).

Koch, C. & Braun, J. The functional anatomy of visual awareness, *Cold Spring Harbor Symposia on Quantitative Biology*, **61**, 49-57 (1996b)

Leopold, D.A. & Logothetis, N.K. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry, *Nature* **379**, 549-553 (1996).

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain, 106*(Pt 3), 623-642 (1983).

Logothetis, N. Object vision and visual awareness, *Curr Opin Neurobiol* **8**, 536-44 (1998).

Luce, R.D. *Response Times* (Oxford University Press, Oxford, 1986).

Milner, A.D. & Goodale, M.A. *The visual brain in action*. Oxford University Press (1995).

Mouchetant-Rostaing, Y., Giard, M-H., Bentin, S., Aguera, P.A. & Pernier, J. Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci* **12**, 303-310 (2000).

Newsome, W.T., Britten, K.H. & Movshon, J.A. Neuronal correlates of a perceptual decision, *Nature* **341**, 52-4 (1989).

Nichols, M.J. & Newsome, W.T. The neurobiology of cognition, *Nature* **402**, C35-C38 (1999).

Oram, M.W. & Perrett, D.I. Time course of neural responses discriminating different views of the face and head, *J Neurophysiol* **68**, 70-84 (1992).

Perrett, D.I., Rolls, E.T. & Caan, W. Visual neurons responsive to faces in the monkey temporal cortex, *Experimental Brain Research* **47**, 329-342 (1982).

Rolls, E.T. & Tovee, M.J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex., *J Neurophysiol* **73**, 713-26 (1995).

Romo, R. & Salinas, E. Sensing and deciding in the somatosensory system, *Curr Opin Neurobiol* **9**, 487-93 (1999).

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *NeuroReport, 11*(1), 69-74 (2000).

Rugg, M. D., Doyle, M. C., & Wells, T. (1995). Word and non-word repetition within- and across-modality: an event-related potential study. *J Cog Neurosci, 7*, 209-227.

Sahraie, A. et al. Pattern of neuronal activity associated with conscious and unconscious processing of visual signals, *Proc Natl Acad Sci USA* **94**, 9406-9411 (1997)

Schall, J.D. & Thompson, K.G. Neural selection and control of visually guided eye movements, *Annu Rev Neurosci* **22**, 241-59 (1999).

Schendan, H. E., Ganis, G., & Kutas, M. Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35*(3), 240-251 (1998).

Seeck, M., *et al.* Evidence for rapid face recognition from human scalp and intracranial electrodes, *Neuroreport* **8**, 2749-54 (1997).

Shadlen, M.N. & Newsome, W.T. Motion perception: seeing and deciding, *Proc Natl Acad Sci U S A* **93**, 628-33 (1996).

Tanaka, K. Neuronal mechanisms of object recognition, *Science* **262**, 685-688 (1993).

Tanaka, K. Inferotemporal cortex and object vision, *Annual Review of Neuroscience* **19**, 109-139 (1996).

Thorpe, S.J., Fize, D. & Marlot, C. Speed of processing in the human visual system, *Nature* **381**, 520-522 (1996).

VanRullen, R. & Thorpe, S.J. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects, *Perception* (submitted).

Vogels, R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study, *Eur J Neurosci* **11**, 1239-55 (1999).

Reprint requests should be sent to:

Rufin VanRullen
Centre de Recherche Cerveau & Cognition, UMR 5549,
133 Rte de Narbonne, 31062 Toulouse Cedex, France
email rufin@cerco.ups-tlse.fr

2.3 Activité précoce: interprétation

L'activité électrique précoce (entre 75 et 150 ms) que nous avons mise en évidence dans cet article correspond vraisemblablement à l'activité enregistrée dans les différentes études présentées en I.5.1, et postulée par leurs auteurs comme le reflet d'une catégorisation ultra-rapide. Nos résultats montrent que l'activité électrique autour de 100 ms reflète en effet la catégorie du stimulus, mais qu'elle n'est pas corrélée à la décision du sujet, et qu'elle n'a donc, à ce niveau du traitement visuel, pas de valeur sémantique.

2.3.1 La métaphore de l'ordinateur

Un exemple simple illustre le type d'erreur conceptuelle que l'on peut être tenté de commettre face à une telle activité précoce. Cet exemple a déjà été présenté dans la discussion de l'article précédent.

Un ordinateur, système informatique de traitement de données, réalise un programme qui lui permet de catégoriser des nombres entiers qui lui sont fournis en entrée. Il doit répondre (1) lorsque l'entier est un nombre premier, et ne doit pas répondre (0) sinon. Ceci est illustré dans la figure 10. Comment l'ordinateur implémente-t-il cette fonction? Quels sont les traitements effectués, i.e. la séquence de transformations de l'entrée, par un ensemble de représentations intermédiaires, jusqu'à la sortie désirée?

Si l'entier à traiter est fourni par un utilisateur externe, il l'est probablement sous une forme décimale (prenons par exemple 172). L'une des premières opérations réalisées par l'ordinateur, avant même de commencer tout traitement effectif, est de convertir cette entrée en une représentation appropriée, i.e. binaire (172 deviendra donc 10101100 en représentation binaire). Ensuite seulement commence le traitement à proprement parler, qui permettra à l'ordinateur de décider si la variable est un nombre premier ou non, et ainsi de générer sa réponse.

Un expérimentateur consciencieux (voir figure 10) décide d'enregistrer l'activité électrique dans l'ordinateur afin de comprendre comment celui-ci réalise la fonction désirée. Par malchance (nous comprendrons très vite pourquoi), il plante son électrode dans la partie de la mémoire de l'ordinateur qui correspond au dernier bit de la variable encodée en représentation binaire. Comme cet expérimentateur est consciencieux, il réalise un très grand nombre d'observations avant de se risquer à toute conclusion: il présente à chaque fois un entier différent, enregistre l'activité électrique au niveau de ce bit d'information, ainsi que la réponse de l'ordinateur (qui ici ne se trompe probablement jamais). Ayant recueilli suffisamment de données, l'expérimentateur arrête l'enregistrement, et commence l'analyse. Il moyenne l'activité enregistrée sur les essais "cible" (nombres premiers) et obtient 0.99 (tous les entiers premiers sont impairs sauf 2). Il moyenne ensuite l'activité enregistrée pour les essais "distracteurs" (nombres non premiers) et obtient 0.47 (les entiers non premiers

sont soir pairs, soit impairs, avec la même probabilité). Le bit d'information enregistré est donc en moyenne 2 fois plus actif sur les entiers premiers que sur les entiers non premiers.

L'observateur pourrait aisément conclure qu'à l'endroit où il a planté son électrode, la représentation de l'information utilisée par l'ordinateur encode explicitement le statut ou la catégorie du stimulus (cible/distracteur, ou premier/non-premier). Pourtant, mis à part cet encodage préliminaire, aucun traitement réel n'a encore été effectué à ce niveau!

Cette erreur conceptuelle semble avoir été commise dans les expériences présentées en I.5.1, qui décrivent une activité très précoce reflétant des propriétés de "haut-niveau" du stimulus visuel. Notons que cette présentation des résultats n'est pas fausse par elle-même: ce qui est erroné est d'en conclure que le système visuel a déjà extrait cette propriété de haut-niveau.

Le fait que l'activité neuronale varie, en moyenne, avec la catégorie du stimulus visuel n'implique pas que la catégorisation a déjà été réalisée.

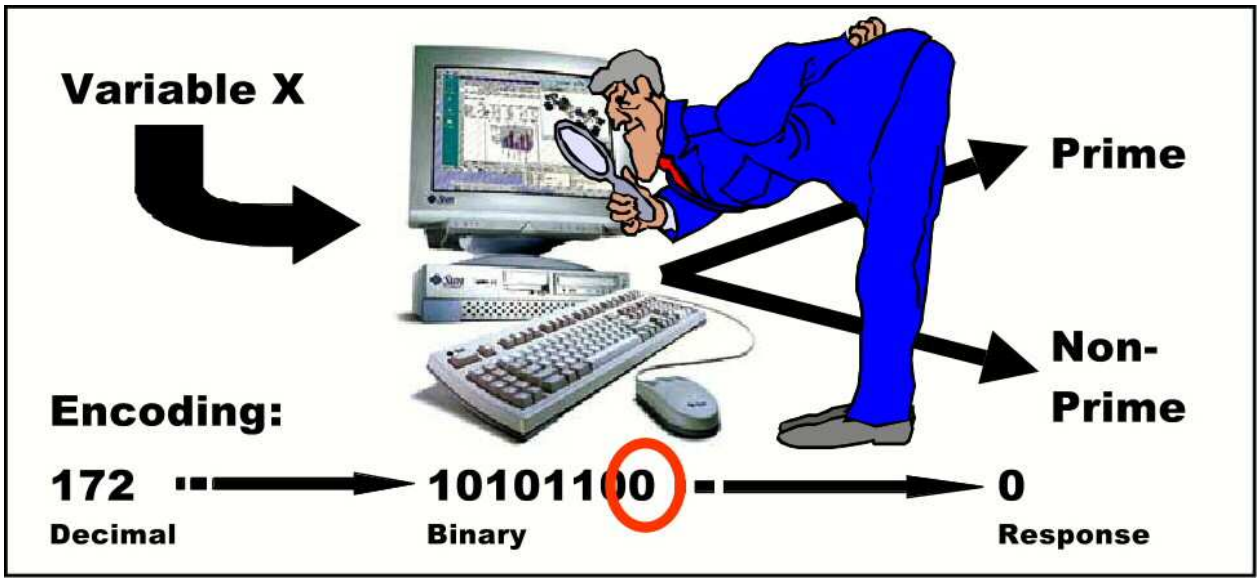


Figure 10. Un expérimentateur qui enregistrerait l'activité électrique dans un ordinateur qui catégorise des entiers selon qu'ils sont premiers ou non trouverait qu'en moyenne, le dernier bit de la variable d'entrée (cercle rouge) reflète la sortie désirée. S'il en déduisait qu'à ce niveau, l'ordinateur a déjà catégorisé la variable, cette conclusion serait bien sûr erronée.

2.3.2. La sortie de la rétine diffère selon la catégorie, mais la catégorisation ne s'effectue pas dans la rétine

L'activité précoce que nous avons enregistrée entre 75 et 150 ms a été localisée aux premières étapes de traitement cortical, dans les aires visuelles extrastriées. Mais si les

EEG pouvaient également refléter l'activité rétinienne, verrait-on apparaître une activité différentielle précoce au niveau de la rétine? Pour renforcer notre argument, nous tentons ici de répondre à cette question.

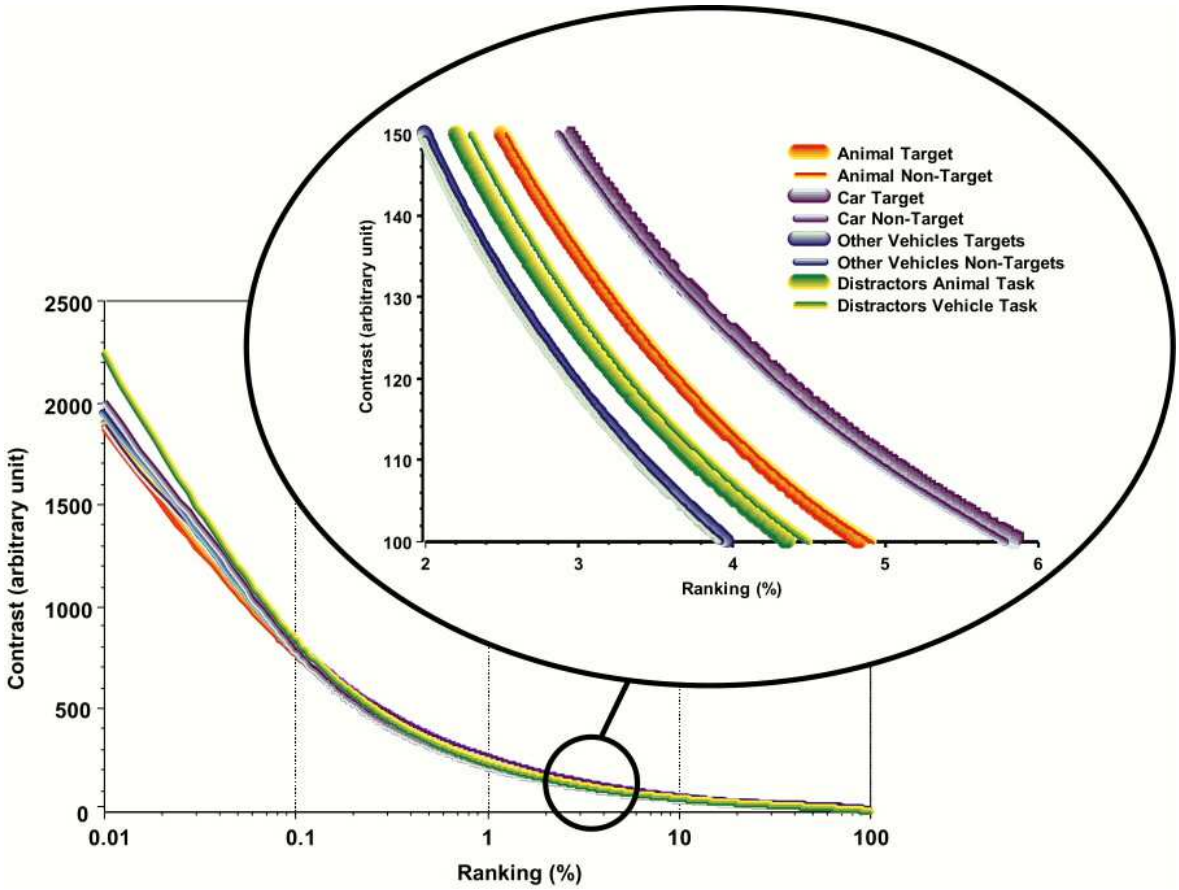


Figure 11. Réponse moyenne d'une population de neurones rétiniens (formels) à différentes catégories d'images (animal, voiture, autre véhicule, distracteur). Les catégories sont séparées en sous-catégories disjointes selon la tâche dans laquelle elles étaient utilisées (catégorisation animal ou moyen de transport), afin de servir de contrôle mutuel (de variance). Le contraste est représenté comme une fonction du quantile d'activation des neurones (en pourcentage du nombre de neurones total, 1 est le premier centile). L'insert représente cette courbe entre les 2ème et 6ème centiles. Les courbes de réponse neuronale de la rétine à différentes catégories sont parfaitement séparables, alors que deux courbes reflétant des images de la même catégorie ne le sont pas.

Un modèle de rétine a été créé, qui sera décrit plus en détail au chapitre III section 2.1. Dans ce modèle, un ensemble de cellules ON- et OFF-center extrait les contrastes présents dans l'image à différentes fréquences spatiales. Le contraste dans le champ récepteur d'un neurone est transformé en un niveau d'activation qui pourra déterminer la réponse du neurone en termes de spike trains. Ce niveau d'activation est une mesure linéaire du contraste.

Pour chaque image utilisée dans l'expérience de l'article 2, préalablement transformée en niveaux de gris, nous avons enregistré l'activité dans notre modèle de rétine en réponse à cette stimulation, sous la forme d'une courbe représentant les valeurs des niveaux d'activation des neurones (donc de contraste), du plus fort au plus faible. Nous avons ensuite moyenné ces courbes de réponse au contraste pour différentes catégories: animaux, voitures, autres véhicules, distracteurs. Nous avons de plus séparé chacune des catégories en fonction de la tâche dans laquelle l'image était utilisée. Ces différentes courbes moyennes sont représentées figure 11. Bien que toutes ces courbes semblent relativement proches à première vue, deux catégories de stimuli différentes donnent en fait deux courbes de réponse parfaitement séparables. Il ne s'agit aucunement de fluctuations d'échantillonnage, car les courbes pour une même catégorie dans une tâche ou dans l'autre sont quasiment identiques, alors qu'elles représentent deux sous-ensembles d'images disjoints définis aléatoirement à partir d'un même ensemble (ces 2 courbes correspondant à une même catégorie d'objets servent en fait d'indication de la variance de la réponse rétinienne sur une catégorie donnée d'objets).

Il apparaît donc que, même au niveau de la rétine, l'activité neuronale globale peut refléter des aspects de haut niveau du stimulus visuel comme sa catégorie. Personne n'oserait cependant interpréter cette observation comme la preuve d'une catégorisation visuelle ultra-rapide... dans la rétine!

2.3.3 Le stimulus diffère selon la catégorie, mais la catégorisation ne s'effectue pas avant la rétine!

En fait, il est possible que le stimulus lui-même diffère, en moyenne, selon sa catégorie (on ne lira pas ici cette phrase au premier degré, auquel cas elle serait triviale). L'analyse que l'on vient de faire a démontré que la statistique des contrastes pouvait refléter la catégorie d'objets à laquelle l'image appartient. Qu'en est-il alors pour la statistique des luminances ?

Lorsqu'on calcule l'image moyenne pour une catégorie donnée (figure 12), on s'aperçoit que ces moyennes sont très différentes pour différentes catégories. Par contre, ici encore, les moyennes pour 2 sous-catégories arbitrairement définies (selon la tâche "animal" ou "moyen de transport" dans laquelle les images étaient utilisées) sont très proches, ce qui signifie que la variance à l'intérieur d'une catégorie est limitée, et que ces moyennes ne reflètent donc pas des fluctuations d'échantillonnage.

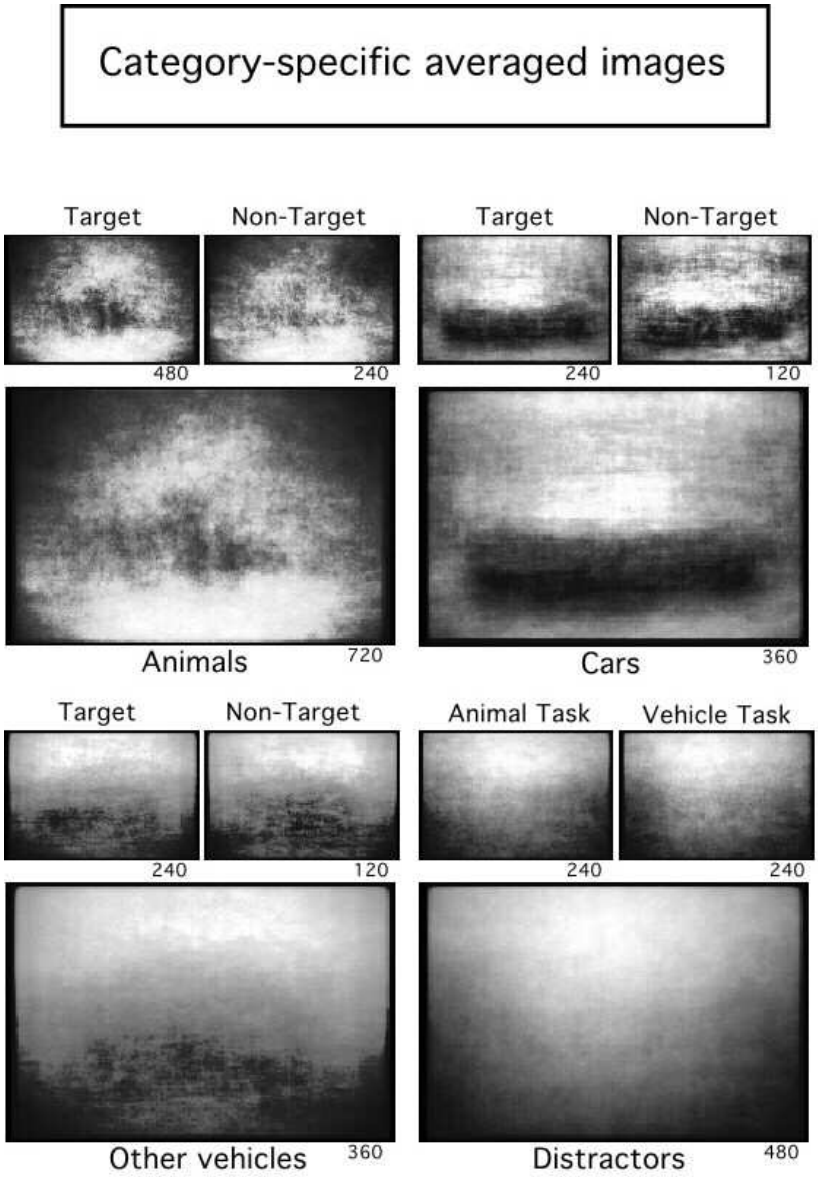


Figure 12. Images moyennes obtenues pour chaque catégorie (renormalisées sur 256 niveaux de gris). Le nombre d'images participant à la moyenne est indiqué sous chaque image. Pour chaque catégorie, les images sont également moyennées pour 2 sous-catégories arbitrairement définies (selon la tâche "animal" ou "moyen de transport" dans laquelle elles étaient utilisées), qui montrent que les fluctuations d'échantillonnage sont limitées.

Le stimulus moyen est donc différent pour chaque catégorie. Pourtant, il n'a à ce jour jamais été proposé que la catégorisation d'un stimulus par un sujet humain puisse avoir lieu avant-même la rétine !

La définition même du mot "catégorie" implique qu'il doit exister **dans l'image** un ensemble d'éléments permettant de la catégoriser. L'image est (trivialement) le reflet de ces éléments ; l'activité rétinienne l'est également, par le fait même qu 'elle transmet la totalité de l'information visuelle au cortex ; probablement toutes les aires visuelles, striées ou extrastriées doivent également refléter ces éléments, car leur rôle est de traiter l'image afin d'en extraire l'information pertinente, notamment sa catégorie ; ce n'est donc que lorsque cette information de catégorie est faite **explicite** dans une représentation neuronale que l'on pourra parler de catégorisation.

2.4 Activité de haut-niveau: interprétation

Après 150 ms, l'activité différentielle que nous avons décrite dans l'article précédent semble refléter ce type de représentation neuronale où l'information de catégorie est faite explicite. En effet, cette activité est corrélée au comportement du sujet. Elle est observée même lorsque l'on compare des essais "cibles" ou "non-cibles" impliquant la même catégorie d'images. Cette activité différentielle peut donc être considérée comme un corrélat de la décision du sujet, indépendemment de la catégorie d'images qui doit être traitée comme "cible" (i.e. indépendemment des détails visuels de la tâche). Or, la décision (cible/non-cible ; génération/inhibition de la commande motrice) requiert nécessairement que la catégorisation ait précédemment eu lieu.

En fait, le signal porté par la population neuronale qui génère cette activité différentielle contient à la fois une information visuelle (la catégorie de l'objet, cette fois encodée explicitement) et une information concernant la tâche que les sujets doivent effectuer (quelle est la catégorie cible ?). Ce signal peut donc constituer pour les sujets une base pour la décision de déclencher ou d'inhiber la commande motrice de réponse.

2.4.1 Relation entre l'activité électrique et la réaction motrice

Ceci est remarquablement illustré par la quasi-parfaite correspondance entre l'activité différentielle (absolue) cible vs non-cible après 150 ms et l'histogramme des temps de réaction dans les deux tâches (catégorisation animal ou moyen de transport).

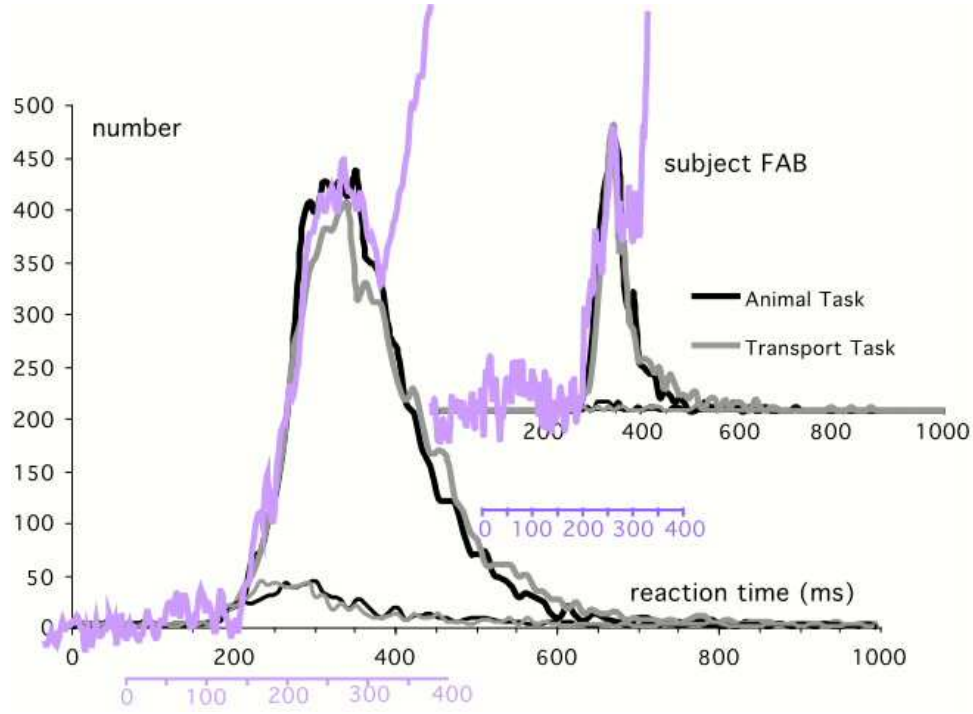


Figure 13. Mapping de l'activité différentielle nette (absolue) sur l'histogramme des temps de réaction pour les 2 tâches (animal, trait noir ; véhicule, trait gris). L'activité différentielle nette "cible vs non-cible" (trait violet) est la somme des valeurs absolues de cette activité différentielle sur les 32 électrodes. La courbe principale montre la moyenne pour les 16 sujets, la courbe insérée à droite montre le même mapping individuellement pour un sujet. Après alignement des lignes de base (sur la période [–100 ; 100 ms], temps EEG), l'axe temporel des ERPs a été aligné pour que le début de l'activité électrique différentielle significative coïncide avec les premiers temps de réaction significatifs (déplacement +70 ms). L'axe vertical (EEG) a ensuite été linéairement normalisé pour correspondre aux temps de réaction lors des 50 ms suivantes. Le fait que les 2 types de courbe (ERP, temps de réaction) ne diffèrent pas significativement avant 390-400 ms (temps comportemental) suggère que cette activité électrique différentielle reflète le mécanisme de décision sur lequel est basée la génération de la commande motrice.

L'énergie ou l'amplitude globale de l'activité différentielle entre cibles et non-cibles peut être mesurée par la somme des valeurs absolues de cette activité différentielle sur les 32 électrodes. Nous appelerons cette mesure "l'activité différentielle nette" (μV net). Nous proposons que cette activité est directement corrélée à la décision du sujet de générer ou d'inhiber une réponse motrice. Nous nous intéresserons donc ici à la relation qui peut exister entre cette activité et la réponse motrice elle-même.

Afin de se débarrasser du bruit ambient présent sur les 32 électrodes, et qui s'est amplifié lors du passage à la valeur absolue, puis lors de la somme sur toutes les

électrodes, nous ramenons la baseline à 0 sur la période [–100 ; 100 ms] (en soustrayant 2.68 μV nets). Nous alignons ensuite le début de l'activité différentielle nette significative (155 ms, t-test, $p<0.05$ pendant plus de 5 échantillons, soit 10 ms) avec le début des temps de réaction discriminant (220-230 ms, voir article 1), ce qui revient à déplacer les ERP de 70 ms. L'activité différentielle nette est ensuite renormalisée de sorte qu'elle corresponde à la distribution des temps de réaction (moyennée pour les 2 tâches) pour les 50 ms suivantes (facteur de normalisation : 1.134 réponses motrices par sujet par μV net). Un test du $\chi 2$ montre que l'histogramme des temps de réaction et l'activité différentielle nette "remappée" ne diffèrent pas significativement ($p<0.05$) avant 390-400 ms (en termes de temps comportemental), soit pour plus de 120 ms après l'intervalle de renormalisation.

Il semble donc que, à une opération de seuillage près ainsi qu'un réalignement temporel, il existe une relation purement linéaire entre la réponse motrice et l'activité différentielle "cible vs non-cible" après 150 ms. Les réponses motrices surviennent 70 ms après cette étape "décisionnelle", avec une probabilité directement proportionnelle à l'intensité de l'activité différentielle observée. Une telle transformation "linéaire" entre activité neuronale et réponse motrice est compatible avec les théories actuelles de la prise de décision et de la génération d'une commande motrice : accumulation (stochastique) d'une certaine "information" (évidence, activité, spikes,□) jusqu'à ce qu'un seuil (stochastique) soit atteint (Luce, 1986 ; Ratcliff et al. 1999).

Il est donc vraisemblable que l'activité différentielle observée entre essais cibles et non-cibles, qui démarre dès 150 ms, reflète l'étape "décisionnelle" sur laquelle est basée la génération (ou l'inhibition) de la réponse motrice.

Cette hypothèse prédit que l'activité différentielle doit être retardée chez les sujets dont les temps de réaction sont les plus lents. Nous avons donc séparé les 16 sujets en 2 groupes de 8 personnes, un groupe de sujets "rapides" et un groupe de sujets "lents" (voir figure 14). Les distributions des temps de réaction ont été recalculées pour chaque groupe, ainsi que les activités différentielles nettes "cible vs non-cible". Ces ERPs ont ensuite été mappés **ensemble** sur les distributions des temps de réaction de la façon suivante : après correction de la baseline (ici on a déduit un niveau de bruit moyen de 3.6 μV nets, supérieur à celui obtenu précédemment car ces activités différentielles sont moyennées sur 2 fois moins de sujets), après décalage de +70 ms (la valeur moyenne obtenue précédemment), l'axe vertical des EEG a été renormalisé de sorte que l'activité différentielle maximale pour les sujets rapides (définie comme la moyenne sur la période [260 ; 280 ms]) corresponde au maximum de la distribution des temps de réaction des sujets rapides, moyennée pour les 2 tâches (facteur de normalisation 1.5μV nets par sujet par réponse motrice, i.e. légèrement supérieur à celui obtenu précédemment, du fait d'un niveau de bruit supérieur). Le mapping

obtenu pour les sujets lents est donc entièrement déterminé par celui réalisé pour les sujets rapides. La figure 14 montre cependant que le facteur de normalisation appliqué dans un cas (sujets rapides) fonctionne parfaitement pour l'autre (sujets lents) : la probabilité d'occurrence d'une réponse motrice pour un sujet lent est proportionnelle à l'activité différentielle cible vs non-cible, avec le même coefficient de proportionnalité que pour un sujet rapide. Cette activité différentielle est d'amplitude moindre chez les sujets lents, d'où un étalement des temps de réaction.

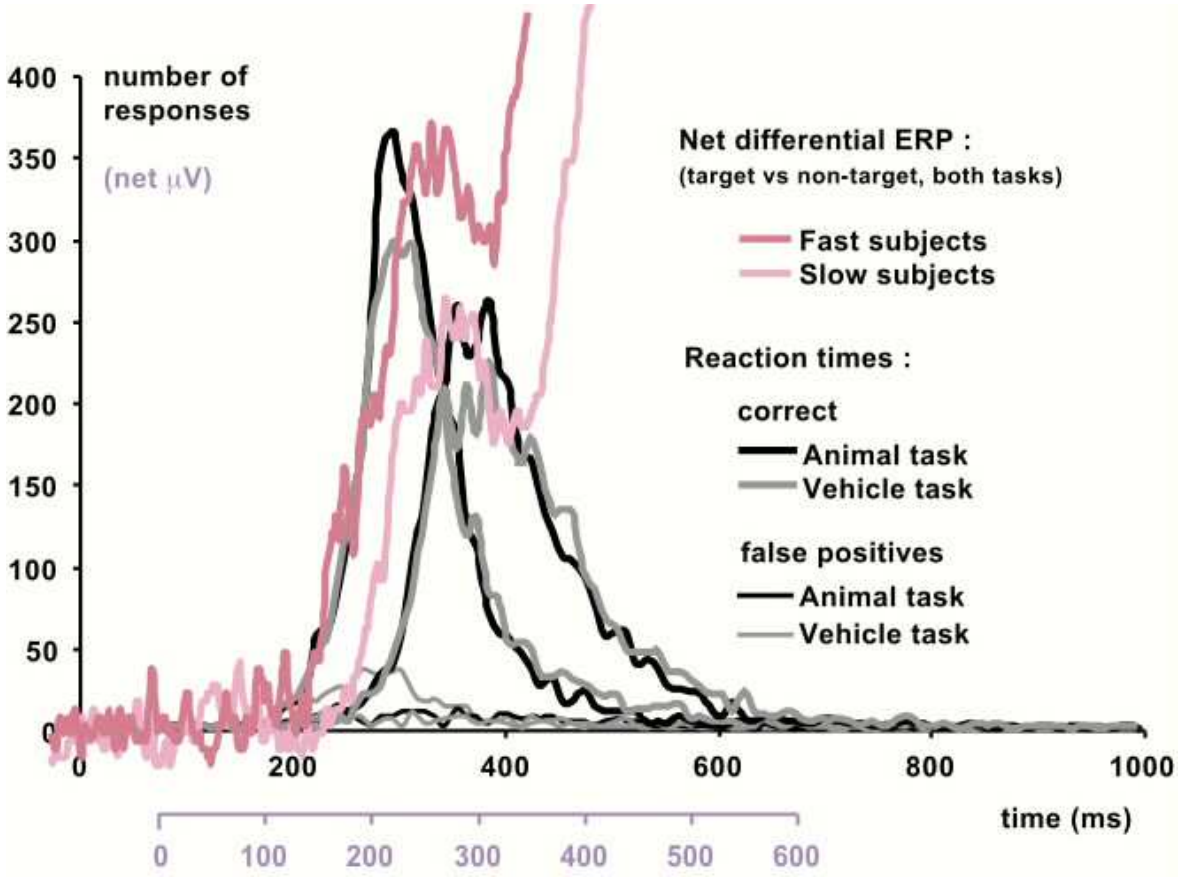


Figure 14. Sujets rapides, sujets lents (notations comme à la figure 13). Le décalage temporel entre ERPs et temps de réaction est de 70 ms. Le facteur de normalisation met en correspondance le maximum de l'activité différentielle nette pour les sujets rapides (ligne fuschia) et le maximum de leur distribution de temps de réaction. La courbe pour les sujets lents (ligne rose) a subi la même transformation. La probabilité d'occurrence d'une réponse motrice est une fonction linéaire de l'activité différentielle "cible vs non-cible" qui précède, avec le même facteur de proportionnalité quel que soit la rapidité du sujet. Seul le délai entre le début de l'étape "décisionnelle" et la réponse motrice est corrélé à la rapidité des sujets (environ 70 ms pour les sujets rapides, et 100 ms pour les sujets lents).

Notons enfin que, si le décalage temporel de 70 ms entre processus "décisionnel" et réponse motrice semble très légèrement sur-estimé dans le cas des sujets rapides, il est sous-estimé (d'environ 30 ms) pour les sujets lents. Ceci peut bien sûr s'interpréter comme une durée supérieure de la réponse motrice (ce qui définirait bien les sujets "lents"), mais il est également possible que ce délai soit dû à un temps d'intégration "décisionnelle" plus long : puisque l'amplitude de l'activité différentielle est plus faible chez ces sujets, elle est donc moins fiable, et il est nécessaire de l'intégrer plus longuement avant de prendre une décision "raisonnable". Les sujets lents ne seraient donc pas définis par une moins bonne "vivacité" dans la réponse motrice, mais par un délai et surtout une faible amplitude de l'activité neuronale qui différencie les essais "cibles" des autres.

Toutes ces observations suggèrent que l'activité différentielle entre essais cibles et non-cibles après 150 ms reflète la décision des sujets, sur laquelle est basée leur réponse comportementale.

2.4.2 Décision ou attention?

Une autre interprétation cependant serait que cette activité différentielle ne reflète pas la décision du sujet, mais l'attention qu'il (ou elle) porte sélectivement aux animaux, ou aux véhicules en fonction de la tâche. Selon les principales théories de l'attention (voir chapitre III section 3.3), lorsque l'attention est portée sélectivement sur une catégorie d'objets (par exemple "animal"), les neurones qui participent à la représentation de cet objet, ou des traits le composant, voient leur gain augmenter. Ainsi l'on s'attend à une différence d'activité neuronale entre les images de la catégorie "animal" dans les 2 tâches de catégorisation, différence qui s'expliquerait uniquement par l'attention requise par la tâche, sans impliquer la décision du sujet.

Cependant, si l'activité neuronale lors d'un essai "cible" diffère, du fait de l'attention, de l'activité générée par un essai "non-cible" impliquant la même catégorie visuelle, cela implique nécessairement que la catégorisation elle-même ait déjà commencé.

En effet, supposons que la modification par l'attention du traitement neuronal à une étape donnée se reflète dans les ERPs. Comme il n'y a pas de différence entre le traitement des distracteurs dans la tâche de catégorisation "animal" et dans la tâche "moyen de transport" (voir figure 15), cette modification par l'attention (si elle existe) doit survenir à une étape du traitement neuronal où les représentations des animaux et des distracteurs sont séparées, et de même, où les représentations des moyens de transport et des distracteurs sont séparées (sinon cette modification affecterait également le traitement des distracteurs). Qu'elle soit le reflet d'une modulation attentionnelle ou non, l'activité différentielle à partir de

150 ms est donc générée à un niveau de traitement où les représentations neuronales sont suffisamment abstraites pour permettre la catégorisation.

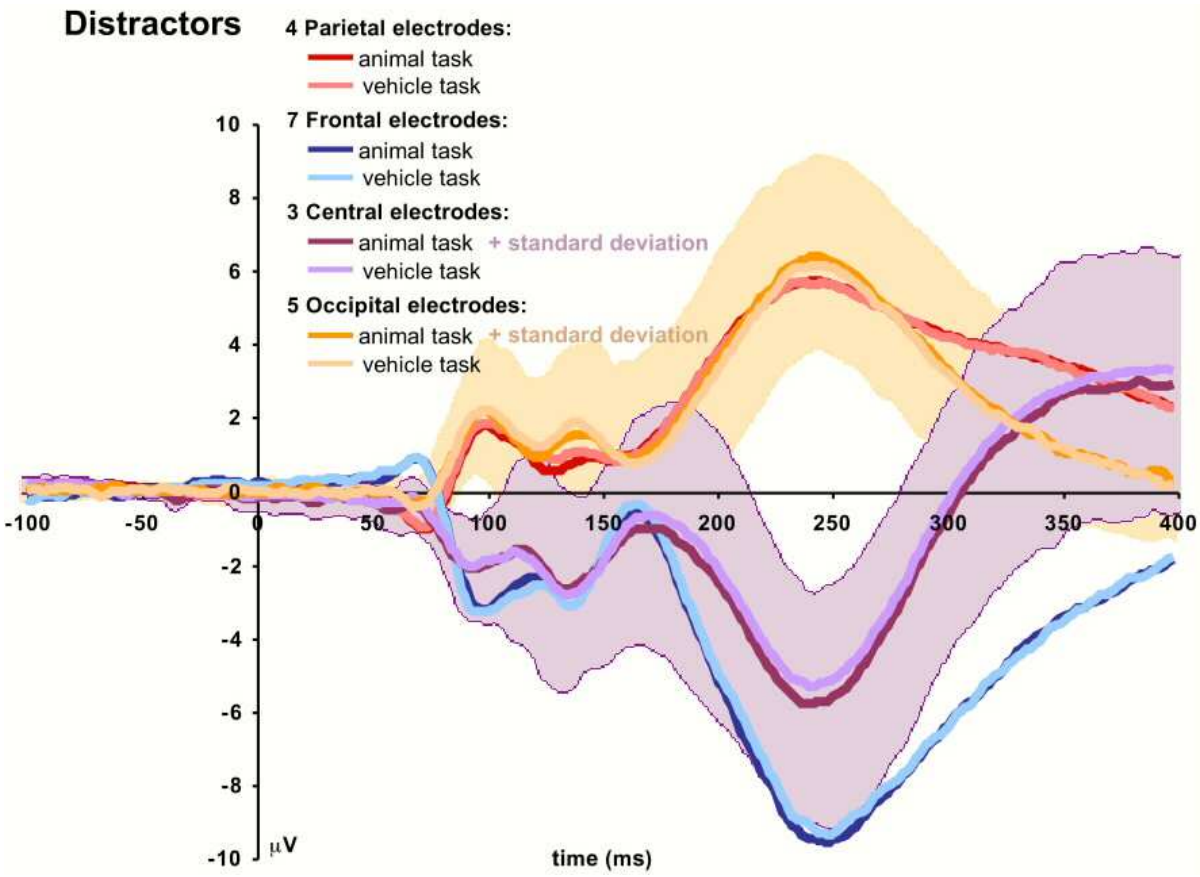


Figure 15. Potentiels évoqués visuels pour la catégorie "distracteur" (ni animal, ni moyen de transport) dans les 2 tâches de catégorisation ("animal", "moyen de transport"), sur différents groupes d'électrodes. Pour illustrer la très forte proximité entre les potentiels evoqués dans ces 2 conditions, l'écart-type est également représenté (aires pleines) pour les électrodes centrales et occipitales, dans la tâche "animal". Le traitement des distracteurs est identique dans les 2 tâches, ce qui implique que si l'activité différentielle à 150 ms peut s'expliquer par une modulation attentionnelle pour les traits de la cible, cette modulation ne peut avoir lieu qu'à un niveau de traitement où les représentations des animaux et des distracteurs sont séparées (de même, où les représentations des moyens de transport et des distracteurs sont séparées), c'est-à-dire un niveau où les représentations sont suffisamment abstraites pour permettre la catégorisation.

2.4.3 Conscience visuelle

Il est important, lorsque l'on met en évidence un mécanisme visuel de haut niveau, de se demander si ce type d'activité neuronale est corrélé ou non à la conscience visuelle, i.e. à

la perception subjective que l'observateur a de la scène visuelle. Bien sûr, il ne s'agit pas de prendre cette question au sens trivial, où l'activité de tout neurone du système visuel est en relation avec la perception visuelle : les cellules ganglionnaires de la rétine, par exemple, sont nécessaires à toute perception visuelle consciente. La recherche du "corrélat neuronal de la conscience" (e.g. Crick et Koch, 1990a,b, 1995 ; Koch et Braun, 1996) considère plutôt le problème inverse : une activité rétinienne n'implique pas forcément (dans des cas pathologiques, notamment) une perception consciente ; 2 stimulations rétiniennes identiques peuvent engendrer deux percepts bien différents (en particulier lors d'une rivalité binoculaire; Blake et al., 1971 ; Logothetis et Schall, 1989; Leopold et Logothetis, 1996 ; Logothetis et al. 1996 ; Kovacs et al. 1996 ; Lee et Blake 1999). La rétine ne peut donc être le siège de la conscience visuelle. Une aire visuelle sera ainsi considérée "intéressante" vis-à-vis de la conscience visuelle si l'activité des neurones de cette aire varie avec la perception consciente, l'image rétinienne étant identique.

Nous ne pouvons utiliser ce paradigme pour corréler avec la conscience visuelle l'activité de haut-niveau que nous avons mise en évidence précédemment, car la tâche que nous avons utilisée ne présuppose pas que les sujets doivent être conscients du stimulus. Dans la plupart des cas, les sujets sont cependant capables de reporter ce qu'ils ont vus. Pourtant, une expérience récente menée par Thorpe et al (1999) où les images d'une tâche de catégorisation "animal" pouvaient être présentées jusqu'à 70° d'excentricité, suggère que la conscience visuelle ne soit pas nécessaire pour réaliser la tâche : bien que les sujets soient incapables de reporter les images présentées à la périphérie extrême du champ visuel, ils peuvent toujours réaliser la tâche au-dessus du niveau de la chance à ce degré d 'excentricité.

Cette corrélation n'a cependant pas été étudiée systématiquement, et il reste à démontrer que l'activité différentielle "cible/non-cible" à 150 ms puisse encore être observée dans des conditions où les sujets ne sont pas "conscients" de la stimulation visuelle.

2.5 Deux mécanismes distincts : bases neuronales

Nous avons mis en évidence, puis isolé dans leur décours temporel 2 mécanismes neuronaux distincts participant à une tâche de catégorisation visuelle rapide. Nous avons pu interpréter ces mécanismes et leur rôle fonctionnel, le premier reflétant un encodage sensoriel, le deuxième une véritable catégorisation, sur laquelle peut être basée la décision du sujet. Quelles sont les populations neuronales (i.e. les aires visuelles) qui génèrent ces 2 activités électriques ? L'analyse de sources de potentiels évoqués permet d'esquisser une réponse à cette question.

2.5.1 Méthode

L'analyse des sources a été réalisée avec le logiciel BESA. L'enregistrement des ERPs fournit une topographie de l'activité neuronale à un moment donné ; c'est en fait la projection sur le crâne du sujet (discrétisée sur l'ensemble des 32 électrodes) de cette activité neuronale. L'analyse de sources résoud le "problème inverse" : étant donnée une topographie crânienne, quelle est la localisation la plus probable de l'activité neuronale qui en est la source ? Pour cela, on fait appel à un modèle de la diffusion de l'activité électrique neuronale à travers le cerveau et la surface du crâne. Le modèle utilisé par BESA implique 4 coquilles ellipsoïdales (4-shell ellipsoidal model).

Pour chaque activité différentielle modélisée, nous avons rejeté les électrodes les plus bruitées (9 pour l'analyse décrite en 2.5.2, et 6 pour celle présentée en 2.5.3). Les grandes moyennes (i.e. moyennes sur l'ensemble des sujets) des activités différentielles ont été filtrées (passe-bas, 40 Hz). Une paire de dipôles (ou éventuellement 2, dans un cas qui sera précisé plus loin) a été placée en position centrale, et contrainte en symétrie. L'algorithme de BESA (utilisant une descente de gradient) déplace ensuite cette paire de dipôles et modifie leur orientation jusqu'à obtenir les sources "les plus probables", i.e. celles qui expliquent le maximum de variance du signal.

Classiquement, la vérification de la résistance, et donc de la validité de ces sources, consiste à répéter la même procédure avec différentes conditions initiales (i.e. position et orientation des dipôles) : les sources sont résistantes si toutes les conditions initiales amènent à la même solution finale. Toutes les sources reportées ici sont résistantes selon ce critère.

2.5.2 Activité précoce : bases neuronales

Comme le montre la figure 16, l'activité précoce (ici modélisée entre 80 et 145 ms) différentiant les animaux des moyens de transport semble être générée dans des régions occipitales. Cette première modélisation repose sur l'hypothèse que l'activité neuronale sur cette période est générée par une population neuronale avec une localisation précise (i.e. une seule aire visuelle), et bilatérale car les images présentées occupent de façon égale les hémichamps visuels droit et gauche.

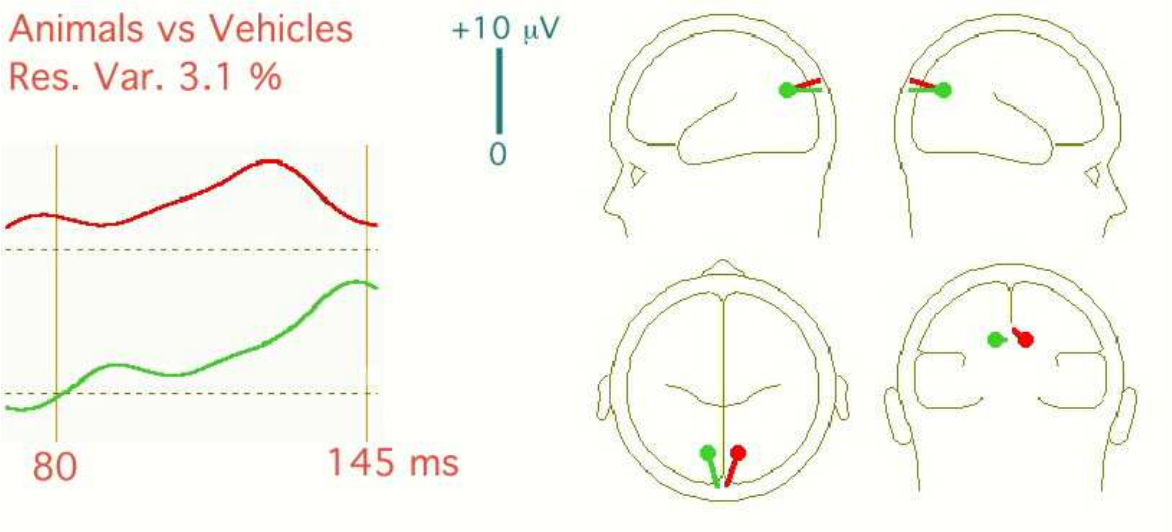


Figure 16. Localisation, orientation et amplitude des sources de l'activité différentielle entre animaux et véhicules, sur la période [80;145ms] post-stimulus. Les dipôles sont contraints en symétrie de position. Ces 2 dipôles localisés dans les régions occipitales expliquent 97% de la variance du signal sur la période considérée.

En fait, si l'on s'intéresse de plus près à cette activité précoce, notamment en observant la figure 3 de l'article 2, il s'avère qu'elle serait plus précisément le reflet de 2 sous-composantes distinctes. En effet, la différence entre animaux et véhicules (non-voitures) présentée à la figure 3c possède clairement 2 pics distincts. De même la différence entre voitures et autres véhicules montre 2 pics de signe opposé (figure 3d), alors que la différence entre animaux et voitures ne présente que le deuxième pic, toutefois plus prononcé. S'il existe 2 pics d'activité distincts, il se peut qu'ils soient générés par 2 sources de localisations distinctes. Un modèle plus exact devra donc utiliser 2 paires de dipôles au lieu d'une. La figure 17 montre ce type de raffinement de l'analyse de sources précédente.

Comme mentionné dans l'article 2, étant donnés la latence de ces activités ainsi que les besoins spécifiques de la tâche (notamment extraction de contours, ségrégation figure-fond), étant donnée par ailleurs la faible précision spatiale des méthodes d'analyse de sources (de l'ordre du centimètre), ces sources sont compatibles avec des activités neuronales générées dans des aires visuelles extrastriées, comme par exemple les parties dorsales de V2 ou V4.

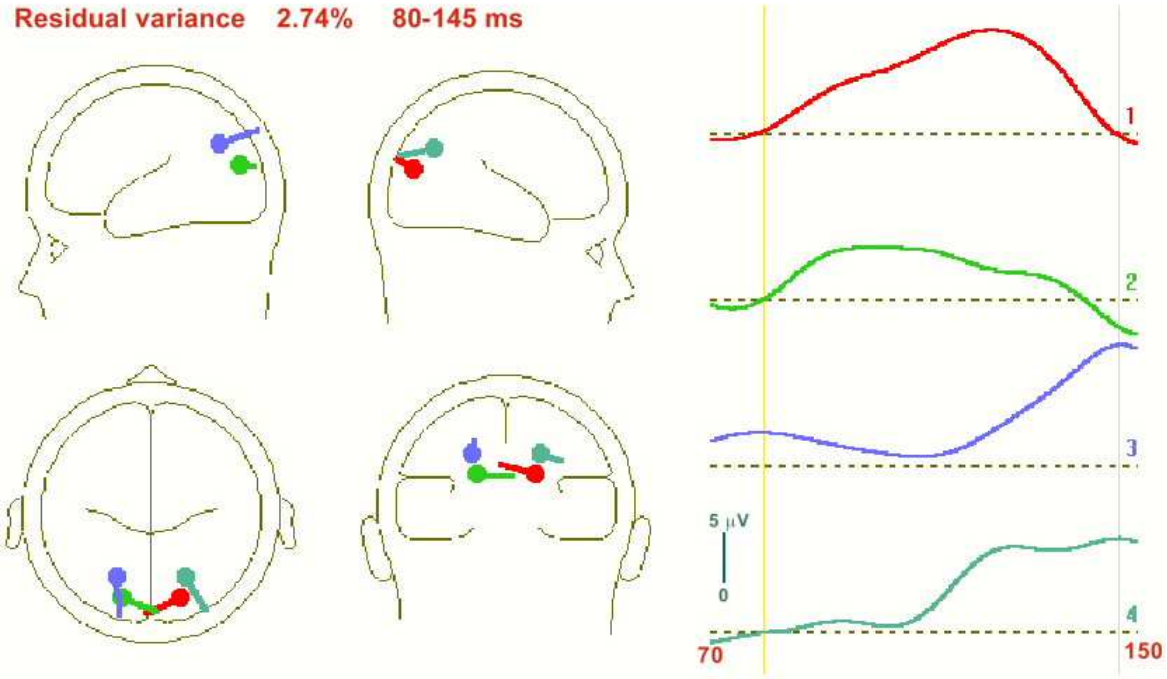


Figure 17. Raffinement de l'analyse présentée figure 16, utilisant 2 paires de dipôles. La première paire a été placée comme pour la figure 16. Ensuite, la première paire étant fixée, la deuxième paire a été positionnée par l'algorithme BESA de façon à expliquer la variance restante. Enfin, partant de ces conditions initiales, les 2 paires ont été repositionnées ensemble par l'algorithme BESA, sur la totalité de la période considérée ([80; 145 ms]). Les amplitudes montrent que l'activité différentielle comporte 2 composantes: l'une plus postérieure avec une latence de 80 ms, la seconde plus pariétale avec une latence de l'ordre de 110 ms.

Cependant, une autre interprétation pourrait également être supportée par la localisation des dipôles présentés figures 16 et 17. En effet, ces dipôles sont situés de façon légèrement dorsale par rapport à la position attendue pour des aires extrastriées comme V2 et V4. Les sources seraient donc compatibles avec des activations sur la voie occipito-pariétale, comme par exemple au niveau du cortex postérieur pariétal (PPC), reconnu pour son implication dans le contrôle et la direction de l'attention spatiale (Lawler et Cowey, 1987; Steinmetz et Constantinidis, 1995), ou dans l'encodage d'une intention de mouvement (Andersen et al 1997; Snyder et al 1997, 2000). Cette activité pourrait également être générée plus tôt dans la hiérarchie des aires pariétales, par exemple dans PO (cortex pariéto-occipital) ou V3a, afférents connus du cortex postérieur pariétal (Cavada et Goldman-Rakic, 1989; Morel et Bullier, 1990; Baizer et al 1991). Une récente étude en MEG a mis en évidence une activité localisée dans le cortex pariéto-occipital à une latence d'environ 110 ms, qui répond sélectivement à différentes formes d'objets (Simo Vanni, séminaire interne CerCo, Toulouse, janvier 2000). Il peut sembler contre-intuitif, à première

vue, qu'une activité différentiant des catégories d'objets (telles que "animal", "véhicule") puisse être générée dans des aires de "bas-niveau" comme V2, V3a, ou dans des aires plus intéressées par les propriétés spatiales des objets (voie "where") telles que PO ou PPC. Il a pourtant déjà été suggéré (Sereno et Maunsell, 1998) que la distinction entre les voies "what" et "where" n'est pas aussi claire que les premières études le laissaient entendre (Mishkin et al. 1983). De plus, comme démontré dans les sections précédentes, le fait que l'activité neuronale dans une région varie, en moyenne, avec la catégorie du stimulus ne signifie pas que la catégorisation est effectuée dans cette région. Il n'est pas impossible que différentes catégories d'objets ne soient pas "saillantes" de la même manière, et qu'elles impliquent donc des différences d'activité électrique, au niveau de l'allocation de l'attention ou de la génération des mouvements oculaires, dans les aires pariéto-occipitales.

2.5.3 Activité à 150 ms : bases neuronales

La seconde activité différentielle mise en évidence dans l'article 2, reflétant le mécanisme de catégorisation sur lequel est basée la décision du sujet, est localisée dans les lobes temporaux (figure 18). Cette localisation est similaire à celle reportée par Fize et al (Cognitive Neuroscience Society, Poster Session, 1998) avec une analyse de sources de l'activité différentielle "animal vs non-animal", et par Fize et al. (2000) avec une méthode d'IRMf évènementielle, là encore pour une tâche de catégorisation rapide "animal vs non-animal". L'interprétation la plus probable est que cette activité serait générée dans l'équivalent humain du cortex inféro-temporal: le gyrus fusiforme. En effet, des enregistrements intracraniaux ont montré que ces régions occipito-temporales ventrales pouvaient répondre à des visages et d'autres objets avec des latences de l'ordre de 150 à 200 ms (Allison et al. 1999). De même, des études en IRM fonctionnelle décrivent des régions du cortex ventral occipito-temporal qui s'activent différemment pour différentes catégories d'objets (par exemple des visages, des animaux, des maisons ou des chaises; Aguirre et al, 1998; Epstein et Kanwisher, 1998; Chao, Martin et Haxby, 1999; Ishai et al, 1999), et qui peuvent s'activer de façon similaire pour des traitements sémantiques de ces catégories, même en l'absence de stimulation visuelle (Chao, Haxby et Martin, 1999).

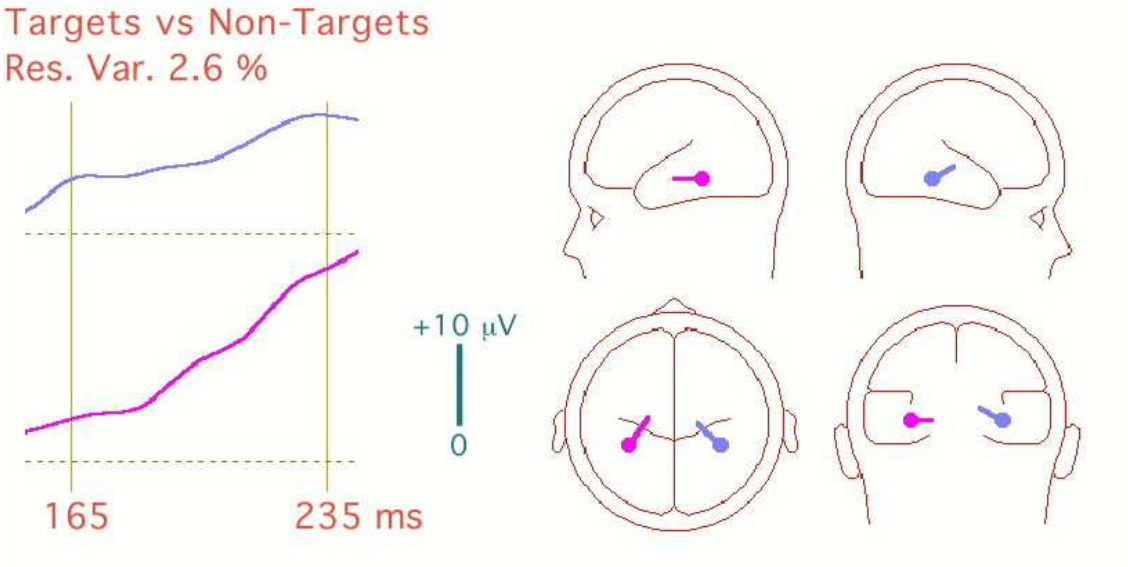


Figure 18. Position, orientation et amplitude des sources de l'activité différentielle "cible vs non-cible" (ici impliquant les 2 catégories visuelles "animal" et "véhicule") entre 165 et 235 ms. Ces 2 dipôles symétriques expliquent plus de 97% de la variance du signal sur cette période. Cette activité semble être générée dans des régions temporales médiales.

Cependant, il ne faut pas oublier que cette activité différentielle reflète principalement, outre la catégorie du stimulus visuel, son statut de cible ou non-cible pour la tâche réalisée. Les assemblées neuronales impliquées dans cette réponse doivent donc avoir "connaissance" de la tâche en cours, de la catégorie visuelle recherchée (cible), afin de comparer cette information au stimulus visuel présenté. Cette activité pourrait provenir de régions temporales médiales, hippocampales ou parahippocampales comme les cortex entorhinal ou perirhinal, connus pour leur participation dans des traitements relatifs à la tâche visuelle, comme par exemple la mémoire de la cible dans des tâches de "match-to-sample" avec délai (Suzuki, 1996, 1999; Higushi et Miyashita, 1996; Buckley et Gaffan, 1998a, 1998b; Buffalo et al, 1998). Comme les aires parahippocampales sont des afférents directs du cortex inféro-temporal postérieur (Baizer et al, 1991), l'effet observé pourrait refléter une collaboration entre ces deux structures: les lobes temporaux medial et inférieur; le dernier apportant l'information concernant le stimulus visuel présenté et son identité, le premier maintenant une "mémoire" ou connaissance de la tâche réalisée. Ces interactions potentielles sont représentées à la figure 19.

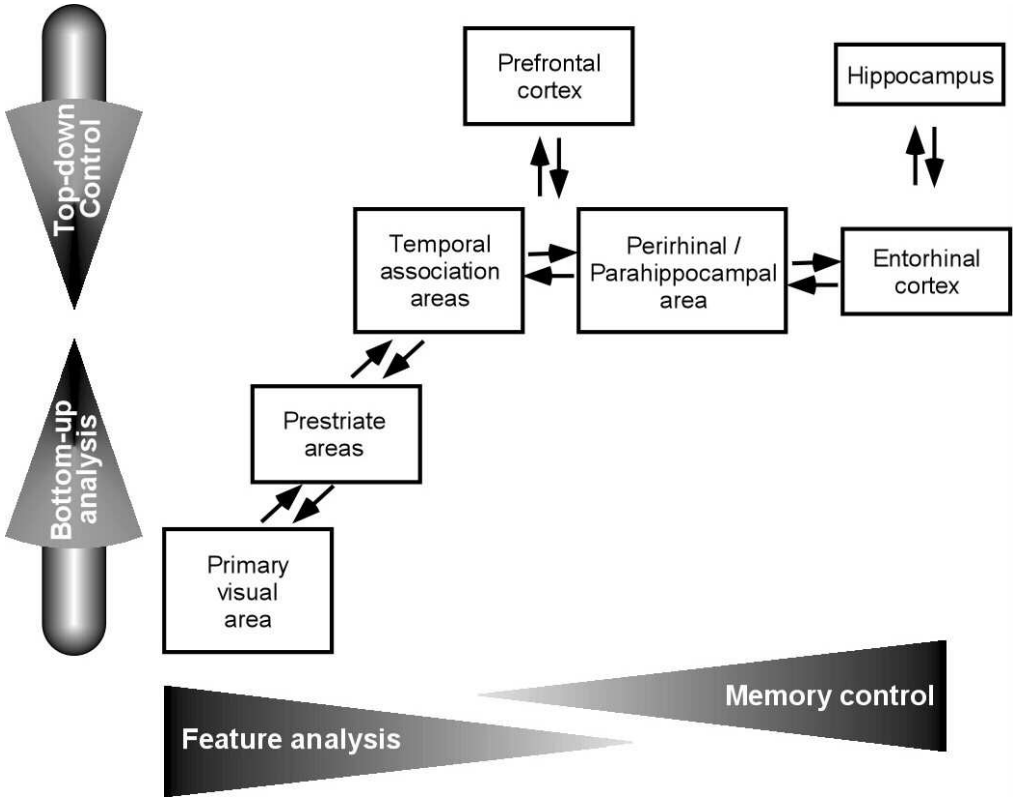


Figure 19. Interactions et collaborations entre les différentes aires du lobe temporal (cortex inféro-temporal et régions parahippocampales). Le stimulus présenté est traité dans les aires visuelles primaire puis extrastriées, avant d'atteindre le cortex temporal où son identité (ou sa catégorie) est faite explicite, et comparée à la catégorie cible recherchée, maintenue en mémoire de travail par les structures parahippocampales. Reproduit d'après Miyashita et Hayashi (2000).

3 Synthèse: la durée du traitement visuel

Dans ce chapitre, nous avons tout d'abord démontré que la durée de traitement de 150 ms mise en évidence par Thorpe et al (1996) n'était pas un cas particulier, dû au choix de la catégorie cible "animal". Il n'est ni plus long, ni plus difficile de répondre à des images d'une catégorie totalement artificielle: les moyens de transport. Cette durée de 150 ms semble donc être généralisable à différentes catégories cibles.

Par contre, les différents résultats expérimentaux présentés au chapitre I section 5.1, et qui suggèrent un traitement visuel de haut niveau beaucoup plus rapide (de l'ordre de 50 ms), ont dû être rejetés. Ils apparaissent en fait être le fruit d'une erreur conceptuelle: le fait que l'activité neuronale varie, en moyenne, avec la catégorie du stimulus visuel, n'implique pas que l'image ait été catégorisée. Les processus de haut niveau liés à la décision sont

bien observés à partir de 150 ms, comme décrit précédemment, alors que les activités précoces (i.e. entre 50 et 100 ms) semblent être le reflet de mécanismes de bas niveau, de différences entre les propriétés visuelles des catégories comparées. Ces résultats, et le décours temporel des différents mécanismes mis en évidence dans la tâche de catégorisation rapide, sont résumés dans la figure 20.

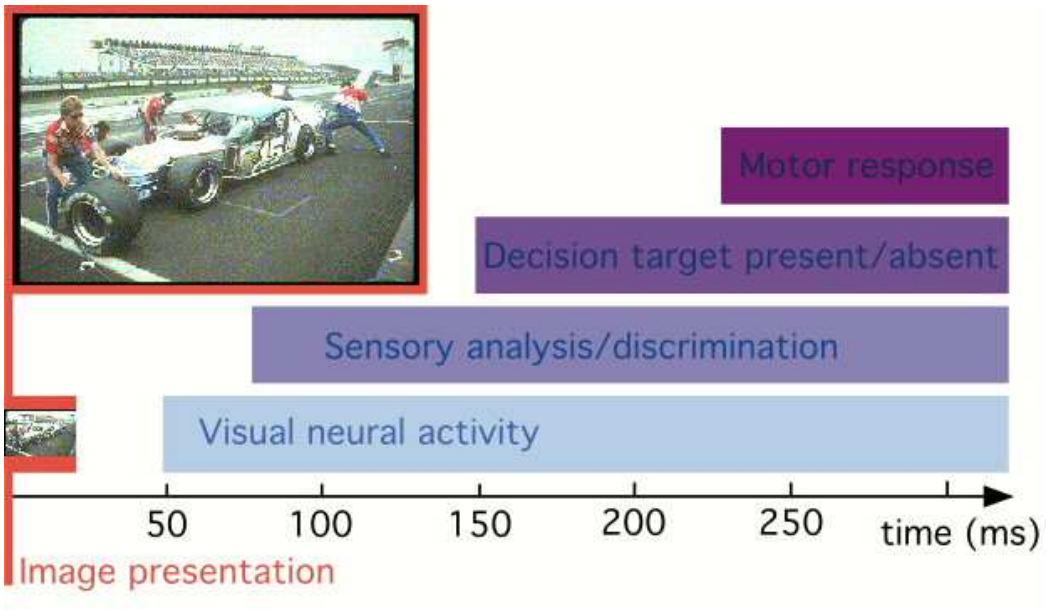


Figure 20. Décours temporel des différents mécanismes neuronaux mis en jeu dans la tâche de catégorisation rapide. L'activité neuronale se reflète dans les ERPs seulement 50 ms après la stimulation visuelle. Après 75 ms, on observe le reflet de l'encodage sensoriel des différentes catégories. A partir de 150 ms, la représentation neuronale mise en jeu est suffisamment abstraite pour permettre la catégorisation, i.e. pour décider qu'une cible est présente et générer une commande motrice de réponse, qui pourra survenir après environ 230 ms.

Bien que 3 fois supérieure à "50 ms" (une durée tout simplement incompatible avec les propriétés du système visuel), cette durée de 150 ms est en fait extrêmement courte, compte-tenu de la complexité des traitements qui doivent être effectués durant ce laps de temps. A partir du moment où une image est présentée à la rétine, le système visuel pourra, en 150 ms, décider s'il s'agit d'un oiseau ou d'un avion, par exemple. Comme souligné dans l'introduction, bien peu de modèles théoriques envisageables pourraient expliquer cette rapidité.

Dans le chapitre suivant, nous nous attacherons à comprendre, d'un point de vue théorique, et à l'aide de simulations et modèles, comment le système visuel des primates

peut réaliser des traitements aussi complexes en si peu de temps. Ainsi, par souci de plausibilité biologique, nous imposerons aux modèles développés les contraintes connues des systèmes biologiques: architecture, connectivité, propriétés physiologiques des neurones (constantes de temps, taux de décharge, ...). De façon générale, nous nous placerons dans une situation où une image est présentée à un système visuel (primate, modèle ou machine), au temps zéro. On se donnera tout simplement 150 ms pour accéder à une représentation de "haut niveau" de cette image c'est-à-dire non plus en termes de pixels et de leur couleur, non plus en termes de contrastes ou d'orientations locales, mais en termes sémantiques: identité, catégorie ("qu'est-ce que c'est?"), importance comportementale ("qu'est-ce que j'en fais?"). Etape par étape, nous définirons des mécanismes de traitement neuronaux, les plus efficaces et rapides possibles, qui permettront de mieux comprendre ce qui peut réellement se passer dans le système visuel humain, durant ces 150 ms.

III. THEORIE: CONTRAINTES, MODELES, SIMULATIONS

Surfing a spike wave down the ventral stream.

1. Contraintes biologiques:

Les données présentées au chapitre précédent démontrent que, à partir du moment où une image est présentée à la rétine, il faut environ 150 ms au système visuel humain pour accéder à une représentation de haut niveau (i.e. sémantique) de cette entrée. Les implications théoriques de ce résultat (présentées au chapitre I section 3.3) sont fondamentales. Nous revenons ici plus en détail sur ces implications, qui constitueront le fondement de la théorie du traitement visuel rapide que nous allons développer dans ce chapitre.

1.1 Traitement feed-forward

Compte-tenu de l'architecture hiérarchique du système visuel, impliquant au moins une dizaine d'étapes synaptiques de la rétine au cortex inféro-temporal où a lieu la reconnaissance des objets, une durée de traitement de 150 ms ne laisse que peu de place à des mécanismes feed-back, i.e. connexions en retour, récurrentes, connexions latérales extensives.

Ceci ne signifie pas que ce type de connexion est inutile pour le traitement visuel. Comme nous l'avons rappelé au chapitre I section 1.3, ces connexions existent bel et bien et constituent une part de la connectivité corticale au moins aussi importante que les connexions feed-forward. Leur rôle est donc indiscutable. Cependant, pour le type de traitement visuel rapide mis en évidence au chapitre II, ces connexions n'ont tout simplement pas le temps d'être mises en jeu. Plus précisément, l'activité neuronale dans le cortex inférotemporal à 150 ms après la présentation du stimulus n'est vraisemblablement pas ou peu influencée par ce type de mécanisme. Par contre, il est très probable que l'information circulant dans ces connexions en retour puisse venir influencer, renforcer ou modifier les réponses neuronales après quelques dizaines de millisecondes. Ce renforcement pourrait même être nécessaire à certains processus tels que la mise en mémoire, l'activation d'une représentation visuelle consciente, ou la modification des poids synaptiques (i.e. l'apprentissage). Une théorie récente (Subramaniam et al. 2000) suggère par exemple que, bien que les 50 premières millisecondes d'activité neuronale dans IT portent la majeure

partie de l'information nécessaire à l'identification d'une image (Tovee et al, 1993; Tovee et Rolls, 1995), c'est uniquement la continuation de cette activité pendant 350 ms supplémentaires qui permettra la mémorisation du stimulus.

Ainsi, nous ne mettons pas ici en doute l'importance des connexions corticales en retour. Nous pensons simplement qu'un modèle biologiquement plausible du traitement visuel devra pouvoir implémenter un mécanisme de reconnaissance ou de catégorisation rapide d'objets où l'information circulera principalement en avant. La mise en jeu de connexions récurrentes ne devra pas être un élément déterminant pour ce traitement rapide. Elle pourra par contre sous-tendre d'autres mécanismes de traitement, plus tardifs, que nous n'aborderons pas ici.

1.2 Un spike par neurone

Bien que nécessaire, une propagation feed-forward de l'information visuelle n'est pas suffisante pour expliquer la remarquable rapidité du système visuel humain. En effet, ce type de propagation permet uniquement de minimiser le nombre d'étapes de traitement que l'information devra traverser. Ce nombre minimal étant fixé (plus d'une dizaine d'étapes synaptiques), il apparaît qu'un neurone participant à ce traitement dispose de 10 à 15 ms seulement (le temps d'intégration rétinienne, de l'ordre de 20-30 ms doit être pris en compte) pour recevoir l'information de ses afférents, intégrer cette information de ses synapses dendritiques distantes vers son soma, générer éventuellement une réponse (spike), et transmettre cette réponse le long de son axone. La seule variable dans ce processus est le nombre de spikes qui constituera la réponse du neurone. Or, la fréquence maximale d'émission de potentiels d'action enregistrée pour les neurones du cortex visuel est de l'ordre de 100 Hz. Chaque neurone ne peut donc générer au maximum qu'un seul spike en réponse à une stimulation. La possibilité d'un codage par fréquence de décharge est donc exclue pour ce traitement visuel rapide, et avec elle la plupart des modèles classiques du traitement visuel (chapitre I section 2), y compris ceux qui faisaient appel uniquement à des connexions feed-forward (par exemple le Néocognitron de Fukushima, 1982).

Ici encore, nous ne mettons pas en doute le fait que les neurones du cortex visuel émettent en général plusieurs spikes en réponse à une stimulation. Le rôle de ces spikes supplémentaires est indiscutable, ne serait-ce que parce que, dans des conditions "écologiques", la stimulation visuelle de notre rétine est quasiment constante. Cependant, pour le type de traitement visuel rapide présenté au chapitre II, seulement un spike pour chaque neurone pourra éventuellement participer à l'activité des neurones temporaux à 150 ms. Un modèle biologiquement plausible du traitement visuel devra donc pouvoir implémenter un mécanisme de reconnaissance ou de catégorisation rapide d'objets où l'information est portée par un seul spike pour chaque neurone. L'utilisation de spikes

supplémentaires ne devra pas être un élément déterminant pour ce traitement rapide, mais pourra par contre participer à d'autres processus, plus tardifs, que nous n'aborderons pas ici.

Dans un premier temps, nous essaierons de définir le type de computation qu'il est possible de réaliser avec un seul spike par neurone. Nous verrons que cette contrainte n'empêche aucunement d'encoder, transmettre et décoder une grande quantité d'information entre 2 populations de neurones. Nous nous attacherons ensuite à définir les possibilités d'un modèle du traitement visuel impliquant uniquement une architecture feed-forward. Enfin, nous intègrerons ces 2 idées dans un modèle biologiquement plausible de la reconnaissance visuelle: un modèle qui permet de réaliser une catégorisation rapide de l'entrée visuelle, tout en respectant les contraintes temporelles des systèmes visuels réels.

2. Coder l'information avec un spike par neurone

L'idée d'utiliser uniquement un spike par neurone à chaque étape semble à première vue extrêmement limitante. Si un neurone ne peut générer que zéro ou un spike, l'information qu'il transmet est a priori binaire (neurone actif/inactif). Cependant, les neurones corticaux ne sont pas des unités de traitement isolées et indépendantes. Chaque neurone appartient à une population, à une aire visuelle, à une colonne corticale, etc Les propriétés physiologiques des neurones réels (intégration d'un potentiel électrique jusqu'à atteindre un seuil; décharge de potentiel d'action si le seuil est dépassé) impliquent qu'à l'intérieur d'une telle population, les neurones les plus activés auront tendance à atteindre leur seuil plus tôt, et donc à émettre un spike plus tôt également (voir figure 9A).

Le moment auquel un neurone donné émet un potentiel d'action porte donc bien plus d'information qu'une simple variable binaire. Dans des conditions où aucun bruit ne viendrait perturber le système, chaque neurone pourrait d'ailleurs encoder précisément son niveau d'activation, valeur analogique, dans sa latence de décharge. Cependant, les systèmes réels sont soumis à une certaine quantité de bruit non négligeable, qui rend l'utilisation de telles méthodes de codage peu probables. De plus, comme mentionné au chapitre I section 4, le décodage de telles valeurs ne peut se faire simplement, et doit par exemple mettre en jeu des lignes à délai, dont la plausibilité biologique laisse à désirer.

Une alternative est de s'intéresser au pattern spatio-temporel des spikes émis par une population de neurones (figure 9B). L'ordre dans lequel les neurones d'une population déchargent reflète leurs niveaux d'activation respectifs. De plus, un neurone peut décoder simplement l'ordre de décharge de ses afférents par un mécanisme de désensibilisation: les spikes reçus se voient attribuer de moins en moins d'importance; le premier participe plus à la réponse du neurone efférent que le suivant, etc (Thorpe et Gautrais, 1997; 1998).

Cette hypothèse permet-elle de rendre compte de l'extraordinaire efficacité et rapidité du système visuel humain? L'article suivant tente de répondre à cette question, en s'intéressant à une des premières étapes du traitement visuel, où la rapidité et l'efficacité sont particulièrement critiques: la transmission de l'information rétinienne au cortex visuel. Plus précisément, nous comparons dans ce contexte les performances obtenues par un codage par l'ordre d'activation tel que nous venons de le décrire, et celles obtenues par des codes plus classiques, basés sur la fréquence de décharge des cellules ganglionnaires de la rétine.

2.1 Article 3 : **VanRullen & Thorpe, 2000.** Rate coding vs. Temporal Order Coding: What the retinal ganglion cells tell the visual cortex. ***Neural Computation*.**

in press in Neural Computation

Rate Coding vs Temporal Order Coding: What the Retinal Ganglion Cells tell the Visual Cortex.

Rufin VanRullen
Simon J. Thorpe
Centre de Recherche Cerveau et Cognition, Faculté de Médecine Rangueil, 31062 TOULOUSE Cedex. France
email: rufin@cerco.ups-tlse.fr thorpe@cerco.ups-tlse.fr

It is often supposed that the messages sent to the visual cortex by the retinal ganglion cells are encoded by the mean firing rates observed on spike trains generated with a Poisson process. Using an information transmission approach, we evaluate the performances of two such codes, one based on the spike count, the other on the mean inter-spike interval, and compare the results with a rank order code, where the first ganglion cells to emit a spike are given a maximal weight. Our results show that the rate codes are far from optimal for fast information transmission, and that the temporal structure of the spike train can be efficiently used to maximize the information transfer rate, under conditions where each cell only needs to fire one spike.

Keywords: neural coding, retina, spike train, information transfer, rank order coding.

1 Introduction

How do neurons transmit information? This question is a central problem in the field of neuroscience (Perkel and Bullock, 1968). While locally, signals can be conveyed by analog and/or electrical mechanisms, over distances information has to be encoded in the spatio-temporal pattern of trains of action potentials generated by a population of neurons. However, the exact features of these spike trains that actually carry information between neurons need to be defined. The most commonly used code is one based on the firing rates of individual cells, but this is by no means the only option. In recent years a strong debate has opposed partisans of codes embedded in the neurons' mean firing rates, and researchers in favor of temporal codes, where the precise temporal structure of the spike train is taken into account (Softky, 1995; Shadlen and Newsome, 1995, 1998; Gautrais and Thorpe, 1998). Here we address this question of neural coding in the context of information transmission between the retina and the visual cortex.

The retina is a particularly interesting place to study neural information processing (Meister and Berry, 1999). First it is relatively easy to stimulate and record retinal cells. Furthermore, the general architecture and functional organization of the retina is remarkably well known (Rodieck, 1998). There is probably no other place in the visual system where one can define more rigorously:

(i) what information needs to be represented.
(ii) how many neurons are available to do it.
(iii) how long the transmission should last.

A widely used simplification states that the information transmitted from the retina to the brain codes the intensity of the visual stimulus at every location in the visual field. While this strong statement can certainly be discussed, it is clear that the aim of retinal coding is to transmit enough information about the image on the retina to allow objects and events to be identified.

One can also consider that the different types of ganglion cells "tile" the entire retina, so that there is little or no redundancy in the number of neurons, which should therefore be kept to an absolute minimum. In particular, in monkeys, the number of ganglion cells is roughly one million. However, the limited redundancy in the number of neurons encoding a given stimulus feature at a particular location does not mean that there is no overlap between ganglion cells receptive fields. In fact, in the cat retina, between 7 and 20 ganglion cells have receptive field centers that share a given common position in the visual field (Peichl and Wässle, 1979; Fischer, 1973). Correlations among the responses of neighboring ganglion cells also demonstrate that they do not operate as independent channels (Arnett and Spraker, 1981; Mastronarde, 1989; Meister, et al, 1995). Nevertheless, it is not clear in the literature how much of that correlation in the output firing pattern can be explained by shared common inputs (either from photoreceptors, bipolar, horizontal or amacrine cells) to the ganglion cells (Brivanlou, et al. 1998; De Vries, 1999; Vardi and Smith, 1996).

Finally, data on the speed of visual processing provide severe limitations to the time available for information transmission through the visual system. First, recorded neuronal latencies can be extremely short: responses start around 20 ms in the retina (Sestokas, et al. 1987; Buser and Imbert, 1992) and approximately 10 ms later in LGN (Sestokas, et al. 1987); the earliest responses in V1 already exhibit selectivity to stimulus orientation around 40 ms post-stimulus (Celebrini, et al. 1993; Nowak, et al. 1995); in IT, face-selective responses begin between 80 and 100 ms after stimulus presentation (Perrett, et al. 1982) and show selectivity to face orientation even at the very start of the response (Oram and Perrett, 1992). Taken together, these data indicate that visual processing should rely on very short transmission times, of the order of 10-20 ms between two consecutive processing stages, and less than 50 ms between the retina and the cortex. The same conclusions can be derived from psychophysical observations on monkeys and humans, in a task where subjects must decide whether a briefly flashed photograph of a natural scene contains a target category such as an animal, food, or a means of transport: monkeys can respond as early as 160 ms after stimulus presentation, and human subjects in around 220 ms (Thorpe, et al. 1996; Fabre-Thorpe, et al. 1998; VanRullen and Thorpe, submitted). Given the large number of synaptic stages involved, it appears here again that information processing and transfer should not last more than about 10 ms at each processing stage, and probably less than 50 ms between the retina and the brain (the delay of transduction in the photoreceptors has to be taken into account). Therefore, further computation should rely on very few spikes per ganglion cell.

As Meister and Berry (1999) argue, computations that use a very restricted number of spikes are difficult to conciliate with the common view stating that retinal encoding uses

the firing frequencies of individual ganglion cells. Classically, ganglion cells are thought to encode their inputs in their output firing frequency (Warland, et al. 1997), and the process of retinal spike train generation is supposed to be stochastic, i.e. subject to a Poisson or pseudo-Poisson noise. As an alternative, Meister and Berry (1999) review a number of arguments for taking into account the temporal information that can be derived from the very first spikes in the retinal spike trains. For example, this information could be represented by synchronous firing among neurons.

Here we introduce another temporal coding scheme, based on the order of firing over a population of ganglion cells. One can consider the ganglion cells as analog-to-delay converters -the most strongly activated ones will tend to fire first, whereas more weakly activated cells will fire later, or not at all. Under such conditions, the relative timing in which the ganglion cells fire the first spike of their spike train can be used as a code (Thorpe, 1990). A more specific version of this hypothesis uses only the order of firing across a population of cells (Thorpe and Gautrais, 1997). This coding scheme has already been proposed to account for the speed of processing in the visual system (Thorpe and Gautrais, 1997, 1998). We have also shown (VanRullen, et al. 1998) that it can be successfully applied to a computationally difficult task such as detecting faces in natural images.

We will compare the performances of this code with two classical implementations of rate coding: one that relies on the spike count, the other on the mean inter-spike interval to estimate the ganglion cells' firing frequencies over a Poisson spike train.

The approach that we use tests the efficiency of the different coding schemes for reconstructing the input image on the basis of the spike trains generated by the ganglion cells (Rieke, et al. 1997). More precisely, we suppose (for simplicity) that a natural input image is briefly presented to the retina, preceded and followed by a dark uniform field. Previous experiments (e.g. Thorpe et al, 1996) suggest that under these conditions, information should be available to the visual cortex as early as 50 ms post-stimulus. We take the position of an imaginary observer "listening" to the pattern of spikes coming up the optic nerve, and trying to derive information about the input image. Of course, we do not consider the role of the visual system in general as being to reconstruct the image in the brain. Rather, this reconstruction should be seen as a form of benchmark, a test of the potential of a particular code.

First, we will introduce a simple model of the architecture of the retina and its functional organization, independent of the way information will be represented. Then we will describe our rank order coding scheme, and show how it can be applied to retinal coding. Finally, we will compare both a "noise-free" and a "noisy" version of this code with two rate-based coding models (one in which the information is embedded in the spike count, the other in the mean inter-spike interval), by estimating the quality of the input image reconstruction that they provide as a function of time.

2 Retinal model

2.1 Wavelet-like transform

We designed a model retina that can be described by the following features :

Our model ganglion cells compute the local contrast intensities at different spatial scales, and for 2 different polarities (ON- and OFF-center cells). We can consider this decomposition as a wavelet-like transform, using Differences of Gaussians as the basic

filters (Rodieck, 1965). The spatio-temporal properties of our model ganglion cells match those of X-type cells: they use linear spatial summation between the center and surround regions of the receptive field and there is no temporal component in the input-output function (Buser and Imbert, 1992). The ganglion cells receptive fields are shown in Figure 1.

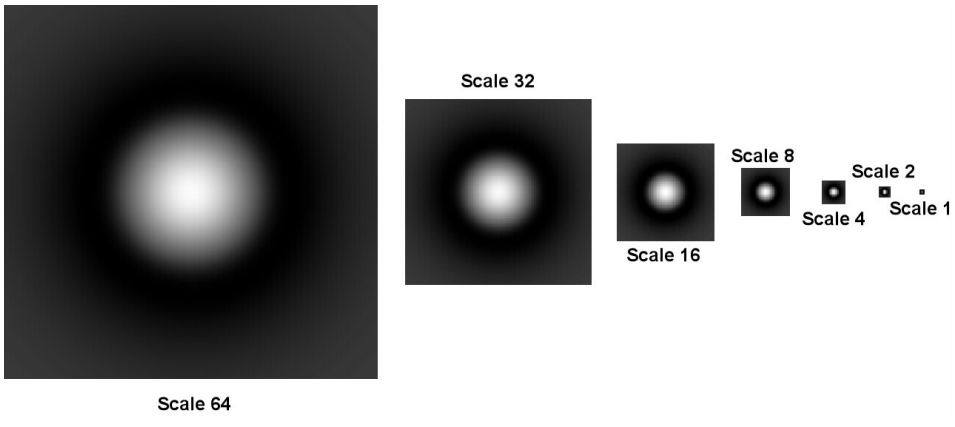


Figure 1. Difference of Gaussians : ON-Center cells receptive Fields. The "Scale 128" filter is not represented, although it was used in our simulations.

We used the simple Difference of Gaussians (DoG) described by Field (1994), where the surround has three times the width of the center. An OFF-center filter was simply an inverted version of the ON-center receptive field. The narrowest filters (at scale 1) were 5x5 pixels in size, and the widest 767x767 pixels (at scale 128). Furthermore, these filters were normalized so that when the input pattern is identical to the filter itself, the result of the convolution at this given scale should be 1. The result of the application of these filters at any position and scale is the output of the wavelet-like transform which produces a set of analog values, corresponding to the activation levels of our model ganglion cells. According to wavelet theory (Mallat, 1989), the wavelet-like reconstruction will simply be obtained by applying on the reconstructed image, for each scale and position, the corresponding ganglion cell's receptive field, multiplied by the corresponding activation value.

More precisely, the contrast at a particular position (x,y) and scale (s) is defined as:

$$Contrast_{\mathrm{Im}}(x, y, s) = \sum_i \sum_j (\mathrm{Im}(i + x, j + y).DoG_s(i, j))$$

where:

DoG_s	denotes the DoG filter at scale s
(i,j)	spans the width and height of the DoG_s filter

Using these contrast values, the reconstruction Im_{Rec} of the image Im is obtained by:

$$\mathrm{Im}_{\mathrm{Rec}}(i,j) = \sum_x \sum_y \sum_s Contrast_{\mathrm{Im}}(x,y,s).DoG_s(x-i,y-j)$$

where:

s	spans the range of spatial scales
(*x*,*y*)	spans the image width and height

2.2 Subsampling

For computational reasons, as well as for biological plausibility, the spatial resolution of the transform varies together with the scale of the filters, so that when the scale is doubled, the resolution is divided by 2. More precisely, the narrowest convolutions (at scale 1) are computed for every pixel in the original image, whereas the filters at scale 2 are applied once every 2 pixels horizontally and vertically. Therefore, the number of neurons per image is no more than 8/3 times the number of pixels in the original image:

let *n* be the number of pixels in the input image, the number of ganglion cells is

$$2.(n + n/4 + n/16 + n/64 + \ldots + n/16384)$$

This organization scheme is detailed in figure 2.

All natural images that will be used in the following simulations are 364x244 pixels in size, and the number of ganglion cells will then be approximately 236,000.

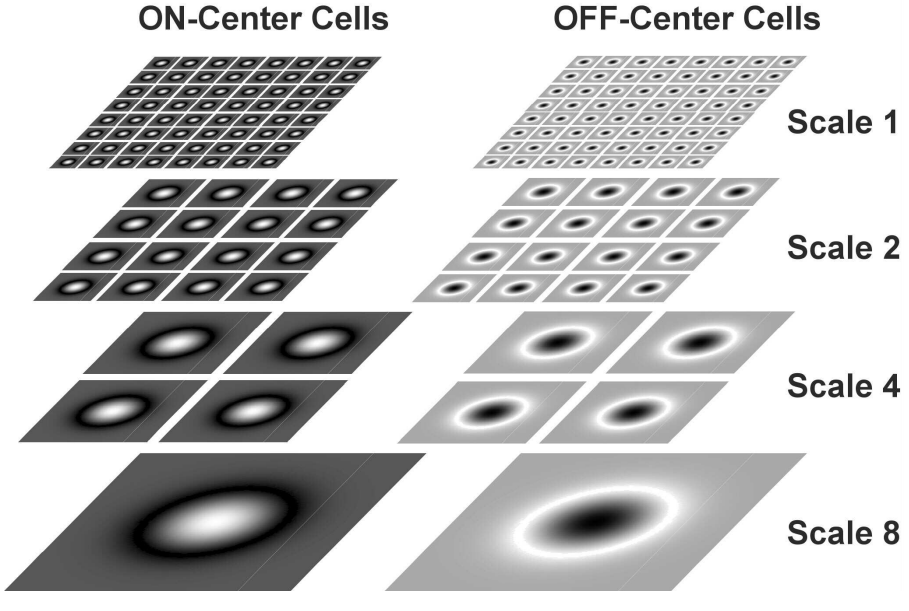


Figure 2. Retinal organization. The image is encoded through a bank of filter maps, with two different polarities (ON- and OFF-center cells) and 8 different scales (only 4 are shown here). The sampling resolution is inversely proportional to the scale.

Of course, we do not claim that this precise architecture is biologically realistic. The real organization of ganglion cells in the mammalian retina has numerous differences with our model. First, the actual recorded receptive fields sizes do not span as many octaves as our model receptive fields do. The ratio between the biggest and the smallest receptive field sizes across the entire retina is less than a hundred, and only around 2 at a given eccentricity (Croner and Kaplan, 1995). However, one could argue that even if there were cells with very large receptive fields, they would be very rare, and therefore difficult to record. Another notable difference is that in biological visual systems, the ganglion cells at each spatial scale are not equally distributed over the retina.

On the other hand, the model we used is one that allows the information in an image to be fully encoded with a wavelet-like transform, and therefore we are able to address the real issue that concerns us here, namely, what would be the most efficient way of transmitting this information to the brain using spiking neurons? Given the architecture, how do these neurons convert an analog intensity value, representing the local contrast in their receptive field, into a succession of firing events, the spike train? What is the optimal way of doing so, in terms of the maximization of information transmission?

In the following sections, we will consider a variety of different coding schemes that could be used to transmit information about the image to the brain. First, we will describe a method that uses an analog-to-delay mechanism coupled with a coding scheme in which the order of firing in the retinal ganglion cells is used to code the information in the image (Thorpe, 1990; Thorpe and Gautrais, 1998). Later we will compare the efficiency of this code with more conventional rate-based codes using either counts of the total number of spikes produced by each cell, or measures of the inter-spike interval.

3 Rank Order coding

The result of the convolution computed by our model ganglion cells on the original image is an analog value, that can be thought of as the neuron's membrane potential. It has to be converted into a spike train that will be transmitted to the cortex via the optic nerve.

3.1 Analog-to-rank conversion

Since we want to show that the relative order in which the ganglion cells will generate their first spike can be used as a code, we can simply assign a rank to each neuron as a function of its activation. The most activated will fire first, and so on. This is supported by the characteristics of integrate-and-fire neurons: the higher the membrane potential, the sooner the threshold will be reached, the sooner a spike will be emitted.

We do not need to model the absolute or relative timing of the spike precisely, since the only relevant variable that will be used for decoding is the neuron's rank. However, it is possible to assign a latency to each neuron, and we will implement such a function in section 4.1 in order to compare the present model with rate coding schemes.

3.2 Rank Order decoding using image statistics: Contrast=f(Rank).

Now that the mechanism responsible for order encoding is described, we need to evaluate the quality of that code. A good and simple way of estimating the information about the visual stimulus that is carried by the spikes along the optic nerve is to use these

spikes to reconstruct the input stimulus.

With the kind of wavelet-like transform that is computed by our model ganglion cells, the reconstruction of the image is simply obtained by applying once again the DoG filters on the result of the previous convolution. Each ganglion cell's receptive field needs to be added to the reconstruction image, at the position corresponding to its center, with a multiplicative value equal to the result of the previous convolution, namely, this neuron's activation level. The problem with our rank order code is that this result has been "forgotten" through the analog-to-order conversion. If the output of the ganglion cells were simple analog values there would not be any such problem, but the only information available in our case is the relative order in which the ganglion cells fired. If neuron *i* fired first, what does it mean in terms of the contrast intensities in the original image? If neuron *j* fired right after neuron *k*, how relevant is the information transmitted by neuron *j* compared to that conveyed by neuron *k*? A simple way of answering these questions is to associate each possible order with the average contrast value that drove the corresponding neuron above threshold. This is roughly equivalent to a sort of reverse correlation analysis.

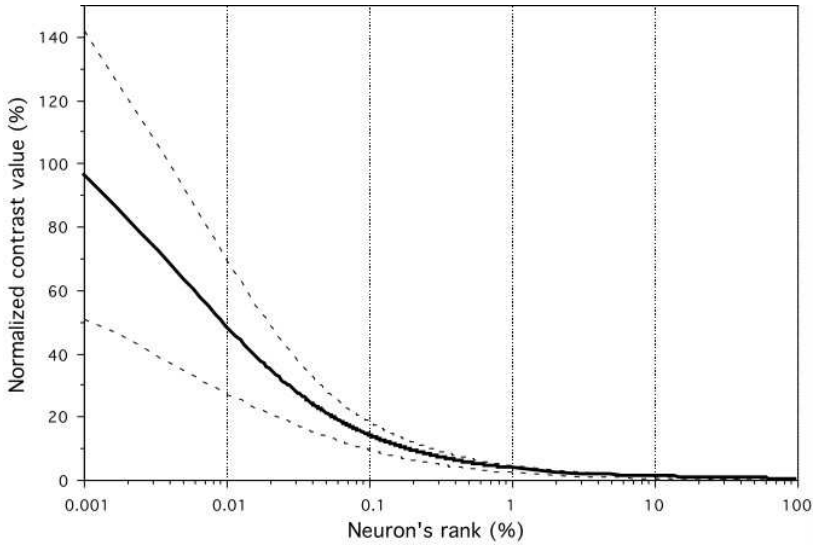


Figure 3. Mean contrast values as a function of the ganglion cell's firing rank (as a percentage of the total number of neurons). X axis in log scale. Y axis as a percentage of the maximum contrast value, averaged and normalized over more than 3000 images. Standard deviation is plotted using dashed lines.

We presented our model ganglion cells with more than 3000 natural images (364x244 pixels), sorted the obtained contrast values for all scales and polarities in decreasing order for each image, and then computed the average contrast value obtained at the location of the first neuron to generate a spike. This location, of course, varied with the different images. The same was done for all second spikes in the images, and so

on, until the last neuron that fired. This procedure is described by the following equation.

The average maximal contrast *MaxC* is defined by:

$$MaxC = \frac{1}{card\{I\}} \sum_{Ime\{I\}} \max_{(x,y,s)} \left\{ \left| \sum_i \sum_j \mathrm{Im}(x+i, y+j).DoG_s(i,j) \right| \right\}$$

where:

{I}	denotes the ensemble of natural images, of cardinal (card{I})
Im	spans the ensemble {I}
s	spans the range of spatial scales
(*x,y*)	spans the image width and height
(*i,j*)	spans the width and height of the DoG_s filter

Note that, because we use a subsample when the scale is greater than 1, the range of possible coordinates (x,y) depends on the scale s. For simplicity, this does not appear in the above equation, nor in the following ones.

The average contrast at rank r is obtained with the same procedure, replacing the function max() (which returns the value of rank 1) by the function $rank_r()$ (returning the value of rank r).

We obtained a list of results that we can consider as a particular kind of look-up table (LUT), that allows to look up the most likely contrast value for a spike with a given rank. The LUT is plotted in figure 3.

Note that the absolute contrast values obtained here depend on the normalization that we applied on the DoG filters, as well as the intensity levels in the input images, that were in the range 0-255 in our simulations. Therefore, these values have been normalized: 100% denotes the average maximal contrast, normalized over the whole set of 3000 images. Furthermore, no preprocessing was applied on the set of input images: different images might thus span a very different range of intensity and contrast levels. Consequently, the variance of contrast values obtained over the image set was relatively high, as shown in figure 3. Note that a high variance is not optimal for rank order decoding, because it means that the contrast value attributed to a ganglion cell firing with a certain rank can be strongly over- or under-estimated.

3.3 Qualitative results: some examples of image reconstruction

Now that we can estimate which contrast value corresponds to a spike arriving with a given rank, it is possible to reconstruct the input image.

The reconstruction is empty at the beginning of the process, i.e. initialized with a gray level of 128. Each time a spike is received from a ganglion cell, the DoG filter of the corresponding scale and polarity is added to the reconstruction, at the corresponding position, and with the multiplicative value corresponding to its rank (the equations describing the reconstruction process with an order code are detailed in section 4.2). Therefore we can stop the spike propagation mechanism at any time, and consider how much information has been received, when a given percentage of the ganglion cells have fired.

Some examples of reconstructed images are presented in figure 4, together with the proportion of neurons that have fired. Note that this percentage cannot be greater than

50%. Since a contrast at a given position and scale in the input image is either positive or negative, a pair of ON- and OFF-center ganglion cells coding for the same scale and position cannot both fire. From figure 4 it appears that, even when less than 1% of the retinal ganglion cells have fired one spike, the identity of objects in the image is often clear.

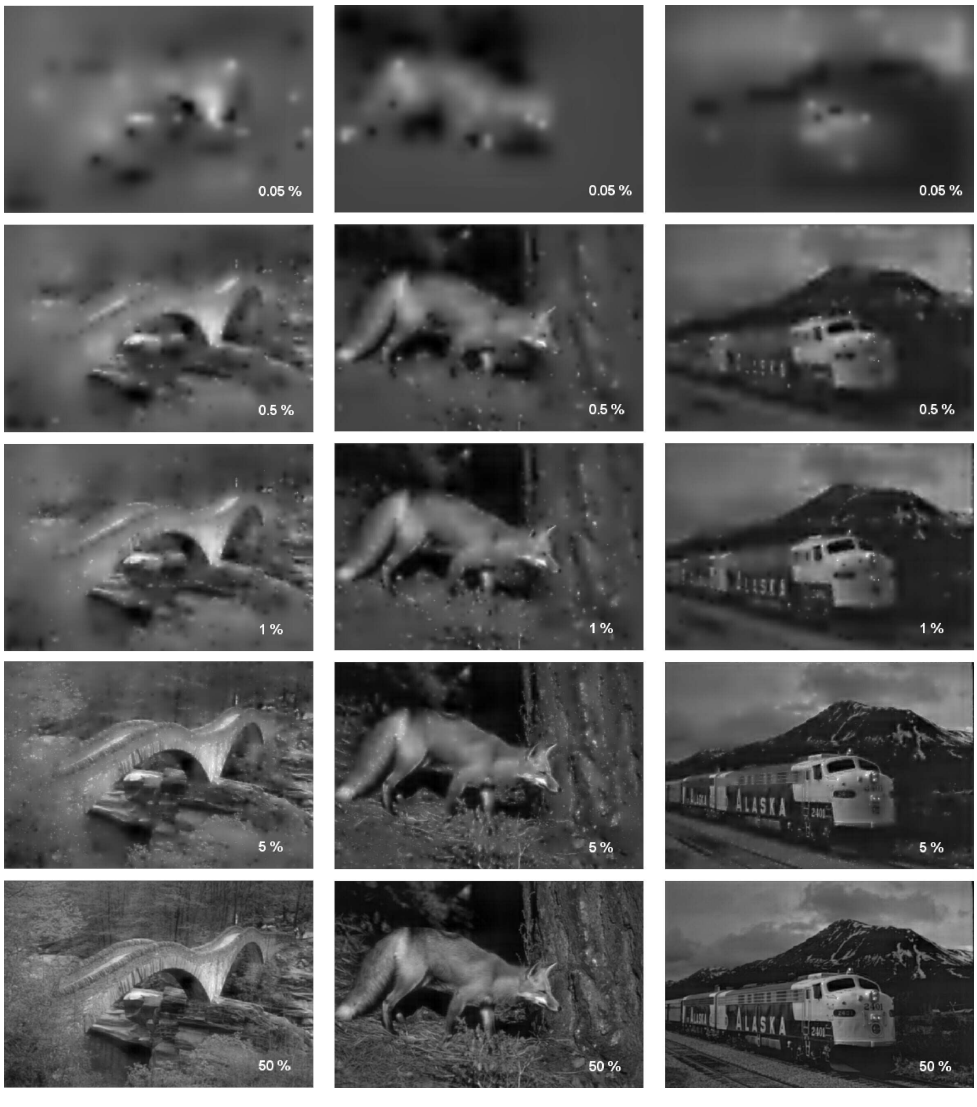


Figure 4. Examples of image reconstruction at different steps. The percentage of neurons that have generated a single spike is indicated for each image.

4 Rate coding models

The code described above seems efficient for rapid information transmission between the retina and the brain. Now we would like to compare it with other classical codes based on neuronal firing rates. Thus we adapted our model to make it suitable for that kind of code.

The architecture of the retina is the same as before, but the neurons are allowed to generate more than one spike, and the whole spike train can then be used for rate decoding. The mechanisms of spike train generation are discussed below.

4.1 Poisson spike train generation

In the rank coding scheme, spikes were propagated with a rank that depends on the activation level of the neurons that generated them. There was no need to specify the precise timing of each spike, since only the relative order was relevant. Now if we assign a delay to the first spikes generated by the ganglion cells, used by the order coding (see 3.1.), as well as for the rest of the spikes in the spike train, used for rate decoding, we will be able to compare the performances obtained with order coding and rate coding under comparable conditions. Note that in the case of our rank order coding scheme, any negative strongly monotonic (decreasing) function would have the same effect. The stronger the input, the shorter the latency. The only thing that would vary would be the amount of time needed to encode a given signal.

The problem here is to choose a function that would transform the contrast intensity in a neuron's receptive field into a firing latency, or a firing frequency. The properties of real neurons can help us decide what function should be chosen.

Croner and Kaplan (1995) proposed that the response of ganglion cells in the primate retina to a given contrast can be described by a Michaelis-Menten function. This kind of function has also been applied to model the response of various sensory neurons (Lipetz, 1971), including horizontal cells in the fish retina (Naka and Rushton, 1966).
The Michaelis-Menten function can be written

$$R(C) = a.C/(b+C)$$

where R is the response (firing rate, Hz) to a given contrast C, a represents the maximum response, and b is the contrast at which the response is $a/2$. Therefore if we consider the mean interspike interval I to be inversely proportional to the neuron's firing rate, we have

$$I(C) = 1/R(C)$$

$$= 1/a + (b/a).(1/C)$$

Here b/a can be thought of as the inverse of the neuron's contrast gain G, and $1/a$ as the neuron's refractory period Ref. Thus, with these new notations, the mean interspike interval, and the ganglion cell's firing rate can be rewritten

$$I(C) = Ref + 1/(G.C)$$

$$R(C) = G.C / (1+Ref.G.C)$$

The neuron's firing rate *R* appears to be the combination of a "free" firing rate *G.C* proportional to the contrast in the neuron's receptive field, and a refractory period *Ref*. The concept of "free" firing rate was introduced by Berry and Meister (1998). It represents the rate that would be produced if there was no refractory period. However, this refractory period has been found to play a major role in the ganglion cell's spike generation mechanism (Lankheet, et al. 1989), and is also thought to be responsible for the reproducibility and precision of the timing of firing events in retinal spike trains (Berry, et al. 1997; Berry and Meister, 1998), under conditions where inter-spike intervals are relatively short.

The values of the parameters *Ref* and *G* must now be chosen. As an example, Croner and Kaplan (1995) give for a "typical" ganglion cells the approximate values 200 for *G* (in Hz . contrast^{-1}) and 10 ms for *Ref*. The refractory period *Ref* acts as a constraint on the observed firing rate *R*, so that *R* cannot be greater than *1/Ref*. Therefore lower values of *Ref* will benefit a count code, because they will allow the neuron to fire at a higher maximum rate. Most ganglion cells in the retina have a firing rate inferior to 100 Hz, thus a refractory period *Ref*=0.005s, restraining the range of possible firing rates to 0-200Hz seems a reasonable value for both the order and count codes. The parameter G, the contrast gain of the ganglion cell, will determine how the "free" firing rate varies with the input contrast. Let *MaxC=1* (i.e. 100%) denote the maximal contrast that can fall into a ganglion cell's receptive field. As an example, if *MaxF*, the maximal "free" firing rate, is chosen equal to 2000 Hz, then $G = MaxF / MaxC = 2000$ (using the previous values) implies that all contrast values above 50% will give a free firing rate *G.C* greater than 1000Hz (i.e. *MaxF/2*), that is to say an observed firing rate *R* greater than 166Hz when *Ref*=5ms (in our case, according to the LUT obtained previously, this should not be more than 0.01% of the neurons). This maximum value of the free firing rate (2000 Hz) is in the same range as the one found by Berry and Meister (1998). In our simulations, we will therefore use a contrast gain *G=2000* and a refractory period *Ref=5*ms. The Michaelis-Menten function transforming the input contrast into a firing rate with the above parameter values is plotted in figure 5, together with the estimated firing rate distribution derived from the distribution of contrast intensities obtained in 3.2. The great majority of neurons appear to have low theoretical firing rates. However, this can be an advantage in 2 different ways: first, the system acts as if it was minimizing the neurons' firing rates and thus the energetic cost for a given amount of transmitted information (Baddeley, et al. 1997); second, most of the neurons are computing in the semi-linear part of their regime, i.e. very similar to what would be expected for an optimal wavelet reconstruction.

Finally, the spike train is obtained by applying a Poisson law on the theoretical firing rate *R* described above. This Poisson process is a probabilistic mechanism known to induce noise on the precise timing of the firing events, but to keep the observed firing rate identical, when considered over relatively large time windows.

Note that a consequence of the application of a Poisson law is that only the ganglion cells with strictly positive firing rates will be allowed to fire. Thus, there will not be any spontaneous activity in the model. Moreover, when an ON-center ganglion cell is activated, the OFF-center cell coding for the same scale and position can not fire. These

are limitations of the stochastic process of spike train generation that we use, that are clearly different from what would be expected in real systems.

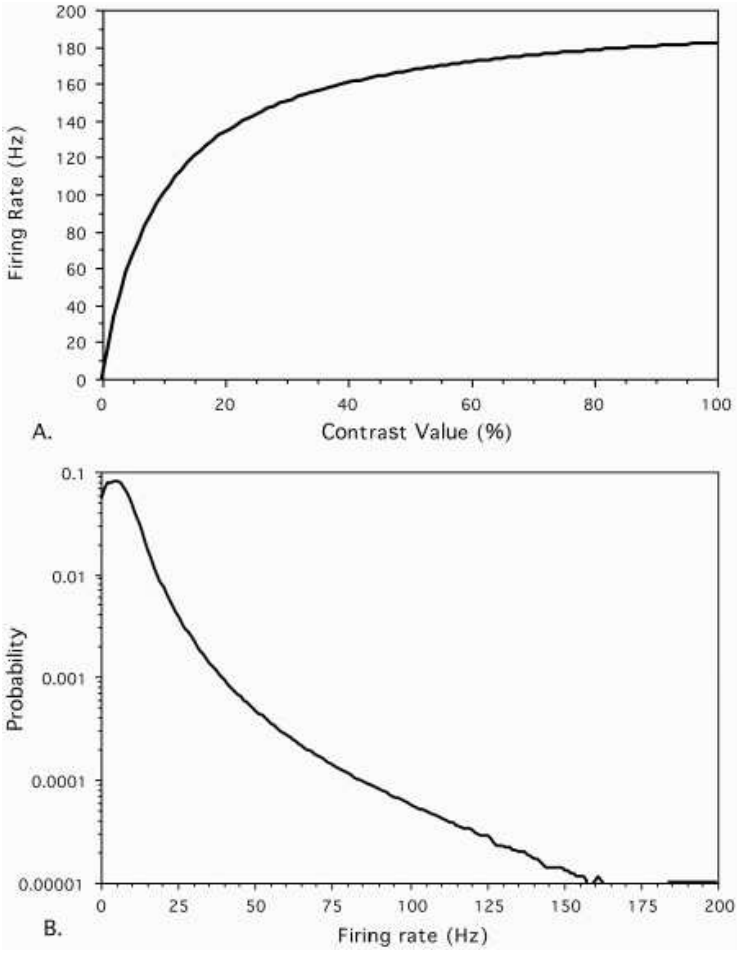


Figure 5. A. Theoretical firing rate as a function of the input contrast (in percentage, notation as in figure 3). Michaelis-Menten function, parameters G=2000, Ref=0.005s (see text for details). B. Distribution of firing rates as derived from the distribution of contrast levels in figure 3, calculated over more than 3000 natural images (Y axis in log scale).

4.2. Different rate codes.

We will test the performances of two different codes that try to use this spike train in order to estimate mean firing rate. The first is a Count Code that assigns to each ganglion cell a value proportional to the number of spikes it has generated. Another code will estimate the mean Interspike Interval (ISI) of the spike train, and assign to a given ganglion cell a firing rate inversely proportional to its mean ISI. Note that this kind of

code requires at least 2 spikes in a spike train to induce non-zero estimated rates. We will call this second code an ISI Code. Clearly, ISI coding is potentially more accurate than the other tested codes, because the interval between 2 spikes is an analog value that can be quickly transmitted and arbitrarily precise.

To keep comparisons with the Order-Coding model as simple as possible, we will let the delay of the first spike emitted by a ganglion cell be the mean interspike interval *I* of the spike train, i.e. the inverse of the neuron's firing rate *R*.

For each code, the decoded value can be interpreted as an estimation of the contrast in a given neuron's receptive field. Thus we can define, for a particular neuron n (described by its position and scale (x_n, y_n, s_n)), the estimated contrast value decoded at time t, by each of the considered codes.

For the order code, we have:

$$\begin{aligned} Contrast_{\text{Order}}(n,t) &= 0 && \text{if} \quad t < \text{Latency}(n) \\ &= \text{sign}(n).\text{LUT}(\text{rank}(n)) && \text{if} \quad t \geq \text{Latency}(n) \end{aligned}$$

where:

sign (*n*) is the sign of neuron n,
-1 if n is an OFF-Center cell, +1 if n is an ON-Center cell

rank (*n*) is the rank of discharge of neuron n

LUT (*r*) is the average contrast value corresponding to rank r (section 3.2)

The estimated contrast value of neuron n at time t with a count code is defined as:

$$Contrast_{\text{Count}}(n,t) = \text{sign}(n).\text{count}(n,t)$$

where:

count (*n,t*) is the number of spikes emitted by neuron n at time t.

Finally, the estimated contrast value derived by an ISI code will be:

$$\begin{aligned} Contrast_{ISI}(n,t) &= 0 && \text{if count}(n,t) < 2 \\ &= \frac{\text{sign}(n).(\text{count}(n,t)-1)}{t_{\text{count}(n,t)} - t_1} && \text{if count}(n,t) \geq 2 \end{aligned}$$

where:

t_k denotes the time at which neuron n fires its k^{th} spike

In the previous equations, the obtained contrast values can be negative. This is only a simplification of the fact that neurons in our model can be of two different types: ON- and OFF-center.

The reconstructed image at time t, with a particular code, can now be defined as:

$$\text{Im}_{\text{Rec}}(i,j,t) = \sum_n Contrast_{\text{Code}}(n,t).DoG_{s_n}(x_n - i, y_n - j)$$

where:

s	spans the range of spatial scales
n	spans the whole set of neurons

This equation is in fact derived from the one introduced in section 2.1. However, here the contrast value applied to a filter at a given position and scale is not the exact value extracted by convolution, but rather an estimation of that value according to a particular code.

Now we can compare the performances of an Order Code, a Count code and an ISI Code under the same conditions. For rank order decoding and image reconstruction, we will simply use the process described above: the reconstructed image is empty at the beginning of the process; when the current timestep corresponds to a neuron's latency, the filter of the corresponding scale and polarity is added to the reconstruction, at the corresponding position and with a multiplicative factor corresponding to the neuron's rank. This multiplicative factor is derived from the LUT obtained previously (3.2).

The spike train decoding and the image reconstruction for the Count Code can be simply obtained: the reconstructed image is empty at the beginning of the process, and every time a spike is received, whatever its position in the spike train, the filter of the corresponding scale and polarity is added to the reconstruction, at the corresponding position. There is no need for a multiplicative factor, since the neuron's activity is supposed to be encoded in the number of spikes emitted (the neuron's observed firing rate R) rather than any other variable.

For the ISI Code, the estimated firing rate of a ganglion cell is recomputed every time the cell generates a spike, starting with the second spike. The change in the estimated firing rate can update the image reconstruction, by adding or subtracting the corresponding filter at the corresponding position with a multiplicative value equal to this change. Here, as for the count code, the analog variable used for reconstruction is the ganglion cell firing rate; one should keep in mind that this rate does not depend linearly on the local contrast, but is transformed through a non-linear Michaelis-Menten function.

Note that for the three different codes, the reconstructed images that are used for result analysis are normalized, so that they all span the whole range of gray levels (here between 0 and 255). Thus there is no need to precisely define the absolute value of the maximum contrast applied to the ganglion cell with the shortest latency (for an order code), the highest count (for a count code) or the smallest inter-spike interval (for an ISI code). Only the relative values attributed to different ganglion cells will have an actual consequence on the reconstruction.

5 Comparison

5.1 Performance estimation: Mutual Information and Mean Square Error

To evaluate the performances obtained with the different models with the different spike train types, we have to define a measure of the quality of the image reconstructions obtained with the different codes.

A widely used measure in image processing is the simple Mean Square Error (MSE), defined by

$$\mathrm{MSE} = \frac{1}{\mathrm{width}\,.\,\mathrm{height}} \sum_i \sum_j (\mathrm{Im}_1(i,j) - \mathrm{Im}_2(i,j))^2$$

Mean Square Error formula.

where Im_1 and Im_2 are the two grayscale images to compare and (i,j) spans the image width and height.

However, a number of problems are associated with this measure. It only takes into account the difference in intensity levels between 2 images. For instance, an image obtained by dividing by a factor of 2 all intensity levels in the original image will often yield a much higher MSE than a "flat" image obtained by assigning to every pixel the mean intensity level of the original image. Yet the former contains all the information present in the original image, whereas the latter contains virtually no information at all.

To overcome these limitations, one can also use the Mutual Information measure (Rieke, et al. 1997). We assume that the original and the reconstructed images are 2 random variables X and Y (respectively) of distributions P(X) and P(Y). The mutual information I is the average information that observations of Y provide about X (and conversely).

$$\mathrm{I} = \iint \mathrm{P}[\mathrm{X},\mathrm{Y}] \log_2 \left(\frac{\mathrm{P}[\mathrm{X},\mathrm{Y}]}{\mathrm{P}[\mathrm{X}]\mathrm{P}[\mathrm{Y}]} \right) [\mathrm{dX}][\mathrm{dY}]$$

Mutual information formula.

The Mutual Information (MI) between 2 images can be thought of as the average information (in bits) that a pixel of one of the images provides about the corresponding pixel in the other. For example, the self MI (MI between an image and itself) of a 8-bit gray level image is 8 if and only if the image actually contains 256 different gray levels and they are equally distributed. The MI is a more accurate measure of the relative information between two images than the MSE, except that it does not account for differences in gray level. At one extreme, the reconstructed image could be the exact negative of the original, and the MI would be maximal at the same time.

Therefore a good coding scheme should lead to a good reconstruction both in terms of MI and MSE.

5.2 Results

50 images were randomly chosen among the 3000 natural images used previously (3.2), and presented to our model retina. The contrasts at different scales and polarities were computed using the DoG filters.

A first step was to estimate the theoretical limit performance of the model retina that we used. For that purpose, we reconstructed the 50 images on the basis of the exact analog contrast values obtained above: for each position in the visual field, and each scale (one should keep in mind that the visual field is subsampled when the scale increases), the corresponding filter (ON-center if positive contrast, OFF-center if

negative) was multiplied by the absolute analog contrast value obtained for that neuron, and applied on the reconstruction. This algorithm is directly derived from the wavelet theory (see equation in section 2.1). On the basis of these reconstructions, we were able to calculate the MSE and MI theoretical limits of the model, with the same image set that will be used for estimating the performance of the different codes.

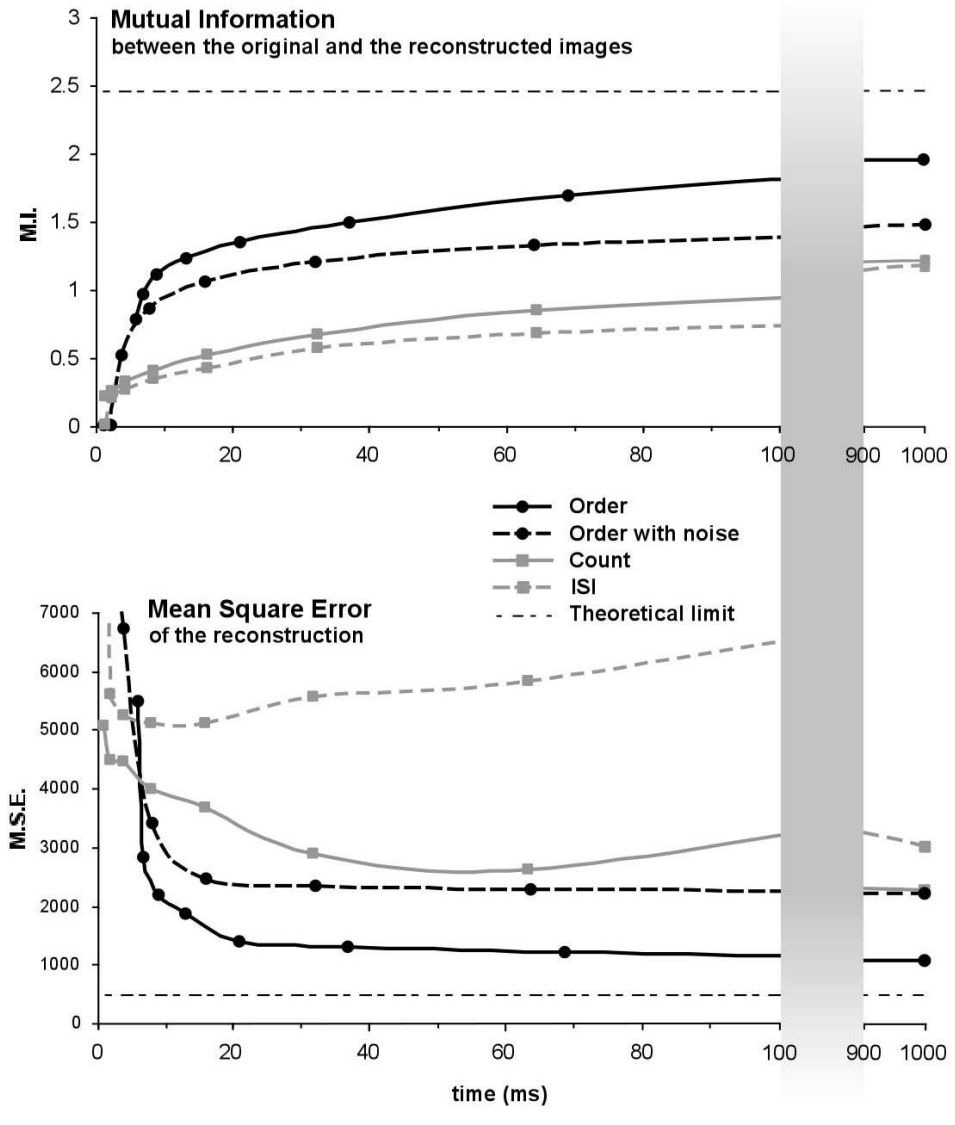


Figure 6. Results. Mutual Information and Mean Square Error of the images reconstructed by the different codes between 0 and 100 ms after stimulus presentation, as a function of time. The theoretical limit performance of the model is plotted as a dotted line. Performance between 900 and 1,000 ms is also indicated on the right.

Then the contrast values were used to compute the firing rate *R* and the first spike latency *I* of all ganglion cells in the retina. Time was divided into 1ms time steps, and for each ganglion cell, a Poisson spike train was generated. A cell could emit one or more spikes at any given time step, with a probability depending on its theoretical firing rate with a Poisson law.

For the Order Code, every cell generated its first and only spike during the time step corresponding to its latency *I*. To make the comparison with (noisy) rate codes as fair as possible, a second version of this order-based reconstruction procedure was applied using latencies subject to a Gaussian stochastic distribution, centered on the theoretical latency *I*, with a standard deviation $\sigma=I/5$. This standard deviation can be thought of as a 20% time jitter applied on a neuron's latency. The value of 20% means that the minimum jitter (corresponding to the most activated cells) will in any case be greater than 1ms. Under these conditions, according to the average neuron's firing frequency distribution (figure 5), the average time jitter observed over an infinite time period would be approximately 60 ms. Note that the stochastic distribution applied can lead to negative latencies with non-zero probability, for any neuron considered: the (very rare) observed negative latencies were replaced by zero latencies. It is worth underlining the fact that the new procedure introduced here is not a different coding scheme than the order code. Rather, it is a test of how the rank order coding scheme can resist a decrease in timing precision.

For each image, the same exact spike trains were used for decoding and reconstruction purposes by the different rate codes described. At different time steps (1,2,4,8, ...1024 ms for the rate codes as well as the noisy order code, *Ref*+1, *Ref*+2, *Ref*+4, …*Ref*+1024 ms for the order code), the reconstructed images derived by all different codes were normalized and saved for further analysis. The average MI and MSE of the reconstructions could then be plotted for each code as a function of time. Note that this renormalization process, i.e. an adjustment of gray levels between 0 and 255, can only have an influence on the estimated MSE, but not on the MI, which is independent of the actual gray level values. The MSE can thus be found to be relatively high with input images that span very few gray levels, i.e. low contrast images (one should keep in mind that there was no preprocessing or normalization on the set of input images). Results are presented in figures 6 and 7.

5.3 Analysis

In terms of MI as well as MSE, the Order Code (i.e. both the "reliable" and "noisy" versions of that code) is better than either of the firing rate based codes (Count Code and ISI Code). Both rate-based codes appear to be limited by the probabilistic process in the Poisson spike train generation as well as the relatively low average firing rate *R*. Nevertheless, it is worth noticing that the Poisson nature of the spike firing in the two rate codes means that some information can indeed be transmitted during the first 5 ms after stimulus presentation, whereas with the ("reliable") order code, the first 5 ms are totally blind. Recent studies support the idea that this advantage explains how rate codes can read out neuronal responses over short time windows (Panzeri, et al. 1999). However, this early information transmitted by the rate codes is clearly not sufficient here to lead to a better reconstruction than the one obtained with the order code. The

saturation of the observed firing rate R at 200Hz is clearly a limiting factor. Indeed, count codes rely on the number of spikes received from each ganglion cell, and therefore give better results with high firing rates. Also, the observed firing rate is not linearly dependent on the input contrast, which is not optimal for a wavelet-like transform.

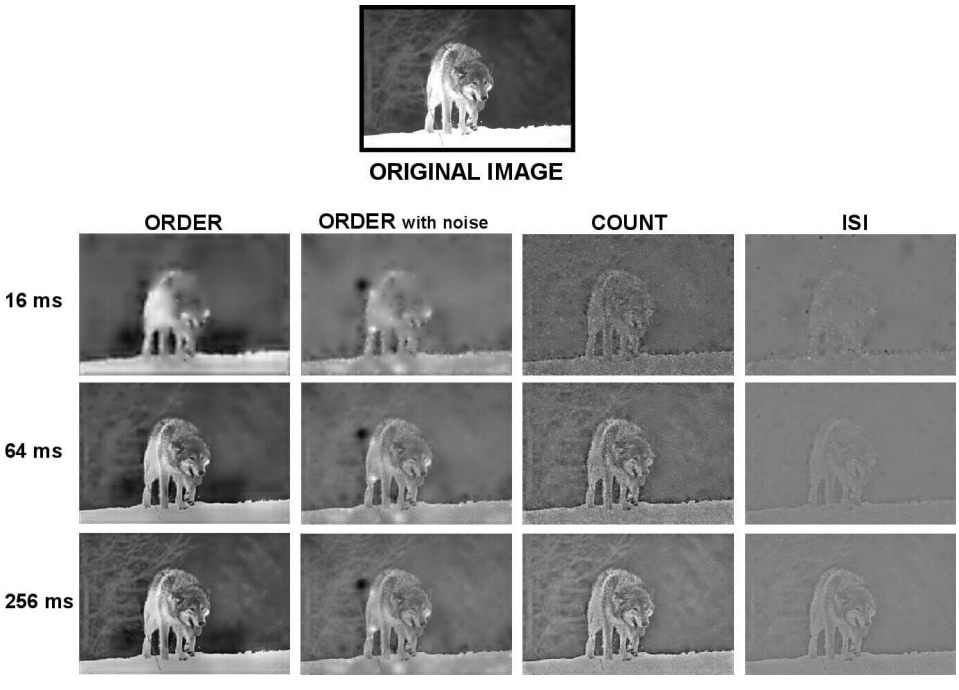


Figure 7. Image reconstructions obtained with the different codes at various time steps after stimulus presentation.

These limitations appear more clearly on the examples of reconstructed images in figure 7. The information transmitted by the two rate codes during the first milliseconds is subject to a noise that cannot be compensated for by the redundancy in the number of spikes. This noise implies spurious high firing rates from ganglion cells not necessarily carrying the most relevant information. As a consequence of the contrast saturation at these points, less importance will be attributed to other regions.

The count code, in the first few milliseconds of computation, i.e. when most of the neurons have not fired at all, and the great majority of the rest has only fired once, acts as a binary code. Even after 100 ms, if the highest observed firing rate is around 200 Hz, which corresponds to a spike count of 20 spikes, the count code can only differentiate between 20 different contrast levels. In the same time with an order code the system can already signal 100 different contrast levels (because the time step of the simulation is 1 ms).

With the ISI code, even after more than 200 ms, a ganglion cell that has generated only 2 spikes, but very close in time (i.e. on the order of 1 ms), will be attributed much

more importance than one that has already generated 10 spikes, but separated by 20 ms on average. This illustrates why this code can only be reliable after every neuron has fired a considerable number of spikes (the more the better in the presence of noise, and at least 2 if the spike train is a noise-free process).

In the case of a rank order coding scheme, one spike per neuron is sufficient to generate the reconstruction. Nevertheless it is not a binary code: the importance given to a spike depends on when it has been emitted relative to the other cells in the population. Since a ganglion cell will fire sooner if it is highly activated, the first information to appear on the reconstruction will be the most relevant. Indeed, more than 90% of the information that can possibly be transmitted with this code is sent during the first 100 ms. Furthermore, with the parameter values that have been chosen (section 4.2), after only 15 ms of computation, the first 1% of the ganglion cells have fired. It can be seen on figure 4 that, most of the time, this information is already sufficient to identify the object in the image. However, it is true that a reconstruction with a rank order coding scheme can not exactly reproduce the original intensity levels, as can be observed on figure 7, because it applies on every image the average statistical distribution of intensities that we derived from thousands of natural images (section 3.2). The specific distribution of intensity levels corresponding to a particular image is "forgotten" by the system during the analog-to-rank conversion. But in many cases, this is actually an advantage: it can be seen as an automatic renormalization of inputs. In particular, this scheme is completely invariant to changes in global contrast or mean luminance level. An image at low contrast will produce a similar result, whereas with a Poisson rate code, the effects of noise would be even greater.

Finally, the resistance of rank order coding to a considerable time jitter applied on the neurons' latencies needs to be stressed. As discussed later, there are numerous arguments for taking into account millisecond-precise firing events. One could argue, on the other hand, that this precision will only hold for relatively high input contrasts. Here we have shown that, with a time jitter of at least one millisecond for the most activated neurons, and of the order of tens of milliseconds for the great majority of neurons, the rank order coding scheme can still outperform rate codes based on Poisson stochastic spike trains. This is even more surprising because the statistical rank distribution of natural images (namely, the average contrast value corresponding to a neuron firing with a given rank) has been computed (section 3.2) under conditions where the firing order was fully reliable. If the ganglion cells firing process actually was unreliable, this would certainly lead to a different statistic over the ensemble of natural images, reflecting the fact that the first neurons to fire are not always the most activated. There is no doubt that using this new statistical distribution instead of the one that we applied here would improve the order reconstruction, under conditions where the neurons' latencies are not fully reliable.

In summary, codes based on the firing probability of the neurons are not optimal for information transmission between the retina and the brain. It is important to make use of the precise temporal information carried by the spike trains, when it is available. Furthermore, assigning a greater impact to the first spikes generated in the retina is a good way of maximizing the information transfer rate.

Under these conditions, the very first spikes occurring after stimulus presentation can carry enough information to allow considerable further cortical processing.

6 Discussion

It has already been argued (Gautrais and Thorpe, 1998) that codes based on the neurons' firing rates are unlikely to be efficient enough for fast information processing. During the first 10 ms of computation (i.e. restricting the number of spikes per neuron to zero or one), n neurons can only transmit $\log_2(n+1)$ bits of information with a count code, whereas $\log_2(n!)$ bits of information can be transmitted with an order code. While ISI codes could transmit an arbitrarily large amount of information with 2 spikes per neuron, no information at all is transmitted by the first spike. It is also worth noticing that these calculations constitute an upper bound on information transmission under noise free conditions. Using a Poisson process for spike train generation just makes things even worse. As Gautrais and Thorpe (1998) argued, to overcome these limitations, rate codes need to be either time-consuming, or they require a high redundancy in the number of neurons encoding a simple analog value (see also Shadlen and Newsome, 1998). In the retina, the limited number of ganglion cells does not allow for such a redundancy. Here we have shown, using an information transmission approach, that rate codes indeed require too much processing time to be efficient under these conditions.

Our analysis of the problems arising as one tries to decode the neurons' firing rates has also shed light on other interesting questions. In particular, the ISI code has not proved as efficient as expected. In our implementation of the ISI code, the estimated firing rate was updated only when a new spike was received. But if a neuron has received 2 spikes in a few milliseconds, and no information at all for the next 20 ms, should the estimated inter-spike interval be updated? Or should the estimated firing rate remain high? In the former case, the loss of computational efficiency is clear, because all values need to be updated at each time step. In the latter, the interpretation of the firing rate is far from optimal, as shown in our simulations. Perhaps one could devise a better decoding mechanism, taking into account some combination of the information in the count and the ISI code. However, this would require additional decoding machinery in the postsynaptic cell.

Another argument with which rate codes can not easily cope is the existence of fast post-synaptic plasticity (Abbott, et al. 1997; Markram and Tsodyks, 1996). A count code is fairly useless in the presence of a fast synaptic depression, that will only attribute some meaning to the first few spikes received. However, this kind of synaptic advantage to the first spikes generated, i.e. those that carry the most relevant information, is exactly what one would need in order to decode a rank order code.

On the other hand, the idea of a rank order coding scheme is also associated with a number of limitations, but as we will argue here, these limitations are more theoretical than practical. First, the order code as described here requires a general "reset" between the processing of two successive images, so that the first spikes emitted in response to a new image are not considered as the last spikes triggered by the previous one. Visual saccades or micro-saccades (Rodieck, 1998; Martinez-Conde et al, 2000) could constitute the functional basis for such a reset mechanism. A much simpler solution however is to restrict our investigation to conditions where the input image is flashed briefly, preceded and followed by a uniform dark field: the response of the retina is a

short wave of asynchronous spikes, and the time between two successive waves can be made long enough for recovery (i.e. reset) to occur. Indeed, the very fast scene processing demonstrated by Thorpe et al. (1996), Fabre-Thorpe et al (1998), or VanRullen and Thorpe (submitted), was achieved under precisely these conditions. Another point worth discussing is the assumed "globality" of the order code. For reconstructing the input image, we have used the order of firing accross the whole neuron population. One could argue that this is neither realistic nor fair, when the rate coding mechanisms described are essentially local. However, as we underlined before, the idea of a global image reconstruction certainly does not correspond to what real cortical neurons do. A cortical neuron receiving retinal information via the LGN will only need to know the order of firing of afferents inside its own receptive field. Such local order decoding schemes have been shown to account for simple-cell oriented (Thorpe and Gautrais 1997) and more complex receptive field properties (VanRullen et al, 1998).Therefore, if global information is required for the whole image reconstruction, more local information is sufficient for cortical processing. In the same way, the rather wide range of spatial scales that were used in our model were necessary to reconstruct the whole image in an efficient way. However, a limited number of spatial scales will be sufficient to fully describe the visual information falling into one cortical neuron's receptive field (for example, only 3 spatial scales will describe the stimulus in a 15x15 pixels cortical neuron's receptive field). It is reasonable to believe that what we have demonstrated here for the whole retinal image will still hold for a single cortical neuron receptive field, which can be considered as a much smaller image. In conclusion, most of the theoretical problems associated with the idea of a rank order code can be overcome with very few assumptions about the simulated visual processing and the underlying cortical organization. The main question remaining, and that we have left open until now, concerns the assumed temporal precision of the neurons' latencies.

There are numerous arguments for taking into account the precise temporal structure of the spike trains generated by the ganglion cells. The timing of firing events has been found to be highly reproducible, with a timing jitter as low as 1ms (Berry, et al., 1997; Reich, et al., 1997). This temporal structure has been shown to carry several times more information than the mean firing rate alone (Berry, et al., 1997; Naka and Sakai, 1991; Reich, et al., 1998). More importantly, the latency of the first spike generated by ganglion cells has been found to vary systematically with the input contrast, a property that could not be explained as a trivial consequence of the change in firing rate amplitude (Sestokas, et al. 1991). We have shown here that it is easy to take advantage of this temporal information and requires few assumptions about the neurons' properties. Furthermore, the integrate-and-fire ganglion cells that we used for that purpose are well known to reproduce the temporal properties of retinal spike trains described above (Reich, et al., 1997, 1998).

However, most studies refer to the mean firing rate as the only relevant variable of a ganglion cell's response (Warland, et al., 1997). This idea is also widely used to model neural computation throughout the visual system (Heller, et al, 1995; Gerstner and van Hemmen, 1992), despite the fact that numerous studies have reported a high temporal precision of individual spikes (millisecond or submillisecond range) in various visual areas such as V1 (Richmond and Optican, 1990; Victor and Purpura, 1996), V2, V3 (Victor and Purpura, 1996), IT (Nakamura, 1998) or MT (Bair and Koch, 1996). Importantly, the latency of the first spike of a spike train has been found to be the most

reliable (Mainen and Sejnowski, 1995). The question of the efficiency of rate codes is the subject of a strong and still open debate (Softky, 1995; Shadlen and Newsome, 1995). In this context, we believe that our approach constitutes a strong argument in favor of fast codes that use the temporal structure of the spike train instead of a simple mean firing rate, estimated from the spike count or the mean ISI.

The idea that the mean firing rate might not be the optimal way to describe neural activity is not new (Perkel and Bullock, 1968; Thorpe and Imbert, 1989). But many of the alternative temporal coding schemes that are currently attracting so much interest (Bair, 1999) involve spike synchronization as a way of increasing information transfer (Meister, et al. 1995; Engel, et al. 1992; Singer, 1999; Eckhorn, 1994). Here we have explored another type of temporal coding scheme that relies on spike asynchrony rather than synchrony, and shown that it could be applied to fast information transfer between the retina and the visual cortex. In particular, we show that many scenes can be recognized when only 1-2% of the neurons have fired one spike. This is important, given the speed with which the visual system is known to operate. A pure rate code would simply not be adequate.

The rank order code, as well as the two rate codes examined here, do not constitute the complete set of possible coding schemes, and further work is needed to evaluate other candidate codes.

For example, although this is hardly plausible from a biological point of view, the rate codes implemented here could be tested under noise-free conditions. Another option would be to choose another spike distribution (e.g. a gamma function) for the spike train instead of the Poisson function that some studies consider as a non-optimal description of the statistical properties of neuronal spike trains (Baddeley, et al. 1997; Reich, et al. 1998). To overcome the limitations induced by the non-linear contrast-to-rate transformation (Michaelis-Menten function), an efficient rate code could also estimate the neuron's "free" firing rate (see section 4.1) instead of its theoretical firing rate, subject to a saturation at high contrast levels (Berry and Meister, 1998).

Other temporal codes could also be investigated. First, the rank order coding scheme could be applied to the first spikes of the Poisson spike trains that were used for rate decoding in our simulations. Nevertheless, previous considerations on the precision of the timing of individual spikes, and on the relevance of the first spike latency, show that it is perfectly reasonable to use millisecond-precise firing events. Another temporal code could extract the neurons' exact firing latencies instead of their rank of discharge. This would allow the fast transmission of analog values. Gautrais and Thorpe (1998) showed that this was indeed a very powerful code. However, it implies that the brain should have an idea of the time of stimulus presentation (zero-time), which is not realistic. Moreover, such a code would be dependent on the contrast and luminance levels in the input image, whereas one of the advantages of the order coding scheme presented here is its invariance to such changes. To be as complete as possible, codes taking into account synchronous firing events among neurons should also be tested. Recordings of neural responses in the LGN have demonstrated that correlated activity can occur with a millisecond precision (Alonso, et al. 1996), and could be used to reinforce the thalamic input to visual cortex or serve as an additional information channel (Dan, et al. 1998).

Finally, it should be possible to evaluate "hybrid" codes, regrouping features from two or more of the codes described above. For example, we could design a code in which the high temporal precision of the first spike is used to encode as much information as

possible on the basis of rank order coding, and afterwards the neuron switches to a more sustained Poisson-like discharge for the rest of the spike train. This latter analog information could then refine the one derived by the order code.

It is important to stress that the model retina that we used in this article is unrealistic for a number of reasons. First, the model neuron's properties do not include an adaptation mechanism, which is known to influence the response of real ganglion cells (Smirnakis, et al. 1997). Second, there were no lateral interactions between our ganglion cells, whereas these interactions have been found to play a major role in retinal processing (Nirenberg and Latham, 1998; Meister, et al. 1995; Brivanlou, et al. 1998). A recent study (Berry, et al. 1999) has also demonstrated that ganglion cells are sensitive to moving stimuli in such a way that they can help predicting the next position of the visual stimulus. To model the ganglion cells population and topography more accurately would also require implementing a fovea and a resolution decrease with eccentricity (Croner and Kaplan, 1994; Curcio and Allen, 1990; Crook, et al., 1988).

But here the idea was to keep the model as simple as possible, in order to show that our hypothesis could lead to efficient results, when compared to other rate-based codes. That is why we did not try to mimic the mammalian retina in detail. At first sight, it seems obvious that any improvement in the architecture of the model would benefit the rank order coding scheme as much as the rate codes, though this statement could be directly investigated by making our model retina more similar to the mammalian retina. Furthermore, there is little doubt that such a scheme, being biologically more plausible, could save a great amount of neurons, spikes, and computational time.

An interesting alternative to the purely theoretical model we described here, would be to apply to real retinal spike trains the different decoding mechanisms compared in the present study. A recent experiment by Stanley et al. (1999) has demonstrated the possibility of reconstructing spatio-temporal visual inputs from an ensemble of spike trains simultaneously recorded in the LGN. In this study however, the variable used to describe the neuron's responses was their time-varying firing rate. Our results suggest that other variables reflecting the precise temporal structure of the spike trains could constitute a powerful way of describing neural responses, and this theory could benefit from direct experimental investigation.

7 Conclusions

We have shown that codes based on the ganglion cells mean firing rate, as derived from the spike count or the mean interspike interval in a Poisson spike train, cannot account for the efficiency of information transmission between the retina and the brain. We have introduced instead a coding scheme based on the relative order in which the ganglion cells emit their first spike in response to a given visual stimulation, and shown that this rank order coding, although computationally very simple, can lead to a very good stimulus reconstruction, even over relatively short time periods, and even when firing latencies are subject to a 20% time jitter. Provided that we take into account the temporal structure of the spike train, and assign a greater impact to the spikes with the shortest latencies, it appears that the very first spikes generated in the retina can carry sufficient information for further cortical processing. In addition, these results

demonstrate that the idea of temporal coding does not only refer to using spike synchrony as a way of multiplexing information with multiple channels; spike ***a****synchrony* is a powerful alternative that should not be neglected.

Moreover, these results do not only apply to retinal coding. They explain how, given a certain architecture, the neurons can maximize their information transfer rate. Thus this scheme could be used for stimulus encoding in the input layer of any kind of artificial neural network, but could also account for information transfer between any 2 layers of such a network. Indeed, we used this code in simulations with SpikeNET (Delorme, et al. 1999), and it has proved powerful, especially when the next layers of processing also use a rank order coding scheme.

Finally, the proximity between the neural coding scheme presented here and wavelet-based compression, as well as MPEG or JPEG technologies, should be pointed out. Millions of years of natural selection have no doubt produced an image compression scheme in the optic nerve that is highly optimized; engineers could well find that the main features of modern image compression techniques were already discovered millions of years ago.

Acknowledgements

This work was supported by the CNRS and the Region Midi-Pyrénées. We wish to thank the anonymous reviewers for helpful comments on the manuscript.

References

Abbott, L.F., Varela, J.A., Sen, K., & Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science, 275*(5297), 220-224.

Alonso, J.M., Usrey, W.M., & Reid, R.C. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature, 383*(6603), 815-819.

Arnett, D., & Spraker, T.E. (1981). Cross-correlation analysis of the maintained discharge of rabbit retinal ganglion cells. *J Physiol (Lond), 317*, 29-47.

Baddeley, R., Abbott, L.F., Booth, M.C., Sengpiel, F., Freeman, T., Wakeman, E.A., & 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.

Bair, W. & Koch, C. (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput*, 8 1185-1202.

Bair, W. (1999). Spike timing in the mammalian visual system. *Curr Opin Neurobiol, 9*, 447-453.

Berry, M.J., Warland, D.K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proc Natl Acad Sci U S A, 94*(10), 5411-5416.

Berry, M.J., & Meister, M. (1998). Refractoriness and neural precision. *J Neurosci, 18*(6), 2200-2211.

Berry, M.J., Brivanlou, I.H., Jordan, T.A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature, 398*(6725), 334-338.

Brivanlou, I.H., Warland, D.K., & Meister, M.(1998) Mechanisms of concerted firing among retinal ganglion cells. *Neuron*, 20, 527-539.

Buser, P., & Imbert, M. (1992). *Vision* (R.H. Kay, Trans.). Cambridge, MA: MIT Press.

Celebrini, S., Thorpe, S.J., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake monkey. *Visual Neurosci*, *10*, 811-825.

Croner, L. J., & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Res, 35*(1), 7-24.

Crook, J.M., Lange-Malecki, B., Lee, B.B. & Valberg, A. (1988) Visual resolution of macaque retinal ganglion cells. *J. Physiol*, 396, 205-224.

Curcio, C.A., & Allen, K.A., (1990) Topography of ganglion cells in human retina. *J Comp Neurol*, 300, 5-25.

Dan, Y., Alonso, J. M., Usrey, W. M., & Reid, R. C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci, 1*(6), 501-507.

Delorme, A., Gautrais, J., Van Rullen, R. & Thorpe, S.J.(1999). SpikeNET : a simulator for modeling large networks of integrate and fire neurons. *NeuroComputing*, 24.

DeVries, S. H. (1999). Correlated firing in rabbit retinal ganglion cells. *J Neurophysiol, 81*(2), 908-920.

Eckhorn, R. (1994). Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features. *Prog Brain Res, 102*, 405-426.

Engel, A.K., Konig, P., Kreiter, A.K., Schillen, T.B., & Singer, W. (1992). Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci, 15*(6), 218-226.

Fabre-Thorpe, M., Richard, G., & Thorpe, S.J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9*(2), 303-308.

Field, D.J. (1994). What is the goal of sensory coding? *Neural Comput, 6*, 559-601.

Fischer, B. (1973). Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Res, 13*(11), 2113-2120.

Gautrais, J., & Thorpe, S.J. (1998). Rate Coding vs Temporal Order Coding: a theoretical approach. *Biosystems, 48*(1-3), 57-65.

Gerstner, W., & van Hemmen, J.L. (1992). Universality in neural networks: the importance of the 'mean firing rate'. *Biological Cybernetics, 67*(3), 195-205.

Heller, J., Hertz, J.A., Kjaer, T.W., & Richmond, B.J. (1995). Information flow and temporal coding in primate pattern vision. *J Comput Neurosci, 2*(3), 175-193.

Lankheet, M.J., Molenaar, J., & van de Grind, W.A. (1989). The spike generating mechanism of cat retinal ganglion cells. *Vision Res, 29*(5), 505-517.

Lipetz, L.E. (1971). The relation of physiological and psychological aspects of sensory intensity. In W. R. Lowenstein (Ed.), *Handbook of sensory physiology* (Vol. 1. Principles of receptor physiology, pp. 191-225). New York: Springer-Verlag.

Mainen, Z. F., & & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science, 268*, 1503-1506.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *Inst. Electrical Electronics Engrs. Trans. on Pattern Analysis and Machine Intelligence.*, *11*, 674-693.

Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature, 382*(6594), 807-810.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nat Neurosci, 3*(3), 251-258.

Mastronarde, D.N. (1989). Correlated firing of retinal ganglion cells. *Trends Neurosci, 12*(2), 75-80.

Meister, M., Lagnado, L. & Baylor, D.A. (1995) Concerted signaling by retinal ganglion cells., *Science*, 270 , 1207-1210.

Meister, M., & Berry, M.J. (1999). The neural code of the retina. *Neuron, 22*(3), 435-450.

Naka, K., & Sakai, H.M. (1991). The messages in optic nerve fibers and their interpretation. *Brain Res Brain Res Rev, 16*(2), 135-149.

Naka, K.I., & Rushton, W.A.H. (1966). S-potentials from luminosity units in the retina of fish (cyprinidae). *Journal of physiology, London, 185*, 587-599.

Nakamura, K. (1998) Neural processing in the subsecond time range in the temporal cortex., *Neural Comput*, 10 567-595.

Nirenberg, S. & Latham, P.E. (1998) Population coding in the retina., *Curr Opin Neurobiol*, 8, 488-493.

Nowak, L.G., Munk, M.H.J., Girard, P., & Bullier, J. (1995). Visual Latencies in Areas V1 and V2 of the Macaque Monkey. *Visual Neurosci, 12*(2), 371-384.

Oram, M.W., & Perrett, D.I. (1992). Time course of neural responses discriminating different views of the face and head. *J Neurophysiol, 68*(1), 70-84.

Panzeri, S., Treves, A., Schultz, S., & Rolls, E.T. (1999). On decoding the responses of a population of neurons

from short time windows. *Neural Comput, 11*(7), 1553-1577.

Peichl, L., & Wassle, H. (1979). Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *J Physiol (Lond), 291*, 117-141.

Perkel, D.H., & Bullock, T.H. (1968). Neural Coding. *Neurosciences Research Program Bulletin, 6* (3), 221-348.

Perrett, D.I., Rolls, E.T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47*, 329-342.

Reich, D.S., Victor, J.D., & 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., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *J Neurophysiol, 77*(5), 2836-2841.

Richmond, B.J., & 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.

Rieke, F., Warland, D., de Ruyter van Steveninck, R.R., & Bialek, W. (1997). *Spikes: exploring the neural code.* Cambridge, MA: MIT.

Rodieck, R.W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res., 5*, 583-601.

Rodieck, R.W. (1998). *The first steps in seeing.* Sunderland, MA.: Sinauer Associates.

Sestokas, A.K., Lehmkuhle, S., & Kratz, K.E.(1987). Visual latency of ganglion X- and Y-cells: a comparison with geniculate X- and Y-cells. *Vision Res, 27*(9), 1399-1408.

Sestokas, A.K., Lehmkuhle, S., & Kratz, K.E.(1991). Relationship between response latency and amplitude for ganglion and geniculate X- and Y-cells in the cat. *Int J Neurosci, 60*(1-2), 59-64.

Shadlen, M.N., & Newsome, W.T.(1995) Is there a signal in the noise?, *Curr Opin Neurobiol*, 5, 248-250.

Shadlen, M.N., & Newsome, W.T.(1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci, 18*(10), 3870-3896.

Singer, W. (1999). Time as coding space? *Curr Opin Neurobiol, 9*(2), 189-194.

Smirnakis, S.M., Berry, M.J., Warland, D.K., Bialek, W. & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale., *Nature*, 386 , 69-73.

Softky, W.R. (1995)Simple codes versus efficient codes., *Curr Opin Neurobiol*, 5, 239-247.

Stanley, G. B., Li, F. F., & Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J Neurosci, 19*(18), 8036-8042.

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.

Thorpe, S.J. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In R. Eckmiller, G. Hartman, & G. Hauske (Eds.), *Parallel processing in neural systems* (pp. 91-94). North-Holland: Elsevier.

Thorpe, S.J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Thorpe, S.J. & Gautrais, J. (1997) Rapid visual processing using spike asynchrony. In M.C. Mozer, M. Jordan and T. Petsche (Eds.), *Advances in Neural Information Processing Systems, Vol. 9*, MIT Press, Cambridge, pp. 901-907.

Thorpe, S.J. & Gautrais, J. (1998) 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.

VanRullen, R., Gautrais J., Delorme A., & Thorpe, S.J. (1998). Face processing using one spike per neuron. *Biosystems, 48*(1-3), 229-239.

VanRullen, R., & Thorpe, S. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, submitted*.

Vardi, N., & Smith, R.G. (1996). The AII amacrine network: coupling can increase correlated activity. *Vision Res, 36*(23), 3743-3757.

Victor, J.D., & Purpura, K.P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol, 76*(2), 1310-1326.

Warland, D.K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol, 78*(5), 2336-2350.

2.2 Complements

L'article précédent demontre que seuls les codages utilisant la structure temporelle précise des spike trains peuvent encoder et transmettre l'information de façon suffisamment rapide pour permettre un traitement ultérieur en seulement 150 ms. Le décodage de l'ordre de décharge des cellules ganglionnaires donne une bien meilleure (en termes d'information mutuelle comme d'erreur quadratique moyenne) reconstruction de l'image d'entrée qu'un décodage basé sur la fréquence moyenne de décharge, que celle-ci soit déterminée par un simple comptage des spikes reçus, ou par une estimation de "l'inter-spike interval" moyen.

Cet avantage peut bien sûr s'expliquer, en partie, par le fait que les codages en fréquence étudiés ici sont soumis à un processus stochastique (loi de Poisson) de génération des spike trains. Ce "bruit" implique que sur une fenêtre temporelle courte, le nombre de spikes générés par un neurone n'est probablement pas le reflet exact de son niveau d'activité. Ce n'est que sur des temps d'intégration plus longs, de l'ordre de plusieurs centaines de millisecondes à quelques secondes, que ce type de codage pourra transmettre l'information de façon fiable. Par contre, le codage par l'ordre d'activation semble résistant à une quantité de bruit non négligeable (bruit gaussien sur la latence de décharge, déviation standard = 20% de la latence moyenne). Ce n'est donc pas simplement la présence de bruit pour les codes en fréquence, ou la quasi-absence de bruit pour le codage par l'ordre d'activation, qui peut expliquer la différence de performance observée.

On pourrait également penser que cette différence provient de la "localité" du codage en fréquence (chaque neurone encode une valeur analogique, correspondant à son niveau d'activation, indépendamment des neurones voisins), alors que le codage par l'ordre d'activation est plus "global", distribué sur l'ensemble de la population des cellules ganglionnaires (le "rang" d'un neurone présuppose évidemment que ce neurone soit considéré comme faisant partie d'une plus large population). Selon cette hypothèse, la performance d'un codage par l'ordre d'activation devrait chuter lorsque l'on considère une population de neurones plus restreinte. Dans l'article précédent, l'image était encodée par environ 236 000 neurones. Qu'advient-il si l'on s'intéresse à une image beaucoup plus petite, correspondant par exemple au champ récepteur d'un neurone de V1?

Afin de rejeter cette hypothèse, les simulations présentées dans l'article précédent ont été reproduites sur une série d'images de taille 22x15 pixels, encodées par 860 neurones, à 3 échelles spatiales différentes. L'apprentissage du niveau de contraste correspondant à un ordre de décharge donné (i.e. une "look-up table" similaire à celle présentée dans la figure 3 de l'article précédent) a été réalisé sur ces images réduites, puis les reconstructions produites par les différents codes ont été comparées, comme dans l'article précédent. Notons que dans ce cas, l'information mutuelle et l'erreur quadratique pour ces reconstructions ont été calculées à partir de versions binaires des images originales et de

leurs reconstructions. En effet, le calcul de l'information mutuelle requiert la connaissance de la distribution de probabilité des niveaux de luminance dans chaque image. Or, avec seulement 330 pixels par image, et 256 niveaux de gris possibles, en moyenne 1 ou 2 pixels seulement seraient associés à un niveau de gris donné. Dans ce cas, l'information mutuelle serait optimale simplement si la reconstruction contenait un grand nombre de niveaux de gris, quelle que soit leur répartition spatiale. Une image constituée d'un bruit blanc conduirait à une information mutuelle maximale! Le fait de travailler sur des versions binaires des images et de leurs reconstructions évite ce problème: une centaine de pixels en moyenne sont associés à chaque niveau de luminance (noir/blanc), ce qui permet d'estimer la corrélation existant entre les pixels d'une image et leurs homologues dans sa reconstruction.

Les résultats obtenus sont présentés dans la figure 21.

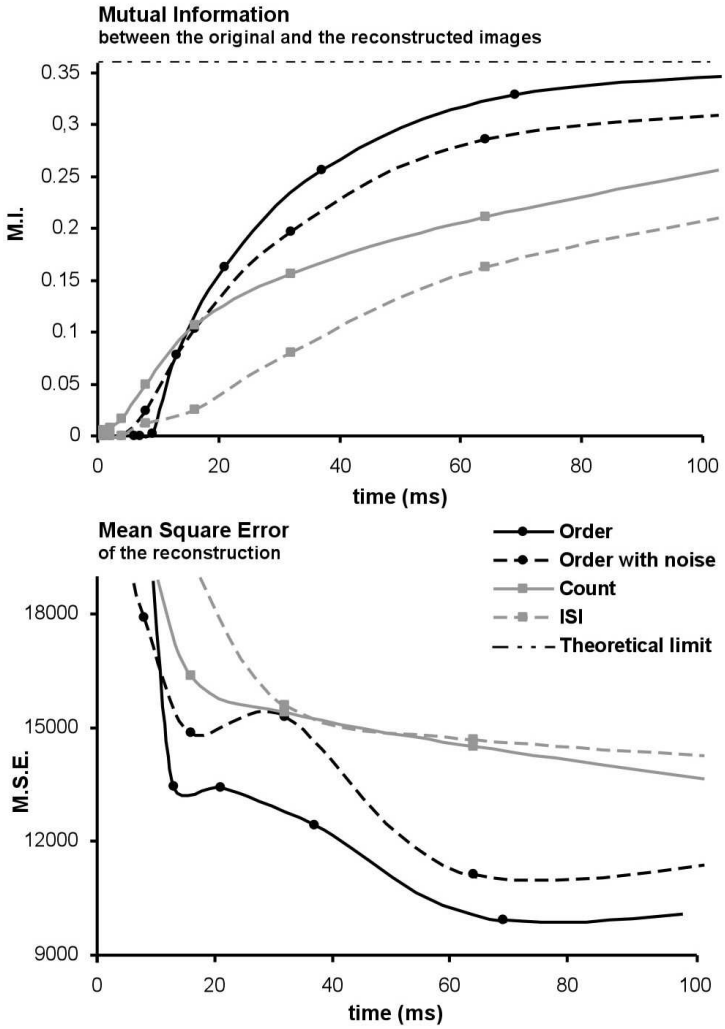


Figure 21. Information mutuelle et erreur quadratique moyenne pour des reconstructions d'images de taille 22x15, encodées par seulement 860 neurones, à 3 échelles spatiales. Pour chaque codage, la performance est indiquée en fonction du temps.

Le codage par l'ordre d'activation, ainsi que la version "bruitée" de ce code, permettent ici encore de reconstruire l'image d'entrée de façon optimale. Un neurone de V1 qui interprèterait l'information en utilisant ce codage pourrait donc estimer au mieux le stimulus présent dans son champ récepteur. L'avantage mis en évidence pour ce type de code par rapport aux codages en fréquence ne semble donc pas dépendre de la taille de l'image qui est encodée. De plus, le nombre d'échelles spatiales mises en jeu est bien plus faible pour ces images réduites que pour celles utilisées dans l'article précédent, ce qui suggère que ce nombre n'est pas non plus un déterminant de la performance du codage par rang. Il en résulte que la puissance de ce type de codage réside certainement dans sa capacité à transmettre une grande quantité d'information avec un nombre restreint de potentiels d'action (un seul spike par neurone, et très peu de neurones activés), et à tirer parti de la précision temporelle de la latence de décharge des cellules ganglionnaires, mise en évidence par de nombreuses études (Berry et al., 1997; Reich et al., 1997; Sestokas et al., 1991).

Ainsi, l'information portée par les toutes premières décharges générées dans la rétine (de l'ordre de 1%) serait suffisante pour permettre aux étapes de traitement suivantes, au niveau du cortex strié puis extrastrié, d'extraire l'information nécessaire à la réalisation d'une tâche de haut niveau, comme la détection de la catégorie cible "animal", ou "véhicule". Dans le modèle présenté ici comme dans la rétine des primates, du fait de la structure des images naturelles ainsi que de l'organisation physiologique de la rétine, l'information correspondant aux basses fréquences spatiales est transmise plus tôt vers le cortex (Beaudot et Herault, 1994; Beaudot, Oliva et Herault, 1995). Or, d'après des études psychophysiques et computationnelles, ces basses fréquences spatiales semblent suffisantes pour permettre une première catégorisation des scènes naturelles (Schyns et Oliva, 1994; Oliva, Herault et Guerin-Dugue, 1997). L'image reconstruite grâce à ces tout premiers spikes (voir article 3) ne laisse, il est vrai, aucun doute sur la quantité d'information transmise.

2.3 Le décodage de l'ordre

L'article précédent suggère que l'information visuelle au niveau de la rétine pourrait être simplement encodée par l'ordre de décharge sur la population de cellules ganglionnaires. Bien que ce type de codage semble attrayant par sa puissance et son efficacité, il soulève le problème du décodage d'une telle information. Lorsque les entrées d'un neurone sont des valeurs analogiques, reflétant le taux de décharge des neurones afférents, le niveau d'activation d'un neurone donné peut être obtenu simplement par le produit scalaire de son vecteur d'entrée avec son vecteur de poids (typiquement, dans des neurones formels de type McCulloch et Pitts, voir figure 5). Mais si l'on suppose que

l'information est encodée par l'ordre de décharge des neurones afférents, et que chacun de ces neurones ne génére que (au plus) un spike, alors le vecteur d'entrée est un vecteur binaire, et les derniers neurones afférents ayant déchargé (portant l'information la moins importante, selon l'hypothèse du codage par l'ordre) se voient attribuer par le neurone post-synaptique autant d'importance que les premiers afférents (portant l'information la plus intéressante, selon l'hypothèse du codage par l'ordre). Prenons le cas de 3 neurones afférents A, B et C. Quel que soit l'ordre de décharge de ces 3 neurones, le produit scalaire du vecteur de poids avec le vecteur (1, 1, 1) donnera toujours le même résultat, et il est donc impossible dans ces conditions de décoder l'ordre d'entrée.

Il existe cependant un moyen simple pour le neurone post-synaptique de décoder l'ordre d'activation des afférents (Thorpe et Gautrais, 1997, 1998; voir également la thèse de Jacques Gautrais, 1997). Il suffit en fait d'invoquer une désensibilisation ou modulation progressive du neurone post-synaptique, à mesure que les spikes afférents lui parviennent. Le neurone est "au départ" (c'est-à-dire, avant qu'une stimulation ne soit présentée au système) maximalement réceptif. Chaque spike lui parvenant fera décroître sa sensibilité. Si l'on considère maintenant l'ordre d'activation des afférents (A,B,C), le niveau d'activation du neurone post-synaptique sera le résultat du produit scalaire du vecteur de poids avec le vecteur (mod(1), mod(2), mod(3)), où mod est la fonction (strictement décroissante) de désensibilisation du neurone. Ainsi, l'ordre d'activation des afférents peut être décodé, car le produit scalaire sera maximal lorsque cet ordre correspondra à l'ordre des poids synaptiques. Notons que n'importe quelle fonction strictement décroissante peut théoriquement être utilisée pour cette désensibilisation. Dans la pratique cependant, et en particulier dans des conditions où le rapport signal sur bruit n'est pas optimal, le choix de cette fonction pourra déterminer le pouvoir discriminant du neurone post-synaptique. Dans toutes les simulations qui suivront dans cette thèse, la fonction choisie pour le décodage de l'ordre sera une suite géométrique de raison r strictement inférieure à 1: la sensibilité est 1 lorsque le premier spike est reçu, r pour le deuxième, r^2 pour le troisième, etc□ et tend donc vers 0 lorsque le nombre d'afférents augmente.

Pour finir, signalons que pour des populations de neurones réels corticaux, cette désensibilisation pourrait être implémentée par un mécanisme de "shunting inhibition" rapide, comme il l'a été récemment démontré par Delorme et Thorpe (en préparation).

2.4 Synchronie et oscillations: un épiphénomène?

Comme mentionné à la section 4 du chapitre I, la fréquence de décharge et le "spike timing" (qui comprend également le codage par l'ordre d'activation) ne sont pas les seuls candidats au titre de codage de l'information neuronale. De nombreux auteurs attribuent un

rôle potentiel à la synchronie des décharges entre 2 ou plusieurs neurones, ou aux oscillations de l'activité neuronale sur des populations à plus grande échelle.

En effet, de nombreuses études expérimentales ont révélé la présence de corrélations entre décharges neuronales, avec bien souvent une précision temporelle de l'ordre d'une milliseconde. Chez le chat anesthésié, Gray et Singer (1989) ont démontré la présence de corrélations entre les décharges des neurones d'une colonne corticale de l'aire 17 (i.e. l'équivalent de V1) et les oscillations électriques locales. L'amplitude de ces oscillations à 40 Hz dépend de la correspondance entre l'orientation du stimulus et l'orientation préférée du groupe de neurones enregistré. Le même type de synchronisation avec un décalage de phase nul (en moyenne) a été également observé entre différentes colonnes corticales (Gray et al, 1989), ou différentes aires corticales (aires 17 et 18) du chat anesthésié, et a montré une dépendance à des propriétés diverses du stimulus comme sa position, son orientation, sa direction de mouvement ou sa vitesse (Eckhorn et al, 1988). Les assemblées de neurones déchargeant de manière synchrone semblent dépendre de la cohérence du stimulus: 2 assemblées synchrones (non corrélées l'une à l'autre) correspondant à 2 stimuli indépendants, peuvent se regrouper en une seule assemblée synchrone si les 2 stimuli sont réunis (Engel et al, 1991). Plus récemment, des résultats similaires ont été obtenus chez le singe éveillé: oscillations à 70 Hz sans décalage de phase (en moyenne) entre V1 et V2 (Frien et al, 1994); décharges synchrones entre neurones voisins de MT, plus prononcées avec un seul stimulus cohérent qu'avec 2 stimuli séparés (Kreiter et Singer, 1996; cette augmentation de la synchronie ne semble pas être dûe à une augmentation de l'amplitude de la réponse, ou à la présence de 2 stimuli au lieu d'un); corrélations entre neurones voisins de IT (Gochin et al, 1991); synchronie de décharges dans le cortex moteur, reflétant aussi bien des évènements externes (stimulus, mouvements) qu'internes (attente d'un signal; Riehle et al, 1997).

Même lorsque la présence de décharges synchrones ou d'oscillations n'est pas démontrée, la très haute précision temporelle des spike trains ("precisely replicating patterns" dans le LGN ou V1, Lestienne 1996, Lestienne et Tuckwell 1998; "precise firing sequences" dans les cortex prefrontal ou premoteur, Prut et al, 1998; encodage de variations de mouvement à la milliseconde près dans MT, Buracas et al, 1998) est souvent interprétée comme un argument en faveur des codages utilisant la synchronie.

Sur la base de ces observations, de nombreux auteurs ont proposé, de façon explicite, que la synchronie des décharges neuronales serait un facteur clé du traitement et du codage de l'information neuronale (Milner, 1974; Von der Malsburg, 1981; Abeles, 1982; Engel et al, 1992; Singer, 1993; Singer et Gray, 1995; Varela, 1995). Cette théorie a engendré une importante quantité d'études expérimentales ou de modélisation, qui ont permis de caractériser plus précisément, bien que non exhaustivement, les propriétés de la

synchronisation et des oscillations de l'activité corticale. Nous décrirons ici ces propriétés, avant de proposer une interprétation alternative de ces mêmes résultats.

2.4.1 Synchronie

La synchronie des décharges neuronales est supposée pouvoir être induite par une présentation synchrone des différents éléments de la stimulation visuelle. C'est en utilisant ce paradigme que différentes études psychophysiques ont pu proposer que l'activité neuronale synchrone serait corrélée à la perception visuelle consciente. Fahle et al (1993), puis Usher et Donnelly (1998) ont montré que des éléments présentés de façon synchrone étaient perçus comme cohérents par les sujets humains, alors que ces mêmes éléments sont perçus comme indépendants si l'on introduit un décalage de phase dans la stimulation visuelle. De même, Alais et al (1998) ont montré que des éléments dont le contraste est modulé de façon synchrone sont regroupés par le système visuel humain. Cependant, ces derniers résultats peuvent également être interprétés par un mécanisme de regroupement par "mouvement commun" (Gegenfurtner, symposium "Neural Binding in Space and Time", Leipzig, Mars 2000). De plus, ces conclusions vont à l'encontre de 2 études plus anciennes, où une asynchronie significative dans la présentation des éléments de la stimulation visuelle était restée sans effet sur la performance de sujets humains (Fahle et Koch, 1995; Kiper, Gegenfurtner et Movshon, 1996). Enfin, notons que cette approche comporte des dangers théoriques non négligeables: si la synchronie de stimulation visuelle induit effectivement une synchronie d'activité neuronale, corréler la performance des sujets ou leur perception de la scène visuelle avec cette activité neuronale synchrone peut s'avérer erroné. Dans le monde réel, en effet, la stimulation visuelle est continue (i.e. ni synchrone, ni asynchrone), et les conclusions obtenues expérimentalement ne sont donc pas forcément vérifiées en dehors du laboratoire. L'étude en MEG de Tononi et al (1998), dont les sujets étaient exposés à une rivalité binoculaire grâce à une stimulation de chaque œil à une fréquence différente, et où les canaux MEG montraient une modulation d'activité importante à la fréquence de l'œil "perçu", est un exemple où la conclusion d'une corrélation entre la perception consciente et l'activité neuronale synchrone est invalide. L'étude de Rager et Singer (1998), démontrant qu'un taux de rafraîchissement de la stimulation visuelle entre 2 et 50 Hz se retrouve dans l'activité du cortex visuel du chat à une fréquence identique, en est un autre.

L'enregistrement de l'activité neuronale dans le cortex visuel est une source plus fiable de données concernant la synchronie de décharge. Outre les études citées plus haut (e.g. Engel et al, 1991; Kreiter et Singer, 1996) démontrant une altération de la synchronie selon la "cohérence" de la stimulation visuelle, certaines expériences mettent en évidence un lien entre perception et synchronie. En utilisant un paradigme de rivalité binoculaire chez le chat, Fries et al (1997) ont montré que les neurones de V1 répondant à l'œil "perçu"

déchargent avec une synchronie accrue, alors que la synchronie décroît pour les neurones répondant à l'œil supprimé. Par contre, aucune modulation liée à la perception n'est observée pour la fréquence de décharge. Les études du groupe de Gilles Laurent sur le système olfactif de l'abeille démontrent qu'une perturbation de la synchronisation des décharges (par antagonie de l'inhibition GABAergique), sans effet sur la fréquence de décharge, empêche l'animal de percevoir des odeurs complexes (Stopfer et al, 1997). Cet effet s'explique par une dégradation de la sélectivité des neurones efférents (MacLeod et al, 1998). Il semblerait donc que la synchronie des décharges soit nécessaire pour l'encodage de certaines informations perceptuelles.

La synchronie de décharge a également été reliée à des processus attentionnels: elle augmente durant la période précédent l'apparition du stimulus dans MT et MST (De Oliveira et al, 1997) ou dans les aires motrices (Riehle et al, 1997); le degré de synchronie dans le cortex sensori-moteur est corrélé à la demande attentionnelle de la tâche (Steinmetz et al, 2000). Ceci a mené certains auteurs à proposer des modèles de l'attention visuelle reposant sur l'activité synchrone des populations neuronales (Niebur et al, 1993; Niebur et Koch, 1994; Borisyuk et al, 1998; Wu et Guo, 1999). La synchronie serait également liée au contexte comportemental dans le cortex frontal (Vaadia et al, 1995). Enfin, différentes simulations suggèrent que la synchronisation des décharges pourrait jouer un rôle dans la réactivation de patterns mis en mémoire (Wang et al, 1990; Ritz et al, 1994; Sompolinsky et Tsodyks, 1994).

Cependant, si ces études, ainsi que de nombreuses autres qu'il ne serait pas possible d'énumérer ici, mettent en évidence un rôle de la synchronie dans divers processus cognitifs, il advient de mentionner plusieurs cas où cette hypothèse a été mise en échec. Schwarz et Boltz (1991) rapportent que la corrélation des décharges neuronales dans le cortex visuel n'est pas influencée par des changements de stimulation visuelle. Young, Tanaka et Yamane (1992) ne purent reproduire chez le singe anesthésié les résultats observés chez le chat: ils ne trouvent que très peu de réponses synchrones entre 30 et 70 Hz, et une absence totale de sélectivité au stimulus. De même, Tovee et Rolls (1992) n'observent aucune synchronie dans le cortex inférotemporal du singe éveillé, et en concluent que l'implémentation du "liage par synchronisation des décharges" ne serait pas une solution généralement utilisée par le système visuel. De Oliveira et al (1997) reportent que la synchronisation des décharges dans MT et MST ne reflète ni la performance, ni les propriétés du stimulus. Enfin, lors d'une tâche de discrimination de texture, Lamme et Spekreijse (1998) ne trouvent aucune relation systématique entre la synchronie de décharge de paires de neurones dans V1 et l'organisation perceptuelle de la scène (cette étude est cependant critiquée pour des raisons méthodologiques par Singer, 1999).

2.4.2 Oscillations

L'observation d'oscillations rythmiques de l'activité électrique neuronale dans la bande Gamma (typiquement de 30 à 80 Hz) est en général associée à la théorie du codage de l'information neuronale par synchronisation des décharges. En réalité, cette association est plus souvent faite par les chercheurs qui mettent en évidence ces oscillations (Tallon-Baudry et Bertrand, 1999), argumentant que de telles oscillations nécessitent certainement l'activité synchrone d'un très grand nombre de neurones (e.g. Tononi et al, 1998), que par les électrophysiologistes (e.g. Singer 1999), qui reportent parfois des décharges neuronales précisément corrélées en l'absence d'oscillations (Frien et al, 1994). Le fait que 2 ou plusieurs neurones déchargent de manière corrélée n'implique pas, en effet, que cette corrélation soit présente sur des populations bien plus larges. Cependant, des oscillations du potentiel électrique local sont fréquemment reportées en même temps que les décharges synchrones (Gray et al, 1989; Engel et al, 1991a,b,c; Frien et al, 1994), en particulier lorsque la synchronie est observée sur des distances relativement importantes (i.e. > 2mm; König, Engel et Singer, 1995). De plus, ces oscillations sont proposées comme une base physiologique potentielle pour la génération des décharges synchrones, d'après des observations expérimentales (Nowak et al, 1997; Volgushev et al, 1998) et théoriques (König et Schillen, 1991; Schillen et König, 1991). Il est donc raisonnable d'assimiler "oscillations" et "synchronisation", comme deux manifestations d'un même phénomène, à des échelles différentes.

De même que pour la synchronie des décharges, les oscillations électriques dans la bande de fréquence Gamma ont été mises en relation avec la perception visuelle consciente. Diverses études en EEG montrent que ces oscillations sont augmentées en réponse à des stimuli visuels cohérents vs incohérents (Lutzenberger et al, 1995; Müller 1996; Tallon-Baudry et al, 1996, 1997; Rodriguez et al, 1999). Ces oscillations peuvent être synchronisées sans décalage de phase entre des aires corticales distantes (i.e. entre les aires visuelles et pariétales et entre les aires pariétales et motrices lors de la performance d'une tâche visuo-motrice; Roelfsema et al, 1997). Elles sont augmentées en réponse à des stimuli auditifs sur lesquels est portée l'attention (Tiitinen et al, 1993), et avant la présentation du stimulus dans une tâche de mémoire visuelle à court terme (Tallon-Baudry et al, 1998), suggérant un rôle potentiel lors de l'accès aux stimuli mémorisés. Enfin, ces oscillations pourraient être liées à l'apprentissage ou l'association: la cohérence de ces oscillations entre les aires visuelles et sensori-motrices est renforcée durant le conditionnement d'une association entre stimuli visuels et tactiles, et disparaît après extinction (Miltner et al, 1999). Ces données semblent suggérer que les oscillations de l'activité électrique neuronale dans la bande Gamma, tout comme la synchronisation des décharges, pourraient refléter divers processus cognitifs de haut niveau. L'expérience

visuelle consciente pourrait ainsi reposer sur une telle dynamique concertée des populations neuronales (Varela, 1999a,b).

Cependant, d'autres études suggèrent que ces oscillations corticales pourraient en fait être induites par l'activité intrinsèque d'autres structures comme le LGN (Ghose et Freeman, 1997), ou par un mécanisme de codage par latences précises de décharge (Parodi et al, 1996), et n'être finalement qu'un simple artefact.

2.4.3 Codage d'information par synchronie

Le fait que 2 ou plusieurs neurones déchargent de manière corrélée, en relation avec des évènements externes (stimulation) ou internes (attention, contexte comportemental) n'implique pas que le système visuel utilise effectivement ces corrélations pour encoder ou traiter l'information. Cette synchronisation des décharges pourrait n'être qu'un épiphénomène d'un tout autre mécanisme de traitement neuronal (cependant, la fréquence de décharge est ici exclue, car des modulations de synchronie peuvent être observées sans changement de fréquence de décharge). Afin de valider l'hypothèse d'un codage neuronal par synchronisation, il est nécessaire de montrer que des corrélations temporelles fines entre les décharges de 2 ou plusieurs neurones peuvent effectivement être décodées et interprétées par d'autres neurones ou ensembles neuronaux.

Prenons l'exemple de la rétine. L'existence de corrélations entre les décharges des cellules ganglionnaires, qui surviennent bien plus fréquemment qu'on l'attendrait compte-tenu de la fréquence de décharge (Meister et al, 1995), a été démontrée chez le chat (Mastronarde, 1989) et la salamandre (Brivanlou, Warland et Meister, 1998). Ces corrélations à 3 échelles temporelles distinctes s'expliquent par l'existence de synapses chimiques en provenance des photorécepteurs (corrélations de l'ordre de 40-100 ms), de jonctions électriques en provenance des cellules amacrines (10-50 ms), et de jonctions électriques réciproques (environ 1 ms). Seul ce dernier type de corrélation peut s'expliquer indépendamment des entrées communes en provenance de la rétine externe. Ce type de corrélation (i.e. < 10 ms) porte une quantité d'information qui n'est pas présente dans le taux de décharge des cellules ganglionnaires (Meister et al, 1995). De même, les cellules du corps genouillé latéral (LGN) déchargent de manière hautement synchrone (Alonso, Usrey et Reid, 1996), et cette corrélation temporelle semble porter une quantité d'information supplémentaire non négligeable (20 % en moyenne; Dan et al, 1998). De plus, les connexions en retour du cortex visuel sur le LGN semblent pouvoir sélectivement augmenter la synchronie des cellules thalamiques, en relation avec certaines propriétés du stimulus (Silito et al, 1994; Contreras et al, 1996).

Comment cette corrélation temporelle précise observée à un niveau donné (rétine, LGN) peut-elle être décodée au niveau suivant (LGN, cortex)? Une proposition spécifique

est que des potentiels d'action afférents synchrones augmenteraient la probabilité de réponse du neurone post-synaptique (Bernander, Koch et Usher, 1994; Murthy et Fetz, 1994). Un tel mécanisme ("paired-spike enhancement") a été observé entre la rétine et le LGN (Usrey, Reppas et Reid, 1998), entre le LGN et le cortex (Alonso, Usrey et Reid, 1996; Dan et al, 1998), ainsi qu'entre neurones corticaux (cellules simples et complexes du cortex visuel du chat, Alonso et Martinez, 1998; cellules hippocampales, Stevens et Zador, 1998). Typiquement, si A et B sont 2 neurones afférents, et C un neurone efférent, les auteurs observent que la probabilité d'une décharge post-synaptique (de C) est augmentée lorsque 2 spikes de A et B arrivent dans la même fenêtre temporelle (par exemple 1 ms), relativement à la somme des probabilités de décharge post-synaptique après un spike de A, et après un spike de B, lorsque ces spikes ne sont pas synchrones (Alonso, Usrey et Reid, 1996). Cet effet peut être attribué à divers mécanismes: propriétés basiques des neurones integrate-and-fire (2 EPSPs simultanés augmentent la probabilité d'atteindre le seuil), processus non-linéaires dans le soma ou les dendrites (e.g. calcium spikes; Fregnac 1999), ou la présence éventuelle d'autres afférences synchrones non enregistrées par l'expérimentateur. En fait, la seule conclusion évidente de ces résultats est que l'intégration temporelle des spikes afférents par un neurone efférent est un processus non-linéaire. Comme nous le verrons par la suite, la synchronie n'est pas le seul code à pouvoir tirer profit de cette constatation.

Une autre proposition est que le décalage de phase entre 2 neurones ou groupes neuronaux synchronisés pourrait porter l'information de manière explicite. Ceci vient de l'observation que ce décalage de phase dépend souvent de façon systématique des propriétés du stimulus visuel: position, orientation, fréquence spatiale, direction de mouvement, vitesse (Eckhorn et al, 1988). La figure 22, tirée de König et al (1995), illustre un cas typique de corrélation entre 2 groupes neuronaux (enregistrés par Multi-Unit Activity, et comprenant en général de 1 à 5 neurones voisins, i.e. ayant des propriétés similaires). Les cross-correlogrammes obtenus pour 3 conditions de stimulation différentes (Figure 22D,E,F) sont en général considérés comme un argument en faveur d'un codage par synchronie. Le décalage de phase relevé dans cette étude dépend de façon linéaire de l'orientation du stimulus (voir figure 22), et de la différence entre les orientations préférées des cellules enregistrées. Des effets identiques sont observés si l'on teste l'influence de la direction de mouvement, ou de la fréquence spatiale des stimuli.

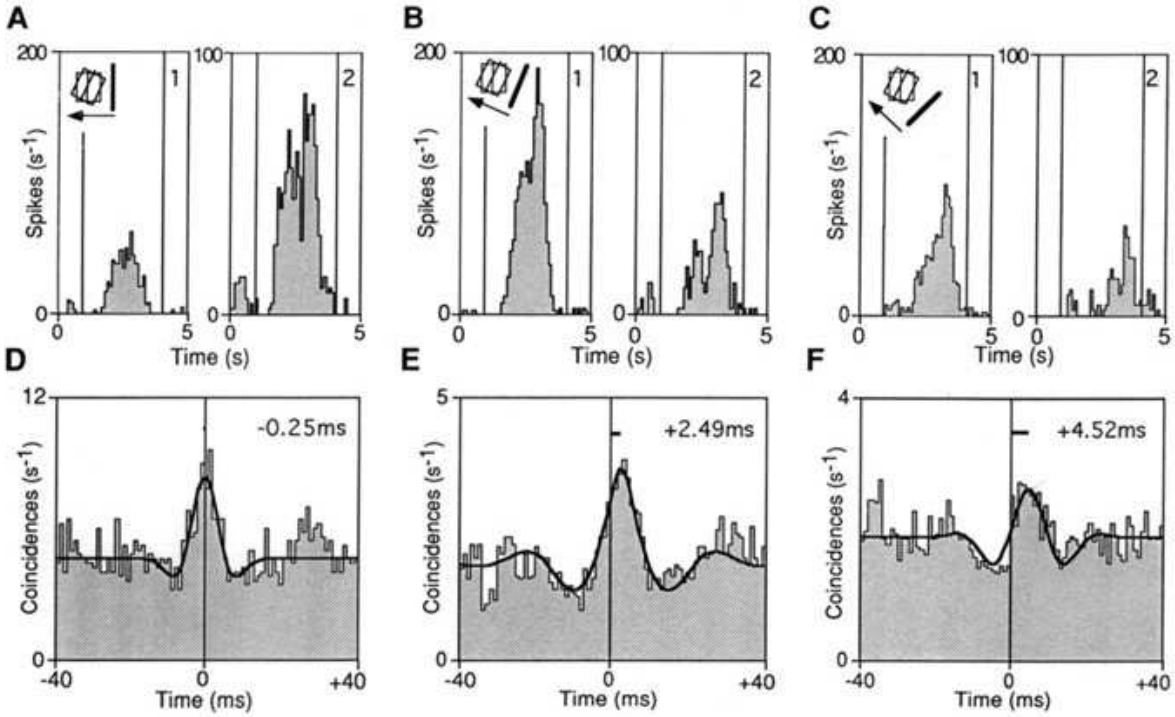


Figure 22. Exemple d'observation de décharges synchrones entre 2 clusters de neurones (Multi-Unit Activity) de l'aire 17 (représentation du champ visuel central) du chat anesthésié (König et al, 1995). Les champs récepteurs correspondant à chaque électrode (1 et 2) sont représentés par des rectangles, et l'orientation préférée par une barre à l'intérieur de ces champs. A,B,C: réponse des 2 clusters à une barre traversant les champs récepteurs, pour différentes orientations. D,E,F: cross-correlogrammes entre les réponses des clusters 1 et 2, pour les 3 conditions décrites en A,B,C. Le décalage de phase (ms) est indiqué sur chaque courbe. Un décalage de phase positif correspond à une précédence temporelle des spikes émis par le cluster 1. Les neurones d'un cluster ont tendance à décharger plus tôt que ceux de l'autre si le stimulus présenté est plus proche de leur stimulus préféré.

2.4.4 La synchronie seule ne peut encoder l'information spatiale dans la rétine

De façon pratique, même si l'on accepte que l'hypothèse du "paired-spike enhancement" peut rendre compte du décodage des décharges synchrones, la manière dont la synchronie pourrait encoder l'information spatiale dans la rétine (ou dans le LGN) n'est pas évidente. 2 décharges synchrones en provenance de 2 neurones différents peuvent être simplement décodées. Un exemple d'interprétation simpliste serait: "les 2 neurones afférents ont le même niveau d'activité". Pour savoir quel est ce niveau, il faudra cependant nécessairement faire appel à un autre code (par exemple la fréquence de décharge). Certains auteurs considèrent le "nouveau" spike train composé des spikes précisément synchrones en provenance de 2 neurones afférents (Ghose et al, 1994) comme un canal d'information supplémentaire, qui pourrait par exemple encoder le champ récepteur

composé de l'intersection des champs récepteurs des 2 neurones en question (Meister et al, 1995; Dan et al, 1998). Cette hypothèse suppose qu'un certaine quantité d'information serait manquante dans les spike trains "classiques" (i.e. non synchrones). Ceci est très certainement vérifié à différents niveaux du système visuel, comme dans la rétine, où il serait illusoire de croire que l'échantillonnage des cellules ganglionnaires est uniforme et exhaustif. Cependant, dans le modèle de rétine utilisé pour l'article 3, l'information spatiale est entièrement encodée dans les spike trains des cellules ganglionnaires simulées; c'est pourquoi ce type de codage par synchronie n'a pu être testé dans cet article. De plus, ceci illustre bien la fragilité de cette hypothèse: le nouveau spike train, résultant des spikes synchronisés, doit lui aussi être décodé, et donc faire appel à un autre type de codage (e.g. fréquence de décharge). Si, comme pour les codes en fréquence, plusieurs spikes consécutifs doivent être considérés dans ce spike train additionnel afin d'encoder l'information de façon fiable, alors le codage par synchronie, tout comme les codes en fréquence, ne permet pas de transmettre l'information assez rapidement pour pouvoir rendre compte de la vitesse de traitement des systèmes visuels biologiques.

Pour résumer, la synchronie ne peut encoder l'information spatiale dans la rétine ou le LGN, sans faire appel à un autre mode de codage. La corrélation des décharges est d'ailleurs en général proposée comme un canal d'information additionnel, mais pas comme un codage propre.

En réalité, la synchronie semble idéale pour encoder des relations d'équivalence entre neurones ou populations, mais peu adaptée à des relations ordinales (codage par rang) ou quantitatives (codages en fréquences, latences précises). La synchronie serait donc plus propice à des problèmes de haut niveau comme le liage des différentes propriétés des objets de la scène visuelle (Treisman, 1996; von der Malsburg, 1995), qui ne requièrent pas de transmettre des informations sur la structure spatiale (rétinotopique) des stimuli visuels.

Ainsi, le problème de la transmission rapide d'informations entre différents niveaux de traitement du système visuel ne peut être résolu par la seule théorie de la synchronisation des décharges, du moins en ce qui concerne la transmission de l'information visuelle de la rétine au cortex. Il est nécessaire de faire appel à un autre type de codage, et les codes qui utilisent la structure temporelle fine de l'information rétinienne (asynchronie, codage par rang) se sont révélés les plus efficaces à cet égard (article 3). Serait-il possible qu'un tel type de codage asynchrone puisse rendre compte des nombreuses observations expérimentales de corrélation temporelle des potentiels d'action, et d'oscillations dans la bande Gamma?

2.4.5 Une succession de vagues de potentiels d'action explique les observations expérimentales de synchronie et oscillations

L'hypothèse d'un codage de l'information neuronale par l'asynchronie des décharges (Thorpe 1990; Thorpe et Gautrais, 1997; voir également article 3) suggère que la réponse de la rétine à une stimulation visuelle est une vague de potentiels d'action à l'intérieur de laquelle les délais ou l'ordre des décharges entre les différents neurones sont le vecteur de l'information. Dans sa version la plus simple, seule la première vague de potentiels d'action, générée en réponse à un stimulation très brève, est utilisée et propagée par les différents niveaux de traitement du système visuel, pour parvenir à construire en seulement 150 ms une représentation de haut niveau de l'entrée visuelle. Dans des conditions naturelles de stimulation, l'environnement visuel est continu dans le temps, et il est raisonnable d'imaginer que dans ces conditions, une succession de telles vagues de potentiels d'action puisse encoder et mettre à jour en permanence le monde visuel et sa représentation dans notre cerveau. Nous ne nous préoccupons pas ici de savoir quels mécanismes sous-tendraient une telle organisation spatio-temporelle des spike trains de la rétine ou du cortex visuel. Nous avons déjà mentionné (chapitre I section 4) un rôle potentiel pour les micro-saccades visuelles ou les oscillations sous-liminaires de l'activité neuronale. Notons simplement que ce problème n'est en rien plus difficile que celui de la synchronisation des décharges sur de larges populations neuronales avec une précision de l'ordre de la milliseconde, alors que ce dernier n'a donné lieu jusqu'à présent qu'à bien peu de controverse.

Le mécanisme de codage par l'asynchronie est simplement cascadable (VanRullen et al, 1998), du fait qu'à n'importe quel niveau de la hiérarchie visuelle, les neurones les plus activés d'une population donnée tendront à décharger les premiers. La réponse d'une population à une vague de potentiels d'action en provenance de la couche afférente est donc une vague de potentiels d'action aux propriétés similaires, mais encodant une information différente (par exemple, en réponse à une vague de spikes du LGN, à l'intérieur de laquelle les premiers spikes correspondent aux régions les plus contrastées, les colonnes d'orientation du cortex visuel primaire génèreront une vague de potentiels d'action à l'intérieur de laquelle les orientations les plus saillantes seront encodées par les premiers spikes). A chaque niveau du système visuel, on observe donc une succession de vagues d'information qui permet d'encoder et de mettre à jour constamment une représentation du monde visuel. D'un point de vue global, l'activité électrique observée au niveau de la population semble osciller de façon rythmique (figure 23). Si l'intervalle moyen entre 2 vagues successives est choisi de manière appropriée (par exemple 25 ms), la fréquence de ces oscillations se trouve dans la bande Gamma (40 Hz).

A un niveau global, les oscillations observées expérimentalement pourraient donc être le reflet d'un mécanisme de codage par l'asynchronie des décharges

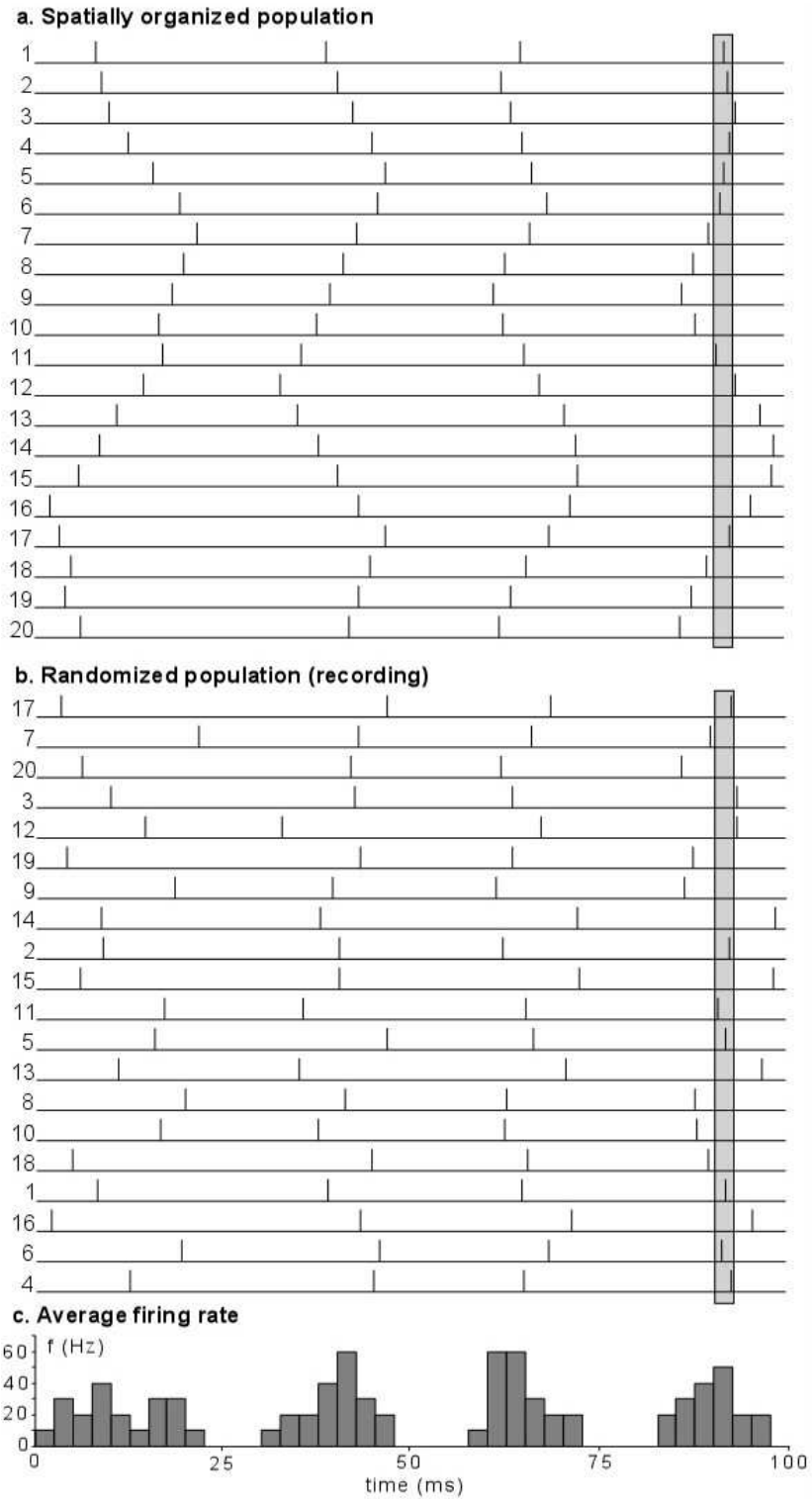


Figure 23. a. L'hypothèse d'un codage d'information neuronal résidant dans l'asynchronie de décharges sur une population de neurones suggère que, dans des conditions normales de stimulation, une population (ici 20 neurones) spatialement organisée répond par une succession de vagues de potentiels d'action. A l'intérieur de chaque vague, l'information est portée par l'ordre ou la latence relative des potentiels d'action. **b.** Lorsque cette même population est organisée différemment, aucune structure temporelle systématique n'apparaît. Ceci correspond à ce qu'observerait un expérimentateur utilisant un multi-electrode array. **c.** Le potentiel électrique local, ici obtenu en regroupant les potentiels d'action sur une fenêtre temporelle de 2.5 ms, correspond à ce qu'observerait un expérimentateur enregistrant l'activité multi-unitaire (MUA). Compte-tenu des paramètres temporels utilisés (en moyenne 25 ms entre 2 vagues de potentiels d'action successives), l'activité oscille à une fréquence de 40 Hz

Cette hypothèse rend compte également de la corrélation temporelle forte entre les décharges de 2 neurones d'une même population. Ces neurones participant aux mêmes vagues de spikes, ils émettront en général des spikes temporellement proches. Compte-tenu de la diversité des stimuli visuels dans le monde réel, et de l'organisation des populations neuronales, qui très certainement reflète cette diversité, un neurone donné n'a que peu de chances de se trouver systématiquement plus activé que son voisin. C'est pourquoi l'hypothèse d'un codage par l'asynchronie peut donner lieu également à des cross-corrélogrammes prononcés, et relativement symétriques (figure 24).

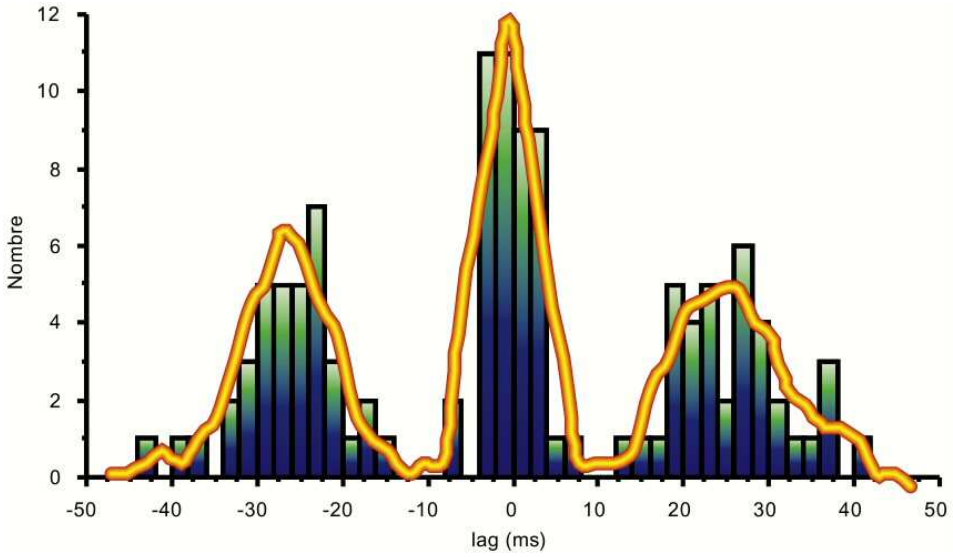


Figure 24. Cross-corrélogramme entre 2 neurones dont l'activité participe à des vagues de potentiels d'action asynchrones. La réponse de chaque neurone est obtenue en regroupant les enregistrements des neurones pairs (respectivement impairs) de la figure 23. La fenêtre temporelle de l'histogramme est de 2.5 ms. La courbe de tendance est une moyenne mobile sur 4 périodes. Ce cross-corrélogramme montre une forte corrélation, avec en moyenne un décalage de phase nul.

Ainsi, bien que les nombreuses observations expérimentales de synchronie et d'oscillations dans le système visuel soient classiquement interprétées comme la preuve d'un rôle de la synchronie dans le traitement de l'information neuronale, ces mêmes observations peuvent également s'expliquer par un codage reposant sur l'asynchronie des décharges à l'intérieur d'une succession de vagues de potentiels d'action.

2.4.6 Codage par la synchronie, ou codage par l'asynchronie?

En réalité, la dichotomie entre synchronie et asynchronie est principalement une question de précision temporelle. Classiquement, les spikes de 2 neurones sont considérés comme synchrones si ils sont séparés par moins de 10 ms. Or, c'est précisément l'intervalle de temps à l'intérieur duquel est généralement réalisé le codage par l'asynchronie (si 150 ms sont suffisantes au système visuel pour parvenir à une représentation de haut niveau de l'entrée visuelle, alors 10 ms doivent suffir pour transmettre l'information entre 2 couches successives; Gautrais et Thorpe, 1998).

Le phénomène de "paired-spike enhancement" suggère que 2 spikes touchant un neurone à environ une milliseconde d'intervalle augmentent la probabilité de décharge post-synaptique, relativement à une situation où ces spikes seraient séparés de plus d'une milliseconde. Pour mettre en évidence ce phénomène, Alonso, Usrey et Reid (1996) enregistrent simultanément l'activité de 2 neurones présynaptiques (A et B) et d'un neurone post-synaptique C. Ils regroupent les spikes de A et B en 3 catégories: les spikes de A et B synchronisés à moins d'une milliseconde (A&B), les spikes de A qui ne sont pas synchronisés avec ceux de B, toujours avec une fenêtre de synchronisation de l'ordre d'une milliseconde (A*), et les spikes de B qui ne sont pas synchronisés avec ceux de A (B*). La probabilité d'occurrence d'un spike de C après un "spike synchrone" (A&B) est environ 50% supérieure à la somme des probabilités d'un spike de C après un spike isolé de A (A*) et d'un spike de C après un spike isolé de B (B*). L'interprétation des auteurs est que les spikes "isolés" sont traités linéairement, alors qu'une non-linéarité augmente la probabilité de décharge post-synaptique dans le cas "synchrone". Cependant, la contraposée ne devrait pas être écartée: les spikes synchrones pourraient être traités linéairement, et les spikes isolés pourraient subir une non-linéarité qui diminuerait la probabilité de décharge post-synaptique. En effet, le terme "isolé" dépend fortement de la fenêtre de synchronisation choisie. Même lorsque les spikes de A et B ne touchent pas le neurone C pendant la même milliseconde, ils adviennent typiquement avec quelques millisecondes d'intervalle (Singer et Gray, 1995; voir par exemple figure 22). Pour une large partie des spikes "isolés" de A, un spike de B a donc été émis quelques milisecondes plus tôt, et de même, pour de nombreux spikes "isolés" de B, un spike de A aura été émis quelques millisecondes plus tôt. Il est possible que le premier spike reçu par C puisse diminuer l'efficacité du second, ce qui expliquerait les résultats obtenus. Selon cette hypothèse, il ne s'agirait donc pas d'un phénomène de "paired-spike enhancement", mais plutôt d'une "second-spike depression", qui correspond précisément au mécanisme proposé par Thorpe et Gautrais (1998) ou VanRullen et al (1998) pour décoder l'ordre d'arrivée des potentiels d'action.

De même, l'idée d'un encodage des propriétés du stimulus par différence de phase entre 2 ou plusieurs neurones est en réalité plus favorable à un codage par l'asynchronie qu'à un codage par la synchronie. Selon de nombreux enregistrements (e.g. Eckhorn et al, 1988; König et al, 1995), le neurone dont la préférence (par exemple à l'orientation, mais aussi à la fréquence spatiale, la direction de mouvement, la vitesse, etc...) correspond le mieux au stimulus présenté déchargera systématiquement avant ses voisins. L'observation de la figure 22 montre que chaque neurone décharge plus ou moins tôt (la date de décharge est ici définie relativement aux décharges d'un autre neurone de la population, et non en relation avec un événement extérieur comme la présentation du stimulus) en fonction de la "distance" entre le stimulus présenté et son stimulus préféré. Les dates relatives de décharge des neurones considérés (i.e. l'asynchronie) plus que leur degré de corrélation (i.e. la synchronie), permettent d'encoder les propriétés du stimulus. Ceci est précisément ce que suggère l'hypothèse d'un codage par l'asynchronie des décharges à l'intérieur d'une vague de potentiels d'action.

Pour résumer, le faisceau d'observations expérimentales classiquement interprété comme la preuve d'un rôle de la synchronie dans l'encodage et le traitement de l'information neuronale, supporte également l'hypothèse d'un codage par l'asynchronie des décharges sur une population neuronale. Les conclusions de Singer (1999) dans une revue récente sur la synchronie illustrent bien cette potentielle double interprétation:

- *"Les réseaux de neurones peuvent manipuler l'information temporelle avec une haute précision"*
- *"Les réseaux de neurones transmettent l'activité synchronisée plus efficacement que l'activité asynchrone"* (ici, synchronisée signifie sur des fenêtres temporelles de l'ordre d'une dizaine de millisecondes, à l'intérieur desquelles l'asynchronie peut bien évidemment être préservée, et jouer un rôle perceptuel).
- *"Grâce à des interactions internes, les réseaux de neurones peuvent générer des relations temporelles précises sur des décharges distribuées, indépendamment de la structure temporelle des stimuli"* (l'encodage par l'asynchronie constitue un exemple de telles relations temporelles distribuées, générées indépendamment de la structure temporelle des stimuli)
- *"Les réseaux de neurones interprètent la synchronisation des décharges comme la signature d'un lien, dans le contexte des processus cognitifs comme de l'apprentissage"* (ici encore, le terme synchronie fait référence à

une précision de l'ordre d'une dizaine de millisecondes, ce qui n'exclut pas un rôle de l'asynchronie à l'intérieur de cette fenêtre temporelle).

2.5 Synthèse

L'article 3 a démontré que, dans le contexte de l'encodage et de la transmission de l'information visuelle de la rétine au cortex, où la rapidité est un facteur particulièrement critique, seuls les codes qui prennent en compte la structure temporelle précise (i.e. asynchronie) de la première vague de potentiels d'action générée en réponse à la stimulation visuelle peuvent rendre compte de l'extraordinaire vitesse de traitement du système visuel, mise en évidence au chapitre II. Par comparaison, les codages en fréquence semblent bien trop lents ou imprécis. De plus, nous avons vu à la section précédente que l'hypothèse d'un codage par synchronie des décharges ne peut satisfaire complètement ces contraintes de rapidité, même si de nombreuses données expérimentales semblent converger en faveur de cette hypothèse. En fait, la plupart de ces observations expérimentales se sont avérées compatibles avec la théorie d'un codage par l'asynchronie des décharges neuronales à l'intérieur d'une vague de potentiels d'action se propageant dans le système visuel. Dans ces conditions, il est parfaitement possible, et même profitable, d'encoder efficacement l'information visuelle avec seulement un spike par neurone.

3 Transmettre l'information en feed-forward

Les contraintes temporelles qui pèsent sur les systèmes visuels biologiques, mises en évidence au chapitre II, n'impliquent pas seulement que l'information transmise entre 2 neurones réels doive être encodée avec un nombre minimal de potentiels d'action. Ces contraintes vont en fait bien au-delà, et suggèrent que, compte-tenu de l'organisation hiérarchique du système visuel, même si un seul spike est transmis par chaque neurone, le flux d'information doit encore se faire principalement vers l'avant. Ceci limite une fois encore les modèles de la vision inspirés des systèmes biologiques, et semble contraindre les architectures et mécanismes implémentables dans de tels modèles.

3.1 Que peut-on faire de façon “feed-forward”?

La plupart des modèles classiques attribuent un rôle primordial aux connexions en retour et aux boucles computationnelles, qui constituent, dans certains cas, le "meilleur" moyen (ou le plus évident) d'implémenter certaines fonctions. En particulier, le

mécanisme de l'intégration des contours, et celui de l'attention spatiale (e.g. Kirkland et Gerstein, 1999), font en général appel à des connexions récurrentes. Doit-on simplement considérer que ces fonctions particulières ne prennent aucune part active dans le type de traitement visuel rapide que le système visuel humain peut réaliser en 150 ms, ou existe-t-il d'autres alternatives pour implémenter ces mécanismes, tout en respectant les contraintes mentionnées? En d'autres termes, peut-on réaliser ces fonctions de manière feed-forward?

Nous verrons dans cette section que l'asynchronie sur laquelle réside le codage par rang permet d'implémenter de façon "feed-forward" des mécanismes qui, "traditionnellement", sont supposés reposer sur un traitement en boucle ou "feed-back". Dans un premier temps, nous définirons précisément les notions de feed-forward et feed-back, et dissocierons la hiérarchie anatomique (figée), d'une hiérarchie fonctionnelle plus dynamique. Ainsi, l'intégration des contours pourra être obtenue par des interactions latérales au niveau du cortex visuel primaire, sans impliquer aucune itération. Nous montrerons ensuite que la relation explicite qui existe entre la latence des décharges neuronales et la saillance d'un stimulus permet d'implémenter un mécanisme d'attention spatiale, de manière entièrement feed-forward.

3.2 Article 4 (en préparation) : VanRullen, Delorme & Thorpe. Feed-forward contour integration based on asynchronous spike propagation. *NeuroComputing*

Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation.

Rufin VanRullen*, Arnaud Delorme & Simon J. Thorpe.

Centre de Recherche Cerveau & Cognition, CNRS-UPS, UMR 5549,
Faculté de Médecine de Rangueil, 133 Route de Narbonne, 31062 TOULOUSE Cedex. France.

Most current models of visual contour integration involve iterative lateral or feedback interactions among neurons in V1 and V2. However, some forms of visual processing are too fast for such time-consuming loops. We propose a model avoiding iterative computation by using the fact that real neurons in the retina or LGN fire asynchronously, with the most activated firing first. Thus early firing V1 neurons can influence processing of their neighbors which are still integrating information from LGN. By limiting the number of spikes to one per neuron, we show that contour integration can be obtained in a purely feed-forward way.

Introduction

Visual contour integration, a modulation of V1 neuron responses by contextual influences outside their receptive field, responsible for the selective enhancement of smooth aligned contours (Fig. 1A, 1B), is thought to be mediated by lateral interactions among V1 neurons (Field et al, 1993; Kapadia et al. 1995; Gilbert et al. 1996) or feed-back inputs from V2 (Von der Heydt et al. 1984). Current models of contour integration involve iterative, recurrent lateral or feed-back projections (e.g. Shashua and Ullman, 1988; Hummel and Zucker, 1983; Yen and Finkel, 1997; Li, 1998; Heitger and Von der Heydt, 1993; Grossberg and Mingolla, 1985; Gove et al. 1995). The underlying idea behind these models is that visual inputs are computed and sent all at once, in a single step, to the primary visual cortex neurons (depending on the model, visual inputs can remain active during the following steps). These V1 neurons then calculate their activity levels, and send it, all at once, to their neighbours. The last step is then repeated until a satisfactory solution or an equilibrium state is reached (Fig. 1C).
These time-consuming mechanisms are in contradiction with psychophysical and electrophysiological studies of rapid visual categorization in man (Thorpe et al. 1996; VanRullen and Thorpe, 2000) and monkey (Fabre-Thorpe et al, 1998; Vogels, 1999). Behavioral responses in tasks where subjects have to detect a target (e.g. animal) in a briefly flashed (20 ms) natural image can be as short as 180 ms in monkey and 250 ms in man, bringing evidence that at least some forms of visual processing are so fast that it must be performed in a single feed-forward pass through the visual system, with probably no more than one spike per neuron between two successive processing stages. This leaves remarkably little time for recurrent loops or feedback to occur. One possibility is that this sort of ultra-rapid scene categorization could be performed without involving much contour integration. However, here we show that in fact, contour integration can occur very rapidly if one makes use of the fact that real neurons fire asynchronously, with the most activated neurons reaching their threshold first. This is a simple and well-known property of integrate-and-fire neurons.

* Corresponding author.

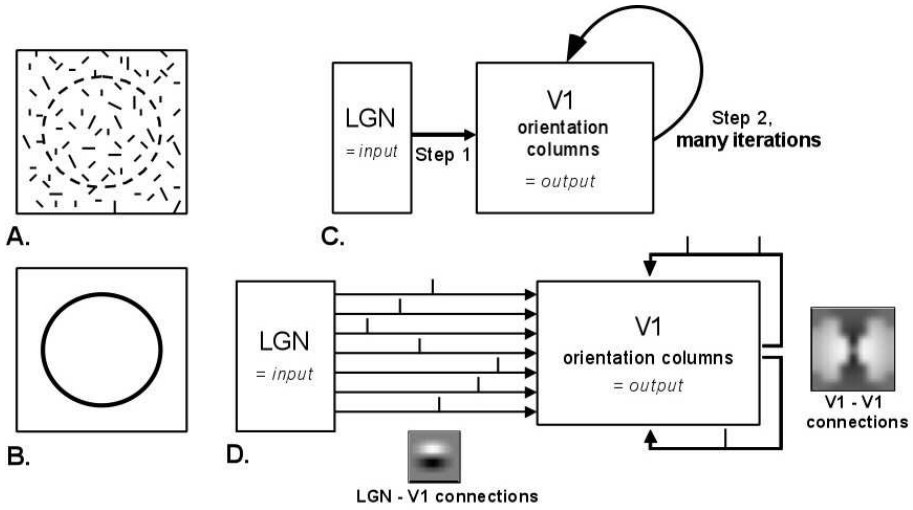


Fig. 1. Contour integration enhances the smooth contour in the retinal image A to yield to the activity pattern B in orientation columns. Classical models suppose recurrent interactions among V1 neurons (C). We propose a model based on asynchronous spike propagation where lateral interactions occur in a "feed-forward" way (D). Patterns of afferent and lateral connectivity for a V1 neuron selective to a horizontal orientation are shown.

Results

We design a two-stage model where the response of the first, input layer (which can be compared to the retina or the LGN) is a wave of asynchronously generated spikes. Neurons in the second layer, modelling the primary visual cortex V1, are selective to a particular orientation in their receptive field. In addition, they are interacting locally with one another through a set of lateral connections, similar to those proposed by Li (1998). The very first neurons to fire in V1 can start influencing the processing of their neighbors when they are still integrating the incoming information from LGN (Fig. 1D). This is compatible with electrophysiological recordings in cats (Volgushev et al. 1995) showing that the delay between the rise of PSPs in a V1 cell and the emission of its first spikes leaves enough time for lateral feed-forward interactions to occur. Neighboring neurons forming a smooth contour will receive excitatory lateral input, whereas neurons in clearly orthogonal directions will be inhibited. This lateral wave of activity modulation in V1 can spread over large distances before the end of the orientation extraction process, i.e. before the last spike has arrived from LGN. To make sure that our implementation did not permit iterations or loops in any way, we limited the number of spikes per neuron to zero or one. Orientation selectivity can still be achieved by making V1 neurons sensitive to the order in which they receive afferent spikes (Thorpe and Gautrais, 1997, 1998). Under these conditions, we were able to show for a variety of examples that contour integration indeed occured in a purely feed-forward way (Fig. 2). More specifically, we compared the activations of V1 orientation colums with and without these lateral feed-forward interactions. Activation in the retinotopic orientation maps decreased at the locations where the visual input had little or no structure, and was recruited at the locations where the contour was smooth. This was verified with artificial input images of broken lines, circles, segments, embedded in a field of random oriented bars, as well as with natural input pictures (Fig. 3).

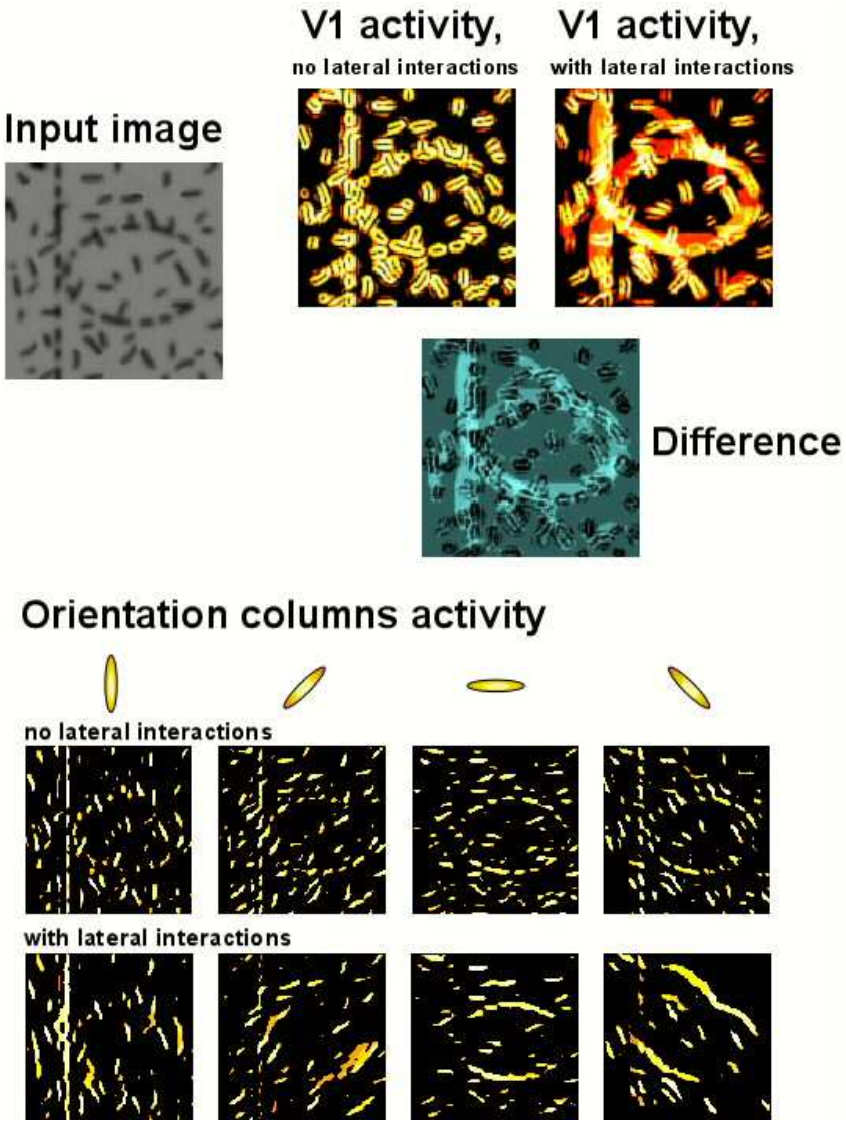


Fig 2. An example of contour integration generated by our model. The input image contains a smooth but broken contour (vertical line, oval shape) embedded in a field of random oriented bars. V1 global activities (sum of activities for 8 different orientation maps) with and without lateral interactions are shown (top). The difference image (middle) shows the regions where activity is decreased (dark spots) and the locations to where activity has moved (light spots). Activities for 4 (out of 8) orientation maps (bottom) demonstrate the selectivity of the orientation columns. These results were obtained with no more than one spike per neuron.

Discussion

The main feature of these results is that they were obtained with no iteration or computational loop, although using the same pattern of connectivity, hence the same functionality as was proposed by Li (1998) for a clearly iterative model. The major difference between these two approaches stems from the concept of information transmission and processing that we used. By simply taking into account the properties of real neurons, i.e. the fact that integrate-and-fire neurons will tend to fire asynchronously, with the most activated cells firing first, we were able to implement a visual contour integration process without any iterations.

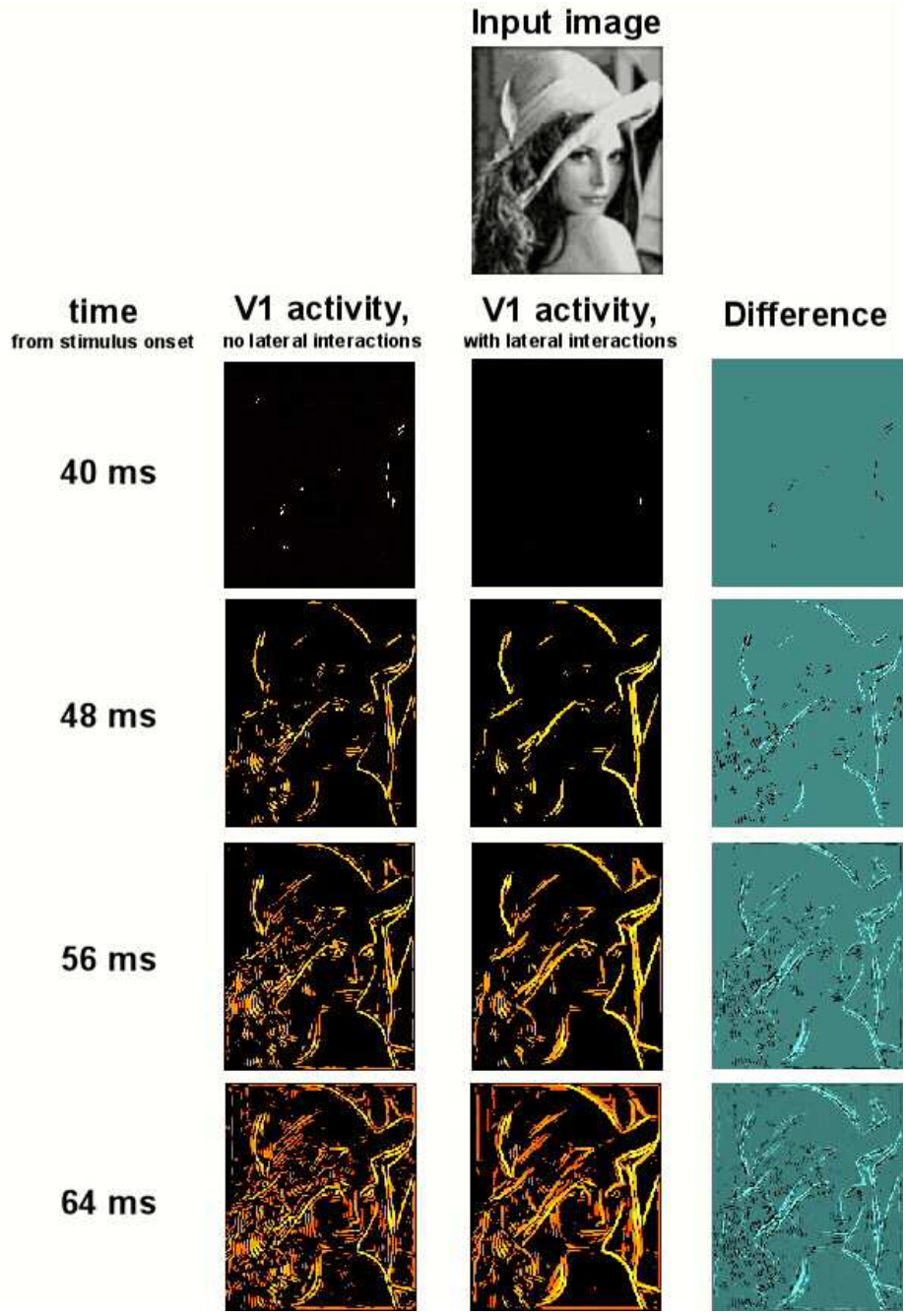


Fig. 3. Example of contour integration obtained with a natural input image. The temporal course of contour integration is simulated by assuming a uniform distribution of spikes in the retina and a 40 ms delay between the retina and V1. Lateral interactions enhance activity at smooth contours (outline of the face, shoulder, or hat), and decrease activity at locations without strong contour coherence (feathers on the hat). The effects of contour integration can be observed in the first milliseconds of processing.

We believe that many other computational mechanisms traditionally thought to involve feed-back, or recurrent lateral interactions, could be implemented in such an asynchronous feed-forward information flow framework. For example, we have already proposed (VanRullen and Thorpe, 1999) a model of feed-forward spatial attention based on the temporal precedence of the attended information under conditions where spikes are propagated asynchronously.

One could argue, on the other hand, that since our model V1 neurons are locally mutually interconnected, there is indeed some kind of recurrent processing in our model. This can not be true however, under conditions where a neuron is only allowed to generate one spike. Indeed, when a neuron fires and influences its neighbors, driving some of them above threshold, the resulting lateral interactions will have virtually no "backwards" influence on this neuron. This raises the question of how to characterize a computational loop, which is well defined in a sequential processing framework, but lacks a clear definition in the context of parallel asynchronous networks. A neural circuit can have *anatomical feedback* even under conditions where *functionally* it operates in a feed-forward mode (Treves et al, 1996). The critical issue is whether or not the circuit is able to compute the desired function even when each neuron only fires at most one spike. We suggest that the notions of iteration, loop and feedback should not depend only on the respective positions of the involved neurons in the visual cortical hierarchy, but rather on the relative time at which they respond to a visual stimulus. This is supported by recent electrophysiological studies showing that activity can occur simultaneously accross multiple hierarchically "successive" visual areas, rather than in a strictly sequential way (Bullier and Nowak, 1995).

References

1. **Bullier, J., & Nowak, L. G.** (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol, 5*(4), 497-503.
2. **Fabre-Thorpe, M., Richard, G., & Thorpe, S. J.** (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9*(2), 303-8.
3. **Field, D. J., Hayes, A., & Hess, R. F.** (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Res, 33*(2), 173-93.
4. **Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G.** (1996). Spatial integration and cortical dynamics. *Proc Natl Acad Sci U S A, 93*(2), 615-22.
5. **Gove, A., Grossberg, S., & Mingolla, E.** (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Vis Neurosci, 12*(6), 1027-52.
6. **Grossberg, S., & Mingolla, E.** (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept Psychophys, 38*(2), 141-71.
7. **Heitger, F., & von der Heydt, R.** (1993). A computational model of neural contour processing: Figure-ground segregation and illusory contours. *Proc. of the 4th Intl. Conf. on Computer Vision*, 32-40.
8. **Hummel, R., & Zucker, S. W.** (1983). On the foundation of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 267-287.
9. **Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G.** (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*(4), 843-56.
10. **Li, Z.** (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput, 10*(4), 903-40.
11. **Shashua, A., & Ullman, S.** (1988). Structural saliency. *Proceedings of the International Conference on Computer Vision*, 482-488.
12. **Thorpe, S. J., Fize, D., & Marlot, C.** (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.
13. **Thorpe, S. J., & Gautrais, J.** (1997). Rapid visual processing using spike asynchrony. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems* (MIT Press ed., Vol. 9, pp. 901-907). Cambridge: MIT Press.
14. **Thorpe, S. J., & Gautrais, J.** (1998). Rank order coding: a new coding scheme for rapid processing in neural networks. In J. Bower (Ed.), *Computational Neuroscience : Trends in Research.* New York: Plenum Press.
15. **Treves, A., Rolls, E. T., & Tovee, M. J.** (1996). On the time required for recurrent processing in the brain. In V. Torre & F. Conti (Eds.), *Neurobiology: Proceedings of the International School of Biophysics, XXIII course, May 1995* (pp. 371-382). New York: Plenum Press.
16. **VanRullen, R., & Thorpe, S.** (1999). Spatial attention in asynchronous neural networks. *NeuroComputing, 26-27*, 911-918.
17. **VanRullen, R., & Thorpe, S.** (2000). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, submitted*.
18. **Vogels, R.** (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur J Neurosci, 11*(4), 1223-38.
19. **Volgushev, M., Vidyasagar, T. R., & Pei, X.** (1995). Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. *Vis Neurosci, 12*(4), 621-8.
20. **von der Heydt, R., Peterhans, E., & Baumgartner, G.** (1984). Illusory contours and cortical neuron responses. *Science, 224*(4654), 1260-2.
21. **Yen, S.-C., & Finkel, L. H.** (1997). Salient contour extraction by temporal binding in a cortically-based network. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems* (MIT Press ed., Vol. 9,). Cambridge: MIT Press.

3.3 Attention: précédence temporelle pour les régions d'intérêt

L'asynchronie des décharges neuronales ne permet pas seulement de repenser en termes dynamiques la connectivité neuronale, afin d'utiliser en mode "feed-forward" des boucles "anatomiques". Cette asynchronie est également le reflet d'une propriété fondamentale des neurones "integrate-and-fire": les neurones les plus activés atteignent leur seuil plus tôt, et déchargent ainsi les premiers (voir figure 9). Dans ces conditions, la "saillance" d'un stimulus est encodée à chaque niveau par les dates relatives de décharge, à l'intérieur d'une population neuronale. Ceci permet d'implémenter un mécanisme d'attention spatiale agissant comme un "biais temporel" pour la région d'intérêt. Les neurones dans cette région déchargeant plus tôt que le reste de la population, ces décharges sont interprétées aux étapes supérieures comme correspondant aux aspects les plus saillants du stimulus. Cette théorie est décrite plus en détail dans l'article suivant. Divers résultats expérimentaux en faveur de cette hypothèse seront ensuite examinés.

3.3.1 L'hypothèse: Article 5. VanRullen & Thorpe, 1999. **Spatial attention in asynchronous neural networks.** ***NeuroComputing***

ELSEVIER

Neurocomputing 26–27 (1999) 911–918

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Spatial attention in asynchronous neural networks

Rufin VanRullen*, Simon J. Thorpe

Centre de Recherche Cerveau et Cognition, Faculté de Médecine de Rangueil, 133, Route de Narbonne, 31062, Toulouse Cedex, France

Accepted 18 December 1998

Abstract

We propose a simple mechanism for spatial visual attention that involves selectively lowering the thresholds of neurons with receptive fields in the attended region. Whereas such a mechanism is of no use in classical artificial neural networks, where all activities for each position in the visual field are computed simultaneously, it can be of great interest in an asynchronous neural network, where the relative order of firing in a population of neurons constitutes the code. Since neurons in the attended region will tend to reach threshold and fire earlier, they will tend to dominate later stages of processing. We illustrate this hypothesis with simulations based on SpikeNET.

Keywords: Attention; Rank-order coding; Spiking neurons; Threshold lowering

1. Introduction

There are numerous different theories and models to explain spatial attention mechanisms in the visual field. But none of them takes account of the asynchrony inherent in real neural networks such as the visual system. Yet it is well known that neurons in a given population fire at different rates, but also at different latencies.

We have already proposed [3,4] that these differences in firing latencies, e.g. the relative order of firings in a population, could be used as a code for transmitting the information from one processing stage to the next. The most strongly activated

* Corresponding author. E-mail: rufin@cerco.ups-tlse.fr.

0925-2312/99/$ – see front matter
PII: S0925-2312(98)00136-2

neurons will tend to fire first, with the result that early processing in later stages will be dominated by the shortest latency inputs. Neurons in later processing stages can be made to be sensitive to the order in which their inputs fire by invoking a mechanism which progressively decreases the post-synaptic neuron's sensitivity as more and more inputs arrive [4]. We have demonstrated that it is perfectly conceivable to produce multi-layered feed-forward architectures based on such principles that are capable of performing complex visual processing tasks that include the localization of faces in natural images [6].

Under such conditions, we can make the hypothesis that spatial attention involves selectively lowering the effective threshold of neurons with receptive fields in the attended region. This means that neurons at this location will tend to fire earlier, giving a temporal precedence to the attended stimuli, and allowing them to dominate processing at later stages.

2. Why the visual system needs spatial attention

The need for spatial attention, as pointed out by Mozer and Sitton [2] stems from the resource limitations of real visual systems.

Consider a neural network performing object recognition. With one neuron selective to a particular object for each spatial location, such a system does not need any attentional mechanism to perform accurately. For example, we have proposed [6] a model for face detection that does not use attention. The problem arises in real networks such as the human visual system, where the amount of resources, namely the number of neurons, is limited.

Clearly, the human visual system cannot afford one "object detector" for each object and each retinotopic location. It is well known that neurons in the visual system have increasing receptive fields sizes, and many neurons in the latest stages, such as the inferotemporal cortex, have receptive fields covering the entire visual field. They can respond to an object independently of its spatial location. Such a system needs far fewer neurons. But how can it deal with more than one object at the same time? With no attentional mechanism, if you present to that network an image containing many objects to detect, it is impossible to decide which one it will choose. Furthermore, there is a risk that features from different objects will be mixed, causing problems for accurate identification. This is an aspect of the well-known "binding problem" [5]

Suppose now we lower the thresholds for neurons with receptive fields in one part of the visual field. Provided that the neurons have dynamical properties such as those observed in real neurons (integrate-and-fire, spiking neurons ...), information concerning the object in this region will tend to propagate more quickly through the network, and so will activate the appropriate output detector before information about the other objects has arrived. The network response will thus correspond to what was in the image at the location of attention.

3. Simulations

Here we follow the argumentation of Mozer and Sitton [2] and translate it in the context of asynchronous neural networks. We have shown [6] that this kind of networks are suitable for complex visual tasks like face detection, provided that the amount of neurons used is not a limiting factor. Here we illustrate the problem of resource limitations in the context of object recognition. Finally, we demonstrate that our hypothesis allows the model to overcome these problems.

More precisely, we have built simple object recognition models to explore the possibility that such a threshold-decrease mechanism could underly the effects of spatial attention. These models were implemented with SpikeNET, our large-scale asynchronous neural network simulation software [1].

Units in SpikeNET are simple integrate-and-fire neurons, which basically generate no more than one spike for each image presented to the network. Furthermore, they can be made to be selective to a particular *order* of their afferent spikes, by a mechanism which decreases the neurons sensitivity as more and more inputs arrive, irrespective of their weight. Therefore, the neurons will be best activated when the order of their inputs matches the order of their synaptic weights [4].

Using this particular neural network scheme, we built 2 different models of object recognition, and compared their performance on a very simple categorization task: 9 views of 9 different objects (1 view per object, Fig. 1) were learned, and had to be recognized at any of 4 different locations, corresponding to the left or right and upper or lower hemifields.

The two models shared the same 6-level hierarchical organization. Units in the first level, corresponding to the retina, responded to a positive or negative local contrast (ON- or OFF-Center cells). At that level, the analog intensity of the input contrast was transformed in a firing latency. Units in the second layer were selective for edges of a particular orientation (8 different orientations separated by 45°), like the simple cells of the primary visual cortex V1, whereas the third layer combined these informations in 4 different maps, in which neurons were selective for an orientation irrespective of its polarity. At the next processing stage, basic features like terminations, T- or L-junctions, at 8 possible orientations, were extracted, and then combined using "complex" cells in the 5th layer. Finally, neurons in the last layer were trained to respond specifically to different objects.

The 2 models differ only by the presence or absence of an attentional mechanism.

3.1. *Limited resources model, without attention*

In the first model, as in the visual system, we wanted the sizes of the neurons receptive fields to increase from one processing stage to the next, so that the object detectors receptive fields, as observed in IT, would cover the entire visual field. In this case there is only one neuron per object category in the final layer. That kind of organization required only 72073 neurons, whereas the same hierarchical model without resource limitations would use up to 1146880 neurons. An example of the propagation of an image through that network is shown in Fig. 2.

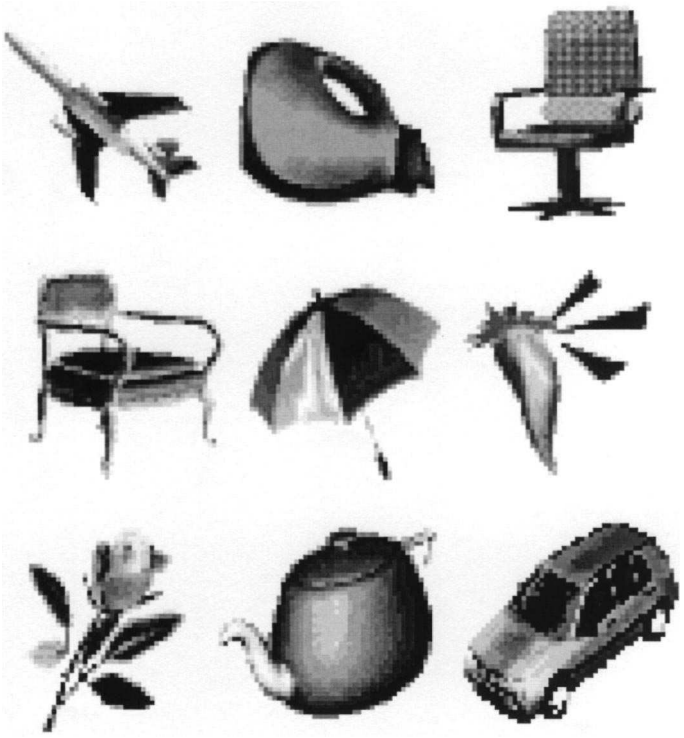


Fig. 1. Objects to be learned and categorized by the different models.

Since we had only one single "object detector" per object, the supervised learning was made as follows: we computed the mean pattern of firing order obtained, in the "complex features layer", for one object presented at each of the four possible locations, and that mean pattern became the order of the weights of the neuron selective for that object.

Furthermore, we introduced a lateral inhibition between the output neurons, so that only the first(s) one(s) to reach their threshold would respond.

Though the computation time was less than 1 s, the performance of that model was really poor. When objects were presented alone, they were always detected, without confusion with other objects. But when the objects were presented by pairs, in only 88% of the images one of the 2 targets was recognized, and in 22% of the trials, a completely different object was detected (see Fig. 2).

As expected, this kind of organization, with increasing receptive fields sizes, is a good way of saving neurons, but makes the model unable to deal with more than one object simultaneously, because features belonging to different objects are likely to be wrongly associated.

Nevertheless, it is well known that this organization scheme is indeed used by the human visual system. We propose that an attentional model in which the thresholds

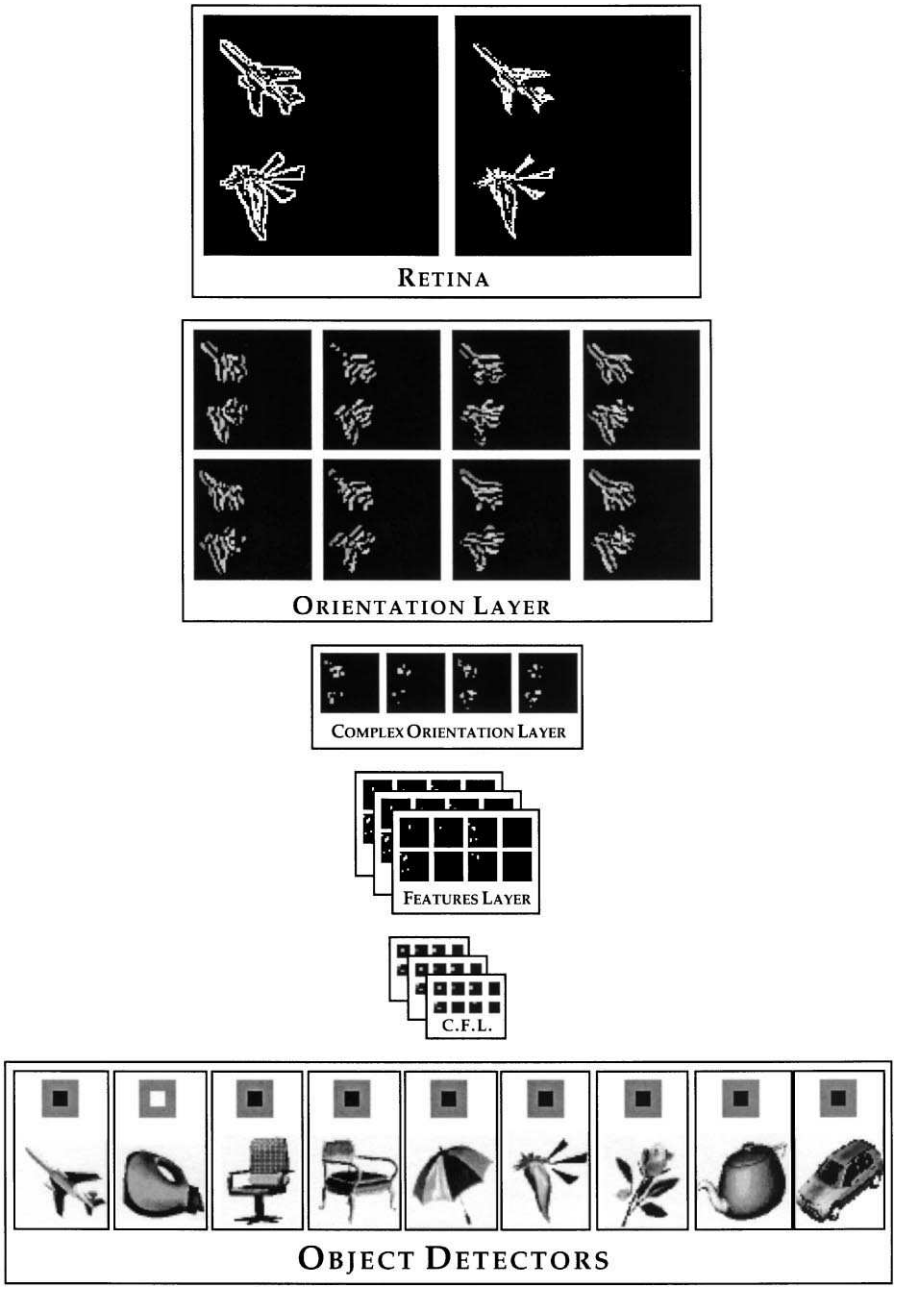


Fig. 2. An example of the propagation of an image in the 1st network. Each pixel in these maps represents a neuron, with white pixels corresponding to activated neurons. The output neurons sizes have been increased. Note that the network outputs a wrong object.

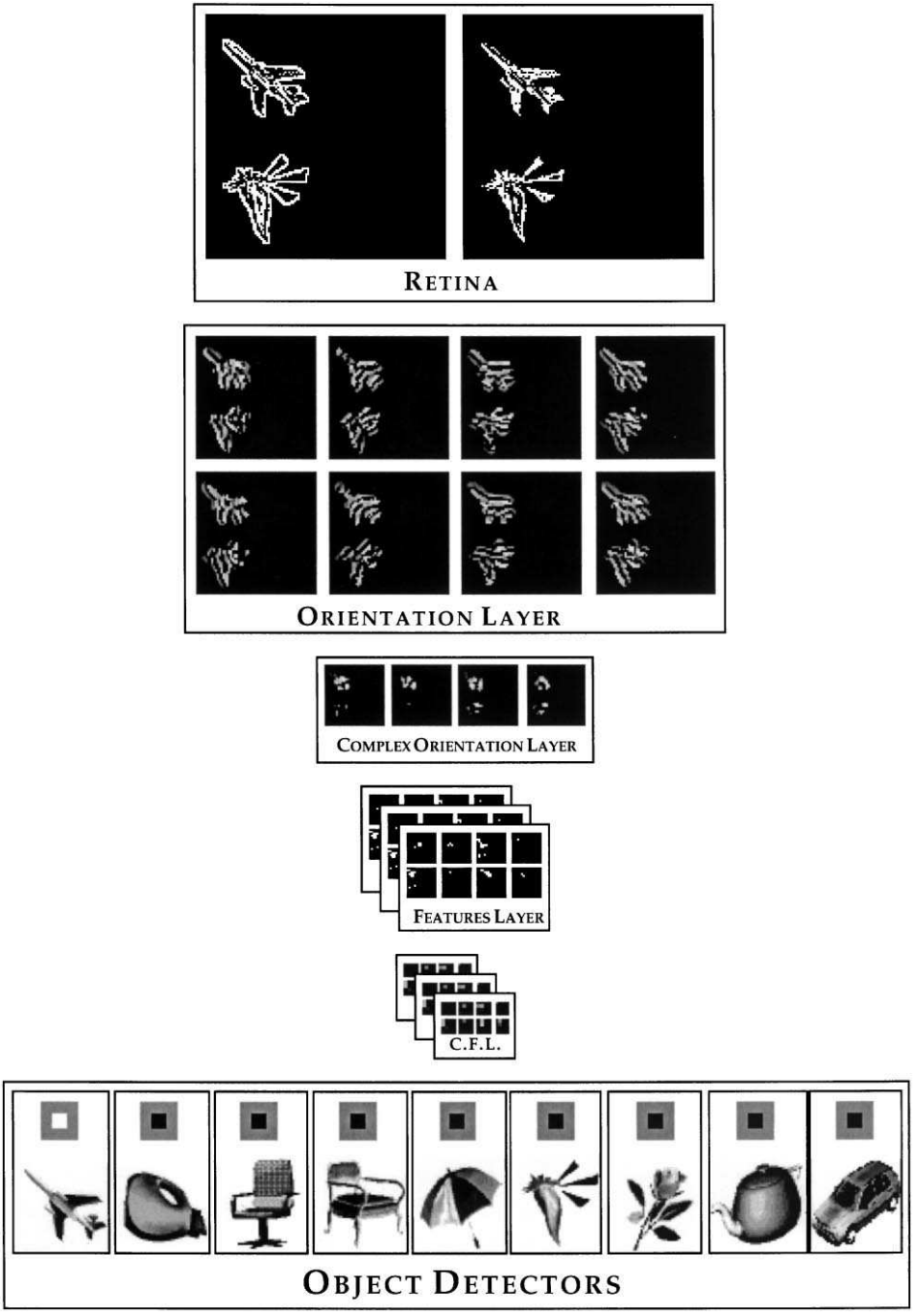


Fig. 3. The 2nd network after propagation of the same input image as in Fig. 2, with attention drawn to the upper-left part of the visual field. Here the attended object is correctly detected.

of the "relevant" neurons would be decreased, giving a temporal precedence to the "relevant information", could account for both computational efficiency and limited resources.

3.2. *Limited resources model, with attention*

In the second model, we kept the preceding model's organization, but we introduced an attentional mechanism, involving a threshold decrease for neurons whose receptive fields fell within a particular region of the visual field. An example of the propagation of an image through that network is shown in Fig. 3.

The computation time for that model was still under 1 s, but the level of performance significantly improved. All possible pairs of objects at all possible locations (different for the two objects) were tested with attention "drawn" (i.e. thresholds decreased) to a region containing one of the 2 targets. In 97% of the images, the network detected one of the targets, which was the attended target in 96% of the images. In contrast, a wrong object was selected for only 2% of the images.

These results seem to indicate that our model of attention constitutes an efficient way to overcome the problems arising with the resource limitations of biological visual systems.

4. Conclusion

An important feature of our results is that they can only be exhibited in a network of asynchronously spiking neurons. Lowering the thresholds for a given location in a classic artificial neural network, say a perceptron (with thresholded neurons), would be of no advantage. Neurons at this location would simply reach threshold when receiving a lower weighted sum (i.e. a less specific input). Hence they would be less selective, and performance would decrease. At the same time there would be no processing speed-up, because in such a network neurons need to compute the weighted sum of *all their inputs* at each time step before outputting their response.

A further point that distinguishes our model from most of the existing ones is that it is not only relevant to spatial attention, but can also explain other forms of attention, like feature-selective attention: attending selectively to a particular stimulus feature, such as its shape, orientation, or color, can be viewed as a global lowering of the thresholds of neurons encoding that particular feature, irrespective of their spatial location.

From a more biological point of view, the precise mechanism by which some cells thresholds could be selectively lowered remains unclear. It could for example rely on a localized neuromodulators release, that would affect the membrane properties of the neurons at that location. This is clearly not the only possibility, and we wish to leave that question open for further investigation.

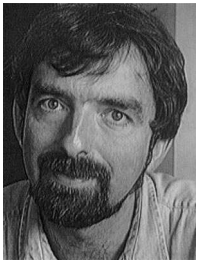
As yet there is no direct physiological evidence for a selective lowering of thresholds for neurons with receptive fields in attended parts of the visual field. Nevertheless, it

seems clear that giving a temporal precedence to the information in some retinotopic location is a good and simple way to explain spatial attention. Whether it involves a localized lowering of threshold, or rather some sort of preactivation is still an open question that merits direct neurophysiological investigation.

References

[1] A. Delorme, R. VanRullen, J. Gautrais, S.J. Thorpe, SpikeNET: a simulator for modelling large networks of integrate and fire neurons. Neurocomputing, submitted.
[2] M.C. Mozer, M. Sitton, Computational modeling of spatial attention, in: H. Pashler (Ed.), Attention, 1998, pp. 341–393.
[3] S.J. Thorpe, J. Gautrais, Rapid visual processing using spike asynchrony, in: M.C. Mozer, M.I. Jordan, T. Pesche (Eds.), Neural Information Processing Systems, MIT Press, Cambridge, 1997, pp. 901–907.
[4] 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, pp. 113–118.
[5] A. Treisman, The binding problem, Current Opinion in Neurobiol. 6 (1996) 171–179.
[6] R. VanRullen, J. Gautrais, A. Delorme, S.J. Thorpe, Face detection using one spike per neurone, Biosystems, 1998, in press.

Rufin VanRullen is a Ph.D. student in Cognitive Neuroscience at the Centre de Recherche Cerveau et Cognition in Toulouse, France. His background is in Mathematics and Computer Science. He is currently working on modeling the processes occurring in the primate visual system, e.g. object and face recognition or visual attention. One goal of this work is to explain the astonishing speed of processing in real visual systems when compared to artificial ones. Therefore, his interest has moved towards networks of asynchronously spiking neurons.



Simon Thorpe (D.Phil) is a Research Director working for the CNRS at the Centre de Recherche Cerveau and Cognition in Toulouse. He studied Psychology and Physiology at Oxford before obtaining his doctorate with Prof. Edmund Rolls in 1981. He joined Michel Imbert's group in Paris in 1982 and moved to Toulouse in 1993. He has used a range of techniques including single unit recording in awake monkeys, as well as ERP and fMRI studies in humans to study the brain mechanisms underlying visual processing.

3.3.2 Les arguments

3.3.2.1 Illustration: modification du transfert d'information dans le LGN par l'attention spatiale

Afin de bien comprendre le fonctionnement du mécanisme attentionnel que nous avons proposé dans l'article précédent, nous illustrons ici ses propriétés dans le cadre du transfert de l'information entre le LGN et le cortex visuel. En effet, l'information encodée dans le LGN, tout comme celle qui circule le long du nerf optique, représente principalement les contrastes locaux à différentes échelles spatiales. Il est donc possible, tout comme nous l'avons fait au niveau de la rétine dans l'article 3, de reconstruire à partir de ces informations de contraste, et à différentes étapes de la transmission neuronale, une image correspondant à la "représentation courante" du stimulus d'entrée, tel qu'il est encodé par la population de cellules géniculées.

Dans des conditions "normales" de transmission, l'image reconstruite est similaire à celle qui aurait été obtenue à partir des décharges des cellules ganglionnaires de la rétine. En effet, de même que dans la rétine, les contrastes les plus forts engendrent les latences de réponse les plus courtes au niveau du LGN, et cette organisation temporelle des décharges sur la population peut être décodée efficacement.

Supposons maintenant qu'un signal attentionnel "top-down" vienne biaiser cette organisation spatio-temporelle en faveur d'une région particulière. L'origine de ce signal peut être attribuée à différentes structures comme le pulvinar (Pashler, 1998) ou le cortex postérieur pariétal (Lawler et Cowey, 1987; Steinmetz et Constantinidis, 1995). Il n'est pas nécessaire ici de prendre part dans ce débat, et il sera suffisant d'assumer qu'un tel signal provient d'une "carte de saillance" (saliency map; Koch et Ullman, 1985; Itti et Koch, 2000) où est encodée explicitement la région d'intérêt vers laquelle l'attention doit être focalisée. Divers arguments convergent en faveur d'un possible "routage" spatial de l'information au niveau du LGN (e.g. Koch et Ullman, 1985; de Carvalho et Roitman, 1995; Vanduffel, Tootell et Orban, 2000). Dans ces conditions, nous supposons que ce signal attentionnel a pour effet soit d'augmenter l'activité, soit de diminuer les seuils des neurones du LGN dont le champ récepteur se situe dans la zone d'intérêt. Ceci résulte en une précédence temporelle pour les décharges des neurones de cette zone. Aux étapes suivantes, ces décharges étant reçues les premières, cette région sera donc interprétée comme la plus contrastée. Ceci se reflète dans les exemples de reconstructions présentés à la figure 25. Alors que seulement 1% des neurones du LGN ont émis un spike, la représentation de l'entrée visuelle qui est encodée par cette population accorde une importance toute particulière à la région d'intérêt. Cette région sera donc traitée en priorité par le système visuel.

Figure 25. Reconstructions simulées de l'information transmise lorsque 1% des cellules du LGN ont généré un potentiel d'action, dans diverses conditions de stimulation. En haut à gauche, dans des conditions non biaisées (i.e. sans attention top-down), la première information transmise est la plus saillante (correspondant aux contrastes les plus forts). L'attention portée sur une région particulière (comme indiqué par le point lumineux dans les 3 autres images réduites) biaise le flux d'information, laissant se propager plus rapidement l'information au niveau de cette région. L'information reçue en premier par les étapes de traitement suivantes, et interprétée comme la plus pertinente, correspondra donc à l'information la plus saillante (dans le cas de l'attention "bottom-up", non biaisée) ou à l'information attendue (dans le cas de l'attention "top-down").

Cette modification par l'attention du transfert de l'information neuronale peut bien sûr être effectuée à chaque étape du traitement visuel (comme nous l'avons fait dans l'article précédent), et pas seulement (ni obligatoirement) au niveau du LGN. L'activité simulée des populations neuronales du LGN présente néammoins l'avantage d'encoder les contrastes locaux de l'image d'entrée, permettant ainsi de reconstruire cette image grâce à la théorie des ondelettes. Dès le cortex visuel primaire, l'information portée par les réponses neuronales encodera en même temps le contraste (comme conséquence des traitements effectués aux niveaux précédents, i.e. rétine et LGN) et l'orientation locale. Aux niveaux suivants, d'autres propriétés participeront à cette représentation distribuée (par exemple la forme, l'organisation spatiale des objets). Une reconstruction du stimulus à partir de ces informations ne pourrait être obtenue simplement, à moins de disposer d'une théorie complète de la représentation distribuée des objets dans le système visuel. C'est pourquoi le type d'illustration présenté ici ne s'applique aisément qu'au codage de l'information dans la rétine ou le LGN. La théorie proposée pour l'attention, cependant, est valable à chaque niveau de traitement: quelle que soit la

nature de l'information encodée à un niveau particulier du traitement visuel (c'est-à-dire pas seulement pour le contraste local), une précédence temporelle pour une région d'intérêt donnée rendra l'information concernant cette région plus "saillante" pour les étapes suivantes. Comme nous le verrons par la suite, ceci est une des propriétés principales de l'attention visuelle, mise en évidence par l'expérimentation (Reynolds et Desimone, 1999). Ainsi, selon l'endroit (ou les endroits) où agira l'attention, ce modèle sera compatible avec les théories de l'attention "précoce" ou "tardive" (early or late selection models; LaBerge et Samuels, 1974; Duncan, 1980; Luck et al, 1990; Hillyard et al, 1998a,b), basées sur la position spatiale ou sur l'objet lui-même (space-based or object-based; Treisman et Gelade, 1980; Duncan, 1984; O'Craven et al, 1999).

3.3.2.2 Résultats psychophysiques.

De nombreux résultats expérimentaux convergent en faveur de l'hypothèse émise dans l'article précédent. Nous passerons ici en revue les résultats "macroscopiques" (i.e. obtenus par des études psychophysiques et/ou en potentiels évoqués), avant d'examiner les données provenant d'enregistrements unitaires chez le singe.

Notre hypothèse suggère que l'attention a pour effet de raccourcir les latences de réponses neuronales à proximité de la région d'intérêt. Ceci est confirmé par l'illusion de la "ligne en mouvement" (line motion illusion; Hikosaka et al, 1991, 1993a,b) présentée à la figure 26. La présentation d'un point 50 ms avant un segment, aligné avec ce point, induit la perception d'un mouvement à l'intérieur du segment, comme si la ligne apparaissait graduellement à partir du point. L'interprétation proposée par les auteurs est que l'apparition du point "capture" l'attention, et diminue les latences de décharge autour de ce point, provoquant un ordre temporel, similaire à celui qu'on obtiendrait si la ligne apparaissait graduellement (voir figure 26). L'attention semble donc induire une précédence temporelle pour la région d'intérêt. Cet effet attentionnel prend en fait la forme d'une différence de gaussiennes, avec une large facilitation temporelle autour du focus attentionnel, accompagnée d'une inhibition modérée sur le pourtour de cette zone (i.e. une organisation "center-surround"; Steinman et al, 1995).

En se basant sur cette illusion de ligne en mouvement, Kirschfeld et Kammer (2000) ont récemment proposé le même type de mécanisme attentionnel que nous avons décrit dans l'article précédent: l'attention agirait comme une modulation sous-liminaire des seuils neuronaux, et raccourcirait ainsi les latences de décharge dans la zone d'intérêt.

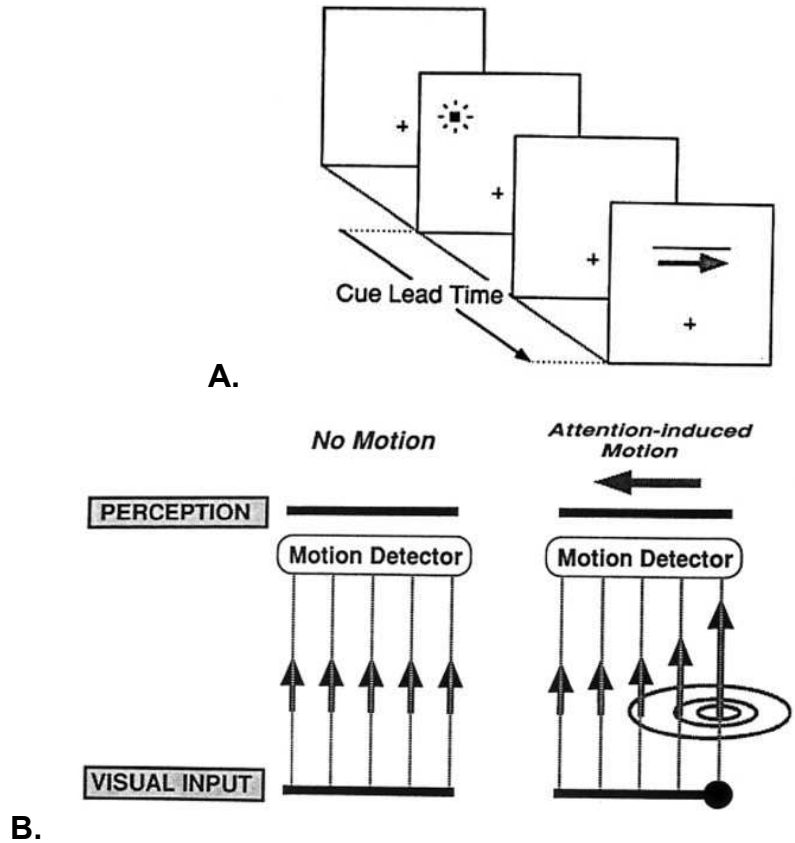


Figure 26. L'illusion de la "ligne en mouvement" (line motion illusion) de Hikosaka, Miyauchi et Shimojo (1993). Lorsqu'un point est présenté brièvement avant un segment, aligné avec ce point, le sujet voit la ligne apparaître graduellement à partir du point (A). Interprétation: l'attention est "capturée" par l'apparition du point, et focalisée sur ce point. Les latences de décharge des neurones situés sur la ligne sont graduellement raccourcies (l'effet est maximal à proximité du point, et diminue avec la distance), induisant une perception équivalente à celle d'une ligne en mouvement (B).

De nombreuses données de psychologie expérimentale supportent en fait cette hypothèse (Pashler, 1998): le "precueing" de la position d'une cible (i.e. l'apparition d'un indice visuel avant celle de la cible) donne lieu à des temps de réaction plus courts, accompagnés d'une augmentation significative des taux d'erreur, dans des tâches de discrimination rapide, ou de type go/no-go (Kingstone, 1992; Posner et al, 1980; Proverbio et Mangun, 1994). Shiu et Pashler (1993) obtiennent le même effet de "precueing", et reportent pour une tâche de discrimination rapide "lettre/chiffre", un raccourcissement d'environ 30 ms des temps de réponse. Ceci semble cohérent avec la durée minimale du traitement visuel, de l'ordre de 150 ms. L'augmentation des taux d'erreur est illustrée par Posner et al (1980) qui rapportent, pour une tâche de catégorisation go/no-go:

"*subjects found it very difficult to withhold responding when a non-target occurred in the expected position*" (p. 169).

Ces données suggèrent bien une diminution des latences de réponse, parallèle à (ou provoquée par) une diminution des seuils neuronaux.

De même, les résultats obtenus en potentiels évoqués sont compatibles avec notre hypothèse. L'attention visuelle permet d'allouer les ressources de traitement dans une région d'intérêt (Luck et al, 1996). Di Russo et Spinelli (1999a) ont explicitement démontré que la latence des potentiels évoqués visuels est raccourcie par l'attention, notamment pour des composantes très précoces comme la N60 ou la P100 (voir figure 27). Cette réduction de latence de l'ordre de 10-15 ms est probablement plus que suffisante pour engendrer une forte différence dans le traitement des stimuli "attendus" ou "non-attendus". Cet effet semble spécifique à la voie magnocellulaire: seule une augmentation par l'attention de l'amplitude de réponse est observée dans le cas d'une stimulation de la voie parvocellulaire (Di Russo et Spinelli, 1999b). Ce résultat corroborre d'autres données psychophysiques (Steinman et al, 1997), suggérant:

"*cues that preferentially excite the M pathway predominantly capture visual attention.*"

Il semble donc que, au moins pour la voie magnocellulaire, le recrutement de l'attention spatiale soit accompagné, au niveau des populations neuronales, par une diminution des latences de réponse, compatible avec une diminution des seuils neuronaux. Ces propriétés, qui supportent notre théorie de l'attention, se retrouvent-elles au niveau cellulaire?

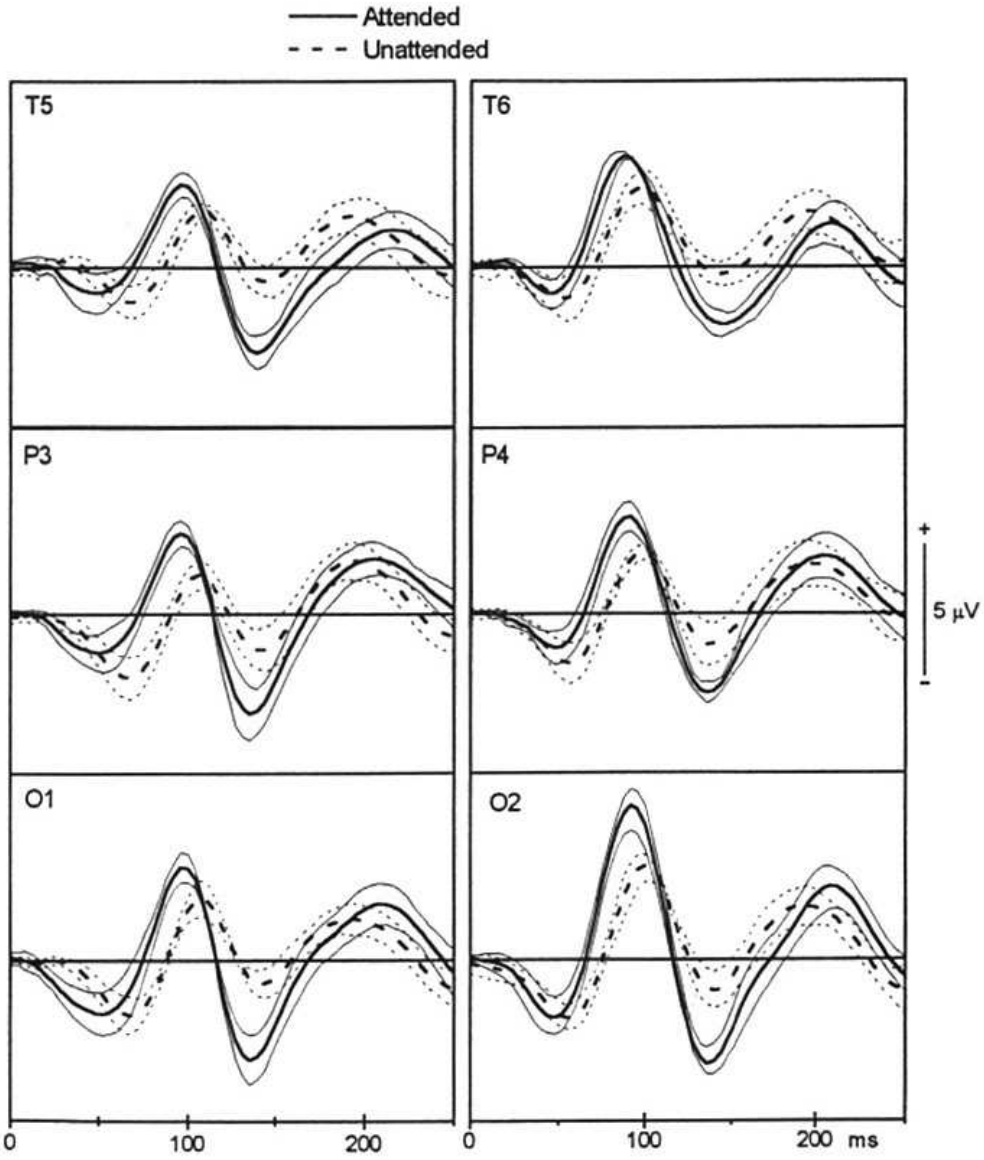


Figure 27. L'attention spatiale affecte l'amplitude et la latence des potentiels évoqués visuels. L'activité électrique en réponse à un grating sinusoïdal (irrelevant pour la tâche) est comparée pour 6 électrodes, dans des conditions où l'attention est portée soit dans l'hémichamp où se situe le grating (condition "attended"), soit dans l'hémichamp opposé (condition "unattended"). La latence des potentiels évoqués visuels est réduite lorsque l'attention est portée sur le stimulus, notamment pour les composantes précoces comme la N60 ou la P100. Reproduit d'après Di Russo et Spinelli (1999a).

3.3.2.3 Résultats électrophysiologiques.

Les données électrophysiologiques sur l'attention spatiale proviennent principalement d'un seul paradigme expérimental. Il s'agit de comparer les réponses d'un

même neurone à un même stimulus lorsque l'attention est portée sur celui-ci, et lorsqu'elle ne l'est pas. En général, 2 stimuli sont présentés à l'intérieur du même champ récepteur, et l'attention est portée alternativement sur l'un ou l'autre.

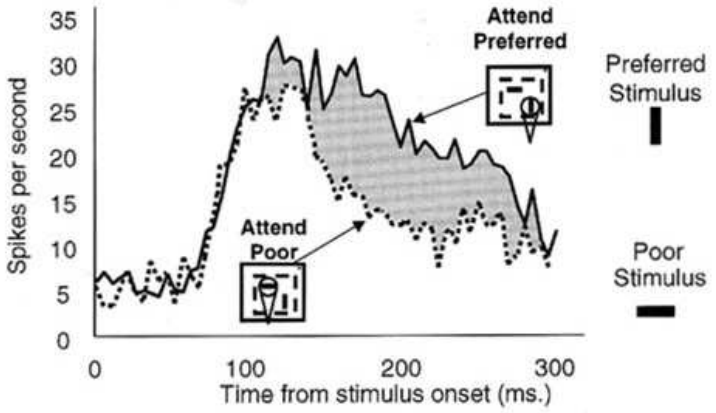


Figure 28. Effet de l'attention sur la sélectivité des neurones corticaux (ici moyenne sur une population de cellules de V4). La réponse est déterminée par le stimulus sur lequel est dirigée l'attention. Reproduit d'après Reynolds et Desimone (1999).

Ce paradigme a été introduit par Moran et Desimone (1985). Ces auteurs ont enregistré la réponse de neurones de IT et V4 lorsque 2 stimuli sont présentés simultanément dans le même champ récepteur. L'un des stimuli est choisi pour être "effectif", c'est-à-dire qu'il engendre une forte réponse de la cellule lorsqu'il est présenté seul dans son champ récepteur, l'autre est "ineffectif" (i.e. engendrant une faible réponse). La réponse du neurone dépend alors du stimulus sur lequel est portée l'attention: forte réponse si le stimulus est effectif, faible réponse s'il est ineffectif. Tout se passe comme si le champ récepteur se contractait autour du stimulus attendu, laissant l'autre stimulus en-dehors de ce "champ récepteur virtuel" (figure 28). De plus, lorsque l'un des 2 stimuli est placé à la même distance de l'autre que précédemment, mais en dehors du champ récepteur, la réponse du neurone n'est plus déterminée par l'attention: le neurone répond maintenant à un stimulus effectif, même si l'attention est portée sur le stimulus ineffectif en dehors du champ récepteur. Classiquement, ce résultat est interprété comme la preuve que l'attention spatiale n'est pas un biais pour une région particulière aux dépens de toutes les autres régions du champ visuel, mais plutôt un biais spécifique à chaque champ récepteur (Luck et al, 1997).

Des résultats similaires ont été obtenus pour des cellules de V2 (Luck et al, 1997; Reynolds et al, 1999), et pour des neurones de la voie dorsale, dans les aires MT et MST (Treue et Maunsell, 1996; voir cependant Seidemann et Newsome, 1999, qui

reportent des effets bien moins prononcés). En ce qui concerne V1, les résultats sont actuellement moins tranchés. Certaines études électrophysiologiques ont reporté des effets attentionnels au niveau du cortex visuel primaire (e.g. Motter, 1993; Vidyasagar, 1998; Roelfsema at al, 1998; McAdams et Maunsell, 1999). En général, les champs récepteurs de ces neurones sont trop réduits pour pouvoir utiliser un paradigme d'activation simultanée avec 2 stimuli (Luck et al, 1997), ce qui pourrait expliquer pourquoi d'autres études reportent une absence d'effet attentionnel dans V1 (e.g. Moran et Desimone, 1985). L'utilisation des potentiels évoqués ou de l'IRMf mène cependant à des résultats tout aussi mitigés (par exemple, *modulation attentionnelle dans V1*: Shulman et al, 1997; Watanabe et al, 1998a,b; Somers et al, 1999; Brefczynski et DeYoe, 1999; Tootell, et al, 1998; Vanduffel, Tootell et Orban, 2000; Smith, Singh et Greenlee, 2000; *Pas de modulation attentionnelle dans V1*: Clark et Hillyard, 1996; Hillyard et Allo-Vento, 1998; Heinze et al, 1994; Luck et Ford, 1998; Martinez et al, 1999). Il est donc aujourd'hui difficile de conclure sur ce point. Cependant, si l'on accepte que l'attention spatiale peut modifier l'activité des cellules du LGN (voir plus haut), alors les neurones du cortex visuel primaire ont également une forte probabilité d'être modulés, ne serait-ce que par l'action de ces afférences thalamiques.

Un certain nombre d'observations expérimentales viennent compléter ces premiers résultats instaurés par Moran et Desimone (1985). Tout d'abord, les données de Connor et al (1996, 1997) suggèrent que la réponse d'un neurone peut en fait être modulée par l'attention dirigée en dehors de son champ récepteur, dans des conditions où le focus attentionnel reste proche du bord du champ récepteur. De plus, cette modulation attentionnelle dépend de la position du stimulus à l'intérieur du champ récepteur: elle augmente lorsque le stimulus se rapproche du focus attentionnel (voir figure 29).

L'attention spatiale est souvent associée avec une augmentation du taux de décharge spontanée des neurones, en l'absence de stimulation. Luck et al (1997) ont rapporté une augmentation de l'ordre de 30 à 40% de l'activité spontanée des neurones de V2 et V4, dûe à l'attention. L'intensité de cette modulation dépend en fait de la distance entre le focus attentionnel et le centre du champ récepteur, ce qui démontre que la résolution spatiale du focus attentionnel peut être plus précise que la taille du champ récepteur. Une augmentation similaire de l'activité spontanée a été démontrée chez l'homme par des techniques d'imagerie (Kastner et al, 1999). Cependant, certaines études électrophysiologiques n'obtiennent aucun effet de l'attention sur le taux de décharge spontanée (McAdams et Maunsell, 1999), et cette divergence de résultats n'est pas expliquée à ce jour.

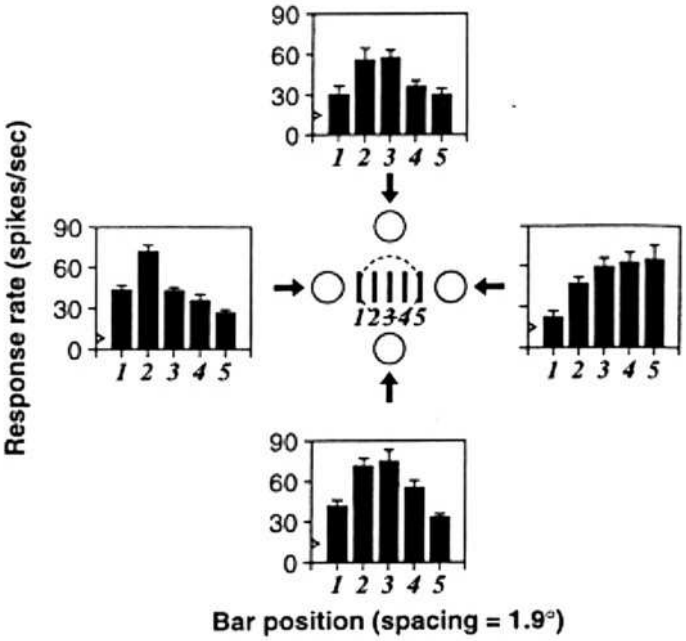


Figure 29. Si le focus attentionnel se trouve à proximité du champ récepteur d'un neurone (de V4), une modulation attentionnelle peut être obtenue pour ce neurone. Chacun des histogrammes montre les réponses de la cellule aux 5 positions possibles du stimulus, avec l'attention focalisée sur l'un des 4 cercles entourant le champ récepteur. Cette modulation attentionnelle est d'autant plus forte que le stimulus (ici, une barre orientée) est proche du focus attentionnel.

En plus du gain de réponse, ou de l'activité spontanée, l'attention pourrait également modifier la sélectivité des neurones corticaux (Haenny et Schiller, 1988; Spitzer et al, 1988). Une étude récente de McAdams et Maunsell (1999) démontre qu'en fait, la sélectivité à l'orientation des neurones de V4 (mesurée par la largeur de la courbe de "tuning" à l'orientation) n'est pas systématiquement modifiée par l'attention, alors que l'amplitude de réponse est augmentée de près de 30%. Leurs données suggèrent que l'effet de l'attention sur la sélectivité de ces neurones prendrait la forme d'un "multiplicative scaling" (voir figure 31A). La modulation attentionnelle est ici observée dès le début de la réponse neuronale.

De même, les interactions entre les objets présentés simultanément dans le même champ récepteur et la position du focus attentionnel apparaissent bien plus complexes que le simple mécanisme "tout-ou-rien" mis en évidence par Moran et Desimone (1985). En utilisant le même paradigme, Reynolds, Chelazzi et Desimone (1999) ont démontré que *(i)* la réponse d'un neurone à 2 stimuli présentés simultanément dans son champ récepteur (dans des conditions où l'attention est portée loin de ce champ récepteur) est une somme pondérée (i.e. intermédiaire) des réponses

de ce neurone à chacun des stimuli présentés séparément (voir également Rolls et Tovee, 1995), et *(ii)* l'attention biaise cette somme pondérée en faveur de l'un ou l'autre des stimuli, selon la position du focus attentionnel (voir par exemple figure 30). La modulation attentionnelle apparaît peu de temps après le début de la réponse dans V2 (i.e. environ 50 ms post-stimulus), et immédiatement dans V4 (i.e. environ 75 ms post-stimulus).

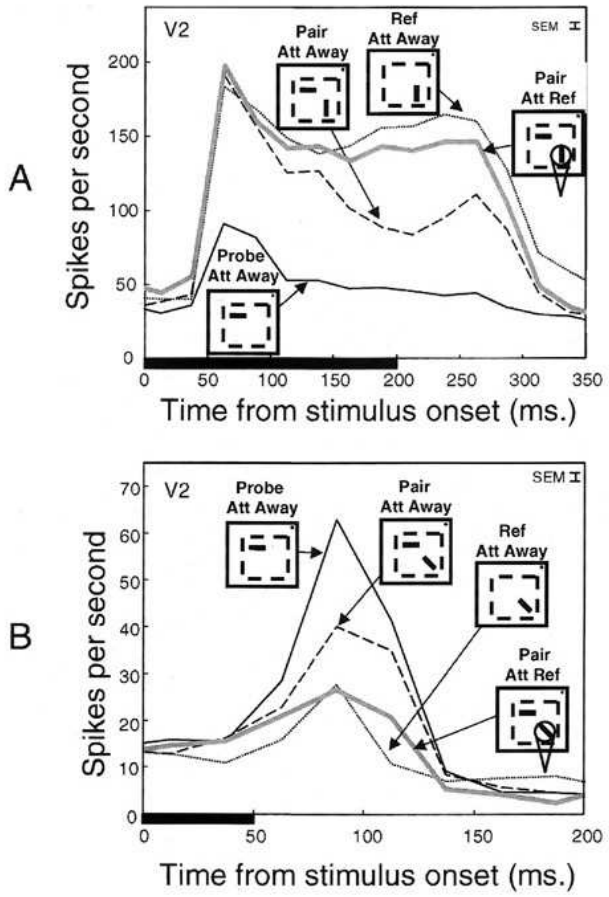


Figure 30. Interactions entre le nombre de stimuli présentés simultanément et la position du focus attentionnel. A. Le neurone répond fortement à un stimulus vertical ("reference") présenté seul, et faiblement à un stimulus horizontal ("probe") présenté seul, l'attention étant dirigée loin du champ récepteur. La réponse aux 2 stimuli présentés simultanément est intermédiaire (une somme pondérée des 2 réponses précédentes). Lorsque l'attention est dirigée vers le stimulus effectif ("reference"), la réponse à la paire de stimuli tend vers la réponse obtenue pour ce stimulus présenté seul. B. Le stimulus "reference" est ici le stimulus ineffectif. De même, la réponse du neurone aux 2 objets présentés simultanément est une somme pondérée des 2 réponses aux stimuli isolés, et l'attention dirigée sur la référence biaise cette somme pondérée en faveur de la réponse à la référence isolée. Reproduit d'après Reynolds et Desimone (1999).

L'effet de l'attention spatiale sur la réponse d'un neurone à un stimulus isolé est moins bien défini que pour une paire d'objets en compétition. Certaines études ont observé une augmentation de la réponse d'un neurone à un stimulus isolé sous l'effet de l'attention (e.g. Connor et al, 1996), augmentation qui peut atteindre 20-25% de la réponse au stimulus non attendu dans V4 (Spitzer et al, 1988; McAdams et Maunsell, 1999). Cependant, d'autres études n'ont trouvé aucune modulation attentionnelle de la réponse à un stimulus isolé, ou une modulation qui pouvait prendre la forme d'une augmentation comme d'une diminution (Moran et Desimone, 1985; Haenny et al, 1988; Maunsell et al, 1991; Motter 1993; Luck et al, 1997). Récemment, une théorie proposée par Reynolds et Desimone (1999) a permis de réconcilier ces résultats divergents. Elle est basée sur l'observation que la sensibilité au contraste des neurones de V4 est accrue lorsque l'attention est portée sur leur champ récepteur (Reynolds et al, 1996; voir figure 31B). En moyenne, sous l'effet de l'attention, les neurones répondent comme si le contraste du stimulus avait été augmenté de 22%. Dans ces conditions, la réponse à un stimulus déjà très contrasté ne peut être modifiée par l'attention (phénomène de saturation), alors que cette réponse sera augmentée pour des stimuli faiblement contrastés. La différence entre les résultats expérimentaux obtenus proviendrait donc d'une différence entre les contrastes des stimuli utilisés dans ces différentes études. Ce résultat permet également d'expliquer la propriété de "multiplicative scaling" mise en évidence par McAdams et Maunsell (1999). En effet, la selectivité à l'orientation est invariante aux changements de contraste (Skottun et al, 1987). Pour des stimuli en-dessous du seuil de saturation au contraste, une augmentation du contraste physique résulte en une multiplication du gain de réponse de la courbe de sélectivité à l'orientation. De même, une augmentation de la sensibilité au contraste sous l'effet de l'attention résulte en un déplacement vers la gauche de la fonction de réponse au contraste (figure 31B), et prédit donc la multiplication du gain de la courbe de sensibilité à l'orientation, démontrée par McAdams et Maunsell (1999; voir figure 31C).

L'attention aurait donc pour effet de rendre les stimuli plus "contrastés". Le même effet a été reporté pour des stimuli définis par une différence de texture (Nicholas et al, 1996): l'attention augmente la réponse à un stimulus de ce type, mais ce phénomène disparaît si le stimulus est trop facilement discriminable du fond. L'attention agit donc comme une augmentation du contraste de texture, dans la limite où ce contraste n'est pas saturé. Si l'on généralise ces 2 résultats (augmentation du contraste de luminance, augmentation du contraste de texture), il semble en fait que l'attention ait, invariablement, pour effet d'augmenter la saillance relative des stimuli attendus

(Reynolds et Desimone, 1999). Nous reviendrons plus tard sur cette propriété qui est un aspect fondamental du modèle de modulation attentionnelle que nous proposons.

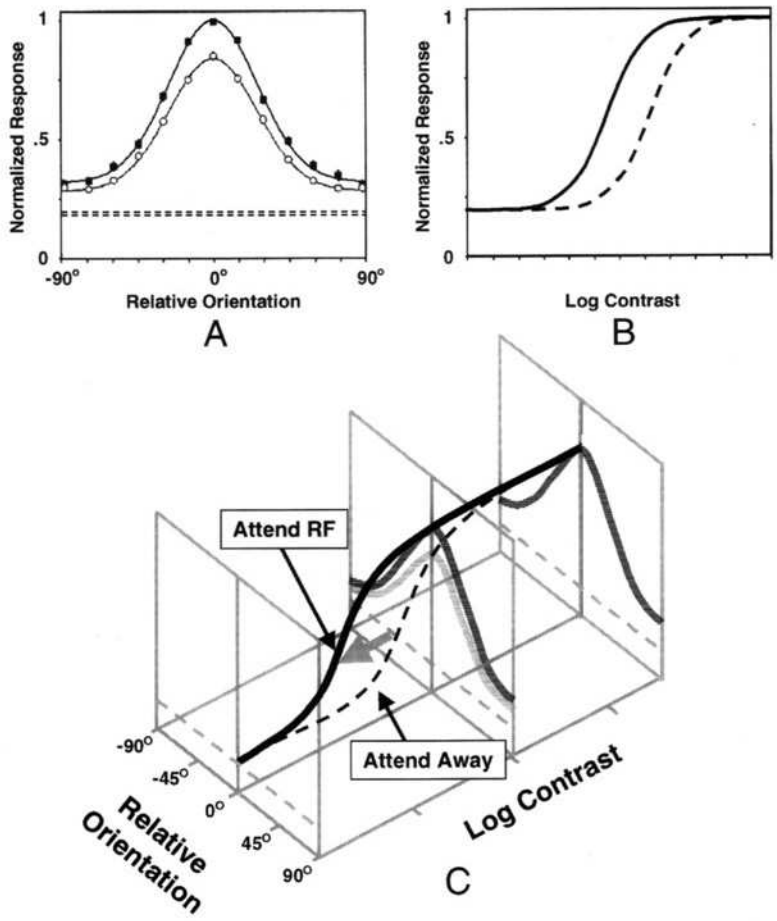


Figure 31. A. L'attention a pour effet de rehausser la courbe de réponse à l'orientation des neurones de V4, sans en changer le pic, la largeur ou l'asymptote (multiplicative scaling; McAdams et Maunsell, 1999). **B.** La sensibilité au contraste des neurones de V4 (ligne pointillée) est accrue par l'attention (ligne continue; Reynolds et al, 1996). **C.** Cette augmentation de la sensibilité au contraste sous l'effet de l'attention (i.e. un décalage vers la gauche de la fonction de réponse au contraste) entraîne une augmentation multiplicative de la courbe de sélectivité à l'orientation, ce qui explique le résultat obtenu en A. Reproduit d'après Reynolds et Desimone (1999).

Sur la base de ces nombreux résultats électrophysiologiques, Reynolds et Desimone (1999) ont récemment défini un ensemble de contraintes biologiques, que tout modèle "plausible" de l'attention visuelle se doit de respecter:

1. Lorsque l'attention est dirigée en dehors du champ récepteur d'un neurone, l'addition d'un second stimulus à l'intérieur de ce champ récepteur a pour effet d'entraîner la réponse du neurone vers la réponse élicitée par ce stimulus isolé. Si l'on change ce second stimulus de sorte qu'il élicite une plus faible réponse lorsqu'il est présenté seul, alors, typiquement, il devient proportionnellement plus suppressif.
2. Les réponses neuronales sont biaisées en faveur du stimulus le plus saillant. Si un stimulus efficient est associé à un stimulus inefficient, la suppression engendrée par ce dernier augmente si l'on augmente son contraste.
3. Lorsque l'attention est dirigée à l'intérieur du champ récepteur, la sensibilité du neurone à des stimuli présentés à cet endroit est augmentée. La cellule peut répondre à des stimuli peu saillants, qui sans attention n'engendreraient aucune réponse.
4. Lorsque plusieurs stimuli sont présentés simultanément, porter l'attention sur l'un d'eux biaise la réponse neuronale en faveur de la réponse qui aurait été engendrée si ce stimulus avait été présenté seul. Typiquement, l'attention portée sur le moins effectif des 2 stimuli réduira la réponse du neurone à la paire d'objets.
5. Le biais engendré par des différences de saillance relative s'ajoute au biais de la modulation attentionnelle. Diriger l'attention sur un stimulus contrasté lui donnera un contrôle encore plus prononcé sur la réponse du neurone, alors que porter l'attention sur un stimulus faiblement contrasté permettra de compenser cette différence de contraste. Ceci permet au neurone de répondre à un stimulus de faible contraste, qui reçoit des ressources attentionnelles, même en présence d'un distracteur très contrasté.
6. L'attention peut moduler la réponse des neurones à une échelle spatiale plus fine que celle d'un champ récepteur individuel. Lorsque plusieurs stimuli sont présents dans le champ récepteur, l'attention permet donc au neurone de traiter sélectivement l'un d'eux, tout en ignorant les autres. Cette haute résolution spatiale se reflète également dans le fait que l'augmentation par l'attention du taux de décharge spontanée dépend de la position précise du focus attentionnel à l'intérieur du champ récepteur.
7. L'effet de l'attention dépend de la séparation spatiale des stimuli. 2 stimuli peuvent être placés assez loin l'un de l'autre pour se trouver dans des champs récepteurs séparés à une étape corticale donnée (par exemple V2), mais rester assez près pour se trouver dans un champ récepteur commun à l'étape de traitement suivante (par exemple V4). Ainsi, à un certain niveau, l'attention

renforcera la réponse évoquée par le stimulus (ceci correspond en fait au point 3), alors qu'au niveau suivant, elle filtrera le signal attendu, et permettra d'ignorer le reste (ceci correspond donc aux points 4 et 6).

Reynolds, Chelazzi et Desimone (1999) ont également proposé un modèle de la modulation attentionnelle qui respecte ces contraintes. Ce modèle dérive de la théorie de "biased competition" (Desimone & Duncan, 1995): les signaux visuels se trouvent constamment en compétition, du fait de la limitation des ressources neuronales (l'agrandissement des champs récepteurs le long de la voie ventrale est une illustration de ces limitations); l'attention permet de biaiser cette compétition permanente en faveur d'un stimulus donné (le plus saillant dans le cas de l'attention bottom-up, celui qui reçoit explicitement les ressources attentionnelles dans le cas de l'attention top-down). Le modèle de Reynolds et al (1999) implémente cette compétition biaisée, en supposant que l'attention permet de jouer sur la balance entre excitation et inhibition, pour un neurone recevant des entrées de 2 (ou plusieurs) sous-populations.

3.3.2.4 Adéquation de notre modèle.

Le modèle que nous avons proposé dans l'article 5 respecte-t-il les contraintes mentionnées plus haut, et peut-il donc prétendre au titre de modèle "biologiquement plausible" de l'attention? Nous examinerons dans un premier temps la fonction de réponse des neurones de ce modèle à une paire de stimuli, sans invoquer l'attention spatiale (contraintes 1 et 2), puis l'effet de notre mécanisme attentionnel sur ces mêmes réponses.

Avant tout, rappelons que les neurones de notre modèle sont sensibles à l'ordre d'arrivée des spikes afférents. A une étape donnée, une vague de potentiels d'action est générée en réponse à une autre vague similaire, provenant de l'étape précédente. Une telle vague contient typiquement le premier spike émis par chaque neurone de la population. Un neurone donné de l'étape suivante intègre les spikes de cette vague, attribuant une importance capitale aux premiers spikes reçus, i.e. aux afférents les plus activés, qui correspondent aux traits les plus saillants du stimulus d'entrée. Un neurone qui reçoit ces spikes afférents dans son ordre "préféré" (i.e. l'ordre de ses poids synaptiques) sera optimalement activé: il aura détecté son stimulus préféré. Si le neurone reçoit les potentiels d'action afférents dans un ordre éloigné de son ordre "préféré", son activation sera faible, et typiquement, ne lui permettra pas d'atteindre son seuil de décharge. Ceci est illustré dans la figure 32.

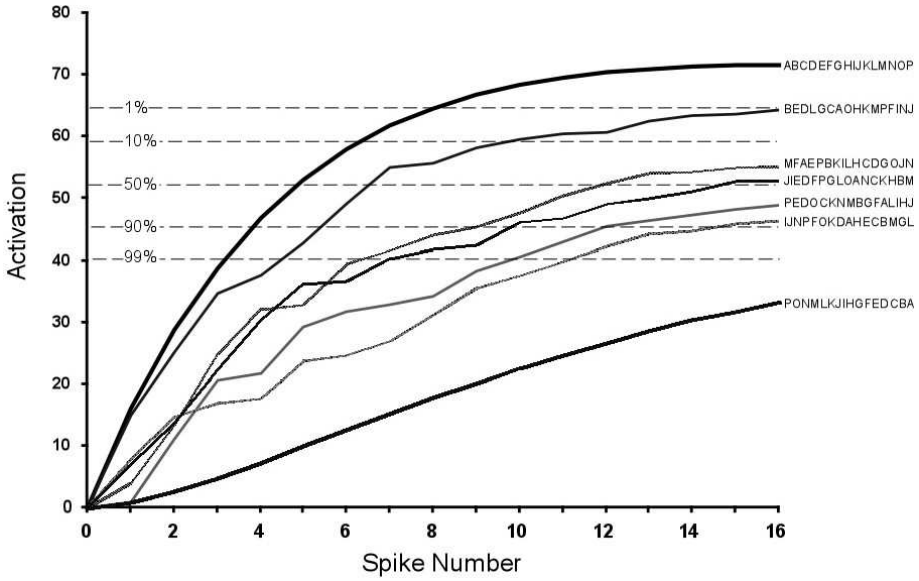


Figure 32. Evolution du niveau d'activité d'un neurone sensible à l'ordre d'activation de ses synapses "ABCDEFGHIJKLMNOP", pour différents ordres d'arrivée des spikes afférents. L'activation est optimale pour l'ordre préféré, i.e. le stimulus pour lequel la cellule est sélective. Pour l'ordre inverse, le niveau d'activité est minimal. La présentation simultanée des 2 stimuli a pour effet d'imbriquer, dans la même vague de spikes afférents, les spikes codant pour le stimulus 1 (dans leur ordre d'origine), et ceux codant pour le stimulus 2 (dans leur ordre d'origine). L'ordre d'arrivée des potentiels d'action sur ce neurone est donc intermédiaire entre les 2 ordres initiaux, de même que son niveau d'activation. La réponse du neurone à une paire d'objets est donc une somme pondérée des réponses individuelles. Si l'on porte l'attention sur le stimulus 1, ou de façon équivalente, si l'on augmente son contraste, l'information codant pour ce stimulus est propagée plus rapidement, et atteint donc plus tôt le neurone considéré. L'ordre d'arrivée de ses afférents se rapproche ainsi de l'ordre induit par le stimulus 1, et le niveau d'activation tend vers celui évoqué par le stimulus 1 seul. On observe donc une facilitation de la réponse si le stimulus 1 est optimal pour ce neurone, ou une suppression s'il est ineffectif.

Supposons maintenant qu'un seul neurone reçoive dans son champ récepteur 2 stimuli simultanément. L'un est effectif pour ce neurone (i.e. proche de son ordre préféré), l'autre est ineffectif (i.e. plus proche de l'ordre inverse). La présentation simultanée des 2 stimuli a pour effet d'entremêler, à l'intérieur de la même vague de potentiels d'action, des spikes codant pour l'un et l'autre des stimuli. Le neurone efférent reçoit donc ces potentiels d'action dans un ordre intermédiaire, et son niveau d'activation reste lui aussi intermédiaire (figure 32). La réponse du neurone à une paire de stimuli est donc une somme pondérée des 2 réponses isolées (contrainte 1). Notons que dans certains cas particuliers, il est possible que l'ordre des spikes afférents résultant de la présentation simultanée de 2 stimuli non-optimaux se rapproche de l'ordre préféré du neurone (typiquement, si le neurone est sélectif à la séquence d'entrées ABCD, si l'objet

1 est codé par la séquence ACDB et l'objet 2 par la séquence BDCA, la présentation simultanée des objets 1 et 2 avec des contrastes appropriés pourra donner lieu à la séquence d'activation ABCD). Ceci pourrait correspondre à un phénomène de "conjonction illusoire": lorsque l'attention est inefficace, par exemple du fait d'une trop grande charge computationnelle, le système visuel peut regrouper des attributs de 2 objets, et percevoir ainsi une conjonction qui n'est pas présente dans la stimulation (Treisman et Schmidt, 1982). Dans notre cas, si l'attention est effectivement portée sur l'un des 2 objets, la conjonction illusoire résultant d'un ordre d'activation "intermédiaire" n'aura pas lieu.

Le stimulus le plus saillant (ou contrasté) se propage le plus rapidement à travers le système (une propriété fondamentale des réseaux asynchrones de neurones integrate-and-fire: les neurones les mieux activés déchargent le plus vite). C'est donc en général l'ordre des potentiels d'action correspondant au stimulus le plus saillant qui parviendra à un neurone. Si le contraste d'un stimulus ineffectif est augmenté, l'ordre d'arrivée des potentiels d'action aura tendance à refléter ce stimulus, et la réponse sera supprimée. Si le contraste d'uin stimulus effectif est augmenté, l'ordre d'arrivée des spikes se rapprochera de l'ordre induit par ce stimulus, et la réponse sera facilitée (i.e. la suppression induite par le deuxième stimulus sera contrebalancée par l'augmentation du contraste pour le premier). Ceci correspond à la contrainte 2 de Reynolds et Desimone (1999).

Lorsque l'attention est portée dans une région donnée de l'espace, les seuils des neurones codant pour cette région sont diminués, et les décharges se produisent plus tôt. Ceci est équivalent à ce qu'on observerait si le contraste (ou la saillance) du stimulus présenté à cet endroit était augmenté. Pour un neurone recevant cette information, tout se passe comme si la sensibilité au contraste était accrue dans la région d'intérêt (contrainte 3).

Si 2 stimuli sont présentés simultanément, la précédence temporelle résultant de l'allocation de l'attention sur l'un des stimuli équivaut à une augmentation de la saillance relative de ce stimulus. Ainsi, comme pour la contrainte 2, la réponse du neurone tendra vers la réponse élicitée par le stimulus attendu seul. Ceci explique les résultats expérimentaux principaux de Moran et Desimone (1985), et la contrainte 4. La figure 33 illustre cette propriété.

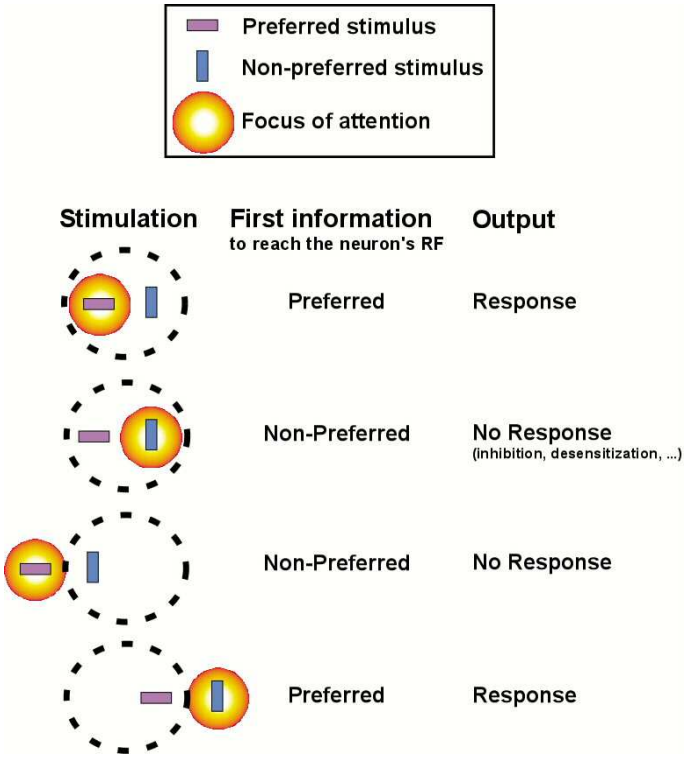


Figure 33. L'hypothèse d'une précédence temporelle pour la région d'intérêt explique les résultats expérimentaux de Moran et Desimone (1985). L'attention allouée à l'un des 2 stimuli présentés simultanément implique que l'information relative à ce stimulus se propage plus rapidement dans le système. Les premières décharges reçues par un neurone déterminent sa réponse: dans le cas où les 2 stimuli se trouvent à l'intérieur de son champ récepteur, il est optimalement activé si l'attention est portée sur le stimulus effectif, et faiblement activé si elle est portée sur le stimulus ineffectif; dans le cas où l'attention est portée sur un stimulus en dehors du champ récepteur, bien que l'information codant pour ce stimulus soit transmise en priorité, elle n'atteint pas le neurone considéré, et sa réponse sera uniquement déterminée par le stimulus se trouvant dans son champ récepteur.

Pour pouvoir porter l'attention sélectivement sur l'un des 2 objets présents dans le même champ récepteur, il faut nécessairement que la résolution du focus attentionnel soit plus fine que la taille du champ récepteur lui-même (contrainte 6). Ceci est réalisé simplement en modulant la réponse des neurones aux niveaux inférieurs, là où les champs récepteurs sont bien plus réduits. Selon la forme (par exemple gaussienne) et l'étendue désirée pour le focus attentionnel, cette modulation pourra être plus ou moins forte, et prendre place à des niveaux plus ou moins élevés de la hiérarchie du cortex visuel. Cette propriété rend compte également de la contrainte 7.

La saillance d'un stimulus se reflète dans l'ordre temporel à l'intérieur d'une vague de potentiels d'action. Selon notre hypothèse, c'est sur cet ordre temporel qu'agit l'attention, simulant en quelque sorte une augmentation de saillance relative pour l'objet attendu. C'est parce que ces deux phénomènes, saillance relative des stimuli et attention, agissent sur le même vecteur d'information, l'asynchronie des décharges d'une population neuronale, que leurs effets s'ajoutent tout simplement: diriger l'attention sur un stimulus contrasté lui donnera un contrôle encore plus prononcé sur la réponse du neurone, alors que porter l'attention sur un stimulus faiblement contrasté permettra de compenser cette différence de saillance (contrainte 5). C'est cette correspondance entre saillance visuelle (souvent interprétée comme le reflet de l'attention "bottom-up") et attention top-down, qui fait à la fois la force et la simplicité du mécanisme attentionnel que nous proposons.

3.3.2.5 Attention non spatiale

Lorsque nous recherchons un ami dans une foule, un produit particulier dans un rayon de supermarché, ou un bijou laissé tomber sur une plage, l'attention spatiale, tout au moins dans un premier temps, n'est d'aucun secours: la position de l'objet à traiter n'est pas connue à l'avance. Par contre, les propriétés de l'objet recherché sont, elles, connues. Il est donc possible de biaiser sélectivement les neurones spécialisés dans le traitement de ces propriétés. Par exemple, Chelazzi et al (1998) rapportent que lorsqu'un singe macaque recherche activement un objet donné, les neurones du cortex inféro-temporal codant pour cet objet voient leur activité spontanée augmenter, avant même l'apparition du stimulus. Après la présentation du stimulus, composé de plusieurs objets, dont la cible, la réponse des neurones de IT à la même stimulation visuelle dépend de la cible recherchée: réponse forte si la cible est "effective" pour la cellule, réponse faible sinon, même si un stimulus effectif est présent, en tant que distracteur. Ceci suggère que l'attention a permis de biaiser, de façon non sélective à la position, la compétition en faveur des neurones participant explicitement à la représentation de l'objet rechérché. Ce biais peut bien sûr avoir lieu au niveau du cortex inféro-temporal, où les cellules peuvent encoder spécifiquement l'identité de la cible, mais également aux niveaux inférieurs, où les populations neuronales peuvent encoder des propriétés caractéristiques de la cible, telles que sa couleur, son orientation, etc

Le mécanisme attentionnel proposé plus haut s'applique bien sûr dans ces conditions d'attention distribuée: une diminution des seuils pour les neurones qui représentent des propriétés de l'objet recherché aura pour effet de laisser l'information se propager plus rapidement pour ces propriétés, quelle que soit la position de l'objet. Tous les objets possédant des propriétés incompatibles avec l'objet recherché verront

leur représentation supprimée dans les derniers niveaux de traitement, où l'information concernant l'objet "cible" sera parvenue en premier.

3.3.2.6 Délai de mise en œuvre de l'attention

L'hypothèse d'une précédence temporelle pour l'information attendue suggère que l'information représentée par les tout premiers spikes arrivant à un neurone donné correspond à l'objet attendu. La réponse de ce neurone, et notamment son premier spike, étant déterminée par cette première information reçue, on s'attendrait à observer l'effet de la modulation attentionnelle dès les premières millisecondes de son activité. Or, selon les études, et selon les cellules étudiées lors d'une même expérience, l'effet de la modulation attentionnelle peut s'observer soit dès le début de la réponse, soit quelques millisecondes plus tard (voir par exemple figure 30). Ce dernier résultat est classiquement interprété comme un rôle du feed-back dans la modulation attentionnelle, et semble contredire notre hypothèse d'un effet de l'attention sur la latence du premier spike. Afin de rejeter cette conclusion, nous traçons ici un parallèle entre la distribution des premiers spikes émis par un neurone donné, et la distribution des temps de réaction des sujets humains lors d'une tâche de catégorisation animal/non-animal, ou véhicule/non-véhicule (article 1).

Pour chaque image présentée, le sujet humain génère (au plus) une réponse motrice et une seule. Les temps de réaction s'étendent pourtant d'environ 200 à 600 ms. De même la latence du premier spike généré par un neurone peut être sujette à une large variabilité. Sur un "post-stimulus time histogram", qui est nécessairement une moyenne sur un nombre élevé d'essais, rien n'indique en général combien des potentiels d'action observés correspondent au premier spike émis par le neurone considéré, mais cette variabilité peut atteindre plusieurs dizaines de millisecondes.

Tout comme le sujet humain, le neurone peut commettre des erreurs (des "false positives"). Typiquement, une erreur reflète une réponse générée alors que l'information accumulée n'était pas suffisante pour pouvoir décider de la catégorie du stimulus présenté. Chez le sujet humain, ces anticipations sont fréquentes lors des 50 premières millisecondes de la distribution des temps de réaction. Pourtant, la tâche est réalisée avec une performance de plus de 95%, et le sujet n'utilise clairement qu'une seule réponse motrice par essai. De même, le neurone peut commettre des anticipations, ou "false positives" sur certains essais, et toujours réaliser la tâche correctement, même si l'on ne prend en compte que les premiers spikes qu'il émet en réponse à chaque stimulus. Tout comme le taux d'anticipations dépend de la tâche comportementale et du sujet humain qui la réalise, il pourra dépendre du neurone individuel considéré: un neurone avec un seuil relativement bas est plus enclin à commettre de telles anticipations qu'un autre. Enfin, il serait certainement intéressant de pouvoir étudier le

lien potentiel entre ces "anticipations du premier spike" au niveau neuronal, et les anticipations motrices du sujet, au niveau comportemental.

3.3.3 Résumé

Nous avons proposé dans cette section un mécanisme de modulation attentionnelle qui permet de renforcer l'information attendue de manière purement feed-forward. Ce mécanisme est basé sur l'idée que la saillance des stimuli visuels se reflète dans l'ordre temporel des décharges des populations neuronales. Ainsi, il suffit de biaiser cet ordonnancement temporel en faveur d'une position ou d'un objet, pour voir s'accroître la saillance relative de ce stimulus. La réponse des neurones situés en haut de la hiérarchie corticale visuelle est déterminée par le stimulus le plus saillant, i.e. par le stimulus attendu. Cette hypothèse, soutenue par des données psychophysiques, permet également d'expliquer les principaux résultats électrophysiologiques. Enfin, elle présente l'avantage de réunir dans un seul cadre théorique, l'asynchronie des décharges neuronales, les notions complémentaires de saillance et d'attention visuelles.

4. Une solution pour le traitement visuel rapide?

Nous avons vu (section 2) qu'une seule vague de potentiels d'action générée en réponse à une stimulation visuelle peut contenir bien plus d'information qu'un ensemble de spike trains encodant l'image d'entrée sur plusieurs centaines de millisecondes. Il suffit pour cela de prendre en compte l'asynchronie des décharges sur une population neuronale. Dans ces conditions, cette vague de potentiels d'action se propageant en mode feed-forward à travers le système visuel permet de réaliser des fonctions (intégration de contours, attention spatiale) que l'on attribue classiquement à des mécanismes itératifs. Les contraintes temporelles pesant sur le système visuel, qui impliquent l'utilisation d'un nombre minimal de spikes, ainsi qu'une propagation vers l'avant, la plus rapide possible, de l'information, s'avèrent en réalité bien moins limitantes qu'on aurait pû le croire. Serait-il finalement possible que le système visuel utilise ce type de stratégie pour parvenir à transformer une image naturelle complexe en une représentation abstraite explicite de sa catégorie, en seulement 150 ms?

Nous tentons dans l'article suivant de répondre à cette question, sur un exemple de tâche visuelle de haut niveau: la détection des visages.

4.1 Article 6 : VanRullen, Gautrais, Delorme & Thorpe, 1998. **Face processing using one spike per neurone. BioSystems.**

ELSEVIER

BioSystems 48 (1998) 229–239

BioSystems

Face processing using one spike per neurone

Rufin Van Rullen *, Jacques Gautrais, Arnaud Delorme, Simon Thorpe

Centre de Recherche Cerveau et Cognition, UMR 5549, 133 route de Narbonne, 31062 Toulouse, France

Abstract

The speed with which neurones in the monkey temporal lobe can respond selectively to the presence of a face implies that processing may be possible using only one spike per neurone, a finding that is problematic for conventional rate coding models that need at least two spikes to estimate interspike interval. One way of avoiding this problem uses the fact that integrate-and-fire neurones will tend to fire at different times, with the most strongly activated neurones firing first (Thorpe, 1990, Parallel Processing in Neural Systems). Under such conditions, processing can be performed by using the order in which cells in a particular layer fire as a code. To test this idea, we have explored a range of architectures using SpikeNET (Thorpe and Gautrais, 1997, Neural Information Processing Systems, 9), a simulator designed for modelling large populations of integrate-and-fire neurones. One such network used a simple four-layer feed-forward architecture to detect and localise the presence of human faces in natural images. Performance of the model was tested with a large range of grey-scale images of faces and other objects and was found to be remarkably good by comparison with more classic image processing techniques. The most remarkable feature of these results is that they were obtained using a purely feed-forward neural network in which none of the neurones fired more than one spike (thus ruling out conventional rate coding mechanisms). It thus appears that the combination of asynchronous spike propagation and rank order coding may provide an important key to understanding how the nervous system can achieve such a huge amount of processing in so little time.

Keywords: Neurone; Face processing; Spike

1. Introduction

Electrophysiological data indicate that some neurones of the monkey temporal lobe respond selectively to complex stimuli such as faces with a latency of 80–100 ms after stimulus onset (Bruce et al., 1981; Perrett et al., 1982; Oram and Perrett, 1992; Jeffreys, 1996). Within this short time, information has to be processed not only by the retina and LGN, but also by several cortical areas, including V1, V2, V4 and PIT, in each of which two synaptic stages at least must be passed through. Hence, information has to run through ten different processing stages, within the 100 ms-window considered.

* Corresponding author.

0303-2647/98/$ - see front matter
PII S0303-2647(98)00070-7

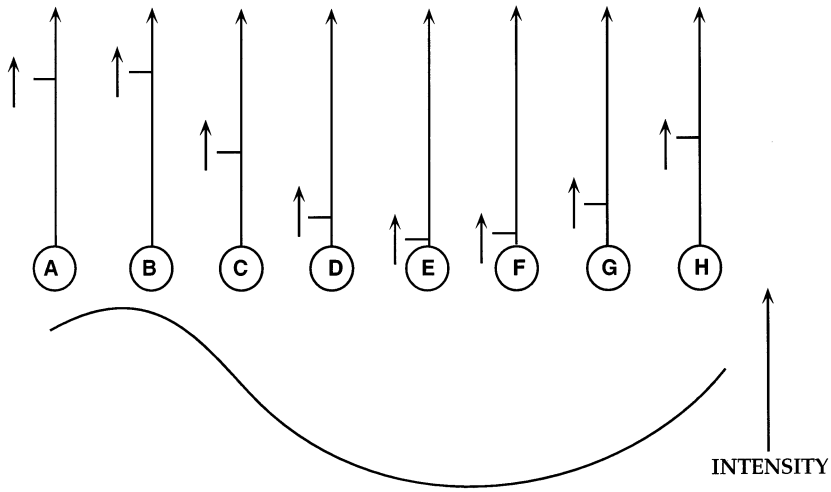


Fig. 1. Neurones can act as analogue-latency converters, with more strongly activated neurones firing first. One can also use the order of firing (B > A > H > C > G > D > F > E) as a code. With eight neurones, there are 8! i.e. 40 320 different possible orderings.

Furthermore, conduction velocities of intracortical fibres are known (Nowak and Bullier, 1997) to be remarkably slow (e.g. < 1 m/s), which leaves less than 10 ms for computation (axonal conduction, synaptic transmission, somatic integration and spike emission) in each processing stage. Such rapid processing poses severe problems for conventional rate coding mechanisms, since very few cells will be able to emit more than one spike in this time, and although it would be possible to calculate firing rates across a population of cells, we have argued elsewhere that this would require very large numbers of redundant cells (Gautrais and Thorpe, 1998).

How might the visual system perform complex tasks like face detection without emitting more than one spike in each processing stage? One option is to use differences between spike latencies across a population of neurones as a code (Thorpe, 1990). This argument is based on the fact that integrate-and-fire neurones, such as those observed in the visual cortex, will tend to fire at different times, with the most strongly activated neurones firing first (Fig. 1). Only one spike per neurone is then required to create a complete representation of the input stimulus.

Under such conditions, one can also use the relative order of firing of units in a particular layer to represent the input information. This sort of coding scheme, which we call Rank Order Coding, is still compatible with the constraint of having only one spike per neurone, whereas it greatly simplifies the computation. Nevertheless, it does not lead to a decrease in representational capacity by comparison with traditional rate coding, since a population of n neurones can actually discriminate $n!$ different stimuli, when in the same time window only $n + 1$ codes could be recognised with rate coding (see Gautrais and Thorpe, 1998, for a theoretical analysis of these issues).

In order to test this new coding scheme, a simulator called SpikeNET was designed, which allows to model very large populations of integrate-and-fire neurones (Thorpe and Gautrais, 1997). We used this simulator to evaluate the performance of a simple model based on Rank Order Coding, and to test whether complex visual processing tasks such as face detection and localisation in a natural image could be performed on the basis of only one spike per neurone.

2. Architecture of the model

2.1. Asynchronous propagation

SpikeNET simulates neurones with simple integrate-and-fire characteristics: afferent spikes in-

crease their activation level, until they reach a threshold and fire a single spike. The response of such a neurone (i.e. the latency of its output spike) can be made to depend upon the relative order of firing of its afferents by progressively desensitizing the neurone each time one of its inputs fires (Thorpe and Gautrais, 1998). This could be achieved by a relatively simple mechanism involving feed-forward shunting inhibition in which all afferents reduce the sensitivity of the target cell, irrespective of their synaptic weights.

More precisely, let $A = \{a_1, a_2, a_3 \ldots a_{m-1}, a_m\}$ be the ensemble of afferent neurones of neurone i, with $W = \{w_{1,i}, w_{2,i}, w_{3,i} \ldots w_{m-1,i}, w_{m,i}\}$ the weights of the m corresponding connections; let $\text{mod} \in [0,1]$ be an arbitrary modulation factor. The activation level of neurone i at time t is given by:

$$\text{Activation}(i,t) = \sum_{j \in [1,m]} \text{mod}^{\text{order}(a_j)} w_{j,i}$$

where order(a_j) is the firing rank of neurone a_j in the ensemble A. By convention, order(a_j) $= +\infty$ if neurone a_j has not fired at time t, setting the corresponding term in the above sum to zero.

Neurone i will fire at time t if (and only if)

$$\text{Activation}(i,t) \geq \text{Threshold}(i)$$

Under such conditions, two important features can be pointed out:

- the better the match between the order in which afferent spikes arrive and the pattern of connectivity, the more strongly the neurone will be activated. Specifically, optimal activation is achieved when the spikes arrive in the order of their weights, with the inputs having the highest weights arriving first.
- he most strongly activated neurones (i.e. those where the order of their inputs best matches their weights) will tend to reach threshold and fire earlier.

These points are of great importance for further computation and learning.

2.2. The face detection model

The most important feature in designing our model was to keep it as simple as possible, so that each step of processing could be fully understood. At the same time we wanted the architecture to be at least inspired by the first stages of processing in the visual system.

Thus, the architecture that we used was a very simple four-layer feed-forward neural network (Fig. 2). Each layer was composed of a set of maps of different selectivity. Each map contained many neurones, each coding the specific information relative to a particular location (pixel) in the input image.

As in the primate visual system, receptive fields properties became increasingly complex as processing progresses. Units in the first layer had concentric ON- and OFF-centre receptive fields, like retinal ganglion cells, responding optimally to a positive or negative local contrast, whereas units in the second layer had orientation selectivity (eight different orientations, separate from each other by 45°) similar to that seen in simple cells in V1. Units in the third layer were trained to respond to the pattern of activation in layer 2 that was characteristic of basic facial features (left eye, right eye and mouth) of a particular size, whereas fourth layer units were designed to respond optimally when these three different components were present in the appropriate locations, i.e. when a face was present. Thus the latter units had a pattern selectivity similar to that found in some neurones of the inferotemporal cortex (Bruce et al., 1981; Perrett et al., 1982; Abbott et al., 1996).

3. Learning method and receptive fields

3.1. Learning method

The principle of the learning method is based on the intrinsic properties of integrate-and-fire neurones and Rank Order Coding. As shown in Fig. 1, the relative order of firing in cells of a population constitutes the code of the input stimulus. A neurone will reach its highest activation level when the relative order of firing in its inputs will best match the order of the corresponding connections (Section 2.1).

Therefore, to detect a specific firing pattern in a population of simulated integrate-and-fire neurones, it is sufficient to use a set of connections

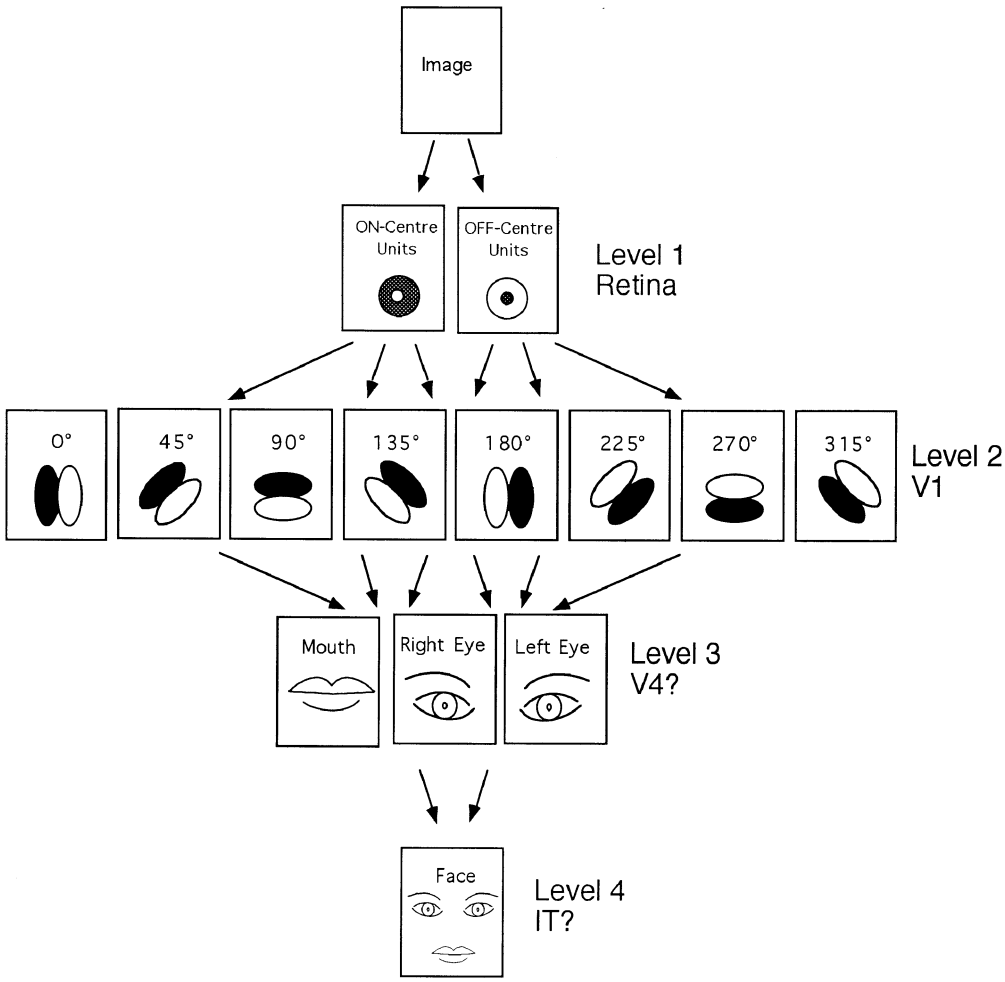


Fig. 2. Architecture of the model. Connections are feed-forward only. The original image is first decomposed in two (positive and negative) local contrast maps, then in eight different orientation maps (each separated by 45°). Units of the third layer respond to the activity in layer two that is characteristic of a mouth, right or left eye. Units at level 4 respond optimally when these three features are present simultaneously in the appropriate locations.

that respect the relative order of this firing pattern. For instance, that order could be obtained by computing the mean order of a set of training examples.

3.2. Receptive fields

3.2.1. Local contrast maps

The first processing step in our model is the decomposition of the input image in local positive or negative contrasts. This can be achieved very easily using sharp Gaussian laplacian filters. However, whereas in a more conventional neural network architecture, the value resulting from this convolution would be sent to later stages, in our model, the resulting value is used to determine at what latency each cell will fire. The earliest firing cells in level 1 will thus correspond to the parts of the input image where the local contrast is highest.

3.2.2. Orientation maps

At the second level of the computation, the input image is segmented in eight different orien-

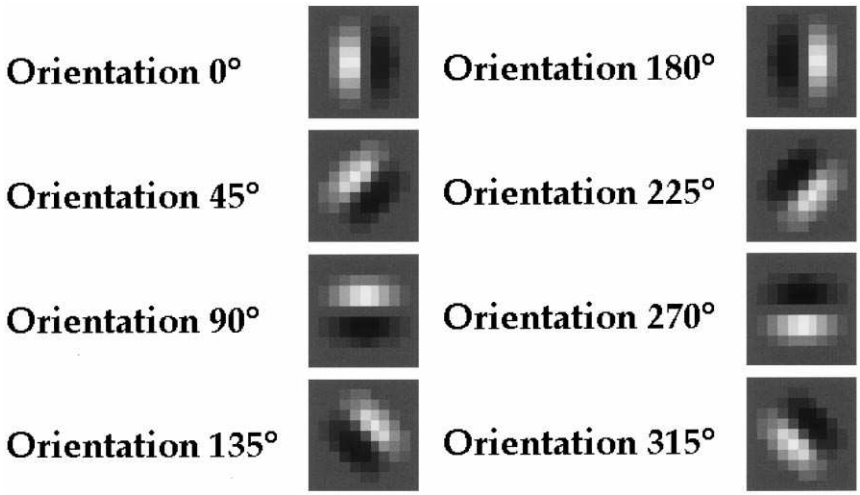


Fig. 3. Receptive field organisation for orientation selective cells in layer two. Bright and dark pixels correspond, respectively to positive and negative weights from ON-centre cells in layer one. Connections from the OFF-centre cells map are identical, but rotated by 180°.

tation maps, each selective to a particular orientation. The receptive fields used for that purpose (Fig. 3) are oriented Gabor filters of the same type as those used by Thorpe and Gautrais (1997).

Neurones at that level show selectivity in that they respond at shorter latencies when the orientation of an edge in the image matches the shape of their receptive fields.

3.2.3. Feature selective maps

The patterns of connections between the eight orientation-tuned maps in layer 2 and the feature detecting cells in layer 3 were set in such a way that the cells responded best when the order of activation in the different maps was close to that seen with a set of training stimuli.

We used a training database of 270 front ($\pm 30°$) views (92×112 grey-level images) of male and female faces (ten views of 27 persons), of which only a small proportion wore glasses (2%) or had a beard (11%). For each image, the precise locations of the right and left eyes and mouth were determined manually. The images were propagated through layers 1 and 2 of the network, and a region of the appropriate size, around the location of the mouth or left or right eyes, was extracted from each orientation map. The size of this region was determined so that it should include not only the feature itself, but also the immediately surrounding area. Thus, the receptive fields of the 'eye-detecting' neurones included the eyebrows whereas the zone of interest for the 'mouth-detecting' cells included part of the nose. For each feature (mouth, left and right eyes) and for each orientation map, the mean order of firing in the corresponding region was computed over the entire database.

The resulting mean order patterns were then used directly to determine the strength of the connections linking the orientation and feature maps (Fig. 4). As a result, neurones in each feature-detection will be strongly activated only if the corresponding feature is present at the appropriate location. Thus, the position of the firing neurones in such a map gives information about the precise location(s) of the feature(s) in the input image.

3.2.4. Face detection map

Neurones in the level 4 (face-detection map) were set up to fire if the three basic facial features

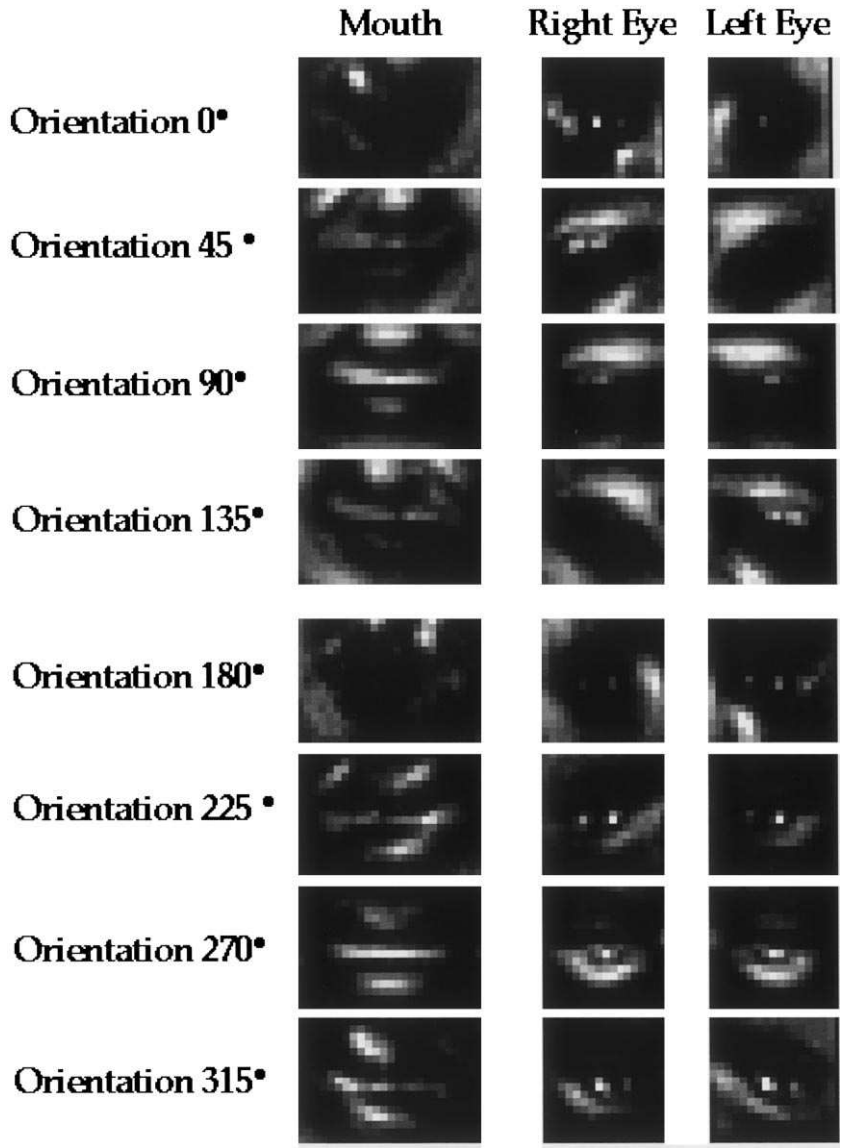


Fig. 4. Receptive field organisation of feature selective neurones. All weights are positive, with black pixels set to zero and white ones corresponding to a maximum weight value. Each orientation map is connected to each feature selective map by the corresponding weight set.

(mouth, left and right eyes) are present in the image, and in the appropriate locations. This 'facial structure', i.e. the relative positions of the component features of a face, can be easily described using a set of three Gaussian filters centred at the appropriate positions in the receptive fields of such a neurone (Fig. 5).

4. Results

4.1. Propagation results

The pattern of firing obtained when an image containing a face is propagated through the network is illustrated in Fig. 6. For clarity, only four

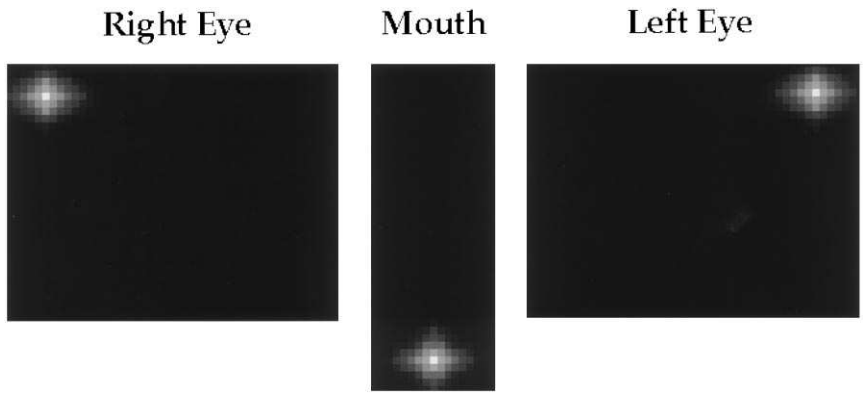


Fig. 5. Connection patterns between feature detecting cells in layer three and the face selective neurones in layer four. Each of these filters is centred at the neurones location.

of the eight orientation tuned maps are shown. Within each map, the brightness of the individual points corresponds to the order in which the neurones fired-bright spots correspond to neurones that were among the first to fire and the grey scale value gets progressively darker for later firing neurones.

It is clear that the network performs accurately. Activity in the three feature level maps is restricted to the places corresponding to the locations of the right eye, the mouth and the left eye, respectively. Similarly the region of activated cells in the face level map corresponds to the centre of the face.

It should be noted that such a network has no problem coping with an input image containing more than one face. The large number of neurones in the face-detection layer means that there is effectively one 'face-cell' for every pixel in the image. Images with multiple targets will simply produce the appropriate number of activated regions in the output map.

Conversely, when an image containing no face is propagated through the network, the first two computation steps are run as described above, leading to a representation of the image in terms of oriented edges. However, none of the feature selective neurones in the following layer should receive an input ordered well enough to let it reach its threshold. As a result, no facial features should be detected and no face-cell activity will be present in the final layer.

4.2. Statistical results

4.2.1. Testing method

To evaluate the performance of the model, we have tested it with a large range of natural images. Two public databases containing many face images were used, together with our own database of natural images that had no face present.

The first database was obtained from the Olivetti Research Laboratory and consisted of 400 frontal ($\pm 30°$) views (92×112 grey-level images) of male and female faces (ten views of 40 persons), that we separated in two groups. The first group, which we will call database 1, included the 270 images that were used for training and contained only a small percentage of people wearing glasses or with a beard. The second group ('database 2') contained the remaining 130 images (ten views of 13 persons), and had a large proportion of people wearing glasses (88%) or a beard (31%). These were used as a set of 'difficult' examples.

The next database, which we will call database 3, contained 300 frontal views (256×171 grey-level images) of different people whose faces were approximately the same size as those used for training. Roughly half of them wore glasses, while 16.6% had a beard, proportions that can be considered as reflecting the 'every-day life' conditions.

The last database ('database 4'), contained 216 (84×104) grey-level images of natural scenes with no faces, but a large range of animals, plants,

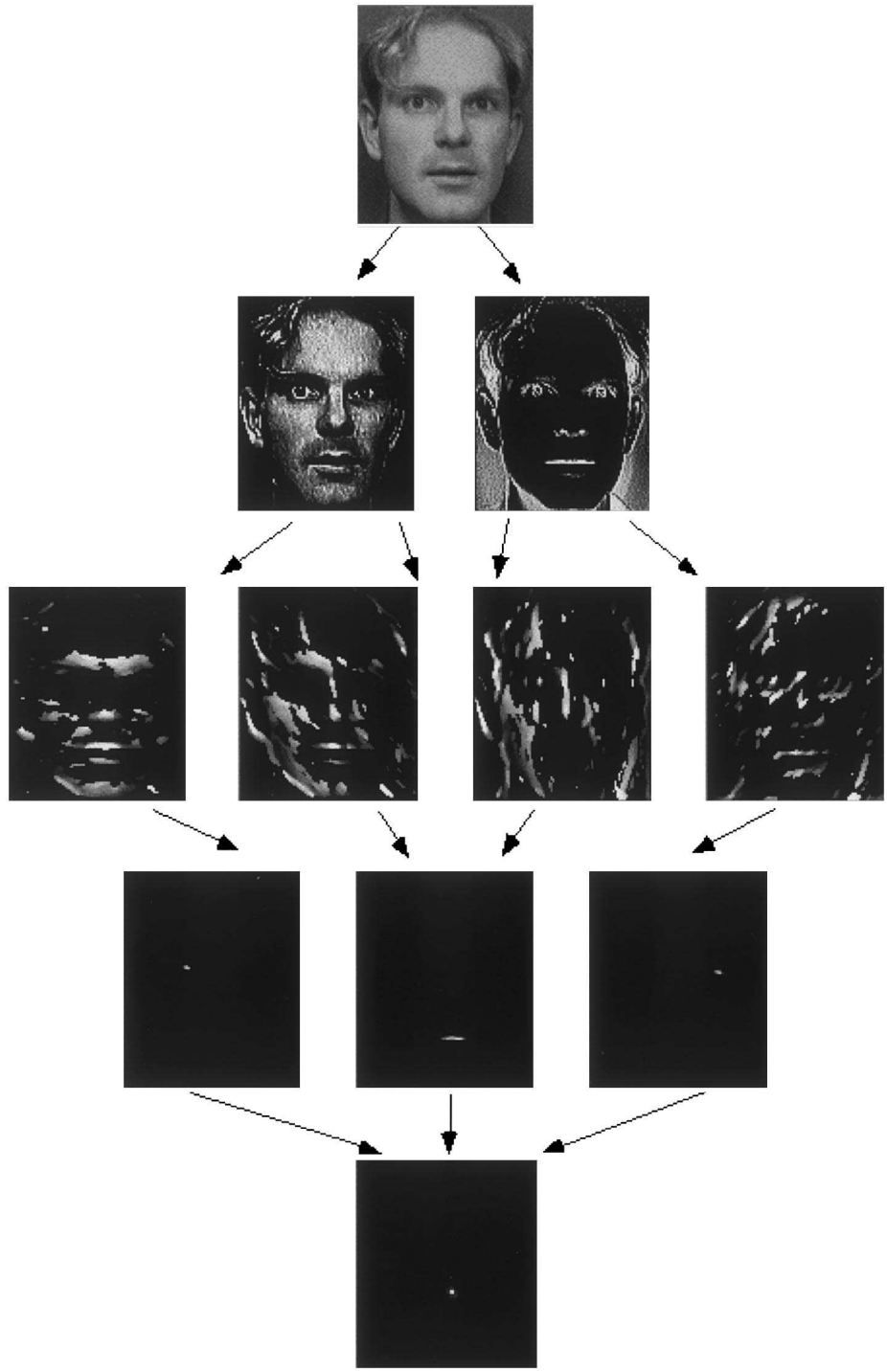


Fig. 6. Result maps obtained after the propagation of an image of a face in the network. From top to bottom: the original image, the ON- and OFF-centre cells maps, four out of eight orientation maps, the three feature-detection maps (right eye, mouth and left eye) and the face detection map. In each map, non-black pixels correspond to firing neurones, with the grey-level intensity representing the firing order in the entire layer (neurones that fired first are brighter). The position of the firing neurones in the last layer gives the explicit location of the face in the original image.

Table 1
Results of the model with four different test databases

Detection map	Mouth	Right eye	Left eye	Face
Test database				
Database 1 (135 images)	92.3% (18)	97.8% (97)	95.6% (100)	96.3% (2)
Database 2 (130 images)	88.5% (27)	83.1 (88)	80% (87)	73.1% (4)
Database 3 (300 images)	91% (89)	92.7 (222)	75% (198)	94% (4)
Database 4 (216 images)	— (14)	— (13)	— (12)	— (1)

For each database and each detection map, the percentage indicates the detection rate, the number in parentheses indicates the false detection number. This number must be compared with the number of images in the database, and the number of neurones per image. Database 2 is the 'difficult' example base. Database 4 contains no face image.

landscapes or objects such as cars, buildings and food… It was used to determine the error rate of the model.

We define the detection rate as

$$\text{Detection rate} = \frac{\text{Number of detected features/faces}}{\text{Number of features/faces present in the database}}$$

and the false detection number as

False detection number

= Number of firing regions where the feature/face was not present

Half of the images of the learning base (database 1) were used to determine optimal values for the various parameters of the system, which include the modulation value and the threshold levels. These were set individually for each feature map in order to maximise the detection rate whilst keeping the mean false detection number at below a maximum of one false detection per image. Coefficients for the face-selective neurones were chosen so that the combination of at least two of the three features should be present in the correct locations to make the neurone fire.

For each of the other databases, each image was propagated through the network. The detection rate and the false detection number were then determined for each detection map.

4.2.2. Testing results

The results obtained are shown in Table 1. Results with database 1 indicate that the faces were accurately located more than 95% of the time, with very few false detections (an estimation of the rate of wrongly firing neurones would be lower than 0.001%). The detection rates of the feature-selective neurones are also very high, whereas the number of false detection is still low, when compared with the number of images (135 in this database) and the number of feature-selective neurones per image (10304 for each feature).

Results with database 2 can be considered as minimal detection results since the faces were particularly difficult to detect: 88% of the people in the database wore glasses, which probably explains the decrease in the detection rate for the eyes, while 31% had a beard, which probably led to a decrease in the mouth detection results. Thus, the 73.1% detection rate can probably be considered as the models minimal performance.

This is corroborated by the results obtained with database 3, which demonstrate a surprisingly good ability to generalise to novel faces, since the face detection rate is approximately the same as that obtained with the learning base. It is worth noting that the image contrast and illumination conditions were very different in these two databases, as can be seen from the marked decrease in the left eye detection rate with database 3. Nevertheless, such differences in image quality do not seem to disrupt the performance of our face detection model. Furthermore, the detection rate was not much influenced by the substantial numbers of people in the database that wore glasses (50%) or had a beard (16.6%).

Finally, results on database 4 clearly demonstrate the specificity of the network responses:

when no face is present in an image, the probability of a false feature detection at any particular location is very low ($\sim 0.002\%$), while the probability of a false face detection is virtually equal to zero.

Thus, the specificity of the simulated face-selective neurones responses can be summarised as follow:

- when a face is present in such a neurones receptive field, the probability of a response is $\sim 95\%$.
- when no face is present in this receptive field, the probability of a response is virtually zero.

5. Discussion

Using a very simple four-layer feed-forward neural network model, we have been able to show that only one spike per neurone is sufficient to perform quite complex processing tasks such as face detection and localisation in natural images.

Furthermore, the results obtained appear to be remarkably good by comparison with more classic image processing techniques. Principal components analysis (Turk and Pentland, 1991), which seems to be one of the most widely used methods for face detection (Valentin et al., 1994), does not lead to significantly better results than ours. For example, Sung and Poggio (1994) and Moghaddam and Pentland (1995) both achieved a 90% detection rate, with a false detection rate that was not significantly better than the one obtained here.

From a purely practical point of view, modelling visual processing with SpikeNET has advantages over more classical methods that stem from its computational efficiency. One of its main features is that it is 'event-driven'—the main task of the simulator is simply to propagate spikes. As a result, if there are no spikes in a particular layer, then there is no computation to do. This is particularly important in multilayer feed-forward networks where higher levels in the hierarchy are not involved at all until relatively late in the propagation.

In addition, once a spike has been emitted in our model's first layer, it is immediately processed by the following layer, whose neurones can fire spikes in their turn, and so on. A given layer does not need to wait until all the preceding layer's neurones activities have been calculated before it can start computing. That is a very original feature of the asynchronous propagation used here and which distinguishes it radically from classical neural networks.

Furthermore, the thresholding mechanism provides another way of minimising computation time. With relatively low thresholds, a neurone can fire when only a relatively small fraction of its inputs have fired (on condition, of course, that the ones that do fire early have high weights). Thus face-selective neurones in the last processing layer can sometimes emit a spike when only 20% or less of the cells in layer 1 have fired. Of course, this will only be possible when the face in the image is particularly clear.

Since the first spikes occurring in any given layer correspond to the neurones that were first to reach their threshold, the asynchronous mode of transmission used here guarantees that the most salient information is computed first.

All these dynamic features, which are based on the observation of information propagation in the visual system, mean that our model tends to be much faster than classical neural networks.

For instance Rowley et al. (1998) describe a neural network model for localising faces in an image, with a detection rate of roughly 90% and a false detection rate of $\sim 0.0002\%$, that took roughly 6 min on a Silicon Graphics Indigo workstation for an image 320*240 pixels. Running on a simple PowerPC 750 at 266 MHz, the network presented here was able to localise the faces in an image with a processing time ranging from 1 to 5 s (depending on the number of firing neurones and the size of the image).

The simple four-level architecture used here is certainly a very poor description of the real visual system, and we certainly would not wish to claim that the perception of faces in humans and monkeys can be realistically modelled in such a primitive way. It is clear that the primate visual system involves extensive feed-back connections at virtually every stage as well as a large number of horizontal connections. We are currently exploring the situations in which such connections could play a role.

One major simplification of the present model is that we assume that both the activation levels and the sensitivities of all the neurones in the network are reset before each stimulus is presented. Clearly, this is not a realistic assumption. Allowing the activation values of neurones in the network to start from random values would certainly increase noise in the rank order code. However, as we have shown elsewhere (Thorpe and Gautrais, 1998), the rank-order coding scheme is remarkably selective and the probability of a neurone responding to a random sequence of inputs can be made very low. Furthermore, it may be that under natural viewing conditions, the suppression of thalamic transmission that occurs during every saccade may mean that there is indeed a form of reset every time a saccade occurs.

Nevertheless, the main claim that we would wish to make on the basis of these simulations is that visual processing based on one spike per neurone is indeed a possibility. In real visual systems, it is obvious that neurones do, in fact, typically generate trains of spikes in response to a given stimulus. As a result, it would be difficult if not impossible to exclude a role for rate coded information. However, in the sort of simulated visual system that can be explored using SpikeNET, restricting each neurone to one spike and one spike only appears perfectly feasible. Under those conditions it becomes possible to demonstrate that rate coding is not required for visual processing.

References

Abbott, L.F., Rolls, E.T., Tovee, M.J., 1996. Representational capacity of face coding in monkeys. Cereb. Cortex 6, 498–505.

Bruce, C.J., Desimone, R., Gross, C.G., 1981. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J. Neurophysiol. 46, 369–384.

Gautrais, J., Thorpe, S.J., 1998. Rate coding vs. temporal order coding: a theoretical approach. Biosystems.

Jeffreys, D.A., 1996. Evoked potential studies of face and object processing. Vis. Cogn. 3, 1–38.

Moghaddam, B., Pentland, A., 1995. Probabilistic visual learning for object detection. In: The Fifth International Conference on Computer Vision. Cambridge, MA.

Nowak, L.G., Bullier, J., 1997. The timing of information transfer in the visual system. In: Kaas, J., Rocklund, K., Peters, A. (Eds.), Extrastriate Cortex in Primates. Plenum, New York (sous presse).

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.

Perrett, D.I., Rolls, E.T., Caan, W., 1982. Visual neurons responsive to faces in the monkey temporal cortex. Exp. Brain Res. 47, 329–342.

Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. 20, 23–38.

Sung, K., Poggio, T., 1994. Example-based learning for view-based human face detection. Proceedings Image Understanding Workshop, II, 843–850.

Thorpe, S.J., 1990. Spike arrival times: A highly efficient coding scheme for neural networks. In: Eckmiller, R., Hartman, G., Hauske, G. (Eds.), Parallel Processing in Neural Systems. Elsevier, Amsterdam, pp. 91–94.

Thorpe, S.J., Gautrais, J., 1997. Rapid visual processing using spike asynchrony. In: Mozer, M.C., Jordan, M.I., Petsche, T. (Eds.), Neural Information Processing Systems, vol. 9. MIT Press, Cambridge, pp. 901–907.

Thorpe, S.J., Gautrais, J., 1998. Rank order coding: A new coding scheme for rapid processing in neural networks. In: Bower, J. (Ed.), Computational Neuroscience: Trends in Research. Plenum, New York.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cogn. Neurosci. 3, 71–86.

Valentin, D., Abdi, H., O'Toole, A., Cottrell, G.W., 1994. Connexionnist models of face processing: a survey. Pattern Recogn. 27, 1209–1230.

4.2 Détection de visages à plusieurs tailles

Les capacités d'un système de traitement de l'information asynchrone feed-forward tel que celui utilisé dans l'article précédent ne se limitent pas à la détection des visages. Tout dépend en fait de l'architecture choisie, mais de nombreuses autres fonctions de reconnaissance visuelle peuvent être implémentées par un tel type de modèle. Par exemple, en raffinant le réseau précédent, il est possible de simuler la détection des visages de différentes tailles. (Ce travail a été présenté sous forme orale au congrès Neurosciences et Sciences de l'Ingénieur, Munster, 1998).

L'architecture choisie pour ce modèle est volontairement simple. En effet, nous avons pu vérifier dans l'article précédent qu'une organisation extrêmement simple pouvait mener à d'excellents résultats, ceci impliquant que c'est dans le codage utilisé (plutôt que dans la complexité du modèle) que réside la puissance du traitement effectué.

Le réseau est constitué de 3 couches successives. A l'intérieur de chaque couche, une carte de neurones organisée de façon rétinotopique traite l'information pour chaque échelle. Une couche est donc composée d'une ou plusieurs "pyramides" de cartes.

Les pyramides de la première couche sont composées de neurones sensibles aux contrastes positifs (cellules ON-Center) ou négatifs (cellules OFF-Center). Les cartes d'une même pyramide codent les contrastes d'une même polarité (ON ou OFF), mais à différentes échelles. Ainsi, tout comme dans la rétine, l'information est décomposée non seulement en termes de contrastes positifs ou négatifs, mais également en termes de fréquences spatiales.

Dans chaque pyramide de la deuxième couche, les neurones répondent sélectivement à une certaine orientation (8 orientations séparées de 45 °). Dans chaque carte d'une de ces pyramides, les neurones répondent à un bord de l'orientation correspondante, mais pour une fréquence spatiale donnée. Ainsi ce deuxième niveau peut être considéré comme un modèle rudimentaire des colonnes corticales de l'aire visuelle primaire V1.

Enfin les neurones du troisième niveau sont entraînés à répondre au pattern d'activité qui est spécifique de la présence d'un visage. En fonction de l'ensemble de cartes de la couche II où ce pattern aura été détecté (c'est-à-dire en fonction de sa fréquence spatiale), c'est dans l'une ou l'autre des différentes cartes de sortie que la réponse aura lieu. La localisation de la réponse dans la carte (qui est organisée de manière rétinotopique) donne l'emplacement précis du visage dans l'image, tandis que la localisation de la carte dans la pyramide indique la taille du visage.

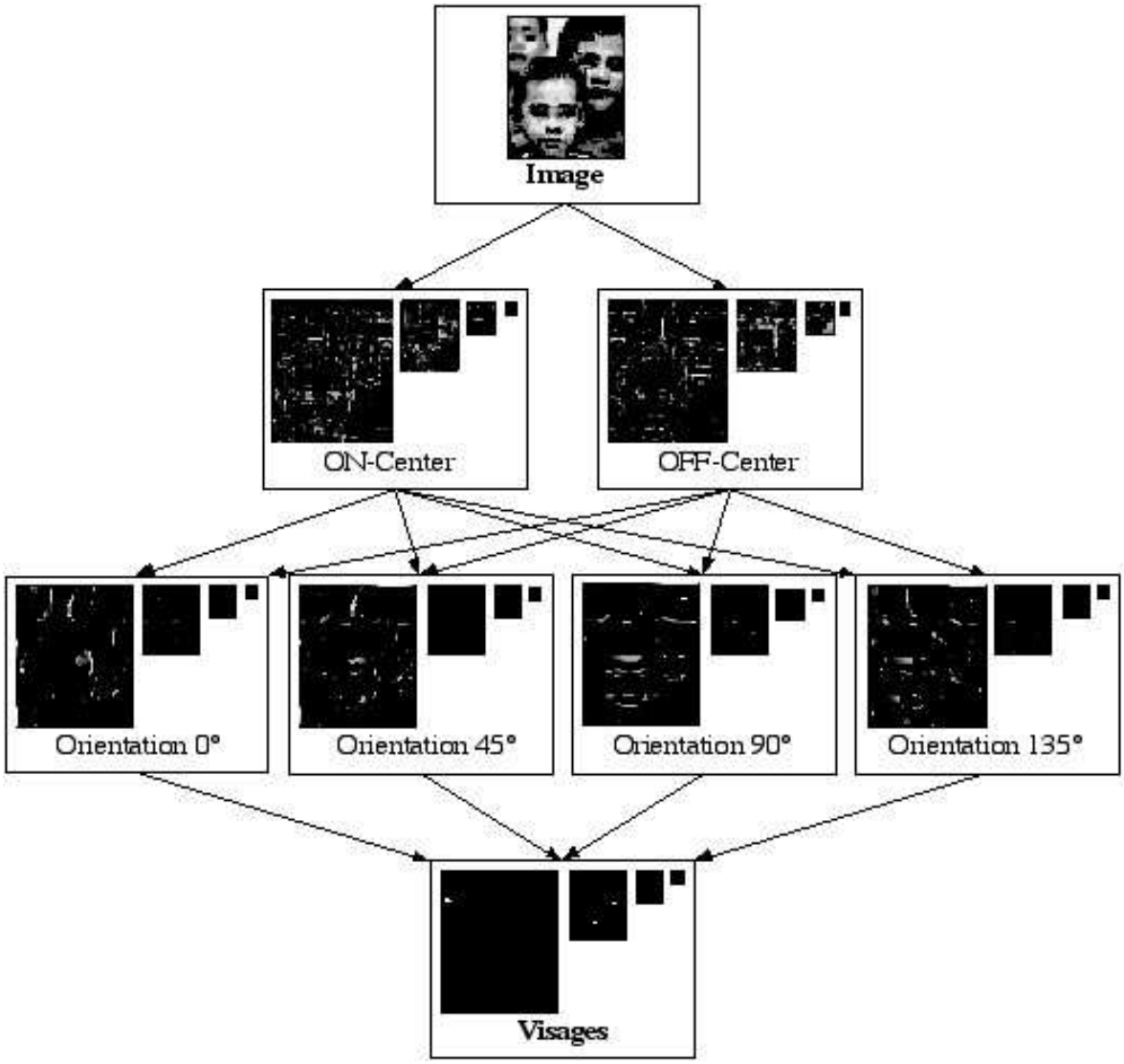


Figure 34. Architecture du modèle. Chaque couche est composée d'un ensemble de pyramides de cartes. Chaque carte y représente une fréquence spatiale donnée. Les neurones de la première couche répondent à des contrastes positifs (cellules ON-Center) ou négatifs (cellules OFF-Center), tandis que ceux de la deuxième couche sont sélectifs à un bord d'une certaine orientation (par souci de clarté, seulement 4 orientations sur 8 sont présentées ici). Les neurones de la dernière couche sont entraînés à répondre au pattern d'activité caractéristique d'un visage. L'emplacement du visage est donné précisément par l'emplacement de la décharge à l'intérieur de la carte, sa taille est déterminée par l'emplacement de la carte dans la pyramide.

Poids synaptiques et apprentissage.

Au niveau de la première couche, l'information analogique (intensité dans l'image) est transformée en un ordre de décharges sur l'ensemble des photorécepteurs. C'est cet

ordre de décharges qui est décodé par les neurones de la deuxième couche, grâce à un mécanisme de désensibilisation permettant de le comparer avec l'ordre des poids synaptiques (les champs récepteurs de ces neurones sélectifs à l'orientation sont obtenus par une fonction de Gabor).

Une carte de la première couche n'est connectée qu'aux cartes du même canal de fréquence spatiale (i.e. à la même position dans la pyramide des échelles) dans la couche suivante. Le pattern de connexions entre deux cartes est *le même quelle que soit l'échelle correspondante*: ainsi les neurones de la première carte d'une pyramide reconnaissent une certaine orientation à l'échelle la plus fine. *Avec les mêmes connexions*, les neurones de la deuxième carte de cette pyramide reconnaîtront des bords de la même orientation, à une échelle plus grossière.

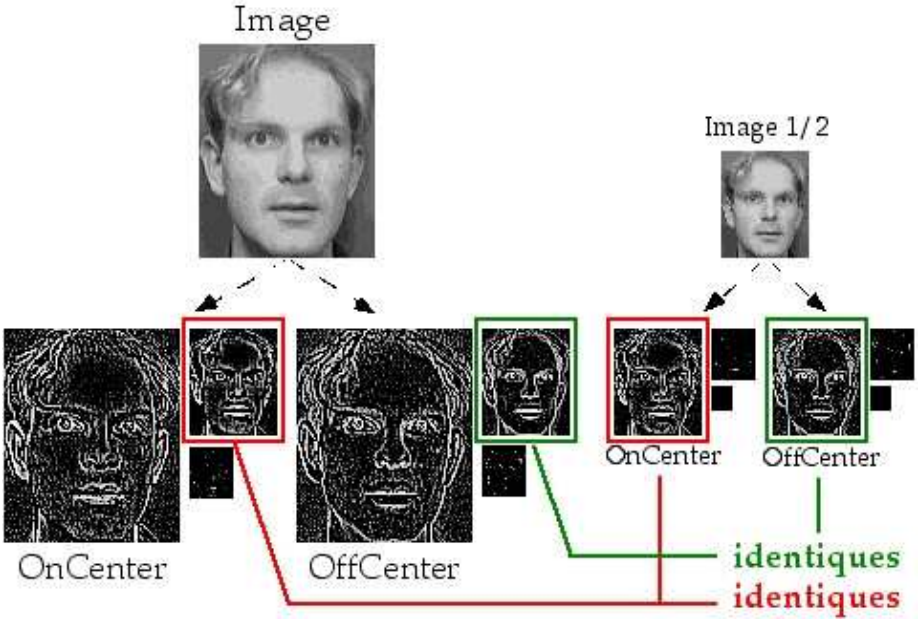


Figure 35. Une seule taille de visages est apprise, et ses connexions sont généralisées aux autres échelles. En effet, le pattern d'activité à l'échelle la plus fine pour une image d'entrée de taille (arbitraire) ½ est identique au pattern d'activité à l'échelle immédiatement plus basse, pour une image d'entrée de taille 1. Un neurone recherchant ce pattern dans les cartes neuronales à l'échelle fine détectera donc des visages de taille ½, alors qu'un neurone recherchant ce même pattern dans les cartes à l'échelle intermédiaire détectera des visages de taille 1.

Pour entraîner les neurones de la dernière couche à répondre sélectivement à la présence d'un visage dans leur champ récepteur, nous avons fait appel à un mécanisme d'apprentissage supervisé. 400 visages d'une taille donnée (la plus petite taille de visage

que le réseau aura à détecter, soit environ 25x25 pixels) ont été présentés au réseau et nous avons récupéré l'ordre des décharges pour chaque image dans les cartes d'orientation à l'échelle la plus fine. A ces ordres ont été associées des valeurs entières (de 0 pour les derniers jusqu'à 255 pour les premiers) qui ont été moyennées sur l'ensemble des images, et sont devenues les poids des connexions reliant les neurones de la couche II à ceux de la couche III. Il est important de noter que, là encore, le pattern de connexions entre 2 cartes situées au même niveau dans leurs pyramides respectives (i.e. une carte d'orientation et une carte de détection des visages à l'échelle correspondante) est *le même, et ce quelle que soit l'échelle correspondante*. En effet, lorsqu'un visage *de petite taille* est présenté au réseau, le pattern d'ordre moyen obtenu précédemment est reconnu dans les cartes d'orientation à l'échelle *la plus fine*. Le neurone correspondant à cette position dans la carte de détection des visages à l'échelle la plus fine sera donc activé. Si c'est un visage *de grande taille* qui est présenté, c'est dans les cartes d'orientation à l'échelle *la plus grossière* que ce pattern moyen sera reconnu, et un neurone de la carte de détection des visages à l'échelle la plus grossière sera activé. Le mécanisme d'apprentissage nous garantit donc qu'un neurone de la dernière couche sera sélectif à la présence d'un visage *d'une taille donnée* à *une position donnée* dans le champ visuel (figure 35).

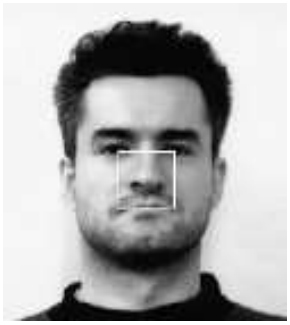
Les patterns de connexions résultant de cet apprentissage sont présentés figure 36. Notons que l'architecture choisie ici est sensiblement plus simple que celle utilisée dans l'article précédent: les neurones détecteurs de visage sont connectés directement à la couche II, sensible à l'orientation, sans passer par une étape intermédiaire de détection d'éléments constitutifs du visage tels que la bouche ou les yeux.

Figure 36. Les patterns de connexions reliant la rétine aux cartes de détection d'orientations (en haut; ici seules les connexions provenant des cellules ON-Center sont représentées), et les patterns de connexions résultant de l'apprentissage, reliant les cartes d'orientation aux cartes de détection de visages (en bas). Ces mêmes connexions sont reproduites dans chaque canal de fréquence spatiale, ce qui permet de détecter des visages de plusieurs tailles.

Résultats

Le réseau obtenu a été testé sur deux ensembles d'images. La base de test 1 est constituée de 1500 visages, i.e. 300 visages différents à 5 tailles possibles: 1 fois, 1.5 fois, 2 fois, 3 fois et 4 fois la taille des visages appris). Notons que les visages de cette base ne sont pas ceux qui ont été appris. La base d'apprentissage et la base de test 1 proviennent de 2 sources différentes, et contiennent donc des photographies de personnes différentes, prises dans des conditions d'illumination différentes. Un visage est détecté si l'un au moins des neurones de la couche de détection des visages décharge dans une zone restreinte au centre du visage (voir figure 37A). Le taux de détection du modèle construit est de 90.13%, et le taux de détection à l'échelle appropriée (avec une tolérance d'un-demi octave) est de 89.53%. Le taux d'erreur, défini comme le nombre de neurones déchargeant en-dehors du centre du visage, sur le nombre total de neurones dans la dernière couche, est de 7.5 10^{-5}. La plupart de ces décharges érronées dépendent néammoins de la présence du visage, comme l'indiquent les résultats obtenus sur la base de test 2. Cette base est constituée de 100 images naturelles, de taille 100x100 pixels, et ne contenant aucun visage (voir les exemples de la figure 36B). Sur cette base de test, seulement 5 erreurs ont été observées, correspondant à un taux de décharge (erronée) de 7.6 10^{-6}.

A.



B.

Figure 37. A. Un exemple de visage utilisé dans la base de test 1, avec le centre du visage indiqué par le cadre blanc. Les décharges neuronales à l'intérieur de ce cadre sont considérées comme des détections correctes. **B.** 3 exemples d'images naturelles de la base de test 2, utilisée pour déterminer le taux d'erreur du système lorsqu'aucun visage n'est présent dans l'image.

Notons pour conclure que ces résultats sont bien meilleurs que ceux obtenus par la plupart des méthodes classiques d'analyse d'images (énumérées dans l'article 6; voir aussi Penev et Atick, 1996; Romdhani, 1996). Ces modèles obtiennent typiquement des taux de détection identiques (souvent pour une seule taille de visages détectée), mais invariablement des taux de fausses détection supérieurs, jusqu'à 1000 fois plus dans certains cas. De même, la durée de traitement nécessitée par ces méthodes est au minimum 2 fois plus élevée, et souvent plus longue d'1 à 2 ordres de magnitude. Cette différence d'efficacité computationnelle peut certainement s'expliquer par la simplicité de notre modèle, mais également par son inspiration biologique. Si le cerveau peut parvenir à une représentation explicite de la catégorie de l'objet présenté en seulement 150 ms, c'est probablement qu'il utilise les mécanismes de transfert d'information les plus efficaces, et ce type de propagation feed-forward d'une vague d'activité asynchrone semble un excellent candidat à ce titre.

4.3 Une représentation de haut-niveau de l'entrée visuelle

Le chapitre II a permis de démontrer que le système visuel humain est capable, en seulement 150 ms, de transformer l'image rétinienne de la scène visuelle en une représentation de haut niveau, abstraite et non ambigüe, encodant explicitement la catégorie ou l'identité des objets présentés, et servant probablement de base à la génération d'une décision et d'une réponse comportementale. Cette observation a soulevé la question des mécanismes neuronaux impliqués dans un tel type de traitement ultra-rapide. Les modèles classiques ne peuvent en effet rendre compte d'une telle rapidité.

La théorie proposée dans ce chapitre s'avère un moyen simple et efficace d'expliquer cette remarquable capacité. La réponse des neurones "détecteurs de visages" présentés aux sections précédentes constitue en effet une représentation de haut niveau de l'entrée visuelle. Ces neurones ne déchargent que si un visage est présent dans leur champ récepteur. Leur activité encode donc explicitement et de façon non ambigüe la catégorie de l'objet présenté, et pourrait servir de base à la réponse comportementale, par exemple lors d'une tâche de catégorisation rapide "visage/non-visage". L'attention portée sélectivement sur les neurones codant pour les éléments constitutifs du visage, ou directement sur les neurones sélectifs aux visages, serait un moyen supplémentaire d'augmenter la sensibilité du système, lui permettant de répondre à un visage si cette catégorie est relevante à la tâche réalisée, ou de supprimer cette réponse si la tâche implique une cible qui n'est pas un visage. Enfin, l'asynchronie sur

laquelle repose le codage ainsi que le transfert de l'information explique comment un tel système peut réaliser cette fonction visuelle de haut niveau en seulement 150 ms, avec pas plus d'un spike par neurone, et une transmission d'information principalement feed-forward.

Ainsi, l'activité des neurones situés dans les niveaux supérieurs de la hiérarchie visuelle d'un tel système de traitement asynchrone pourrait bien correspondre à l'activité de haut niveau que nous avons enregistrée après 150 ms dans le système visuel humain.

IV. SYNTHESE ET PERSPECTIVES

It would be no surprise if, in a biased competition framework, the race for neural resources was first won in the temporal domain.

Dès que nous ouvrons les yeux ou que nous déplaçons le regard, lorsque nous changeons de chaîne de télévision ou qu'un angle de caméra bascule soudainement, lorsque nous ouvrons une porte ou allumons la lumière dans une pièce obscure, une toute nouvelle scène visuelle se présente à nos yeux. Cette situation se répète probablement plusieurs centaines, voire plusieurs milliers de fois par jour. Et à chaque fois, nous parvenons presque instantanément à interpréter ce nouvel environnement visuel, à en extraire les informations utiles pour les actions dans lesquelles nous sommes engagés. Cette capacité est si naturelle qu'elle ne nous surprend jamais. D'un point de vue computationnel, pourtant, une telle compétence est tout-à-fait remarquable. Même les ordinateurs les plus sophistiqués ne peuvent prétendre aujourd'hui égaler les performances du système visuel humain. Et jusqu'à nos jours, les nombreuses tentatives de systèmes artificiels qui ont cherché à s'inspirer des propriétés biologiques de notre cerveau sont malheureusement restées inefficaces. Quel est donc ce secret qui fait du système visuel des primates l'une des technologies les plus avancées de notre époque?

S'il ne s'agit ni de magie, ni de forces surnaturelles, alors c'est peut-être dans l'asynchronie des décharges neuronales que réside la solution de ce problème fondamental. En termes psychophysiques comme en termes électrophysiologiques, il n'est pas de propriété des systèmes visuels biologiques qui ne s'explique dans ce contexte. De même, un système neuronal qui encode l'information par une vague de potentiels d'action asynchrone se propageant à travers ses différents niveaux, possède les principaux avantages de la plupart des modèles de la vision proposés durant ces dernières 50 années. Bien souvent, l'implémentation de ces propriétés se voit même grandement simplifiée par l'utilisation de l'asynchronie.

Par exemple, le modèle de reconnaissance d'objets invariante à certaines transformations (changements de taille, d'angle de vue, etc□) de Sandon (1990), qui possède également une composante attentionnelle, utilise une compétition entre stimuli, implémentée par des inhibitions latérales fortes. Le modèle de Riesenhuber et Poggio

(1999a,b) réalise ce type de reconnaissance d'objets invariante aux transformations en faisant appel à un mécanisme hautement non-linéaire: l'opération MAX (qui extrait le maximum des réponses sur une population neuronale), qui est approximée par l'implémentation d'un algorithme "softmax" (Nowlan et Sejnowski, 1995). Cette stratégie permet, entre autres, de router l'information en fonction de sa pertinence. Les auteurs suggèrent que l'attention visuelle pourrait agir directement sur ce mécanisme. De même, Lee et al (1999) démontrent, sur la base de données expérimentales et théoriques, que l'attention pourrait agir spécifiquement pour augmenter un mécanisme non-linéaire de type "Winner-Take-All".

Le codage par l'asynchronie des décharges est en lui-même un mécanisme non linéaire, similaire à une opération de type softmax: le maximum d'activité sur une population est simplement représenté par la première décharge de cette population. La désensibilisation du neurone post-synaptique permet de traiter sélectivement ce maximum de façon implicite, sans avoir recours à des inhibitions latérales fortes, ou même à un niveau de traitement supplémentaire dédié spécifiquement à la réalisation de la fonction MAX. Selon notre hypothèse, c'est bien directement sur ce mécanisme non-linéaire, l'asynchronie des décharges, qu'agirait l'attention, en donnant une précédence temporelle à l'information la plus saillante, ou à l'information d'intérêt dans le cas d'un processus attentionnel top-down. Ceci revient en fait à augmenter l'intensité d'un mécanisme non-linéaire de type MAX ou Winner-Take-All.

Les modèles de type "shifter circuits" ou de réalignement (Anderson et Van Essen, 1987; Olshausen et al, 1993) suggèrent que l'invariance à la taille, la position ou l'angle de vue lors de la reconnaissance d'objets s'obtient en réalignant une fenêtre de traitement sur l'information d'intérêt (dans le cortex visuel primaire V1). Ce réalignement s'effectue sous le contrôle d'un ensemble de neurones dédiés à cette finalité, qui pourraient se situer dans le pulvinar. En mode "préattentionnel", une carte de saillance guide ces neurones vers l'information la plus pertinente, alors qu'en mode "attentionnel", c'est un signal "top-down" qui vient moduler l'action de ces neurones de contrôle. L'avantage d'un système utilisant une propagation asynchrone dans le cadre d'un tel modèle est immense: l'information de saillance étant encodée dans la structure temporelle des décharges d'une population permet de router l'information de manière implicite, sans faire appel à des neurones de contrôle ou à une carte de saillance. De même, l'attention "top-down" vient biaiser directement cette structure temporelle, et aucun neurone supplémentaire n'est nécessaire à cette fin. Cette stratégie découle en fait simplement de la propriété mise en évidence récemment par Reynolds et al (1996; voir Reynolds et Desimone, 1999): l'attention agit comme une augmentation du gain (de contraste) des cellules touchées. Un modèle utilisant cette propriété a démontré que le champ récepteur des neurones touchés par l'attention se réaligne avec la fenêtre

d'intérêt, permettant un routage sélectif de l'information, similaire à celui proposé dans les modèles de "shifter circuits" (Salinas et Abbott, 1997). Or, dans le cadre théorique que nous avons présenté au chapitre précédent, l'attention agit bien comme une augmentation du contraste d'entrée des stimuli, et permet donc également de rendre compte de ce routage dynamique, d'une façon simple et efficace.

En réalité, aucune des caractéristiques du modèle que nous proposons ici n'est foncièrement novatrice, si elle est considérée de façon isolée. L'idée que les traitements dits "attentionnel" et "préattentionnel" pourraient agir sur les mêmes mécanismes de base (idée dérivant de la théorie de "biased competition"; Desimone et Duncan, 1995) a déjà été avancée, par exemple par Borisyuk et al (1998) qui proposent que ce mécanisme commun pourrait résider dans l'activité synchrone des populations. La correspondance entre latences de décharge et saillance des stimuli a récemment été utilisée dans un modèle hiérarchique du cortex visuel, par Nakamura (1998): les premières décharges d'une population inhibent les neurones voisins par un jeu de connexions latérales, et l'information qui parvient aux derniers niveaux du système est par construction la plus saillante. Notre modèle ne requiert pas ces connexions latérales, car la compétition entre représentations est effectuée par un mécanisme de désensibilisation post-synaptique. Ce dernier mécanisme lui-même, associé à l'asynchronie de décharge, correspond en fait à une opération non-linéaire de type MAX, softmax ou Winner-Take-All; ainsi notre modèle exhibe donc également les propriétés des modèles utilisant ce type de mécanisme (Riesenhuber et Poggio, 1999a,b; Lee et al, 1999). Enfin, une relation causale entre synchronie (ou oscillations) et asynchronie a déjà été proposée dans le cadre d'un modèle du traitement visuel, mais cette relation causale était dirigée dans le sens opposé: l'asynchronie des décharges en provenance de la rétine permettrait de séparer les informations concernant les régions de même contraste, afin de les répartir dans différents cycles d'oscillations (Wörgötter et al, 1996). L'originalité de la théorie proposée ici provient probablement du regroupement, dans un contexte computationnel unique, d'une diversité d'approches souvent supposées être mutuellement exclusives.

1. Une théorie du traitement visuel rapide

Sur la base des observations et des études faites aux chapitres précédents, une tentative de théorie du traitement visuel rapide, basée sur la propagation d'une vague de spikes asynchrone, est ici proposée.

1. Chaque événement transient dans la stimulation visuelle (e.g. saccade) induit dans la rétine la génération d'une réponse sous la forme de trains de potentiels d'action. La prise en compte de la première décharge de chaque cellule ganglionnaire permet de considérer cette réponse comme une vague de spikes sur l'ensemble de la population. A l'intérieur de cette vague, les premiers spikes portent typiquement l'information la plus saillante, i.e. concernant les neurones les plus activés, et donc les régions les plus contrastées.
2. Cette vague de décharges asynchrone est suffisante pour activer sélectivement les neurones des étapes de traitement suivantes. Ces neurones répondent préférentiellement si la structure spatio-temporelle des spikes leur parvenant correspond à la sélectivité de leur champ récepteur. Cette structure étant fixée, ils répondent également plus ou moins tôt selon la date de décharge "moyenne" de leurs afférents. Au niveau de la population, la réponse au stimulus peut donc être considérée comme une nouvelle vague de potentiels d'action, à l'intérieur de laquelle l'asynchronie porte l'information: les premières décharges de cette vague correspondent *(i)* aux neurones les plus activés de la population, donc à l'information la plus saillante selon le critère de cette étape de traitement, et *(ii)* aux neurones touchés le plus tôt, donc à l'information la plus saillante selon les critères des étapes de traitement précédentes. Les points *(i)* et *(ii)* ne sont bien sûr pas mutuellement exclusifs.
3. A chaque étape, des mécanismes d'interactions latérales (ou éventuellement "en arrière") asynchrones peuvent venir compléter ce type de propagation "en avant", pour renforcer, filtrer ou nettoyer sélectivement le signal. Ces interactions agissant sur le premier spike de chaque réponse neuronale, peuvent modifier la structure temporelle de notre vague de potentiels d'action, sans remettre en question l'idée d'une propagation "en avant" de cette vague asynchrone.
4. Au niveau du système global, le mécanisme décrit aux points précédents se reproduit en cascade. Une vague de potentiels d'action en provenance de la rétine traverse donc le système. Cette vague est régénérée à chaque niveau de traitement: l'information des niveaux précédents est conservée; l'information extraite par les neurones de cette étape vient s'y ajouter, et modifier la structure temporelle de cette vague. L'information la plus saillante est toujours représentée par les premières décharges de la vague. C'est en fait l'idée de saillance elle-même qui se transforme au cours de la propagation, définie simplement par le contraste de luminance dans la rétine, auquel viendra par exemple s'ajouter l'orientation dans V1, puis différentes propriétés, pouvant aller jusqu'à la pertinence comportementale dans des aires de haut niveau

comme le cortex inféro-temporal ou le cortex préfrontal. Il en découle qu'aucune "saliency map", encodant explicitement les régions de l'espace les plus saillantes, n'est nécessaire, puisque cette information est déjà présente dans la structure spatio-temporelle de notre vague de spikes asynchrone.

5. L'attention agit comme un biais temporel sur la propagation de cette vague. En provoquant une réponse plus précoce chez les neurones qui représentent explicitement l'information attendue (par exemple, ceux dont le champ récepteur se situe dans la fenêtre d'attention spatiale), ce mécanisme attentionnel permet d'augmenter la saillance relative de cette information pour les étapes de traitement suivantes.
6. Dans les niveaux les plus élevés du système, comme le cortex inféro-temporal, un neurone sélectif à un type d'objet particulier s'activera normalement si cet objet est présent dans la scène visuelle, c'est-à-dire si les traits caractéristiques de cet objet ont été détectés dans les niveaux précédents. La probabilité d'activation de ce neurone sera tout particulièrement augmentée si l'objet ou les traits le composant sont relativement saillants, ou si, du fait de la tâche en cours, l'attention est portée sélectivement sur cet objet, ou sur les traits le composant.
7. Ce processus d'organisation temporelle en une succession de vagues de spikes asynchrones, décrit ici dans le cas où un événement transient déclenche la formation d'une telle vague dans la rétine, pourrait également agir dans des conditions plus générales de stimulation visuelle. En effet, certains mécanismes comme les oscillations sous-liminaires de l'activité corticale, ou la présence d'afférences rythmiques dans des structures thalamiques telles que le LGN, permettraient de modeler la structure des réponses neuronales à différents niveaux, n'autorisant les populations à répondre que sur certaines fenêtres temporelles courtes. Cette régulation "interne" donnerait lieu à la génération et la propagation de vagues de spikes asynchrones, similaires aux vagues déclenchées dans la rétine par des transients visuels, avec l'ensemble des propriétés qui en découlent, énumérées dans les points précédents.

Il semble nécessaire de préciser que le type de traitement visuel qui pourra être effectué selon ce schéma n'englobe évidemment pas l'ensemble des capacités du système visuel des primates. Notamment, l'existence des voies "feed-back" et des traitements "top-down" est indiscutable, et il est vraisemblable que ces mécanismes jouent un rôle primordial pour nos capacités visuelles. Nous pouvons donc émettre l'hypothèse que le type de traitement visuel rapide que nous venons de décrire dans

cette thèse pourrait en quelque sorte servir de premier "allumage" ou "amorçage" pour la mise en jeu de mécanismes plus tardifs, et plus coûteux en temps de traitement.

2. Questions en suspens

La théorie proposée dans cette thèse n'est à ce stade qu'une simple hypothèse, même si elle rassemble en sa faveur une quantité d'observations expérimentales et théoriques. Certaines questions nécessitent cependant une réponse directe avant de pouvoir affirmer que cette hypothèse est valide. En particulier, le dernier point de la section précédente mérite un intérêt tout particulier.

En effet, l'idée d'un codage par la structure temporelle d'une vague de potentiels d'action implique nécessairement une remise à zéro du système, permettant de ne pas confondre les derniers spikes d'une vague donnée avec les premiers de la suivante. Ceci est particulièrement critique lorsque les premiers spikes portent l'information la plus pertinente. Si l'on peut considérer que les saccades et micro-saccades visuelles (Martinez-Conde et al, 2000) pourraient participer à un tel mécanisme, de sorte que l'analyse de la scène recommence à chaque saccade, il n'en reste pas moins que la saccade ou microsaccade doit toujours s'accompagner d'un signal pour la remise à zéro. Un tel signal pourrait avoir été observé dans le corps genouillé latéral du chat. Lal et Friedlander (1989; 1990a,b) ont enregistré les réponses des neurones du LGN du chat à une stimulation visuelle, à différents moments avant ou après l'induction d'un mouvement oculaire passif. Ces réponses se trouvent modulées par le mouvement oculaire, ainsi que par la position de l'œil dans l'orbite. Cette modulation facilite la transmission immédiatement après le mouvement oculaire, et l'inhibe lorsque l'œil se fixe à une nouvelle position. Ceci pourrait expliquer pourquoi la perception visuelle disparaît purement et simplement en l'absence de mouvements de l'image rétinienne (Riggs et Ratcliff, 1952; Ditchburn et Ginsborg, 1952). Un tel signal de remise à zéro, qui prendrait la forme d'une oscillation rythmique dans le LGN, pourrait être généré de façon intrinsèque, ou par des connexions en provenance du cortex (McClurkin et al, 1994; Contreras et al, 1996; Castelo-Branco et al, 1998). De la même façon, les oscillations de l'activité corticale, qu'elles soient induites par des connexions en provenance du LGN (Ghose et Freeman, 1997) ou par des mécanismes corticaux intrinsèques, pourraient être impliquées dans un remodelage actif de la structure temporelle des réponses neuronales (Nowak et al, 1997; Volgushev et al, 1998). Par exemple, l'intervalle de temps nécessaire entre 2 flashs lumineux pour qu'ils soient perçus comme successifs dépend fortement de la phase relative au rythme cortical (alpha) à laquelle ces stimuli sont présentés (Varela et al, 1981; Gho et Varela, 1988). Parodi et al (1996) ont postulé

que dans le cadre d'un codage de l'information de mouvement par les dates d'arrivée des spikes, la nécessaire remise-à-zéro pourrait être à l'origine des oscillations corticales observées dans la bande de fréquences Gamma. Ainsi cette remise à zéro n'aurait pas uniquement lieu dans le LGN, de façon dépendante des saccades ou microsaccades oculaires, mais pourrait également prendre place au niveau de chaque population neuronale, afin de regrouper dans une fenêtre temporelle adéquate les décharges de la même vague, et de séparer efficacement 2 vagues de potentiels d'action successives. Comme proposé au point 7 de la section précédente, ce processus d'organisation temporelle pourrait être déclenché de façon interne, indépendamment des transients de la stimulation visuelle, ou des saccades et mouvements oculaires.

Bien que les données expérimentales sur la durée du traitement visuel ne laissent guère d'alternative, il n'est pas encore directement démontré aujourd'hui que la première vague de potentiels d'action générée par la rétine puisse suffir à déclencher une séquence d'évènements corticaux menant à la réalisation d'une tâche de haut niveau. Notamment, le phénomène de masquage "en arrière" (backward masking; Breitmeyer, 1984) jette le doute sur cette hypothèse: lorsqu'un stimulus est suivi d'un autre (le masque) avec un délai très court (de l'ordre de 20 ms), alors le premier n'est tout simplement pas perçu par le sujet. Les 20 premières millisecondes de la stimulation visuelle (grossièrement, la première vague de potentiels d'action) ne seraient donc pas suffisantes pour donner lieu à une perception visuelle consciente. Cependant, des enregistrements unitaires ont montré que dans de telles conditions expérimentales, les neurones du cortex inféro-temporal du singe macaque pouvaient s'activer de manière sélective au stimulus pendant environ 20 ms; ensuite, leur activité est interrompue par l'activité générée par le masque (Rolls et Tovee, 1994; Rolls et al, 1994; Kovacs et al, 1995). La première vague de potentiels d'action serait donc suffisante pour activer sélectivement les neurones aux derniers niveaux de la hiérarchie de la voie corticale ventrale. Cette brève activité pourrait servir de base à la génération de la réponse motrice, comme nous l'avons vu au chapitre II section 2.4.1. Ceci implique (*i*) que la perception consciente d'un stimulus visuel nécessite une activité prolongée des neurones, par exemple dans le cortex inféro-temporal (Libet et al, 1991) et (*ii*) que les tâches de catégorisation visuelle rapide telles que celle utilisée par Thorpe et al (1996) ou celles décrites au chapitre II pourraient, dans une certaine mesure, être effectuées sans (ou avant) la perception consciente des stimuli présentés. La première vague de potentiels d'action générée dans la rétine serait néammoins suffisante pour activer une représentation de haut niveau de la scène visuelle.

Il ne fait aucun doute que, au moins dans la plus grande majorité des cas, les neurones les plus activés d'une population déchargent effectivement les premiers. L'information visuelle est donc bien présente dans la première vague de potentiels

d'action générée dans la rétine, par exemple. Cependant, de même que la présence indiscutable d'oscillations et de corrélations temporelles précises dans l'activité corticale ne constitue pas une preuve du rôle de ces évènements dans la représentation neuronale, le fait que l'information visuelle soit présente dans l'asynchronie des décharges neuronales n'implique pas que cette asynchronie soit effectivement décodée par les neurones du système visuel. Bien que certaines études (e.g. Gawne et al, 1996; Celebrini et al, 1993) aient démontré que la latence de décharge des neurones de V1 pouvait encoder explicitement le contraste ou l'orientation d'un stimulus, il manque encore aujourd'hui une généralisation de ce résultat à d'autres propriétés de la scène visuelle extraites par les populations neuronales. En termes psychophysiques, une amélioration de la performance visuelle pour un stimulus présenté une poignée de millisecondes avant le reste de la scène visuelle, constituerait un indice supplémentaire en faveur de notre hypothèse. En termes électrophysiologiques, une dépendance de la réponse post-synaptique à la séquence précise d'activation des synapses d'un neurone constituerait une preuve indiscutable du rôle de l'asynchronie dans le codage et la représentation de l'information neuronale.

Pour résumer, s'il était possible de prouver indiscutablement que:

1. la première vague de potentiels d'action générée dans la rétine suffit à la réalisation de tâches perceptuelles de haut niveau
2. à l'intérieur d'une telle vague se propageant à travers le système, l'information est représentée par l'asynchronie des décharges, les premières décharges de la vague étant interprétées au niveau suivant comme portant l'information la plus saillante
3. le système visuel est capable de modeler activement la structure temporelle des réponses neuronales, de manière à regrouper ces réponses en vagues asynchrones, et à sous-tendre un éventuel mécanisme de remise à zéro, ceci indépendamment de la structure temporelle de l'entrée visuelle

alors la théorie avancée dans cette thèse se verrait définitivement validée. Ces questions en suspens constitueront sans aucun doute l'objet des prochaines expérimentations ménées en relation avec ce travail de thèse.

3. Conclusion

Lorsqu'une nouvelle image se présente à nos yeux, les cellules ganglionnaires de la rétine répondent sous la forme d'une vague spatio-temporelle de potentiels d'action qui est propagée dans le système. A chacune des différentes étapes de traitement qui jalonnent ce cheminement en direction du cortex inféro-temporal, les populations neuronales extraient, localement, une certaine caractéristique ou propriété physique de l'image d'entrée, à partir de l'information transmise par la ou les étapes inférieures, par les neurones voisins, et par des populations "centrales", de "haut niveau", qui guident l'attention et préparent le système à recevoir l'information sensorielle dans les meilleures conditions, compte-tenu de l'expérience passée ou de la tâche réalisée. Le résultat de ce traitement complexe, non-linéaire, est une nouvelle vague spatio-temporelle, qui à son tour se propage vers les niveaux suivants du système. Cette succession de vagues n'est pas un processus séquentiel. Au contraire, la vague de spikes correspondant à la réponse d'une couche donnée se déclenchera typiquement bien avant la fin de la vague correspondant à la couche précédente (Nowak et Bullier, 1998). Les propriétés extraites le long de la voie ventrale du système visuel sont de plus en plus complexes et abstraites, de sorte que des caractéristiques de haut niveau des objets de la scène visuelle telles que leur identité, leur catégorie, ou leur statut par rapport à une tâche donnée, peuvent être encodées explicitement aux derniers niveaux de cette hiérarchie, dans le cortex inférotemporal. Cette séquence d'évènements, transformant une scène visuelle projetée sur notre rétine en une représentation abstraite des objets la composant, peut être réalisée par le système visuel humain en seulement 150 ms (articles 1 et 2).

L'une des clés de cette extraordinaire efficacité pourrait bien résider dans l'asynchronie des décharges neuronales à l'intérieur d'un tel type de vague spatio-temporelle traversant le système. L'asynchronie de décharge sur une population neuronale permet de coder efficacement l'information visuelle (article 3), et peut être aisément décodée par un neurone efférent. L'information de saillance, qui par définition correspond aux éléments de la scène visuelle qui présentent un intérêt particulier pour le système visuel, ou pour l'animal lui-même, est naturellement représentée dans ce contexte de codage par l'asynchronie: les propriétés les plus saillantes donnent lieu aux décharges les plus précoces. Ce sont donc les éléments les plus saillants qui pourront en priorité moduler les réponses des neurones d'une couche donnée, par le biais d'interactions latérales (article 4). De plus, cette saillance relative des différents éléments de la scène visuelle peut être augmentée ou diminuée sélectivement par l'attention, de façon extrêmement simple (article 5): les neurones codant pour la région (ou la

propriété) d'intérêt sont encouragés à décharger plus tôt, et le système interprète cette région (ou propriété) comme la plus saillante, lui attribuant ainsi des ressources de traitement supplémentaires. Un système neuronal possédant ces caractéristiques est capable de transformer efficacement et rapidement (i.e. sur une durée compatible avec la durée de 150 ms observée chez l'homme) l'entrée visuelle en une représentation abstraite de la catégorie des objets qui la composent (e.g. visage/non-visage; article 6).

Cette théorie est encore à ce jour une hypothèse, même si de nombreuses données psychophysiques ou électrophysiologiques la supportent fortement. Elle est cependant l'une des rares théories actuelles à pouvoir rendre compte de l'extraordinaire rapidité des systèmes visuels biologiques. Si certaines des prédictions formulées dans cette thèse s'avéraient erronées, on peut néammoins espérer que ce travail participera, dans un futur proche, à la compréhension détaillée des mécanismes neuronaux qui sous-tendent les premières étapes de la perception visuelle, de l'encodage rétinien à la formation d'une première représentation abstraite de la scène visuelle. Il ne serait pas surprenant que, dans un système biologique dont les capacités computationnelles sont limitées par l'architecture physique, la course aux ressources neuronales se gagne avant tout dans le domaine temporel.

REFERENCES

1. Abbott, L. F., Rolls, E. T., & Tovee, M. J. (1996). Representational capacity of face coding in monkeys. *Cerebral Cortex, 6*(3), 498-505.
2. Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science, 275*(5297), 220-224.
3. Abeles, M. (1991). *Corticonics*. Cambridge: Cambridge University Press.
4. Adrian, E. D. (1926). The impulses produced by sensory nerve endings: Part I. *J Physiol (London), 61*, 49-72.
5. Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings: Part II: The response of a single end organ. *J Physiol (London), 61*, 151-171.
6. Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings: Part III: Impulses set up by touch and pressure. *J Physiol (London), 61*, 465-483.
7. Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron, 21*(2), 373-383.
8. Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex, 9*(5), 415-430.
9. Alonso, J. M., Usrey, W. M., & Reid, R. C. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature, 383*(6603), 815-9.
10. Alonso, J. M., & Martinez, L. M. (1998). Functional connectivity between simple cells and complex cells in cat striate cortex. *Nat Neurosci, 1*(5), 395-403.
11. Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu Rev Neurosci, 20*, 303-330.
12. Anderson, C. H., & Van Essen, D. C. (1987). Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proc Natl Acad Sci U S A, 84*(17), 6297-6301.
13. Arnett, D., & Spraker, T. E. (1981). Cross-correlation analysis of the maintained discharge of rabbit retinal ganglion cells. *J Physiol (Lond), 317*, 29-47.
14. Atick, J. J. (1992). Could Information theory provide an ecological theory of sensory processing? *Network: computation in neural systems., 3*(2), 213-251.
15. Aubertin, A., Fabre-Thorpe, M., Fabre, N., & Geraud, G. (1999). Fast visual categorization and speed of processing in migraine. *C R Acad Sci III, 322*(8), 695-704.
16. Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & 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), 1775-83.
17. Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation, 8*(6), 1185-1202.
18. Bair, W. (1999). Spike timing in the mammalian visual system. *Current Opinion in Neurobiology, 9*, 447-453.

19. Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J Neurosci, 11*(1), 168-190.

20. Bar, M., & Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. *Proc Natl Acad Sci U S A, 96*(4), 1790-1793.

21. Barlow, H. B., Hill, R. M., & Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology, 173*, 377-407.

22. Barlow, H. B., & Levick, W. R. (1965). The mechanisms of directionally selective units in the rabbit's retina. *Journal of Physiology (London), 178*, 477-504.

23. Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception, 1*(4), 371-394.

24. Barlow, H. B., Narasimhan, R., & Rosenfeld, A. (1972). Visual pattern analysis in machines and animals. *Science, 177*(49), 567-575.

25. Beaudot, W. H. A., & Herault, J. (1994). A neurobiological and psychophysical consistent model of the vertebrate retina. *Perception, 23, supplement (abstract)*, 25a.

26. Beaudot, W. H. A., Oliva, A., & Herault, J. (1995). Retinal model of the dynamics of X and Y pathways: a neural basis for early coarse-to-fine perception. *Perception, 24, supplement (abstract)*, 93b.

27. Bernander, Ö., Koch, C., & Usher, M. (1994). The effects of synchronized inputs at the single neuron level. *Neural Comput, 6*(622-641).

28. Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proc Natl Acad Sci U S A, 94*(10), 5411-6.

29. Berry, M. J. n., & Meister, M. (1998). Refractoriness and neural precision. *J Neurosci, 18*(6), 2200-11.

30. Berry, M. J., 2nd, Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature, 398*(6725), 334-8.

31. Bialek, W., & Rieke, F. (1992). Reliability and information transmission in spiking neurons. *Trends Neurosci, 15*(11), 428-34.

32. Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science, 252*(5014), 1854-7.

33. Biederman, I. (1987). Recognition by components : a theory of human image understanding. *Psychological Review, 94*, 115-145.

34. Blake, R. R., Fox, R., & McIntyre, C. (1971). Stochastic properties of stabilized-image binocular rivalry alternations. *J Exp Psychol, 88*(3), 327-332.

35. 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.

36. Borisyuk, R. M., Borisyuk, G. N., & Kazanovich, Y. B. (1998). The synchronization principle in modelling of binding and attention. *Membr Cell Biol, 11*(6), 753-61.

37. Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nat Neurosci, 2*(11), 947-57.

38. Botzel, K., Schulze, S., & Stodieck, S. R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res, 104*(1), 135-143.

39. Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci, 2*(4), 370-4.

40. Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Res, 15*(12), 1411-1412.

41. Breitmeyer, B. G. (1984). *Visual masking: an integrative approach*. New York: Oxford University Press.

42. Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci, 13*(1), 87-100.

43. Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat Neurosci, 1*(1), 59-63.

44. Brivanlou, I. H., Warland, D. K., & Meister, M. (1998). Mechanisms of concerted firing among retinal ganglion cells. *Neuron, 20*(3), 527-39.

45. Bruce, C. J., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46*, 369-384.

46. Bruce, V., & Green, P. R. (1990). *Visual Perception. Physiology, Psychology and Ecology. 2nd edition.*: Lawrence Erlbaum Associates.

47. Buckley, M. J., & Gaffan, D. (1998). Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia, 36*(6), 535-46.

48. Buckley, M. J., & Gaffan, D. (1998). Perirhinal cortex ablation impairs visual object identification. *J Neurosci, 18*(6), 2268-75.

49. Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus, 8*(4), 330-9.

50. Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol, 5*(4), 497-503.

51. Bullier, J., Schall, J. D., & Morel, A. (1996). Functional streams in occipito-frontal connections in the monkey. *Behavioural and Brain Science, 76*, 89-97.

52. Bulthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cereb Cortex, 5*(3), 247-260.

53. Buonomano, D. V., & Merzenich, M. (1998). A neural network model of temporal code generation and position-invariant pattern recognition. *Neural Comput, 11*(1), 103-116.

54. Buracas, G. T., Zador, A. M., DeWeese, M. R., & Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron, 20*(5), 959-69.

55. Buser, P., & Imbert, M. (1992). *Vision* (R.H. Kay, Trans.). Cambridge, MA: MIT Press.

56. Campbell, N. W., Thomas, B. T., & Troscianko, T. (1997). Automatic segmentation and classification of outdoor images using neural networks. *Int J Neural Syst, 8*(1), 137-144.

57. Castelo-Branco, M., Neuenschwander, S., & Singer, W. (1998). Synchronization of visual responses between the cortex, lateral geniculate nucleus, and retina in the anesthetized cat. *J Neurosci, 18*(16), 6395-410.

58. Cavada, C., & Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: I. Parcellation of areas based on distinctive limbic and sensory corticocortical connections. *J Comp Neurol, 287*(4), 393-421.

59. 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.

60. Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci, 2*(10), 913-919.

61. Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport, 10*(14), 2945-50.

62. Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol, 80*(6), 2918-40.

63. Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visually evoked potential generators by retinotopic and topographic analyses. *Hum Brain Map, 2*, 170-187.

64. Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *J Cog Neurosci, 8*, 387-402.

65. Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu Rev Neurosci, 22*, 319-49.

66. Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol, 75*(3), 1306-8.

67. Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *J Neurosci, 17*(9), 3201-14.

68. Contreras, D., Destexhe, A., Sejnowski, T. J., & Steriade, M. (1996). Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. *Science, 274*(5288), 771-4.

69. Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. *Proc Natl Acad Sci U S A, 81*(14), 4586-90.

70. Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neuroscience., 2*, 263-275.

71. Crick, F., & Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harb Symp Quant Biol, 55*, 953-62.

72. Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex?. *Nature, 375*(6527), 121-123.

73. Croner, L. J., & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Research, 35*(1), 7-24.

74. Crook, J. M., Lange-Malecki, B., Lee, B. B., & Valberg, A. (1988). Visual resolution of macaque retinal ganglion cells. *Journal of Physiology, 396*, 205-224.

75. Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *J Comp Neurol, 300*(1), 5-25.

76. Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J Neurosci, 16*(10), 3351-3362.

77. Dan, Y., Alonso, J. M., Usrey, W. M., & Reid, R. C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci, 1*(6), 501-7.

78. 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*(15), 3349-3353.

79. de Carvalho, L. A., & Roitman, V. L. (1995). A computational model for the neurobiological substrates of visual attention. *Int J Biomed Comput, 38*(1), 33-45.

80. Delorme, A., Gautrais, J., Van Rullen, R., & Thorpe, S. J. (1999). SpikeNET: a simulator for modeling large networks of integrate and fire neurons. *NeuroComputing, 24*.

81. Delorme, A., Richard, G., & 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.

82. Delorme, A., & Thorpe, S. J. (in preparation). Modelling early cortical orientation selectivity using fast shunting inhibition and rank order coding.

83. de Oliveira, S. C., Thiele, A., & Hoffmann, K. P. (1997). Synchronization of neuronal activity during stimulus expectation in a direction discrimination task. *J Neurosci, 17*(23), 9248-60.

84. Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci, 4*(8), 2051-2062.

85. Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol, 57*(3), 835-868.

86. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience., 18*, 193-222.

87. DeVries, S. H. (1999). Correlated firing in rabbit retinal ganglion cells. *J Neurophysiol, 81*(2), 908-20.

88. Diesmann, M., Gewaltig, M. O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature, 402*(6761), 529-533.

89. Di Russo, F., & Spinelli, D. (1999a). Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vision Res, 39*(18), 2975-85.

90. Di Russo, F., & Spinelli, D. (1999b). Spatial attention has different effects on the magno- and parvocellular pathways. *Neuroreport, 10*(13), 2755-62.

91. Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *J Comp Neurol, 334*(1), 125-150.

92. Ditchburn, R. W., & Ginsborg, B. L. (1952). Vision with a stabilized retinal image. *Nature, 170*, 36-37.

93. Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J Neurophysiol, 65*(6), 1329-1345.

94. Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J Neurophysiol, 65*(6), 1346-1359.

95. Duffy, C. J. (1998). MST neurons respond to optic flow and translational movement. *J Neurophysiol, 80*(4), 1816-1827.

96. Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychol Rev, 87*(3), 272-300.

97. Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General., 113*, 501-517.

98. Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol Cybern, 60*(2), 121-30.

99. Eckhorn, R. (1994). Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features. *Prog Brain Res, 102*, 405-26.

100. Edelman, S., & Poggio, T. (1991). Models of object recognition. *Curr Opin Neurobiol, 1*(2), 270-273.

101. Ejima, Y., & Ohtani, Y. (1987). Simple reaction time to sinusoidal grating and perceptual integration time: contributions of perceptual and response processes. *Vision Res, 27*(2), 269-276.

102. Engel, A. K., Konig, P., & Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proc Natl Acad Sci U S A, 88*(20), 9136-40.

103. Engel, A. K., Konig, P., Kreiter, A. K., & Singer, W. (1991). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science, 252*(5010), 1177-1179.

104. Engel, A. K., Kreiter, A. K., Konig, P., & Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc Natl Acad Sci U S A, 88*(14), 6048-6052.

105. Engel, A. K., Konig, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends Neurosci, 15*(6), 218-26.

106. Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

107. Eskandar, E. N., Richmond, B. J., & Optican, L. M. (1992). Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. *J Neurophysiol, 68*(4), 1277-95.

108. Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9*(2), 303-308.

109. Fabre-Thorpe, M., Fize, D., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of extrafoveal natural images: Implications for biological models. In J. Bower (Ed.), *Computational Neuroscience: Trends in Research* (pp. 7-12). New York: Plenum Press.

110. Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (in press). A limit to the speed of processing in Ultra-Rapid visual categorization of novel natural scenes. *J. Cog. Neurosci*.

111. Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proc R Soc Lond B Biol Sci, 254*(1341), 199-203.

112. Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Res, 35*(4), 491-4.

113. Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex, 1*(1), 1-47.

114. Ferster, D., & Spruston, N. (1995). Cracking the neuronal code. *Science, 270*(5237), 756-7.

115. Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Res, 33*(2), 173-93.

116. Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation, 6*, 559-601.

117. Fischer, B. (1973). Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Res, 13*(11), 2113-20.

118. Fischer, B., & Rogal, L. (1986). Eye-hand-coordination in man: a reaction time study. *Biol Cybern, 55*(4), 253-261.

119. Fize, D., Boulanouar, K., Chatel, Y., Ranjeva, J. P., Fabre-Thorpe, M., & Thorpe, S. J. (2000). Brain areas involved in rapid categorization of natural images: an event-related fMRI study. *NeuroImage, in press.*

120. Fize, D. (2000). *Bases cérébrales de la catégorisation visuelle rapide. Etudes chronométriques et fonctionnelles*, thèse de doctorat, EHESS.

121. Fregnac, Y., Burke, J. P., Smith, D., & Friedlander, M. J. (1994). Temporal covariance of pre- and postsynaptic activity regulates functional connectivity in the visual cortex. *J Neurophysiol, 71*(4), 1403-21.

122. Fregnac, Y., & Shulz, D. E. (1999). Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning. *J Neurobiol, 41*(1), 69-82.

123. Fregnac, Y. (1999). A tale of two spikes [news]. *Nat Neurosci, 2*(4), 299-301.

124. Frien, A., Eckhorn, R., Bauer, R., Woelbern, T., & Kehr, H. (1994). Stimulus-specific fast oscillations at zero phase between visual areas V1 and V2 of awake monkey. *Neuroreport, 5*(17), 2273-7.

125. Fries, P., Roelfsema, P. R., Engel, A. K., Konig, P., & Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc Natl Acad Sci U S A, 94*(23), 12699-704.

126. Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360*(6402), 343-346.

127. Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition, 15*, 455-469.

128. Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259*(5091), 100-103.

129. Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol, 76*(4), 2718-2739.

130. Gautrais, J. (1997). *Théorie et simulations d'un nouveau type de codage impulsionnel pour le traitement visuel rapide: le codage par l'ordre d'activation.* These de doctorat, EHESS.

131. Gautrais, J., & Thorpe, S. J. (1998). Rate Coding vs Temporal Order Coding : a theorical approach. *Biosystems, 48*(1-3), 57-65.

132. Gawne, T. J., Kjaer, T. W., & Richmond, B. J. (1996). Latency: another potential code for feature binding in striate cortex. *J Neurophysiol, 76*(2), 1356-60.

133. Geisler, W. S., Albrecht, D. G., Salvi, R. J., & Saunders, S. S. (1991). Discrimination performance of single neurons: rate and temporal-pattern information. *J Neurophysiol, 66*(1), 334-62.

134. George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport, 8*(6), 1417-1423.

135. Gerstner, W., & van Hemmen, J. L. (1992). Universality in neural networks: the importance of the 'mean firing rate'. *Biological Cybernetics, 67*(3), 195-205.

136. Gerstner, W., Kempter, R., van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature, 383*(6595), 76-81.

137. Gerstner, W., Kreiter, A. K., Markram, H., & Herz, A. V. (1997). Neural codes: firing rates and beyond. *Proc Natl Acad Sci U S A, 94*(24), 12740-1.

138. Gho, M., & Varela, F. J. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J Physiol, 83*(2), 95-101.

139. Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1994). Receptive-field maps of correlated discharge between pairs of neurons in the cat's visual cortex. *J Neurophysiol, 71*(1), 330-46.

140. Ghose, G. M., & Freeman, R. D. (1997). Intracortical connections are not required for oscillatory activity in the visual cortex. *Vis Neurosci, 14*(6), 963R-979R.

141. Ghose, G. M., & Ts'o, D. Y. (1997). Form processing modules in primate area V4. *J Neurophysiol, 77*(4), 2191-6.

142. Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc Natl Acad Sci U S A, 93*(2), 615-22.

143. Gochin, P. M., Miller, E. K., Gross, C. G., & Gerstein, G. L. (1991). Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp Brain Res, 84*(3), 505-16.

144. Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L., & Gross, C. G. (1994). Neural ensemble coding in inferior temporal cortex. *J Neurophysiol, 71*(6), 2325-37.

145. Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391*(6666), 481-4.

146. Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Vis Neurosci, 12*(6), 1027-52.

147. Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc Natl Acad Sci U S A, 86*(5), 1698-702.

148. Gray, C. M., Konig, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338*(6213), 334-337.

149. Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol, 35*(1), 96-111.

150. Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol Cybern, 23*(3), 121-134.

151. Grossberg, S. (1976). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biol Cybern, 23*(4), 187-202.

152. Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol Rev, 92*(2), 173-211.

153. Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept Psychophys, 38*(2), 141-171.

154. Haenny, P. E., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. *Exp Brain Res, 69*(2), 225-44.

155. Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp Brain Res, 69*(2), 245-59.

156. Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. (2000). Cognitive Response Profile of the Human Fusiform Face Area as Determined by MEG. *Cereb Cortex, 10*(1), 69-81.

157. Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology, 121*, 400-415.

158. Hartline, H. K. (1940). The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. *American Journal of Physiology, 130*, 700-711.

159. Hartline, H. K. (1940). The receptive fields of optic nerve fibers. *American Journal of Physiology, 130*, 690-699.

160. Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B., & Rapoport, S. I. (1991). Dissociation of object

and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci U S A, 88*(5), 1621-1625.

161. Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Munte, T. F., Gos, A., Scherg, M., Johannes, S., Hundeshagen, H., & et al. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature, 372*(6506), 543-6.

162. Heitger, F., & von der Heydt, R. (1993). A computational model of neural contour processing: Figure-ground segregation and illusory contours. *Proc. of the 4th Intl. Conf. on Computer Vision*, 32-40.

163. Heller, J., Hertz, J. A., Kjaer, T. W., & Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision. *J Comput Neurosci, 2*(3), 175-93.

164. Hendry, S. H., & Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annu Rev Neurosci, 23*, 127-53.

165. Higuchi, S., & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proc Natl Acad Sci U S A, 93*(2), 739-43.

166. Hikosaka, O., Miyauchi, S., & Shimojo, S. (1991). Focal visual attention produces motion sensation in lines. *Investigative Ophtalmology and Visual Science, 32*(Suppl.), 716.

167. Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993a). Focal visual attention produces illusory temporal order and motion sensation. *Vision Res, 33*(9), 1219-40.

168. Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993b). Visual attention revealed by an illusion of motion. *Neurosci Res, 18*(1), 11-8.

169. Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci U S A, 95*(3), 781-7.

170. Hillyard, S. A., Teder-Salejarvi, W. A., & Munte, T. F. (1998). Temporal dynamics of early perceptual processing. *Curr Opin Neurobiol, 8*(2), 202-10.

171. Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1257-70.

172. Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature, 376*, 33-36.

173. Horn, D., & Levanda, S. (1998). Fast temporal encoding and decoding with spiking neurons. *Neural Computation., 10*, 1705-1720.

174. Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate visual cortex. *J Physiol, 148*, 574-591.

175. Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *J Physiol, 160*, 106-154.

176. Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of the monkey striate cortex. *J Physiol (London), 195*, 574-591.

177. Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychol Rev, 99*(3), 480-517.

178. Hummel, R., & Zucker, S. W. (1983). On the foundation of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 267-287.

179. Humphreys, G. W., Price, C. J., & Riddoch, M. J. (1999). From objects to names: a cognitive neuroscience approach. *Psychol Res, 62*(2-3), 118-130.

180. Hupe, 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*(6695), 784-7.

181. Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *J Exp Psychol [Hum Learn], 6*(1), 1-12.

182. Intraub, H. (1999). Understanding and Remembering Briefly Glimpsed Pictures: Implications for Visual Scanning and Memory. In V. Coltheart (Ed.), *Fleeting Memories: cognition of brief visual stimuli.* (pp. 47-70): MIT Press.

183. Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A, 96*(16), 9379-9384.

184. Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

185. Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition, 3*, 1-38.

186. Johnson, J. S., Guirao-Garcia, J. J., & Olshausen, B. A. (1999,). *Early visual processing of natural images.* Paper presented at the ARVO.

187. Kalaska, J. F., & Crammond, D. J. (1992). Cerebral cortical mechanisms of reaching movements. *Science, 255*(5051), 1517-1523.

188. Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*(4), 843-56.

189. Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-61.

190. Keysers, C., Xiao, D.-K., Foldiak, P., & Perrett, D. I. (in press). The speed of sight. *Nature*.

191. Kingstone, A. (1992). Combining expectancies. *Quarterly Journal of Experimental Psychology., 44*, 69-104.

192. Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Res, 36*(4), 539-44.

193. Kirkland, K. L., & Gerstein, G. L. (1999). A feedback model of attention and context dependence in visual cortical networks. *J Comput Neurosci, 7*(3), 255-67.

194. Kirschfeld, K., & Kammer, T. (2000). Visual attention and metacontrast modify latency to perception in opposite directions. *Vision Res, 40*(9), 1027-1033.

195. Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol, 71*(3), 856-867.

196. Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*(4), 219-227.

197. Koch, C., & Braun, J. (1996). Towards the neuronal correlate of visual awareness. *Curr Opin Neurobiol, 6*(2), 158-164.

198. Koch, C., & Braun, J. (1996). The functional anatomy of visual awareness. *Cold Spring Harb Symp Quant Biol, 61*, 49-57.

199. Koch, C. (1997). Computation and the single neuron. *Nature, 385*, 207-210.

200. Koch, C., & Laurent, G. (1999). Complexity and the nervous system. *Science, 284*(5411), 96-8.

201. Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *J Neurophysiol, 60*(2), 580-603.

202. Konig, P., & Schillen, T. B. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. synchronization. *Neural Comput, 3*, 155-166.

203. Konig, P., & Engel, A. K. (1995). Correlated firing in sensory-motor systems. *Curr Opin Neurobiol, 5*(4), 511-519.

204. Konig, P., Engel, A. K., & Singer, W. (1995). Relation between oscillatory activity and long-range synchronization in cat visual cortex. *Proc Natl Acad Sci U S A, 92*(1), 290-4.

205. Konig, P., Engel, A. K., Roelfsema, P. R., & Singer, W. (1995). How precise is neuronal synchronization? *Neural Comput, 7*(3), 469-85.

206. Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc Natl Acad Sci U S A, 92*(12), 5587-91.

207. Kovacs, I., Papathomas, T. V., Yang, M., & 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.

208. Kreiter, A. K., & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J Neurosci, 16*(7), 2381-96.

209. Kuffler, S. K. (1953). Discharge patterns and functional organization of mammalian retina. *J Neurophysiol, 16*, 37-68.

210. Kulikowski, J. J., Walsh, V., McKeefry, D., Butler, S. R., & Carden, D. (1994). The electrophysiological basis of colour processing in macaques with V4 lesions. *Behav Brain Res, 60*(1), 73-78.

211. LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 292-323.

212. Lagae, L., Maes, H., Raiguel, S., Xiao, D. K., & Orban, G. A. (1994). Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol, 71*(5), 1597-1626.

213. Lal, R., & Friedlander, M. J. (1989). Gating of retinal transmission by afferent eye position and movement signals. *Science, 243*(4887), 93-6.

214. Lal, R., & Friedlander, M. J. (1990). Effect of passive eye movement on retinogeniculate transmission in the cat. *J Neurophysiol, 63*(3), 523-38.

215. Lal, R., & Friedlander, M. J. (1990). Effect of passive eye position changes on retinogeniculate transmission in the cat. *J Neurophysiol, 63*(3), 502-22.

216. Lamme, V. A. F., Super, H., & Spekreijse, H. (1998). Feed-forward, horizontal, and feed-back processing in the visual cortex. *Curr Opin Neurobiol, 8*, 529-535.

217. Lamme, V. A., & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature, 396*(6709), 362-6.

218. Lankheet, M. J., Molenaar, J., & van de Grind, W. A. (1989). The spike generating mechanism of cat retinal ganglion cells. *Vision Res, 29*(5), 505-17.

219. Lawler, K. A., & Cowey, A. (1987). On the role of posterior parietal and prefrontal cortex in visuo- spatial perception and attention. *Exp Brain Res, 65*(3), 695-698.

220. Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat Neurosci, 2*(4), 375-81.

221. Lee, S. H., & Blake, R. (1999). Rival ideas about binocular rivalry. *Vision Res, 39*(8), 1447-1454.

222. Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Res, 38*(15-16), 2429-54.

223. Lennie, P. (1981). The physiological basis of variations in visual latency. *Vision Research, 21*, 815-824.

224. Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature, 379*(6565), 549-553.

225. Lestienne, R. (1996). Determination of the precision of spike timing in the visual cortex of anaesthetised cats. *Biol Cybern, 74*(1), 55-61.

226. Lestienne, R., & Tuckwell, H. C. (1998). The significance of precisely replicating patterns in mammalian CNS spike trains. *Neuroscience, 82*(2), 315-36.

227. Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc Inst Radion Engrs N.Y., 47*, 1940-1951.

228. Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput, 10*(4), 903-40.

229. Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain, 106*(Pt 3), 623-642.

230. Libet, B., Pearl, D. K., Morledge, D. E., Gleason, C. A., Hosobuchi, Y., & Barbaro, N. M. (1991). Control of the transition from sensory detection to sensory awareness in man by the duration of a thalamic stimulus. The cerebral 'time-on' factor. *Brain, 114*(Pt 4), 1731-57.

231. Lipetz, L. E. (1971). The relation of physiological and psychological aspects of sensory intensity. In W. R. Lowenstein (Ed.), *Handbook of sensory physiology* (Vol. 1. Principles of receptor physiology, pp. 191-225). New York: Springer-Verlag.

232. Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science, 245*(4919), 761-3.

233. Logothetis, N. K., Pauls, J., Bulthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Curr Biol, 4*(5), 401-414.

234. Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr Biol, 5*(5), 552-563.

235. Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry?. *Nature, 380*(6575), 621-624.

236. Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annu Rev Neurosci, 19*, 577-621.

237. Logothetis, N. (1998). Object vision and visual awareness. *Curr Opin Neurobiol, 8*(4), 536-544.

238. Luce, R. D. (1986). *Response Times*. Oxford: Oxford University Press.

239. Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalogr Clin Neurophysiol, 75*(6), 528-42.

240. Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? *J Exp Psychol Hum Percept Perform, 22*(3), 725-37.

241. 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. *Journal of Neurophysiology, 77*(1), 24-42.

242. Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. *Proc Natl Acad Sci U S A, 95*(3), 825-30.

243. Lutzenberger, W., Pulvermuller, F., Elbert, T., & Birbaumer, N. (1995). Visual stimulation alters local 40-Hz responses in humans: an EEG-study. *Neurosci Lett, 183*(1-2), 39-42.

244. Maass, W. (1996). Lower bounds for the computational power of networks of spiking neurons. *Neural computation., 8*, 1-40.

245. Maass, W. (1997). Fast sigmoidal networks via spiking neurons. *Neural Comput, 9*(2), 279-304.

246. MacLeod, K., Backer, A., & Laurent, G. (1998). Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature, 395*(6703), 693-8.

247. Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science, 268*, 1503-1506.

248. Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *Inst. Electrical Electronics Engrs. Trans. on Pattern Analysis and Machine Intelligence., 11*, 674-693.

249. Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature, 382*(6594), 807-10.

250. Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science, 275*(5297), 213-5.

251. Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization ofthree-dimensional shapes. *Proceedings of the Royal Society of London, B, 200*, 269-294.

252. Marr, D. (1982). *Vision : a computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

253. Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., & Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat Neurosci, 2*(4), 364-9.

254. Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nat Neurosci, 3*(3), 251-258.

255. Mastronarde, D. N. (1989). Correlated firing of retinal ganglion cells. *Trends Neurosci, 12*(2), 75-80.

256. Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annu Rev Neurosci, 10*, 363-401.

257. Maunsell, J. H., Sclar, G., Nealey, T. A., & DePriest, D. D. (1991). Extraretinal representations in area V4 in the macaque monkey. *Vis Neurosci, 7*(6), 561-73.

258. McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci, 19*(1), 431-41.

259. McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol 2: Psychological and Biological Models*. Cambridge, MA: MIT Press/Bradford Books.

260. McClurkin, J. W., Gawne, T. J., Optican, L. M., & Richmond, B. J. (1991). Lateral geniculate neurons in behaving primates. II. Encoding of visual information in the temporal shape of the response. *J Neurophysiol, 66*(3), 794-808.

261. McClurkin, J. W., Optican, L. M., Richmond, B. J., & Gawne, T. J. (1991). Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science, 253*(5020), 675-7.

262. McClurkin, J. W., Optican, L. M., & Richmond, B. J. (1994). Cortical feedback increases visual information transmitted by monkey parvocellular lateral geniculate nucleus neurons. *Vis Neurosci, 11*(3), 601-17.

263. McClurkin, J. W., & Optican, L. M. (1996). Primate striate and prestriate cortical neurons during discrimination. I. simultaneous temporal encoding of information about color and pattern. *J Neurophysiol, 75*(1), 481-95.

264. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115-133.

265. Meister, M., Lagnado, L., & Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science, 270*(5239), 1207-10.

266. Meister, M., & Berry, M. J., 2nd. (1999). The neural code of the retina. *Neuron, 22*(3), 435-50.

267. Merigan, W. H. (1996). Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. *Vis Neurosci, 13*(1), 51-60.

268. Merigan, W. H., & Pham, H. A. (1998). V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Vis Neurosci, 15*(2), 359-367.

269. Miller, J. P. (1994). Neural coding. Neurons cleverer than we thought? *Curr Biol, 4*(9), 818-20.

270. Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*: Oxford University Press.

271. Milner, P. M. (1974). A model for visual shape recognition. *Psychol Rev, 81*(6), 521-35.

272. Miltner, W. H., Braun, C., Arnold, M., Witte, H., & Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature, 397*(6718), 434-6.

273. Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci, 6*(10), 414-417.

274. Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: encoding and top-down activation. *Curr Opin Neurobiol, 10*(2), 187-94.

275. Moghaddam, B., & Pentland, A. (1995). *Probabilistic visual learning for object detection.* Paper presented at the The Fifth International Conference on Computer Vision, Cambridge, MA.

276. Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784.

277. Morel, A., & Bullier, J. (1990). Anatomical segregation of two cortical visual pathways in the macaque monkey. *Vis Neurosci, 4*(6), 555-578.

278. Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology., 70*, 909-919.

279. Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci, 12*(1), 303-310.

280. Mouchetant-Rostaing, Y., Giard, M.-H., Delpuech, C., Echallier, J.-F., & Pernier, J. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *NeuroReport, 11*(11), in press.

281. Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern Recognition Mechanisms* (pp. 117-151). Vatican: Vatican Press.

282. Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341-393). Psychology Press.

283. Muller, M. M., Bosch, J., Elbert, T., Kreiter, A., Sosa, M. V., Sosa, P. V., & Rockstroh, B. (1996). Visually induced gamma-band responses in human electroencephalographic activity--a link to animal studies. *Exp Brain Res, 112*(1), 96-102.

284. Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biol Cybern, 65*(2), 135-45.

285. Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern, 66*(3), 241-51.

286. Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *J Exp Psychol Learn Mem Cogn, 15*(4), 572-586.

287. Murthy, V. N., & Fetz, E. E. (1994). Effects of input synchrony on the firing rate of a three-conductance cortical neuron model. *Neural Comput, 6*, 1111-1126.

288. Naka, K. I., & Rushton, W. A. H. (1966). S-potentials from luminosity units in the retina of fish (cyprinidae). *Journal of physiology, London, 185*, 587-599.

289. Naka, K., & Sakai, H. M. (1991). The messages in optic nerve fibers and their interpretation. *Brain Res Brain Res Rev, 16*(2), 135-49.

290. Nakamura, K. (1998). Neural processing in the subsecond time range in the temporal cortex. *Neural Comput, 10*(3), 567-595.

291. Natschlager, T., & Ruf, B. (1998). Spatial and temporal pattern analysis via spiking neurons. *Network, 9*(3), 319-32.

292. Neven, H., & Aertsen, A. (1992). Rate coherence and event coherence in the visual cortex: a neuronal model of object recognition. *Biol Cybern, 67*(4), 309-322.

293. Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*(6237), 52-54.

294. Nicholas, J., Reynolds, J., & Desimone, R. (1996). Texture segmentation with attention in area V4 of the macaque. *Society for Neuroscience Abstracts, 22*, 1614.

295. Nichols, M. J., & Newsome, W. T. (1999). The neurobiology of cognition. *Nature, 402*(Supp), C35-C38.

296. Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research, 33*(18), 2789-2802.

297. Niebur, E., & Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience, 1*(1-2), 141-158.

298. Nirenberg, S., & Latham, P. E. (1998). Population coding in the retina. *Curr Opin Neurobiol, 8*(4), 488-93.

299. Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual Latencies in Areas V1 and V2 of the Macaque Monkey. *Visual Neurosci, 12*(2), 371-384.

300. Nowak, L. G., Sanchez-Vives, M. V., & McCormick, D. A. (1997). Influence of low and high frequency inputs on spike timing in visual cortical neurons. *Cereb Cortex, 7*(6), 487-501.

301. Nowak, L., & Bullier, J. (1998). The timing of information transfer in the visual system. In J. H. Kaas, K. Rockland, & A. Peters (Eds.), *Cerebral Cortex* (pp. 205-241). New York: Plenum Press.

302. Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *J Neurosci, 15*(2), 1195-214.

303. O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584-7.

304. Oliva, A., Herault, J., & Guerin-Dugue, A. (1997). Real-world scene categorization by a self-organizing neural network. *Perception, 26, supplement (abstract)*, 19b.

305. Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience., 13*(11), 4700-4719.

306. Olshausen, B. A., & Field, D. J. (1996). Natural Image statistics and efficient coding. *Network: Computation in neural systems., 7*(2), 333-339.

307. Optican, L. M., & Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol, 57*(1), 162-78.

308. Panzeri, S., Treves, A., Schultz, S., & Rolls, E. T. (1999). On decoding the responses of a population of neurons from short time windows. *Neural Comput, 11*(7), 1553-77.

309. Parodi, O., Combe, P., & Ducom, J. C. (1996). Temporal coding in vision: coding by the spike arrival times leads to oscillations in the case of moving targets. *Biol Cybern, 74*(6), 497-509.

310. Pashler, H. (1998). *The Psychology of Attention.* Cambridge, MA: MIT Press.

311. Peichl, L., & Wassle, H. (1979). Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *J Physiol (Lond), 291*, 117-41.

312. Penev, P. S., & Atick, J. J. (1996). Local feature analysis : a general statistical theory for object representation. *Network : Computation in Neural Systems, 7*(3), 477-500.

313. Perkel, D. H., Gerstein. G. L, & Moore, G. P. (1967). Neuronal spike trains and stochastic point processes I. The single spike train. *Biophysical Journal, 7*, 391 - 418.

314. Perkel, D. H., Gerstein. G. L, & Moore, G. P. (1967). Neuronal spike trains and stochastic point processes II. Simultaneous spike trains. *Biophysical Journal, 7*, 419-440.

315. Perkel, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin, 6*(3), 221-348.

316. Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47*, 329-342.

317. Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86*(1), 159-73.

318. Pitts, W., & McCulloch, W. W. (1947). How we know universals. *Bull Math Biophys, 9*, 127-147.

319. Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*(6255), 263-266.

320. Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General., 109*, 160-174.

321. Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J Exp Psychol, 81*(1), 10-15.

322. Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965-966.

323. Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253*(5491), 437-438.

324. Potter, M. C. (1976). Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn], 2*(5), 509-522.

325. Potter, M. C. (1999). Understanding Sentences and Scenes: The role of Conceptual Short-Term Memory. In V. Coltheart (Ed.), *Fleeting Memories: cognition of brief visual stimuli.* (pp. 13-46): MIT Press.

326. Proverbio, A. M., & Mangun, G. R. (1994). Electrophysiological and behavioral "costs" and "benefits" during sustained visual-spatial attention. *Int J Neurosci, 79*(3-4), 221-33.

327. Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., & Abeles, M. (1998). Spatiotemporal structure of cortical activity: properties and behavioral relevance. *J Neurophysiol, 79*(6), 2857-74.

328. Rager, G., & Singer, W. (1998). The response of cat visual cortex to flicker stimuli of variable frequency. *Eur J Neurosci, 10*(5), 1856-77.

329. Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., Lagae, L., & Orban, G. A. (1997). Size and shape of receptive fields in the medial superior temporal area (MST) of the macaque. *Neuroreport, 8*(12), 2803-2808.

330. Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychol Rev, 106*(2), 261-300.

331. Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *J Neurophysiol, 77*(5), 2836-41.

332. Reich, D. S., Victor, J. D., & Knight, B. W. (1998). The power ratio and the interval map: spiking models and extracellular recordings. *J Neurosci, 18*(23), 10090-104.

333. Reynolds, J. H., Pasternak, T., & Desimone, R. (1996). Attention increases contrast sensitivity of cells in macaque area V4. *Society for Neuroscience Abstracts, 22*, 1197.

334. Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci, 19*(5), 1736-53.

335. Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*(1), 19-29, 111-25.

336. Richmond, B. J., Optican, L. M., Podell, M., & Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol, 57*(1), 132-46.

337. Richmond, B. J., & Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J Neurophysiol, 57*(1), 147-61.

338. Richmond, B. J., Optican, L. M., & Spitzer, H. (1990). Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol, 64*(2), 351-69.

339. Richmond, B. J., & 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), 370-80.

340. Riehle, A., Grun, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science, 278*(5345), 1950-3.

341. Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: exploring the neural code.* Cambridge, MA: MIT.

342. Riesenhuber, M., & Poggio, T. (1999). Are cortical models really bound by the "binding problem"? *Neuron, 24*(1), 87-93, 111-25.

343. Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025.

344. Riggs, L. A., & Ratcliff, F. (1952). The effects of counteracting the normal movements of the eye. *J Opt Soc Am, 42*, 872-873.

345. Ritz, R., Gerstner, W., Fuentes, U., & van Hemmen, J. L. (1994). A biologically motivated and analytically soluble model of collective oscillations in the cortex. II. Application to binding and pattern segmentation. *Biol Cybern, 71*(4), 349-58.

346. Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research., 5*, 583-601.

347. Rodieck, R. W. (1998). *The first steps in seeing.* Sunderland, MA.: Sinauer Associates.

348. Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature, 397*(6718), 430-3.

349. Roelfsema, P. R., Engel, A. K., Konig, P., & Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature, 385*(6612), 157-61.

350. Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376-81.

351. 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*(1348), 9-15.

352. Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp Brain Res, 101*(3), 473-84.

353. Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol, 73*(2), 713-726.

354. Rolls, E. T., & Tovee, M. J. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Exp Brain Res, 103*(3), 409-20.

355. Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp Brain Res, 114*(1), 149-62.

356. Rolls, E. T., Treves, A., Tovee, M. J., & Panzeri, S. (1997). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J Comput Neurosci, 4*(4), 309-33.

357. Romdhani, S. (1996). *Face recognition using principal components analysis.* Unpublished MSc. Thesis, University of Glasgow.

358. Romo, R., & Salinas, E. (1999). Sensing and deciding in the somatosensory system. *Curr Opin Neurobiol, 9*(4), 487-493.

359. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognit. Psychol., 8*, 382-439.

360. Rosenblatt, F. (1958). The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 368-408.

361. Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *NeuroReport, 11*(1), 69-74.

362. Rowley, H. A., Baluja, S., & Kanade, T. (1995). *Human face detection in visual scenes.* (Internal Report): School of Computer Science, Carnegie Mellon University, Pittsburg.

363. Ruf, B. (1997). *Computing functions with spiking neurons in temporal coding.* (NeuroColt Technical Report Series NC-TR-97-026): NeuroCOLT.

364. Ruf, B., & Schmitt, M. (1997). *Hebbian learning in networks of spiking neurons using temporal coding.* (NeuroCOLT Technical Series. NC-TR-97-027): NeuroCOLT.

365. Rugg, M. D., Doyle, M. C., & Wells, T. (1995). Word and non-word repetition within- and across-modality: an event-related potential study. *J Cog Neurosci, 7*, 209-227.

366. Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books.

367. Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C., & 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.

368. Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol Rev, 75*(1), 107-54.

369. Salinas, E., & Abbott, L. F. (1997). Invariant visual responses from attentional gain fields. *J Neurophysiol, 77*(6), 3267-72.

370. Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annu Rev Neurosci, 22*, 241-259.

371. Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35*(3), 240-251.

372. Schillen, T. B., & Konig, P. (1991). Stimulus-dependent assembly formation of oscillatory responses: II. desynchronization. *Neural Comput, 3*, 167-178.

373. Schiller, P. H. (1995). Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature, 376*(6538), 342-344.

374. Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *J Neurophysiol, 79*(6), 3272-3278.

375. Schwarz, C., & Bolz, J. (1991). Functional specificity of a long-range horizontal connection in cat visual cortex: a cross-correlation study. *J Neurosci, 11*(10), 2995-3007.

376. Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5*, 195-200.

377. Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69*(3), 243-265.

378. 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*(12), 2749-2754.

379. Seidemann, E., & Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *J Neurophysiol, 81*(4), 1783-94.

380. Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.

381. Sereno, A. B., & Maunsell, J. H. (1998). Shape selectivity in primate lateral intraparietal cortex *Nature, 395*(6701), 500-503.

382. Sestokas, A. K., Lehmkuhle, S., & Kratz, K. E. (1987). Visual latency of ganglion X- and Y-cells: a comparison with geniculate X- and Y-cells. *Vision Res, 27*(9), 1399-408.

383. Sestokas, A. K., Lehmkuhle, S., & Kratz, K. E. (1991). Relationship between response latency and amplitude for ganglion and geniculate X- and Y-cells in the cat. *Int J Neurosci, 60*(1-2), 59-64.

384. Shadlen, M. N., &, & Newsome, W. T. (1995). Is there a signal in the noise? *Curr Opin Neurobiol, 5*(2), 248-50.

385. Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proc Natl Acad Sci U S A, 93*(2), 628-633.

386. Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci, 18*(10), 3870-96.

387. Shashua, A., & Ullman, S. (1988). Structural saliency. *Proceedings od the International Conference on Computer Vision*, 482-488.

388. Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci U S A, 94*(7), 3408-13.

389. Shiu, L.-P., & Pashler, H. (1993). *Spatial precuing in single-element displays: noise-reduction or signal enhancement ?* Paper presented at the 34th annual meeting of the Psychonomic Society, Washington, DC.

390. Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M., & Petersen, S. E. (1997). Top-down modulation of early sensory cortex. *Cereb Cortex, 7*(3), 193-206.

391. Sillito, A. M., Jones, H. E., Gerstein, G. L., & West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature, 369*(6480), 479-82.

392. Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annu Rev Physiol, 55*, 349-374.

393. Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci, 18*, 555-586.

394. Singer, W. (1999). Time as coding space? *Curr Opin Neurobiol, 9*(2), 189-94.

395. Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J Neurophysiol, 57*(3), 773-86.

396. Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature, 386*(6620), 69-73.

397. Smith, A. T., Singh, K. D., & Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. *Neuroreport, 11*(2), 271-7.

398. Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature, 386*(6621), 167-170.

399. Snyder, L. H., Batista, A. P., & Andersen, R. A. (2000). Intention-related activity in the posterior parietal cortex: a review. *Vision Res, 40*(10-12), 1433-1441.

400. Softky, W. R. (1995). Simple codes versus efficient codes. *Curr Opin Neurobiol, 5*(2), 239-47.

401. Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc Natl Acad Sci U S A, 96*(4), 1663-8.

402. Sompolinsky, H., & Tsodyks, M. (1994). Segmentation by a network of oscillators with stored memories. *Neural Comput, 6*, 642-657.

403. Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240*(4850), 338-40.

404. Stanley, G. B., Li, F. F., & Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J Neurosci, 19*(18), 8036-42.

405. Steinman, B. A., Steinman, S. B., & Lehmkuhle, S. (1995). Visual attention mechanisms show a center-surround organization. *Vision Res, 35*(13), 1859-69.

406. Steinman, B. A., Steinman, S. B., & Lehmkuhle, S. (1997). Transient visual attention is dominated by the magnocellular stream. *Vision Res, 37*(1), 17-23.

407. Steinmetz, M. A., Connor, C. E., Constantinidis, C., & McLaughlin, J. R. (1994). Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. *J Neurophysiol, 72*(2), 1020-1023.

408. Steinmetz, M. A., & Constantinidis, C. (1995). Neurophysiological evidence for a role of posterior parietal cortex in redirecting visual attention. *Cereb Cortex, 5*(5), 448-456.

409. Steinmetz, P. N., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K. O., & Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature, 404*(6774), 187-90.

410. Stevens, C. F., & Zador, A. M. (1998). Input synchrony and the irregular firing of cortical neurons. *Nat Neurosci, 1*(3), 210-7.

411. Stopfer, M., Bhagavan, S., Smith, B. H., & Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature, 390*(6655), 70-4.

412. Subramaniam, S., Biederman, I., & Madigan, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cognition, 7*(4), 511-535.

413. 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), 869-73.

414. Sung, K., & Poggio, T. (1994). Example-based learning for view-based human face detection. *Proceedings Image Understanding Workshop, II*, 843-850.

415. Suzuki, W. A. (1996). The anatomy, physiology and functions of the perirhinal cortex. *Curr Opin Neurobiol, 6*(2), 179-86.

416. Suzuki, W. A. (1999). The long and the short of it: memory signals in the medial temporal lobe. *Neuron, 24*(2), 295-8.

417. Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J Neurosci, 16*(13), 4240-9.

418. 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*(2), 722-34.

419. Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *J Neurosci, 18*(11), 4244-54.

420. Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci, 3*(4), 151-162.

421. Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *Neuroreport, 10*(4), 829-835.

422. Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262*, 685-688.

423. Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

424. Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. *Curr Opin Neurobiol, 7*(4), 523-529.

425. Tarr, M. J., & Bulthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *J Exp Psychol Hum Percept Perform, 21*(6), 1494-1505.

426. Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1-2), 1-20.

427. Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nat Neurosci, 1*(4), 275-7.

428. 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.

429. Thorpe, S. J. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In R. Eckmiller, G. Hartman, & G. Hauske (Eds.), *Parallel processing in neural systems* (pp. 91-94). North-Holland: Elsevier.

430. Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

431. Thorpe, S. J., & Gautrais, J. (1997). Rapid visual processing using spike asynchrony. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems* (MIT Press ed., Vol. 9, pp. 901-907). Cambridge: MIT Press.

432. Thorpe, S. J., & Gautrais, J. (1998). Rank order coding: a new coding scheme for rapid processing in neural networks. In J. Bower (Ed.), *Computational Neuroscience : Trends in Research.* . New York: Plenum Press.

433. Thorpe, S. J., Gegenfurtner, K., Fabre-Thorpe, M., & Bülthoff, H. H. (1999). Categorisation of complex natural images in extreme peripheral vision. *Perception, 28 (supplement)*, 61.

434. Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., & Naatanen, R. (1993). Selective attention enhances the auditory 40-Hz transient response in humans. *Nature, 364*(6432), 59-60.

435. Tononi, G., Srinivasan, R., Russell, D. P., & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proc Natl Acad Sci U S A, 95*(6), 3198-203.

436. Tootell, R. B., Dale, A. M., Sereno, M. I., & Malach, R. (1996). New images from human visual cortex. *Trends Neurosci, 19*(11), 481-489.

437. Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I., & Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *J Neurosci, 17*(18), 7060-7078.

438. Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron, 21*(6), 1409-22.

439. Tovee, M. J., & Rolls, E. T. (1992). Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli. *Neuroreport, 3*(4), 369-72.

440. Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol, 70*(2), 640-54.

441. Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition, 2*, 35-39.

442. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognit Psychol, 12*(1), 97-136.

443. Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognit Psychol, 14*(1), 107-41.

444. Treisman, A. (1996). The binding problem. *Curr Opin Neurobiol, 6*(2), 171-178.

445. Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382*(6591), 539-41.

446. Treves, A., Rolls, E. T., & Tovee, M. J. (1996). On the time required for recurrent processing in the brain. In V. Torre & F. Conti (Eds.), *Neurobiology: Proceedings of the International School of Biophysics, XXIII course, May 1995* (pp. 371-382). New York: Plenum Press.

447. Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: the role of texture and shape in complex static stimuli. *Vision Res, 39*(2), 353-366.

448. Trotter, Y., & Celebrini, S. (1999). Gaze direction controls response gain in primary visual-cortex neurons. *Nature, 398*(6724), 239-242.

449. Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3*(1), 71-86.

450. Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition, 32*(3), 193-254.

451. Ullman, S. (1990). Three-dimensional object recognition. *Cold Spring Harb Symp Quant Biol, 55*, 889-898.

452. Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb Cortex, 5*(1), 1-11.

453. 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 Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

454. Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol, 4*(2), 157-165.

455. Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature, 394*(6689), 179-82.

456. Usrey, M., Reppas, J. B., & Reid, R. C. (1998). Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. *Nature, 395*, 384-387.

457. Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature, 373*(6514), 515-8.

458. Valentin, D., Abdi, H., O'Toole, A., & Cottrell, G. W. (1994). Connexionnist models of face processing : a survey. *Pattern Recognition, 27*, 1209-1230.

459. Vanduffel, W., Tootell, R. B., & Orban, G. A. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cereb Cortex, 10*(2), 109-126.

460. Van Essen, D. C. (1979). Visual areas of the mammalian cerebral cortex. *Annu Rev Neurosci, 2*, 227-263.

461. Van Essen, D. C., Felleman, D. J., DeYoe, E. A., Olavarria, J., & Knierim, J. (1990). Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. *Cold Spring Harb Symp Quant Biol, 55*, 679-696.

462. Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science, 255*(5043), 419-423.

463. VanRullen, R., Gautrais J., Delorme A., & Thorpe, S. J. (1998). Face Processing using One Spike Per Neuron. *Biosystems, 48*(1-3), 229-239.

464. VanRullen, R., & Thorpe, S. (1999). Spatial attention in asynchronous neural networks. *NeuroComputing, 26-27*, 911-918.

465. VanRullen, R., & Thorpe, S. (2000). Rate coding vs temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation, submitted*.

466. VanRullen, R., & Thorpe, S. (2000). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, submitted*.

467. VanRullen, R., & Thorpe, S. J. (2000). The time course of visual processing: from early perception to decision-making. *J Cog Neuroscience, submitted*.

468. VanRullen, R., Delorme, A., & Thorpe, S. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing, (submitted)*.

469. Vardi, N., & Smith, R. G. (1996). The AII amacrine network: coupling can increase correlated activity. *Vision Res, 36*(23), 3743-57.

470. Varela, F. J., Toro, A., John, E. R., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia, 19*(5), 675-86.

471. Varela, F. J., & Singer, W. (1987). Neuronal dynamics in the visual corticothalamic pathway revealed through binocular rivalry. *Exp Brain Res, 66*(1), 10-20.

472. Varela, F. J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol Res, 28*(1), 81-95.

473. Varela, F. J. (1999). Present-time consciousness. *J Consc Studies, 6*(2-3), 111-140.

474. Varela, F. J. (1999). The specious present: The neurophenomenology of time consciousness,. In F. J. V. J.Petitot, B.Pachoud and J.M.Roy (Ed.), *Naturalizing Phenomenology* (pp. 266-314): Stanford University Press.

475. Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3D object recognition: invariance to imaging transformations. *Cereb Cortex, 5*(3), 261-269.

476. Victor, J. D., & Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol, 76*(2), 1310-26.

477. Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *Neuroreport, 9*(9), 1947-52.

478. Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur J Neurosci, 11*(4), 1223-1238.

479. Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur J Neurosci, 11*(4), 1239-1255.

480. Volgushev, M., Vidyasagar, T. R., & Pei, X. (1995). Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. *Vis Neurosci, 12*(4), 621-8.

481. Volgushev, M., Chistiakova, M., & Singer, W. (1998). Modification of discharge patterns of neocortical neurons by induced oscillations of the membrane potential. *Neuroscience, 83*(1), 15-25.

482. von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*(4654), 1260-2.

483. Von der Malsburg, C. (1981). *The correlation theory of brain function* . Gottingen, Germany: Max Planck Institute for Biophysical Chemistry.

484. von der Malsburg, C. (1995). Binding in models of perception and brain function. *Curr Opin Neurobiol, 5*(4), 520-526.

485. Von Neumann, J. (1958). *The computer and the brain*. New Haven: Yale University Press.

486. Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog Neurobiol, 51*(2), 167-194.

487. Walsh, V., Butler, S. R., Carden, D., & Kulikowski, J. J. (1992). The effects of V4 lesions on the visual abilities of macaques: shape discrimination. *Behav Brain Res, 50*(1-2), 115-126.

488. Wang, D. L., Buhmann, J., & Von der Malsburg, C. (1990). Pattern segmentation in associative memory. *Neural Comput, 2*, 94-106.

489. Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science, 272*(5268), 1665-1668.

490. Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol, 78*(5), 2336-50.

491. Watanabe, T., Harner, A. M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., & Mukai, I. (1998). Task-dependent influences of attention on the activation of human primary visual cortex. *Proc Natl Acad Sci U S A, 95*(19), 11489-92.

492. Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., Takino, R., & Miyakawa, S. (1998). Attention-regulated activity in human primary visual cortex. *J Neurophysiol, 79*(4), 2218-21.

493. Worgotter, F., Opara, R., Funke, K., & Eysel, U. (1996). Utilizing latency for object recognition in real and artificial neural networks. *Neuroreport, 7*(3), 741-4.

494. Wu, Z., & Guo, A. (1999). Selective visual attention in a neurocomputational model of phase oscillators. *Biol Cybern, 80*(3), 205-14.

495. Yen, S.-C., & Finkel, L. H. (1997). Salient contour extraction by temporal binding in a cortically-based network. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems* (MIT Press ed., Vol. 9,). Cambridge: MIT Press.

496. Young, M. P., Tanaka, K., & Yamane, S. (1992). On oscillating neuronal responses in the visual cortex of the monkey. *J Neurophysiol, 67*(6), 1464-74.

497. Zeki, S. M. (1969). The secondary visual areas of the monkey. *Brain Res, 13*(2), 197-226.

498. Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, 274*(5670), 423-428.

499. Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J Physiol (Lond), 277*, 273-290.

500. Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature, 284*(5755), 412-418.

501. Zeki, S. (1983). Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience, 9*(4), 741-765.