HAL
open science

Bases cérébrales de la catégorisation visuelle rapide -Etudes chronométriques et fonctionnelles

Denis Fize

To cite this version:

Denis Fize. Bases cérébrales de la catégorisation visuelle rapide -Etudes chronométriques et fonctionnelles. Neurosciences [q-bio.NC]. Ecole des Hautes Etudes en Sciences Sociales (EHESS), 2000. Français. NNT : . tel-00069826

HAL Id: tel-00069826

https://theses.hal.science/tel-00069826v1

Submitted on 19 May 2006

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

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

Ecole des Hautes Etudes en Sciences Sociales

Thèse

présentée pour l'obtention du doctorat de

Sciences Cognitives

Spécialité :

Neurosciences computationnelles

BASES CEREBRALES DE LA CATEGORISATION VISUELLE RAPIDE
ETUDES CHRONOMETRIQUES ET FONCTIONNELLES

par

Denis Fize

M	**O. Bertrand**	
Mme	**M. Boucart**	Rapporteur
M	**M. Imbert**	
M	**B. Mazoyer**	Rapporteur
M	**B. Renault**	
M	**S. 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

Résumé

Après un rapide rappel des principaux résultats de la psychologie et des neurosciences de la vision, illustrés par le schéma de Kosslyn, le parcours de six modèles computationnels de reconnaissance nous amène à discuter des principales alternatives élaborées pour décrire le traitement visuel – généralement compris comme complexe et récurrent. Le temps requis par ce traitement apparaît comme un critère crucial de décision sur son fonctionnement et d'affinement de notre compréhension.

Nous constatons que les données d'électrophysiologie disponibles ne permettent pas de disposer clairement de ce critère. Nous mettons alors en œuvre une tâche expérimentale visant à mesurer le temps nécessaire au système visuel humain pour analyser des scènes naturelles contenant ou non un animal. Les résultats montrent que ce traitement peut être extrêmement rapide, d'une durée inférieure à 150 ms.

Cette première mesure est complétée par quatre expériences visant à mieux cerner cette contrainte temporelle, en variant les positions des images, leurs couleurs et la tâche. Cette vitesse du traitement visuel des scènes naturelles se montre particulièrement robuste et constante : lors de présentations par hémichamps parafovéaux, lorsque l'attention n'est pas focalisée sur le lieu d'apparition du stimulus, et en l'absence de couleur comme indice de reconnaissance. Les résultats attenants montrent aussi que les catégorisations d'images contenant des formes simples et la détection de la présence de couleurs ne sont pas plus rapides. La catégorisation « animal » semble d'autant plus résulter d'un mécanisme automatique que sa trace électrophysiologique est encore présente lorsqu'une autre tâche occupe les sujets.

Les bases cérébrales de la tâche ont été recherchées à l'aide de modèles dipolaires ainsi qu'avec la création d'un protocole événementiel d'imagerie cérébrale RMN analogue à celui mis en oeuvre en électrophysiologie. Nous montrons que cette tâche de catégorisation implique de manière différentiée les aires visuelles extrastriées 19 et 31, le gyrus fusiforme et les cortex cingulaires postérieurs. Dans les aires visuelles, un effet de suppression d'activité neuronale lié à la présence d'une cible semble mettre en évidence le mécanisme de compétition postulée dans certains modèles.

Ces résultats plaident en faveur de mécanismes directs et rapides de la reconnaissance visuelle : traitement essentiellement ascendants (sans boucles) sans recentrage des stimuli latéralisés ; l'attention focalisée, la couleur et une forte acuité ne sont pas nécessaires à la reconnaissance d'objets dans des scènes complexes.

La reconnaissance visuelle postulée comme mécanisme nécessitant des traitements récurrents et des représentations complexes semble ainsi céder la place à de simples détections parallèles de traits visuels, en eux-mêmes suffisants à la représentation mentale des scènes. Dans ce cadre, la décision visuelle - le stimulus présent est adéquat à la tâche prévue – pourrait être l'extraction de ces représentations au moyen de l'inhibition des assemblées neuronales non sélectionnées.

TABLE DES MATIERES

VISION ET MODELES 4

1 le système visuel et sa modélisation 4
1.1 perception visuelle 4
a) psychologie 4
b) neurosciences 9
c) une proposition de Kosslyn 12
1.2 construction de modèles 16

2 modèles de la reconnaissance visuelle 18
2.1 reconnaissance par composants 19
2.2 réseau de régularisation 22
2.3 processus d'alignement 25
a) modèle du flux d'information 25
b) le principe d'alignement 28
2.4 routage dynamique 30
2.5 propagation asynchrone 33
2.6 encodage invariant à la vue 36

3 les solutions proposées 39
3.1 construction des représentations 39
3.2 liage et invariance à la position 40
3.3 accès aux catégories 40

4 le temps : une contrainte essentielle 41

VITESSE DU TRAITEMENT VISUEL 45

1 quelques indices de mesure 46
1.1 temps de réaction comportementaux 46
1.2 potentiels évoqués de surface 47
a) visages 47
b) autres catégories 48
c) effets d'amorce 49
d) attention visuelle 50
1.3 latences de réponse des neurones 51

2 nécessité d'une expérience dédiée à la mesure 53
2.1 la chronométrie et les potentiels évoqués 53
a) les questions posées en électrophysiologie 53
b) des traitements de données incompatibles avec notre problème 54
2.2 méthodes 55
a) le choix de la tâche 55
b) l'analyse des potentiels évoqués 56
c) l'analyse des réponses comportementales 57
d) les sources dipolaires 58

3 l'expérience animal / non-animal 59
3.1 protocole et hypothèses 59
3.2 article publié 61
3.3 sources principales de l'activité différentielle 67

4 discussions et conséquences **68**
4.1 la rapidité du traitement visuel 68
4.2 l'activité à 150 ms et autres indices de traitement 70
a) la détection de cible et la P3 70
b) l'inhibition de réponse et la N2 70
c) la sélection de traits (SN) 71
d) la focalisation de l'attention sur une cible (N2pc) 72
e) la saillance des catégories visuelles connues (les P150) 73

VARIATIONS AUTOUR DE LA CATEGORISATION VISUELLE RAPIDE **75**

1 présentations alternées par hémichamps **75**
1.1 expérience et résultats 75
a) protocole et hypothèses 75
b) résultats comportementaux 76
c) résultats électrophysiologiques 77
d) modèles dipolaires 89
1.2 discussions et conséquences 91
a) pas de supériorité hémisphérique « animal » 91
b) peu de mobilisation attentionnelle 92
c) la reconnaissance ne nécessite pas une forte acuité visuelle 94
d) vision latérale : traitement parallèle plutôt que recentrage 95

2 l'absence de couleur dans les scènes naturelles **97**
2.1 expérience et résultats 98
a) protocole et hypothèses 98
b) résultats comportementaux 98
c) résultats électrophysiologiques 99
d) modèles dipolaires 99
2.2 discussions et conséquences 107
a) le peu d'information nécessaires pour réussir la tâche 107
b) la rapidité de la reconnaissance minimise le rôle de la couleur 108
c) le rôle prépondérant de la voie magnocellulaire 110
d) une activité 'Animal' avant 150 ms 111

3 détecter la couleur dans les scènes naturelles **112**
3.1 expérience et résultats 113
a) protocole et hypothèses 113
b) résultats comportementaux 114
c) résultats électrophysiologiques 114
d) modèles dipolaires 115
3.2 discussions et conséquences 125
a) l'encodage visuel et l'adéquation à la tâche 125
b) une reconnaissance implicite de la catégorie animal 127

4 images contenant des formes simples **128**
4.1 expérience et résultats 128
a) protocole et hypothèses 128
b) résultats comportementaux 129
c) résultats électrophysiologiques 130
d) modèles dipolaires 130
4.2 discussions et conséquences 137
a) l'encodage précoce des formes simples et des scènes 137
b) décision et comportement 138

ETUDE EN IRMf **140**
1.1 à la recherche d'un protocole adapté 140
1.2 analyses et traitements des données 143
a) modélisation par système linéaire et statistique de corrélation 143
b) comparaisons de moyennes 145
1.3 article sous presse 146
1.4 conséquences 159
a) un protocole très proche des études électrophysiologiques 159
b) aires cérébrales activées et sources dipolaires 159
c) la suppression d'activité : compétition entre traits visuels ? 160

BASES CEREBRALES DE LA CATEGORISATION **164**

1 bilan des résultats **164**

2 contraintes de modélisation **166**

3 la reconnaissance visuelle : propositions **168**

4 Conclusion **173**

ANNEXE **176**

COMMUNICATIONS **179**

BIBLIOGRAPHIE **180**

VISION ET MODELES

1 le système visuel et sa modélisation

La puissance de notre vision laisse notre prétention à la comprendre bien démunie. Sa quotidienneté masque une myriade de situations diverses : reconnaissance d'objets dans des scènes mal éclairées, tâches de catégorisations, recherche de cibles au milieu de distracteurs, perception soudaine d'une forme là où on ne croyait voir que désordre… Si l'on compte l'ensemble d'objets que l'on peut reconnaître et les nombreuses représentations que notre système visuel pourrait mettre en jeu, on mesure les difficultés auxquelles se confrontent les tentatives de modéliser le simple fait de voir.

Les images que nous pouvons reconnaître sont nombreuses, au vu des quelques estimations effectuées par I. Biederman et S. Thorpe : de 30 000 pour l'un à 60 000 ou 100 000 pour l'autre, les nombres atteints sont déjà amplement suffisants pour qu'on imagine des stratégies subtiles en œuvre dans notre mémoire ou dans la façon d'y accéder. On peut classer les images en catégories de formes, construire un puzzle dont l'arrangement de sa trentaine de pièces suffit à décrire les formes de notre environnement, ou encore collecter suffisamment d'indices sur l'objet que nous regardons pour qu'à son souvenir il nous rappelle bientôt. Mais avant d'admirer la fertilité du problème visuel pour la modélisation par réseaux de neurones, il nous faut considérer ce que la psychologie et les neurosciences nous apprennent d'immédiat à propos de notre capacité à voir.

1.1 perception visuelle

a) psychologie

représentations multiples et modularité

Un point de vue assez répandu en psychologie suggère que la perception visuelle n'est pas unifiée, c'est-à-dire qu'un seul mécanisme n'est pas suffisamment général pour rendre compte de toutes les fonctions de notre système visuel. Pour les auteurs qui insistent sur ce point (Logothetis et Sheinberg 1996; Treisman et Kanwisher 1998), cette modularité fonctionnelle découle de la multiplicité des représentations mises en jeu lors du phénomène de reconnaissance, comme le montrent les études de diverses pathologies : déficit de la reconnaissance des visages (prosopagnosie), déficit spécifique de reconnaissance des expressions faciales sans prosopagnosie (un cas particulier de métamorphopsie), ou déficits de reconnaissance spécifiques à d'autres catégories que les visages (objets animés ou vivants, nourriture, objets familiers). Ainsi pour ces auteurs, à la

distinction bien établie entre un encodage de l'espace séparé de celui de la forme - et probablement sous-tendu par des mécanismes de traitement neuronaux bien différents (Wallis et Rolls 1997), se rajouteraient des mécanismes d'encodages spécifiques à la représentation des catégories visuelles de base, à l'individualisation détaillée des objets, à la représentation des informations métriques pour les manipulations, aux réponses émotionnelles (Treisman et Kanwisher 1998), et aux objets animés (Logothetis et Sheinberg 1996).

Nous ne discuterons pas plus avant de cette modularité des traitements, en nous limitant à constater d'abord l'encodage de l'espace et de la forme par des mécanismes séparés, et ensuite la mosaïque de représentations des catégories visuelles. Le fait que ces représentations soient séparées ne nécessite pas un traitement différent pour chacune, et peut simplement refléter un compromis entre sélectivité neuronale et code distribué, comme nous le verrons dans certains modèles à la deuxième et troisième partie de ce chapitre.

catégories visuelles

Regrouper les objets en catégories semble naturel quand on observe les régularités qui nous entourent. L'existence de ces catégories est bien établie, notamment par les études mesurant le temps pris pour nommer des objets observés, ou celles qui examinent l'effet d'amorce d'une vue sur la suivante. Elles montrent que ces catégories ont une base perceptuelle, reflétant la présence de certains indices visuels les plus probables dans ces classes. Certaines études ont aussi montré qu'une simple moyenne des formes des objets d'une classe est perçue comme étant membre de cette classe (Katz 1953). A cette existence de classes répond l'existence de prototypes, c'est-à-dire d'objets qui accumulent les indices d'appartenance à la catégorie, et d'objets atypiques, appartenant à la catégorie mais dont la forme possède des différences prononcées avec les prototypes. Les catégories de base (basic-level) regroupent les objets dont les traits possèdent les indices les plus pertinents pour décider de l'appartenance, s'ils sont très différents des indices décisifs des autres catégories. C'est pour ces catégories de base que les tests d'appartenance sont les plus rapides et les plus réussis.

A l'inverse, reconnaître rapidement un objet ne suppose pas qu'on accède à une catégorie de base : les objets atypiques sont généralement reconnus aussi rapidement que la classe pour laquelle ils sont atypiques - reconnaître un pingouin est même plus rapide que le reconnaître en tant qu'oiseau. Chaque objet possède son propre « point d'entrée » dans une catégorie : la catégorie de base si l'objet est typique (Rosch 1975 ; Jolicoeur et al. 1984), ou une catégorie dite sous-ordonnée (peu de membres, mais indices très significatifs) pour des objets plus particuliers (voir Boucart 1996). Quant aux catégories plus générales dites super-ordonnées, qui regroupent davantage de membres, elles sont considérées comme plus abstraites, plus « langagières », c'est-à-dire partageant très peu d'indices visuels). Les tâches d'appartenance à ces catégories enregistrent cependant des

taux de réussite élevés, même avec des temps de présentation courts (Intraub 1980). Cette multiplication d'indices nécessaires pour décider de l'appartenance à ces classes moins visuelles est une première indication d'une architecture parallèle de traitement, même sans passer par les catégories de base.

Ainsi, s'il est clair que la catégorisation d'un objet repose sur les capacités du système visuel, et que classer les objets au niveau de base est une tâche de reconnaissance fondamentale (Boucart et Bruyer 1991), reconnaître un objet ne nécessite pas de mettre en jeu une représentation de la classe de base. D'autres stratégies peuvent être mises en œuvre, d'autant plus que les représentations mises en jeu sont certainement dépendantes de la tâche demandée. Concernant les catégories, certains modèles modulaires de reconnaissance proposent même de séparer les représentations d'objets individualisés de leur classe dans chacun des hémisphères cérébraux (à gauche pour les catégories, Marsolek 1995).

supériorités hémisphériques

Une somme importante de littérature est consacrée à la mise en évidence de supériorités d'un hémisphère cérébral sur l'autre dans diverses tâches : perception visuelle, auditive, langage. Autant l'évidence montre une supériorité hémisphérique gauche dans le traitement de la parole, chez l'homme droitier (e.g. Tzourio et al. 1998), autant la supériorité hémisphérique dans la reconnaissance des objets demeure ténue et soumise à controverses. Par contre, sans être réellement spécialisé ni modulaire, il semble qu'un hémisphère cérébral puisse se révéler « supérieur » à l'autre dans l'exécution de certains traitements de l'entrée visuelle. Sans établir une revue exhaustive des différentes expériences menées, le court résumé ci-dessous cherche à préciser les principales caractéristiques de la perception visuelle que l'on trouve être différenciées entre hémisphères visuels gauche et droit.

L'hémisphère gauche se révèle plus rapide dans un traitement analytique et détaillé de l'entrée visuelle. Cet aspect est mis en évidence dans sa supériorité à traiter les hautes fréquences spatiales (Warrington et James 1986; Sergent 1987), dans la facilitation de détection d'objets par une préactivation de détails (Farah 1984), dans la classification de la fonction des objets, l'identification et le contrastes entre vues prototypiques (Hellige et Michimata 1989 ; Kosslyn et al. 1989; de Mattos Pimenta et Tiedemann 1990), ainsi que dans le jugement des relations spatiales haut / bas et droite / gauche (Kosslyn et al. 1989).

L'hémisphère droit se révèle plus rapide dans les traitements globaux, nécessitant moins d'information détaillée ou pouvant traiter des informations partielles et incomplètes. Sa supériorité est mise en évidence lors de tâches de reconnaissance de vues atypiques (Warrington et Taylor 1978), de perception des relations entre parties de l'objet et ensemble du contexte, par exemple en vue de la planification des actions (Nettleton et Bradshaw 1983), dans l'intégration de la posture, des jugements de perspective et de distances.

on mémorise des vues

On pourrait supposer que notre capacité à reconnaître un objet réel (tridimensionnel) de n'importe quel point de vue nécessite d'avoir mémorisé ses formes et contours en 3 dimensions ; lorsqu'on mesure le temps pris pour reconnaître les objets usuels, on constate dans une large mesure une certaine invariance à l'angle de vue. Cependant, les nombreux auteurs ayant étudié ce problème s'accordent pour mettre cette invariance sur le compte de notre familiarité à ces objets. Il semble au contraire que voir (apprendre) un nouvel objet implique un traitement basé sur la vue présente, et que sa mémorisation y est attachée ; reconnaître ce nouvel objet d'un autre angle de vue (généraliser) met en œuvre des processus de transformation et nécessite un temps de reconnaissance plus long que pour la vue originale ; cet effet s'estompe à partir de quelques nouvelles présentations de vues de l'objet qui devient familier, dont le temps de reconnaissance devient invariant à la vue (Bulthoff et al. 1995; Srinivas 1995).

De manière analogue aux catégories visuelles décrites ci-dessus, l'objet est plus particulièrement représenté par une ou des vues canoniques dérivant des vues les plus souvent observées (Palmer et Bucher 1981). Certaines expériences présentant un objet (un dessin hors contexte) sous un angle de vue rare (non-canonique) rapportent que le temps de reconnaissance en est notablement affecté, comme lors de l'apprentissage (Kosslyn et al. 1994).

scènes

En quoi la reconnaissance d'objets est-elle affectée par son contexte ? Le point de vue classique postule que la signification d'une scène prototypique est rapidement perçue, et qu'elle aide en retour à l'identification des objets qui y sont présents. Le fait qu'une représentation mémorisée d'une scène est activée plus vite que les représentations d'objets est soutenu par les études de détection d'objets dans des scènes manipulées (Biederman 1972 ; Biederman et al. 1982) et des mouvements oculaires lors d'exploration libre (De Graef et al. 1990). Ces expériences montrent que la globalité de la scène affecte la reconnaissance d'objet, avant d'avoir même perçu les relations spatiales de la scène. Ce schéma est cependant critiqué par Henderson (Henderson et al. 1987 ; Henderson 1992) qui propose que cet effet provienne d'un biais occasionné par les tâches expérimentales, précisément dû à la difficulté d'intégrer en mémoire de travail l'objet indentifié dans la scène manipulée, difficulté se traduisant dans le temps de réponse comportemental. Il suggère au contraire que l'analyse globale des scènes (leur catégorisation) n'interfère pas avec les processus d'identification d'objets, que les deux processus sont séparés du point de vue des informations qu'ils manipulent. De ce point de vue, l'effet de contexte constaté lors de l'amorçage d'une vue sur une suivante ou lors de la présence d'une scène autour d'un objet est dû à un même phénomène de traitement local, particulier au processus de reconnaissance ; les objets peuvent être identifiés plus rapidement lorsqu'ils sont vus en situation à cause de la proximité d'autres objets dans la région sélectionnée par l'attention

(proximité dans le temps ou en position). Ainsi Henderson refuse explicitement l'idée que des informations de haut niveau soient disponibles pour le processus d'identification. Dans une revue plus récente sur la perception des scènes, Henderson montre que l'identité des objets en vision périphérique n'a pas d'effets sur la programmation du mouvement des yeux (Henderson et Hollingworth 1999).

attention visuelle

L'attention visuelle se manifeste par deux phénomènes distincts (Treisman 1969; Treisman et Gelade 1980; Kahneman et al. 1983; LaBerge 1995) : un mécanisme de préparation avant que le stimulus apparaisse, décrit comme un pré-traitement sur les indices des objets et sur les lieux où doit s'effectuer la tâche : c'est un processus descendant. Après l'apparition du stimulus, un autre mécanisme, de sélection, filtre dans la scène les traits pertinents pour la tâche à accomplir dans la région spatiale privilégiée. Ce mécanisme de sélection est vu comme une propriété fondamentale du traitement visuel par les auteurs qui considèrent l'attention comme un mécanisme de compétition biaisée (Walley et Weiden 1973; Neisser 1978 ; Broadbent et Broadbent 1990 ; Desimone et Duncan 1995). En effet, les traits appréhendés en parallèle par le système visuel se trouvent en concurrence dans les champs récepteurs des neurones qui les intègrent ; cette concurrence, ascendante, privilégie les traits saillants tandis qu'elle privilégie dans son mécanisme descendant les traits/objets/lieux pertinents pour la tâche. Pour ces auteurs la compétition ascendante est complètement « câblée » dans le système visuel, donc indissociable des mécanismes de reconnaissance par exemple. L'attention focalisée est considérée comme requise pour la reconnaissance, pour assurer la combinaison des traits visuels. De manière classique, savoir où se trouve un objet est réalisé via un processus parallèle, tandis que savoir ce qu'est l'objet requiert une recherche sérielle prenant du temps (Treisman et Paterson 1984; Sagi et Julesz 1985 ; Kwak et Egeth 1992).

Accroître l'attention sur un lieu du champ visuel accroît aussi la quantité d'information extraite de ce lieu (Downing 1988 ; van der Heijden 1993; Lavie et Tsal 1994). Les ressources attentionnelles peuvent être attribuées à une petite partie du champ visuel avec une très haute résolution perceptive, ou être allouées sur un champ plus large mais avec un coût dans la résolution perceptuelle (Eriksen et Yeh 1985). La région privilégiée par l'attention focalisée se caractérise par des frontières nettes dans le cas d'une grande mobilisation à la préparation, et une taille variable jusqu'à un maximum de 4° ; l'efficacité du mécanisme de sélection, en terme de rapidité du temps de réaction comportemental, est inversement proportionnelle à la taille de cette région. Une seule région de l'espace peut être sélectionnée à la fois. L'exploration attentive de l'espace est décrite comme un mécanisme sériel, et la sélection peut se déplacer en temps constant quelle que soit la distance (Sperling et Melchner 1978; Sagi et Julesz 1985 ; Eriksen et Murphy 1987 ; Kwak et al. 1991). L'allocation d'attention en une région de l'espace semble prendre 30-50 ms (c'est l'avantage donné par un indice présenté avant la cible par rapport à la situation sans indice, dans les expériences de Eriksen et Hoffman 1972 ;

Posner et al. 1980 et de Saarinen et Julesz 1991); le changement d'orientation (désengagement, déplacement, ré-engagement) semble prendre environ 50-150 ms (le délai occasionné par un indice non valide, expériences de Bachinski et Bachrach 1980 ; Posner et al. 1980 ; Downing 1988) ; la durée d'attention soutenue sur un stimulus semble être de l'ordre d'une seconde au maximum, pouvant être aussi court que 50 ms dans le cas particulier des tâches de recherche visuelle.

b) neurosciences

Les résultats des neurosciences sont trop nombreux pour les résumer en quelques pages. Nous décrivons ici les principales données disponibles au sujet de la reconnaissance des objets, notamment la sélectivité des neurones le long de la voie ventrale. La latence de réponse de ces neurone sera précisée au deuxième chapitre.

architecture du système visuel

Les neurosciences de la vision nous apprennent que le système visuel est constitué de couches de traitement parallèle hiérarchiquement organisées ; la rétinotopie des premières couches s'atténue progressivement, tandis que deux voies séparées s'impliquent plus précisément dans l'identification des objets (voie ventrale V1-V2-V3-V4-PIT-CIT-AIT) et dans l'analyse de la perception spatiale de ces objets, comme la direction et la vitesse (voie dorsale V1-V2-V3-MT-MST-FST-LIP-VIP-7a) (pour revues, voir Van Essen 1979 ; Logothetis et Sheinberg 1996 ; Tanaka 1996).

premier encodage

Dans la rétine les cellules ganglionnaires répondent à l'apparition ou à l'extinction de lumière au centre ou à la périphérie de leur champ récepteur, sélectivement à la couleur. En sortie de la rétine, deux voies principales se dessinent : la voie magnocellulaire rapide véhicule une information phasique (codant les incréments et décréments lumineux) de faible résolution spatiale, tandis que la voie parvocellulaire plus lente véhicule une information tonique à plus forte résolution spatiale. Il a été admis depuis Wiesel et Hubel 1966 que l'information de couleur était portée par la voie parvocellulaire ; quelques études récentes insistent cependant sur la contribution de la voie koniocellulaire (plus lente) à l'information de couleur (Calkins et Sterling 1999).

Après le relais thalamique, le cortex visuel primaire V1 possède des neurones localement sélectifs à l'orientation, à la direction du mouvement et aux contrastes de couleur. D'autre part les réponses des neurones de V1 sont sensibles aux informations de profondeur et de direction du regard (Trotter et al. 1996; Trotter et Celebrini 1999). A mesure de la hiérarchie corticale, les neurones deviennent de plus en plus sélectifs et possèdent des champs récepteurs plus larges.

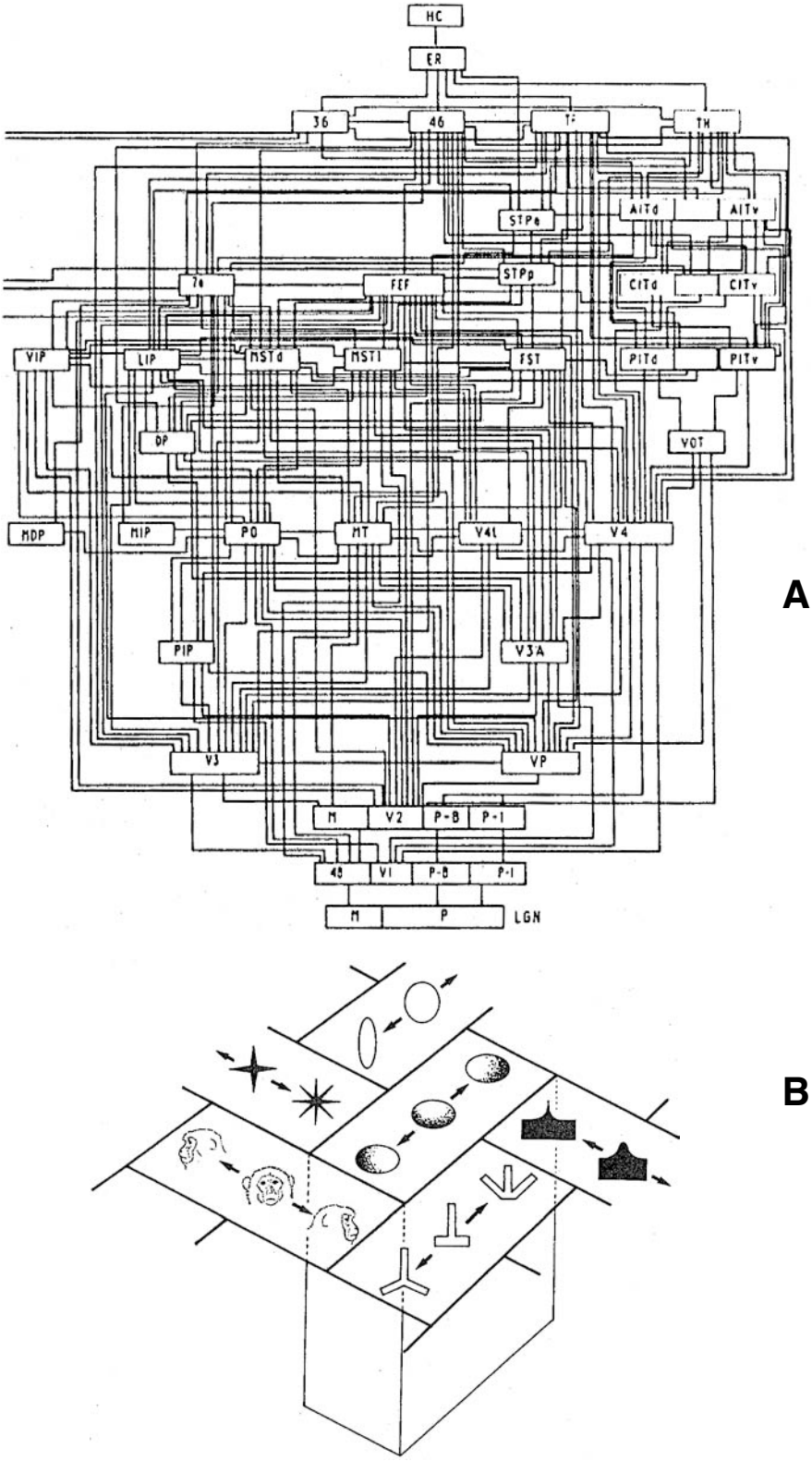


Figure 1 : Organisation du système visuel du primate.

A. Architecture des aires visuelles corticales (d'après van Essen 1990).
B. Colonnes corticales et sélectivités de neurones de l'aire inférotemporale (d'après Tanaka 1996) .

voie dorsale

Chez le primate, la voie dorsale prend essentiellement ses informations de la voie magnocellulaire. En V2, des neurones sélectifs au flux et à la direction du mouvement ont pu être enregistrés. En MT, les neurones se montrent très sélectifs à la disparité binoculaire, à la vitesse et direction de déplacement du stimulus (Mikami et al. 1986; Newsome et al. 1986; Colby et al. 1993) ; ils ne sont que très peu sensibles à la couleur (Gegenfurtner et al. 1994). D'autre part, ils modifient leur réponse en fonction de l'attention portée au stimulus (Maunsell 1995; Treue et Maunsell 1996) ; les décharges de ces neurones sont corrélées avec les réponses comportementales (Britten et al. 1996). La plupart des neurones pariétaux sont en relation avec l'activité visuo-motrice : ils sont sensibles aux paramètres du stimulus qui déterminent la posture de la main et des doigts pour un mouvement de saisie, lors des phases de fixation ou de manipulation, ou les deux. Ces neurones sont aussi sensibles à la forme (Sereno et Maunsell 1998).

voie ventrale

Après V1 et V2, les aires V3 et V4 possèdent des neurones se projetant vers le cortex inféro-temporal, voie principale de la reconnaissance des objets (Livingstone et Hubel 1987). Il est généralement admis que cette voie prenne ces informations de la voie parvocellulaire (Desimone et al. 1985). Après la sélectivité à l'orientation des neurones de V1, les neurones de V2 possède des sélectivités plus élaborées et des champs récepteurs plus larges (Hubel et Livingstone 1987) : certains neurones répondent préférentiellement à des angles d'orientation précise, des ruptures de continuité, et même à des contours illusoires (von der Heydt et Peterhans 1989). L'aire V3 semble jouer un rôle d'intermédiaire entre les aires de la voie ventrale et de la voie dorsale : nombre de neurones répondent préférentiellement à l'orientation, mais aussi à la disparité binoculaire, à la direction du mouvement et à la couleur (Felleman et Van Essen 1987). Les neurones de l'aire V4 sont sélectifs à des formes géométriques simples, mais aussi à la direction du mouvement (Desimone et al. 1985). Ces réponses sont souvent sensibles aux informations de couleur (Schein et Desimone 1990; McClurkin et Optican 1996), et aux tâches effectuées (Mountcastle et al. 1987).

La région temporale inférieure du primate peut être divisée en premier abord en deux aires : TEO postérieure (ou PIT) et TE antérieure (CIT et AIT) ; les champs récepteurs s'accroissent entre les deux aires, et autant TEO possède encore une vague rétinotopie de tout l'hémisphère controlatéral, TE n'est pas rétinotopique et possède des cellules dont les champs récepteurs incluent l'un ou l'autre hémichamp, ou les deux. Les neurones inférotemporaux sont sélectifs à la couleur, l'orientation, la texture, la direction du mouvement, et presque systématiquement à la forme, comme les visages, les mains, partie ou tout le corps animal et humain (Tanaka et al. 1991, Figure 2). Nombre de ces neurones montrent un certain degré d'invariance à la taille et à la position (quelques-uns sont sensibles à la taille), et présentent une forte sensibilité à la polarité du contraste (ombre)

(voir Logothetis et Sheinberg 1996; Tanaka 1996 pour revue). Les neurones répondant aux visages peuvent se montrer sélectifs à des traits particuliers (yeux, bouches), ou ne répondre qu'à la présentation simultanée de ces traits ; d'autres ne répondent qu'au visage entier, et sont sélectifs à la direction du regard, direction de la tête ou posture du corps (Desimone et al. 1985). Certains neurones voient leur réponses se modifier en fonction de la familiarité à des stimuli (Rolls et al. 1989). Les réponses des neurones sont généralement invariantes à la taille et à la position du stimulus, sensibles à la rotation (fort biais pour les visages verticaux), et à la vue (face, profil). Il semble que les neurones répondant à l'expression faciale et la direction du regard soient plutôt localisés dans le STS. Certains neurones de TE sont sensibles aux interactions figure-fond : ces neurones répondant à des formes précises voient leurs réponses altérées quand d'autres formes sont présentes dans le fond, et qu'elles les recouvrent (Missal et al. 1997). C'est aussi en TE encoder plusieurs objets appartenant à une même catégorie (Vogels 1999).

Ainsi le traitement de l'information visuelle suit une progression du local vers le global, ce qui permet un encodage de formes de plus en plus élaborées au fur et à mesure de la hiérarchie des étapes. D'autre part, ces aires forment un réseau densément interconnecté, et forme une structure fonctionnelle pouvant être à la fois ascendante, descendante ou en boucles récurrentes.

c) une proposition de Kosslyn

le schéma

Les principaux résultats de psychologie expérimentale et les contraintes d'architecture du système visuel peuvent être résumés dans le schéma général proposé par S. Kosslyn dans *Image and Brain* (Kosslyn 1996). Ce livre peut être considéré comme la revue la plus approfondie et la plus détaillée sur le fonctionnement du système visuel jusqu'à ce jour, faisant mention d'un grand nombre d'études expérimentales (plus de mille références). Le schéma que l'auteur propose est un modèle d'architecture fonctionnelle du traitement visuel construit pour rendre compte de ses capacités, dont le but est de détailler les modules qui le composent et de préciser les principes qui gouvernent leurs interactions. Cette approche est représentative de la psychologie cognitive, dans le sens où elle divise le modèle en un nombre minimal de sous-systèmes de traitement ou « boîtes noires » pour rendre compte des phénomènes dont l'expérience montre qu'ils sont distincts, par leurs entrées, leurs sorties, ou leurs mécanismes. Du proto-modèle proposé pour rendre compte des contraintes neuro-anatomiques de base, Kosslyn achève un schéma (voir Figure 1) qui intègre un maximum de données provenant des mesures comportementales, de pathologies et de résultats des activations cérébrales par les techniques d'imagerie.

comment un objet est-il identifié ? scénario

A partir de ce schéma d'organisation du système visuel, Kosslyn dérive une proposition pour l'identification d'objets dans une scène naturelle (ibid., pp 247-251). Ce scénario nous permettra de cerner les problèmes posés et les solutions couramment rencontrées dans la plupart des modèles de reconnaissance, que nous discuterons en fin de chapitre. La scène qui nous est proposée en exemple est composée d'un renard se tenant couché derrière une barrière, la tête tournée vers le spectateur et le corps morcelé dans l'image par la présence de l'obstacle.

Tout d'abord, un premier tampon (*buffer*) organise en unités perceptuelles les informations provenant des traitements de bas niveau : luminance, couleur, texture, déplacement et profondeur. Le fait qu'un nouvel objet apparaisse amène le *système de déplacement attentionnel basé sur le stimulus* à déplacer la fenêtre attentionnelle pour segmenter l'ensemble des pattern contigus, et plus précisément les yeux, la tête et le corps, à partir d'indices saillants et d'homogénéités de texture et de couleur. Le contenu de la fenêtre attentionnelle est ensuite repris simultanément par les systèmes ventral et dorsal.

Le système ventral code les propriétés des objets, à travers le *sous-système de pré-traitement* qui extrait les traits parallèles, les intersections et les colinéarités. Simultanément, le *sous-système de codage des relations à partir du mouvement* extrait aussi des caractéristiques de formes, mais ici l'animal est immobile. Le *sous-système d'activation d'exemplaires* tente ensuite de mettre en correspondance (matching) l'entrée avec des patterns mémorisés dépendants de l'angle de vue ; la vue n'est ici pas familière, et n'active que faiblement certains patterns *du sous-système d'activation des catégories*, comme des représentations de chien et un peu mieux de renard, à cause de la couleur. Cette activation faible engendre un pattern de retour (feed-back) vers le *buffer*, comme une image mentale (imagery), déplacée et tournée jusqu'à ce qu'elle corresponde le mieux possible à l'image d'entrée. Le recouvrement étant maintenant un peu meilleur, le pattern de catégorie est transmis à la *mémoire associative*, qui reconnaît le code comme un symbole du pattern 'renard'.

Le système dorsal code le lieu, la taille et l'orientation de toutes les unités perceptuelles, en particulier celles incluses dans la fenêtre attentionnelle transmises au *sous-système de cartographie spatiale* (spatiotopic mapping) qui encode les lieux, taille et orientation de l'objet pris comme un tout. Ces informations sont transmises au système ventral et permet la prise en compte des trois dimensions ; elles sont aussi transmises au *sous-système d'encodage des catégories spatiales* qui fournit à son tour à la mémoire associative des indications sur la vue : de face, largeur moyenne, axe principal horizontal, et au *sous-système d'encodage des coordonnées spatiales* pour les informations métriques de taille, lieu et d'orientation, ainsi disponibles pour le déplacement du regard.

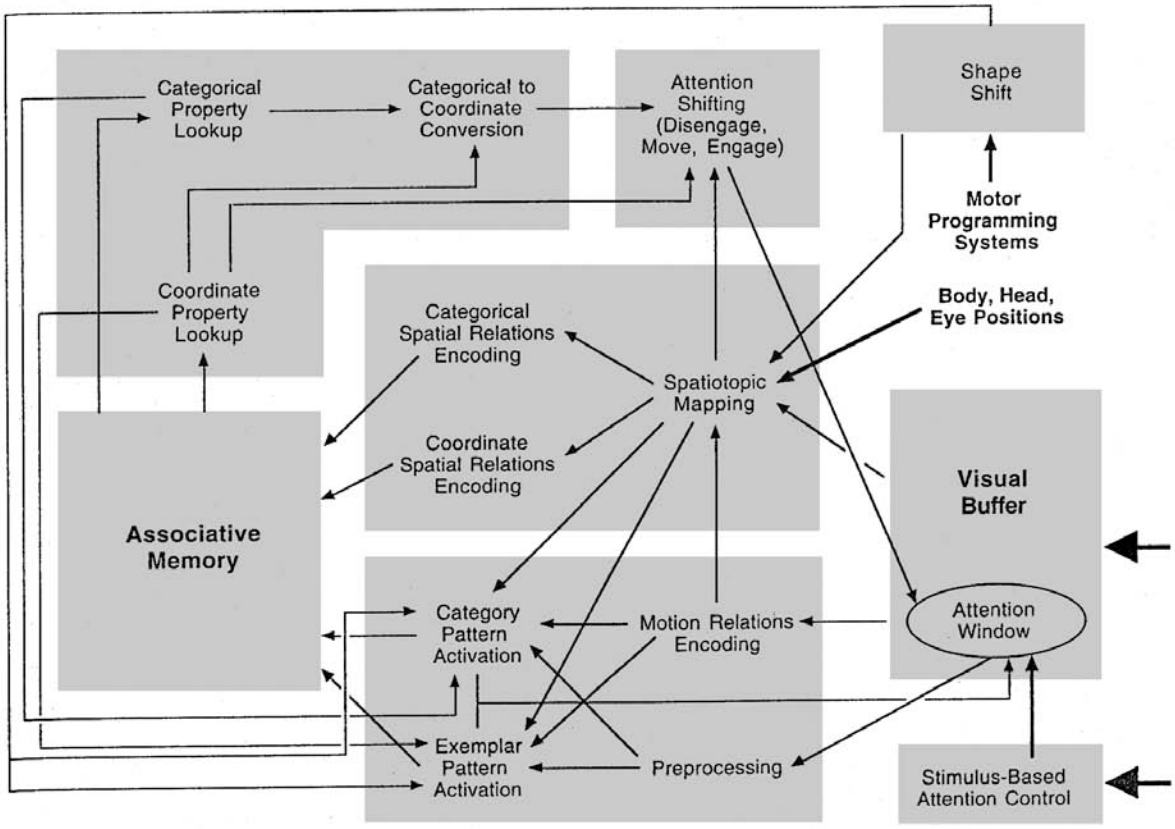


Figure 2 : Modèle de Kosslyn (1996).

L'architecture nécessaire à la reconnaissance visuelle et l'imagerie. Le schéma comprend les systèmes de traitements compatibles avec les données de la psychologie et de la pathologie. Chaque boîte noire est organisée en sous-systèmes fonctionnels (voir texte).

La mémoire associative intègre les informations provenant des deux systèmes précédents, qui pourraient être suffisantes pour l'identification de l'objet si la reconnaissance de sa forme avait été correcte ; mais dans le cas présent, étant donné l'absence d'informations de contexte venant aider à l'activation en mémoire associative de la représentation d'un renard, cette activation reste en dessous d'un seuil de significativité. Cette représentation, plus suggérée que décisive, est alors prise comme hypothèse pour des traitements descendants que nous allons maintenant résumer rapidement.

Les deux *sous-systèmes de parcours (« look-up ») des propriétés de catégories* et *de coordonnées* activent les propriétés les plus fortement encodées : il s'agit ici des relations spatiales entre composantes (taille et orientation de la tête). Ces propriétés sont transmises au *sous-système de conversion catégories-coordonnées,* qui précise les zones du champ visuel vers lesquelles *le sous-système de déplacement attentionnel* va focaliser l'attention. Dans le même temps, les représentations de ces composantes sont amorcées en retour dans les sous-systèmes d'activation des patterns, inhibant les représentations des autres formes. Le système s'engage alors dans un second cycle de reconnaissance, pour lequel la représentation de la tête d'un renard sera privilégiée, et l'attention focalisée sur cette partie de l'image. Ce genre de cycle est répété autant de fois qu'une identification satisfaisante le nécessite, générant à chaque boucle une hypothèse à tester.

les choix du modèle

Qu'apporte Kosslyn en supplément des premiers indices de fonctionnement que nous avons énoncés plus haut ? Nous pouvons préciser deux mécanismes nécessaires au schéma qu'il nous propose : le déplacement de la fenêtre attentionnelle pour chaque objet à reconnaître, et le rôle important de l'imagerie mentale pour compléter l'identification.

Notre capacité à identifier des objets à plusieurs endroits de la rétine réside pour Kosslyn dans le déplacement de la fenêtre attentionnelle dans le *buffer*, par un mécanisme dirigé par les traits saillants du stimulus (p.114, 233) - le modèle d'Olshausen (cf. Partie 2.4) en décrit un mécanisme neurobiologique. Cette proposition vient du choix d'organisation des représentations (p.79) : soit elles sont multiples et rétinotopiques, soit il existe un mécanisme de déplacement qui centre le pattern d'entrée pour qu'il soit apparié à l'unique représentation de l'objet en mémoire. Kosslyn indique qu'il peut s'agir d'un continuum, mais les alternatives qu'il propose ne sont que des conjugaisons des deux extrêmes : quelques représentations (une pour chaque hémichamp, ou chaque cadran), et des déplacements en conséquence. Cette alternative nous semble un peu abrupte, et marque sans doute une limite de l'approche des « boîtes noires » ; mais nous avons vu que postuler un déplacement attentionnel pour la reconnaissance invariante en position se traduit en temps de traitement par un retard de 30 à 50 ms par déplacement : nous voyons apparaître l'importance du temps comme paramètre de la reconnaissance visuelle, paramètre que nous déclinerons tout au long de cette étude.

De manière plus critique, Kosslyn donne dans son schéma une grande place aux processus descendants chargés à plusieurs reprises d'amorcer des représentations lors de boucles successives de traitement. La reconnaissance des formes d'après son modèle nécessite souvent une représentation en retour de l'objet le plus probable (imagery feed-back of the best-matching object), parfois ajusté en taille, lieu et orientation au pattern d'entrée par des sous-systèmes localisés dans les lobes frontaux (cortex dorsolatéral préfrontal, FEF). Par principe, ce mécanisme d'imagerie n'est utilisé que lorsqu'une forme n'est reconnue « qu'en dessous d'un certain seuil », et si les autres sous-systèmes ascendants n'ont pu fournir d'information complémentaire concordante (l'extraction de la forme à partir du déplacement). Mais l'auteur a recours à cette imagerie dans bien des cas : lorsque des objets sont perçus d'un nouveau point de vue (p. 259), lorsqu'un objet est composé de parties dont les formes ou leurs relations spatiales peuvent varier (p. 235 et 260), lorsque certaines parties ne sont pas visibles (p. 122), lorsque l'image est dégradée en luminance ou vers le flou (p. 225).

Si l'imagerie mentale fait partie intégrante des capacités du système visuel, Kosslyn l'utilise dans son modèle comme mécanisme privilégié de la reconnaissance, au point que cette mise en œuvre des processus descendants résout la plupart des problèmes sur lesquels butent beaucoup de modèles de la vision. Cette solution nous semble ne pas correspondre aux constats de rapidité du système visuel dans les tâches d'identification et de catégorisation en neurosciences. Mais avant de discuter de cette rapidité, nous devons interroger les solutions proposées par divers modèles neuronaux de reconnaissance.

1.2 construction de modèles

réseaux de neurones formels

Depuis les années 80 où les réseaux de neurones formels sont utilisés par les techniques d'ingénierie, on sait que les réseaux multicouches sont capables de résoudre tout un ensemble de problèmes d'analyse de données, de statistiques et de traitement du signal. Les tâches abordées par ce genre de réseau vont de la classification à l'optimisation et contrôle, en passant par l'analyse de la parole, des images et des séries chronologiques. On a pu voir ainsi se développer un ensemble d'applications puissantes (séparations de sources, discrétisation, simulation de fonction inverse, reconnaissance de formes) qui ont toutes en commun de réaliser par réseau ce qui est communément appelé en analyse de données des analyses en composantes principales (ACP) et des analyses discriminantes.

Dans le domaine de la vision artificielle, ces réseaux sont basés sur le schéma de Kohonen 1982, où l'information se propage de couche en couche en modifiant à chaque étape l'espace de représentation, et qui est capable de conserver ou non la structure topologique de l'entrée par un mécanisme d'interactions latérales (après les travaux de Marr 82 et les premiers systèmes artificiels en réseaux de neurones comme le Cognitron de Fukushima 1975). L'espace de représentation d'une couche est caractérisé uniquement par

les poids synaptiques associés aux neurones de cette couche, et peut être décrit mathématiquement par un ensemble de fonctions de base. Le calcul effectué sur chaque couche est une simple projection de la fonction d'entrée sur l'ensemble de fonctions de base, et fournit donc comme résultat une distance entre les deux fonctions. De ce schéma simple ont pu être déclinés une grande variété de réseaux et d'algorithmes, dont on connaît bien les capacités de généralisation à partir d'exemples appris et les paramètres de modélisation.

Les systèmes industriels qui réalisent des tâches de reconnaissance d'objets se limitent à résoudre le problème de reconnaissance dans des situations particulières. Leur fonctionnement a inspiré un certain nombre de modélisateurs qui ont adapté ces techniques selon une double exigence : restreindre les algorithmes et architectures à des mécanismes biologiquement plausibles, et en même temps étendre les capacités de ces systèmes au plus grand nombre de capacités connues du système visuel, comme recensées au début de ce chapitre. Il s'agit donc pour ces démarches d'une « ingénierie inversée », supposant des allers-retours entre capacités connues, données neurophysiologiques et modèles.

quel niveau de description ?

De la proposition de Kosslyn d'une certaine architecture et d'un traitement faisant large place à l'imagerie mentale, nous pouvons retenir (outre les prédictions propres à ce modèle) la démarche de modélisation : les modèles du système visuel doivent mimer son comportement et rendre compte de ses capacités. Son modèle fait des hypothèses fortes sur le fonctionnement global et sur les bases cérébrales des sous-systèmes impliqués ; sa validation est donc cherchée dans la mise en œuvre de tâches de psychologie expérimentale, avec le soutien des techniques d'imagerie. Mais quelles précisions peut-on attendre des modélisations par réseaux de neurones ?

D'après Marr, on peut analyser le fonctionnement cérébral en termes algorithmiques, et distinguer, comme le fait l'informatique, le niveau des traitements, de l'algorithme et de l'implémentation : un traitement peut être mis en œuvre par plusieurs algorithmes, et il y a de multiples façons de réaliser un même algorithme par des réseaux de neurones. Que peut donc apporter un réseau de neurone, si ce n'est une façon de plus de réaliser la fonction de la « boîte noire » du schéma général ? Cette assertion est vraie de l'art de l'ingénieur, qui connaît précisément l'outil et ne sera pas surpris des fonctions de son processeur. Or il semble que nous observons tous les jours qu'il n'est pas si aisé de distinguer les questions du « quoi », du « comment » et du substrat par lequel ils se réalisent ; cette distinction fait plutôt partie des buts à atteindre...

La modélisation nécessite de prendre en compte tous les niveaux, étant donné que les représentations utilisées contraignent les calculs qui peuvent être effectués - et inversement. De même on peut supposer, au vu des résultats expérimentaux, que certaines opérations existent et les modéliser, mais on n'en saura rien de plus tant qu'on n'aura pas trouvé des neurones qui réalisent effectivement des opérations analogues, par leur

sensibilité et leur connexions. Cette construction de modèle est illustrée par le « triangle des neurosciences cognitives » de Kosslyn (capacités, cerveau, traitement) : chaque précision apportée à l'un des trois sommets a un retentissement direct sur les deux autres, dans le sens qu'elle nécessite d'adapter la façon de les concevoir. Cette satisfaction de contrainte conduit à la construction de modèles dont la puissance d'explication vient plutôt du « grain » avec lequel ils décrivent le système, que du type de procédure qu'ils modélisent.

Le grain du modèle de Kosslyn est au niveau de l'architecture globale et de ses articulations ; le grain des modèles par réseaux de neurones arrive au niveau des connexions, des types de signaux véhiculés et leur durée, ainsi que du codage des représentations. La mise en œuvre de modèles par ces réseaux exige donc des précisions sur chacun de ces critères, et ne peut se contenter de calquer les solutions provenant des techniques d'ingénierie.

2 modèles de la reconnaissance visuelle

L'exigence et la précision que nous venons d'énoncer sont loin d'être réalisées dans tous les modèles neuronaux de la vision. Ceux-ci sont donc diversement ambitieux, et se contentent parfois de modéliser un fonctionnement global (Rueck, Cave et Kosslyn, 1989) (Vetter, Hurlbert et Poggio, 1995) : ces modèles ne sont pas tous d'un grand intérêt concernant leurs prédictions. D'autres s'attachent plutôt à proposer des mécanismes précis de traitement (modèles de Biederman, Grossberg, Thorpe-Gautrais, Wallis-Rolls) ou des types de représentations (modèles de Ullman, Biederman, Mel). Parmi les modèles présentés ci-dessous, certains proposent de détailler leurs solutions qui puisent leur inspiration dans divers domaines : la psychologie expérimentale (Biederman), l'ingénierie de la vision artificielle (Vetter, Hurlbert et Poggio, Ullman), ou sont construit pas à pas en intégrant les données des neurosciences (Thorpe-Gautrais, Wallis-Rolls).

Les plus anciens modèles sont ceux construits d'après les idées de Marr : ces modèles ont mis en œuvre des représentations 3D de manière à permettre une reconnaissance indépendante de l'angle de vue (Marr et Nishihara 1978 ; les primitives volumiques de Biederman forment aussi une description 3D), ou des représentations 2D analogues à des vues canoniques incluant des informations de profondeur mais devant être normalisées (l'alignement de Ullman 1989). Les modèles plus récents mettent en œuvre des représentations 2D multiples, utilisées soit par combinaison linéaire (Ullman 1992), soit par interpolation via des fonctions de base (Vetter, Hurlbert et Poggio). Les approches les plus récentes utilisent simplement la projection de l'entrée sur des collections de vues 2D, sans nécessiter de transformation ou d'extraction d'information de profondeur (Wallis-Rolls, Thorpe-Gautrais).

2.1 reconnaissance par composants

principe

Hummel et Biederman 1992 proposent un réseau de neurones qui reconnaît les objets à partir de la description explicite de leur structure. Il s'agit du modèle de référence en psychologie de la vision, qui a pu représenter en quelque sorte l'état de l'art de la reconnaissance visuelle jusqu'à ces dernières années. Le modèle est construit à partir de la théorie de la reconnaissance par composants (RPC, Biederman 1987), qui postule que les objets sont représentés par une trentaine de primitives volumiques, les géons, et de leurs relations spatiales. Ces géons sont codés dans le réseau par un ensemble d'attributs (traits d'angles et de contours, axe principal, surface, position) liés dynamiquement par synchronie ; les relations spatiales entre les géons sont représentées par des neurones codant explicitement les rapports de position, taille et orientation relatifs à deux géons, calculés à partir de leurs coordonnées. L'ambition du modèle est de montrer que la théorie RPC permet un mécanisme de reconnaissance plausible et réalisable si le modèle implante un liage dynamique.

architecture

Le réseau comporte 7 niveaux, les deux derniers n'étant implantés que pour classer les objets à partir de la description réalisée progressivement dans les cinq premiers. Les deux premières couches sont rétinotopiques, extrayant les contours, sommets, axes principaux et surface du dessin entré. La couche 3 comporte des neurones représentant un ensemble d'attributs dont chaque arrangement code un géon particulier : forme de l'axe droit ou courbe, orientation de l'axe, section droite ou conique, parallélisme des côtés, allongement, taille, position en abscisse et ordonnée. Ces attributs sont calculés simplement à partir des neurones de la couche 2 ; les attributs codant un géon déchargent de manière synchrone, permettant aux neurones de la couche suivante de discerner chaque géon. Les couches 4 et 5 réalisent le codage des relations entre géons. Les attributs de position ou de taille de chaque géon convergent et se répartissent les dimensions (couche 4) des relations codées en couche 5 : plus grand / plus petit qu'un autre géon, à côté de, devant / derrière, perpendiculaire à, oblique à. Les couches 3 et 5 fournissent donc une description de structure du dessin présenté au réseau en termes de géons et de relations entre eux.

liage par synchronie

Comment les attributs codant un géon peuvent-ils décharger de manière synchrone ? Hummel et Biederman proposent qu'un neurone décharge en fonction de l'activité de ses entrées et d'une entrée particulière seuillée (analogue à la base d'un transistor). Cette entrée particulière suppose l'existence d'un signal se propageant le long de liens entre neurones, ce signal ayant comme principale propriété de déclencher immédiatement la

décharge de tout neurone ayant l'activité suffisante. Deux neurones partagent ce signal si l'une des trois conditions suivantes est remplie : s'ils codent localement la même orientation, si la conjonction de leur projection définit un neurone sélectif à un sommet (Figure 3), ou si ces deux neurones, distants, codent des terminaisons opposées et colinéaires (contour illusoire) en couche 2. La solution proposée est donc un signal qui (1) synchronise les neurones qui suivent un contour dessiné, et (2) permet la distinction de chaque géon (non lié par un sommet).

capacité et plausibilité du modèle

Ce modèle propose donc une reconnaissance invariante à la position, taille, image miroir ou rotation en profondeur pour des objets dessinés. Il propose deux hypothèses : la segmentation des géons s'effectue par un mécanisme de liage synchrone ; il existe des neurones qui codent les relations spatiales entre composantes. La théorie RPC postule des baisses de performances sensibles lors de rotations dans le plan de vue analogues à celles mesurées en psychophysique. De nombreuses expérimentations ont indiqué la plausibilité d'un codage à base de composantes, notamment les études des effets d'amorçage d'une vue sur l'autre (Biederman et Cooper 1991). Concernant le liage par synchronie, Gray et al. 1989 ont montré que les activités dendritiques de neurones de l'aire 17 du chat répondaient aux contours illusoires avec des fréquences corrélées, mais cette expérience ne permet pas d'en savoir plus sur une éventuelle synchronisation de phase.

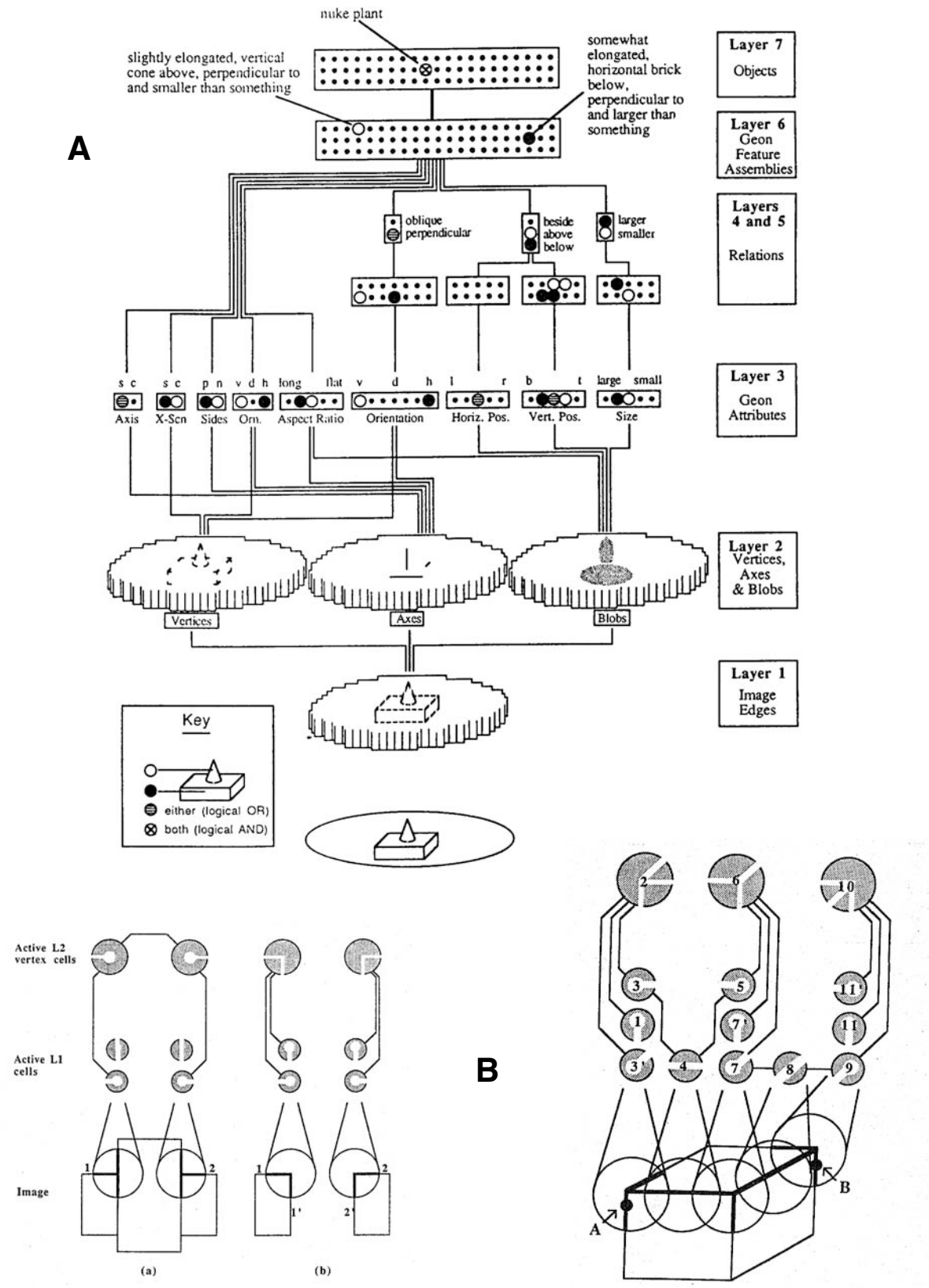


Figure 3 : Modèle de Hummel et Biederman (1992).

A. Architecture. Les couches 3 et 4 codent explicitement les attributs et les positions relatives des géons.
B. Groupement des attributs par liage dynamique.

2.2 réseau de régularisation

un réseau classique

Le modèle proposé par Vetter, Hurlbert et Poggio pour la reconnaissance d'objets (Vetter et al. 1995) est issu directement de l'ingénierie des réseaux de neurones. Ce réseau est classique quant au fonctionnement et aux capacités : son architecture est inspirée de la classe des réseaux dits de régularisation qui résolvent les problèmes d'interpolation-approximation d'images ou de signaux ; l'ensemble des fonctions de base est choisi dans la famille des gaussiennes, hypothèse habituelle pour la modélisation des poids synaptiques. Ce modèle a peu d'ambition explicative : il « ne cherche pas à rendre compte de structures corticales précises mais tente de préciser les propriétés générales que l'ont peut attendre dans une architecture biologique pour la reconnaissance d'objets ». Les auteurs ont plus précisément cherché à démontrer qu'une reconnaissance par vue est possible, sans postuler de description de structure comme le modèle précédent.

module de reconnaissance

Le cœur du modèle est un module de base à trois couches (Figure 4), utilisé en fonctionnement (vs apprentissage) comme un classifieur de vues : chaque vue apprise active un neurone particulier. Lorsqu'on présente au classifieur une nouvelle vue d'un objet appris (en généralisation), une combinaison propre de ces cellules est activée qui peut être considérée comme une vue intermédiaire : l'objet est codé par l'activité simultanée de toutes ces cellules. Le point important dans ce modèle est que ces représentations intermédiaires ou combinaisons restent très sélectives de l'objet 3D et ne sont pratiquement pas activées par des vues d'objets distracteurs : il s'agit de cellules très sélectives, ne réalisant pas à proprement parler un codage par population. L'apprentissage de ce réseau est supervisé (rétro-propagation du gradient).

architecture d'un système complet

Les auteurs proposent de concevoir un modèle du système visuel basé sur des classifieurs ; ce modèle, non implanté, posséderait une architecture constituée de trois étapes : localisation et ségrégation des autres objets réalisées par un classifieur à partir de vues à basse résolution et filtrées ; cette première étape permet d'accéder à la classe de l'objet. Une deuxième étape réaliserait des transformations de classe (ex. visages), notamment une normalisation en translation, en taille, en rotation, calculées à partir de traits spécifiques de la classe (ex. yeux, bouche) ; L'hypothèse est que cette étape génère autant de vues virtuelles que d'opérations effectuées. La troisième étape est la classification de ces vues ainsi normalisées au niveau de l'exemplaire.

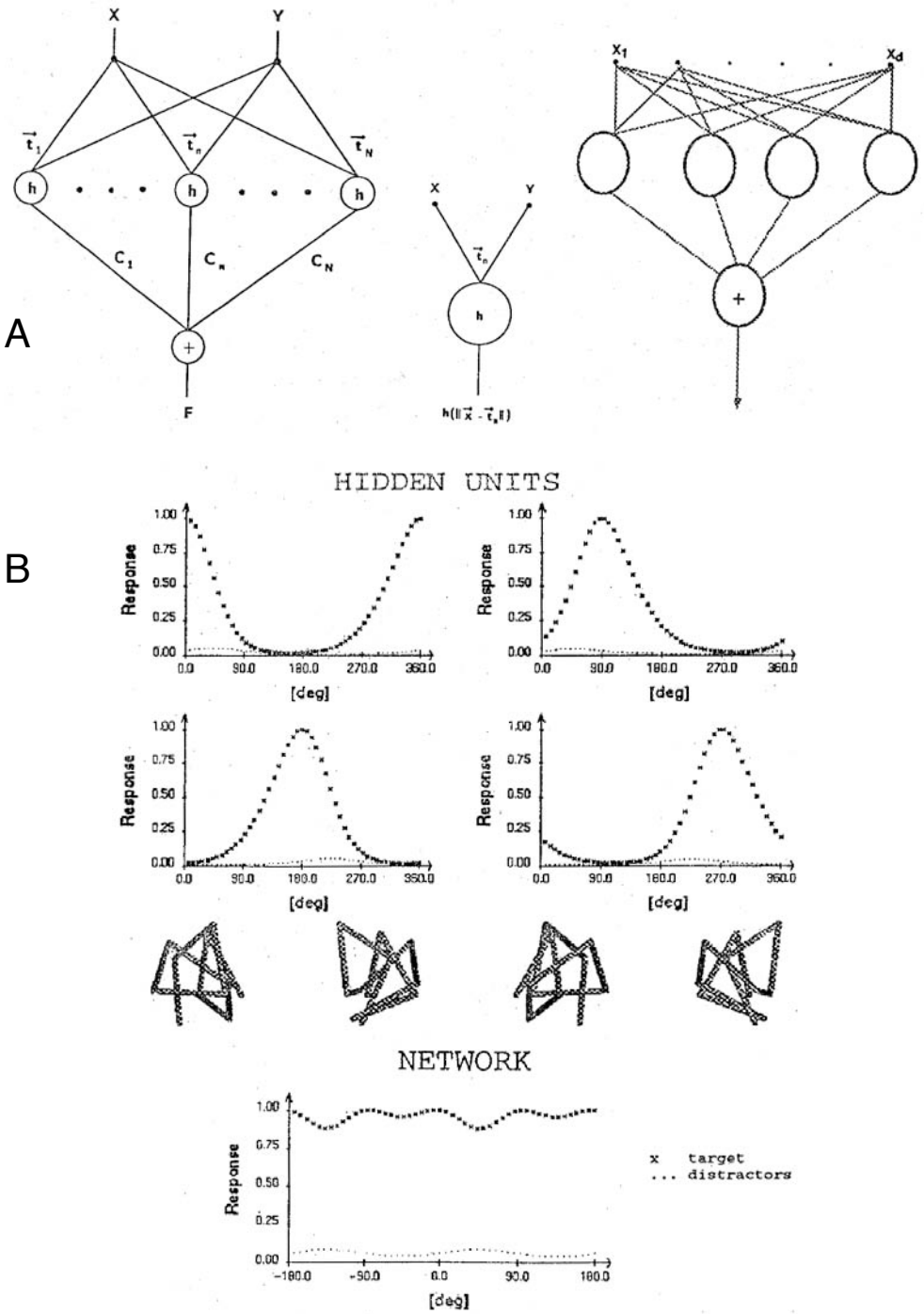


Figure 4 : Modèle de Vetter, Hurlbert et Poggio (1995).

A. Classifieur à deux et à plusieurs entrées, totalement connectées à la couche cachée.
B. Sélectivité des neurones de la couche cachée aux vues d'un même objet. Les couches de sorties sont spécifiques à l'objet et invariantes à la vue (en bas).

mécanisme

La classification est ici le principal mécanisme de reconnaissance. Le modèle est basé sur l'idée d'un grand nombre de vues mémorisées, notamment pour la reconnaissance au niveau de la classe suivant un mécanisme ascendant. Quand peu de vues mémorisées sont disponibles pour reconnaître des exemplaires, la classification est rendue possible par l'extraction de traits caractéristiques à l'objet s'ils existent, ou en générant des vues virtuelles issues de transformations spécifiques à la classe. C'est par ce mécanisme qu'un objet non-familier appartenant à une classe usuelle peut être reconnu : des vues virtuelles peuvent être générées à partir d'objets familiers de la même classe.

capacités et prédictions du modèle

Les auteurs montrent ainsi qu'un classifieur est capable de reconnaître des nouvelles vues de visages appris, sans qu'il soit nécessaire de coder explicitement des informations spatiales. Parmi les « propriétés générales » prédites par le modèle, les auteurs décrivent des cellules des couches intermédiaires centrées sur la vue et des cellules de sortie qui en sont indépendantes ; des cellules sélectives à la pose, à l'expression, à l'illumination, aux vues d'objets complets (rares) et aux vues de composantes. Un système visuel implantant ce principe réaliserait un codage sélectif à une vue pour un nouvel objet, et un codage invariant pour des objets familiers.

Concevoir la reconnaissance visuelle comme une classification de vues amène une distinction entre traitement de vues familières et non-familières. Cette distinction était aussi présente dans le modèle de Biederman, dans lequel le problème des vues non-familières est identique à celui des composantes altérées par masquage ou par modifications importantes (une rotation par exemple). La différence entre ces deux modèles vient plutôt du choix d'élaborer progressivement une représentation, ou de comparer directement, comme un tout, les traits qui la composent (*template matching*).

Ce modèle n'apporte pas vraiment de prédictions sur le type de fonctionnement du système visuel. La solution d'ingénierie proposée est simplement compatible avec les propriétés physiologiques des neurones, et montre qu'une reconnaissance globale de la vue est possible. Les auteurs ont cependant besoin de l'hypothèse supplémentaire des vues virtuelles nécessaires à la reconnaissance ; une autre hypothèse, implicite, est que le réseau possède une connectivité totale, condition qui semble loin d'être réalisée dans le cortex visuel.

2.3 processus d'alignement

Le modèle de Ullman pour la reconnaissance visuelle comprend deux composantes assez distinctes : un processus de reconnaissance proprement dit, pour lequel l'auteur propose une réalisation analogue à un alignement entre le pattern d'entrée et une vue mémorisée, et un modèle de flux d'information dans les voies visuelles développé plus récemment (Ullman 1995) qui cherche à rendre compte de la connectivité des aires corticales en proposant un type de traitement bidirectionnel.

a) modèle du flux d'information

Le modèle de Ullman (Figure 5) étend et modifie l'idée d'étape de pré-traitement (détection d'angle, extraction de traits de différente complexité) en vue de mettre en œuvre deux stratégies principales. La première est d'employer une recherche bidirectionnelle dans laquelle l'appariement du pattern d'entrée et de l'objet mémorisé peut se faire aux niveaux intermédiaires (plutôt qu'en fin de hiérarchie des couches) ; la seconde est d'explorer plusieurs alternatives de traitement en parallèle. Le modèle pose comme préalable que les représentations mémorisées ont une structure d'image ou de vue comme le modèle précédent - une « description picturale », et que les étapes de traitement sont autant de transformations géométriques effectuées sur ces représentations.

traitement bidirectionnel

Ullman propose que le système visuel transforme à la fois l'entrée visuelle et les représentations d'objets mémorisées. Le pattern d'entrée subit des transformations standard comme la translation ou la normalisation de la taille, tandis que les représentations mémorisées subissent des transformations spécifiques à leur classe, par exemple la façon dont un objet 3D apparaît sous plusieurs angles de vue, ou comment un visage peut être déformé par des expressions typiques. Ces deux types de traitements empruntent respectivement les voies ascendantes et descendantes du cortex, et se réalisent finalement dans la même zone corticale lorsqu'il y a appariement entre entrée et modèle mémorisé.

explorer plusieurs alternatives en parallèle

Les processus ascendants extraient les propriétés comme la couleur, la texture, le mouvement, les indices spatiaux, la position et l'orientation, et cela parallèlement à plusieurs échelles, ainsi que les parties et composantes. Dans le même temps, des vues différentes d'une même représentation mémorisée seront manipulées en de multiples orientations et positions, conditions d'illumination, et soumises aux transformations spécifiques à la classe de l'objet représenté. Tout ces patterns intermédiaires générés sont représentés au niveau des étapes de calcul : les nœuds de ce réseau schématique sont autant de populations de neurones impliqués dans le traitement.

Lien entre les voies ascendantes et descendantes

Ce schéma repose sur deux particularités : (1) les voies ascendantes et descendantes sont deux réseaux séparés et complémentaires ; (2) un traitement effectué dans une voie laisse une trace d'amorçage dans l'autre voie, la rendant plus excitable. Un lien est ainsi créé lorsqu'un nœud (un groupe de neurones) de la voie ascendante et son nœud complémentaire sont activés dans un intervalle de temps limité (une centaine de ms). Cette activité peut alors « remonter » jusqu'à la représentation mémorisée, le long des nœuds déjà parcourus ; les processus descendants préparent donc la voie du processus ascendant sélectionné. C'est donc à n'importe quelle étape de traitement que peut s'établir l'appariement de l'entrée et du pattern mémorisé.

Activation des modèles mémorisés

Les modèles candidats à la source des traitements descendants doivent être activés dès l'origine ; Ullman propose deux façons de répondre à ce problème de sélection initiale. Quelques modèles peuvent être sélectionnés à la suite d'analyses ascendantes partielles, par le biais de « voies rapides » correspondant aux liaisons directes entre aires de bas niveau et de haut niveau (comme V4-AIT). Un autre mécanisme d'initiation des modèles pourrait être fourni selon Ullman par les effets d'attente et de contexte.

plausibilité

Les connexions entre aires corticales peuvent être classifiées en ascendantes, descendantes et latérales, sur la base de la distribution laminaire des sources et destinations des projections des neurones : par exemple la hiérarchie ascendante V1-V2-V4-PIT-AIT pour la voie magnocellulaire est établie à partir des projections de la couche corticale 4 vers une population de neurones de la couche superficielle, qui projette elle-même vers la couche 4 de l'aire suivante ; de même, les connexions descendantes sont connues pour traverser une autre population des couches superficielles et joindre une sous-population des couches infra-granulaires (souvent couche 6). Le modèle de flux nécessite au moins 5 couches laminaires, 2 pour chaque flux et au moins une pour une couche de « contrôle » : ce nombre du même ordre de grandeur que le nombre de couches corticales existantes. Le modèle prend en compte le fait que les connexions entre les couches corticales 4 et 6 soient essentiellement excitatrices, et l'existence de connexions de la couche 4B vers les couches 1 et 3.

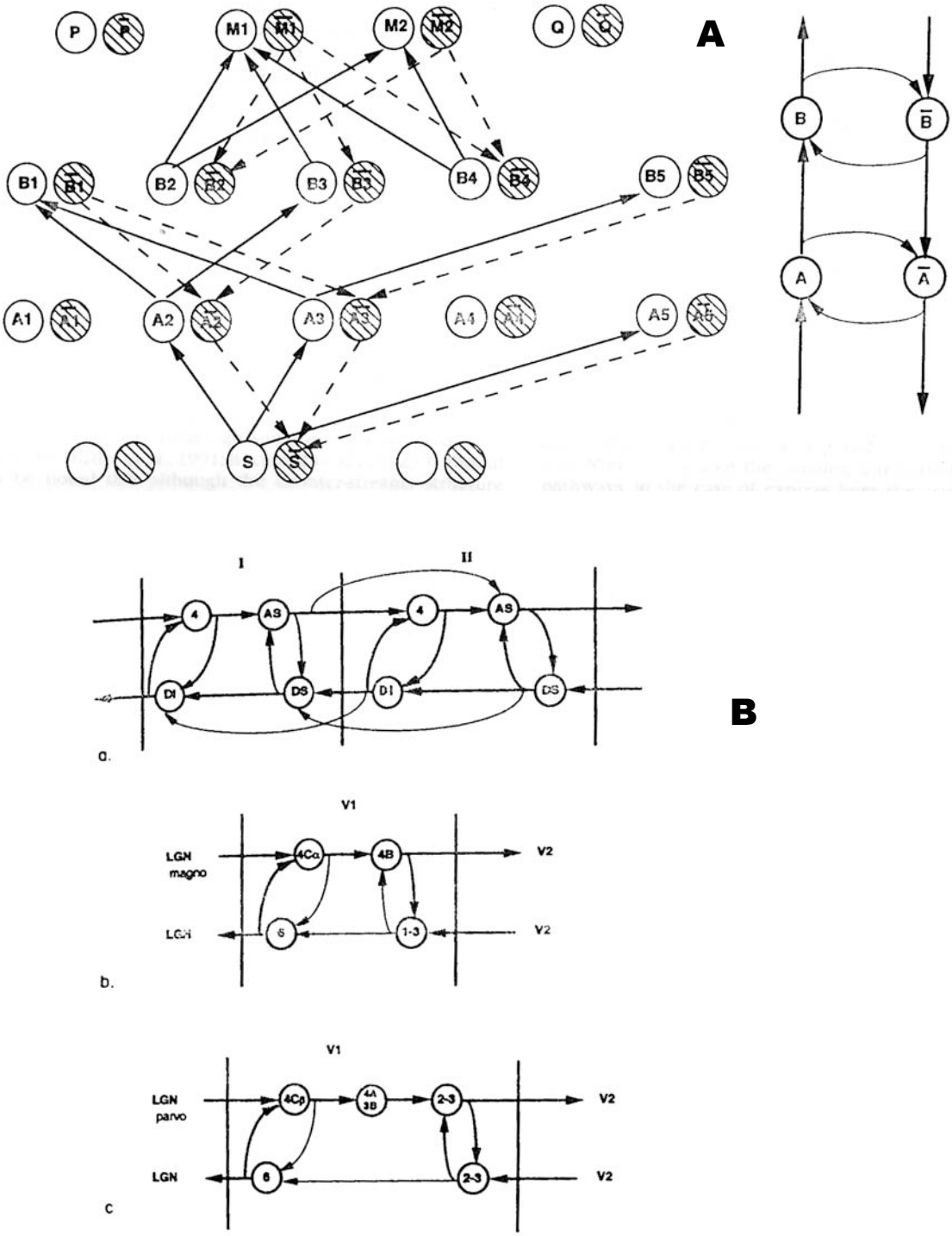


Figure 5 : Modèle de Ullman (1995).

A. Processus ascendants et descendants le long de voies parallèles complémentaires. Chaque activation d'un nœud amorce l'activation de son complémentaire.

B. Présentation des couches corticales de V1 permettant l'analogie. Les flèches fines sont les prédictions du modèle ; les flèches épaisses les connexions établies.

prédictions du modèle

L'auteur dérive plusieurs prédictions sur la connectivité des neurones de l'aire V1 : il prévoit l'existence de projections des couches 1 et 3 vers la couche 4B, de la couche 4Cα vers la couche 6 magnocellulaire, et l'existence de projections de la couche 4Cβ vers la couche 6 parvocellulaire. Physiologiquement, le modèle prévoit un effet réciproque d'amorçage entre les neurones des voies ascendante et descendantes dans la même aire, réalisé sur deux populations différentes de neurones. D'autre part, le modèle prévoit dans son principe des traitements descendants spécifiques aux classes d'objets mémorisées (compenser les effets d'illumination, de distorsion spécifique à la classe, comme dans le modèle de Vetter, Hurlbert et Poggio 1995) et de traitements ascendants de discrimination, filtrage et segmentation. Contrairement au réseau de régularisation, ces traitements descendants sont nécessaires au processus de reconnaissance.

spécificité du modèle de flux

Les voies descendantes jouent un rôle crucial dans ce schéma de reconnaissance en manipulant les objets mémorisés et en préparant la voie aux processus ascendants. Dans d'autres modèles ces voies sont plutôt employées dans des rôles d'attention sélective, de groupement et de ségrégation figure-fond, apprentissage, ou synchronisation neuronale. L'appariement entre entrée et modèle mémorisé peut se faire ici à tous les niveaux, et ne nécessite pas que des parties d'objets, si le modèle en inclut, soient précisément définies comme dans le modèle de Biederman.

b) le principe d'alignement

combinaison de vues

L'auteur propose que l'encodage d'objets réels se réalise par un encodage de quelques vues de l'objet sous différents angles. Mais contrairement à l'approche de Vetter, Hurlbert et Poggio, pour lesquels un très petit nombre de neurones réalise l'encodage d'une vue, Ullman suggère des représentations plus distribuées, des neurones pouvant être sélectifs à des parties spécifiques (composantes) ou des formes plus simples qui sont partagées par des représentations différentes. Par « vue mémorisée », Ullman entend donc un pattern d'activation d'un nombre important de neurones, pattern spécifique à une vue de l'objet, mais qui partage avec les autre vues de cet objet des activations communes.

alignement

Dans l'hypothèse d'alignement, reconnaître un objet équivaut à manipuler le pattern d'entrée pour qu'il « colle » à un pattern mémorisé, en extrayant des clés d'alignement comme les angles ou axes d'élongation ; si le pattern d'entrée est un pattern proche de celui mémorisé, il sera reconnu sans traitement supplémentaire : il est déjà « aligné » avec

lui. Mais si l'objet est vu sous un autre angle, sous des conditions d'éclairage, de taille différentes, le pattern d'entrée nécessitera des manipulations compensatrices qui généreront autant de patterns à comparer avec les vues mémorisées.

spécificité

Le processus est donc très différent d'une élaboration progressive d'une description de l'objet. Sans pour autant être une classification de vues par un réseau entièrement connecté, cette reconnaissance est comparable à l'approche de Poggio quant aux vues virtuelles considérées comme autant d'hypothèses à tester, comme le prévoit le schéma de Kosslyn. Ces représentations sont picturales et peuvent inclure des composantes comme dans l'approche structurale ; mais ces parties sont associées à un lieu sans le secours d'une catégorie de position (à côté de).

plausibilité

Schiller 1995 a montré que des lésions en V4 ou PIT affectent la capacité à reconnaître des images modifiées en taille, orientation ou illumination, alors que les images originales continuent à être facilement reconnues : en ce sens, les neurones des aires touchées peuvent être considérés comme effectuant la transformation compensatrice permettant d'apparier pattern d'entrée et vue mémorisée. De ce point de vue, les modifications de sélectivité des neurones au cours du temps (Richmond et Optican 1990 ; Oram et Perrett 1992; Heller et al. 1995) peuvent refléter le résultat de ce type de transformations : après les premières décharges (immédiatement sélectives, Celebrini et al. 1993), un raffinement du codage 80 ms après pourrait être la conjonction de boucles en retour d'informations issues de la voie magnocellulaire rapide avec l'activité ascendante des voies parvo- ou konio-cellulaires, beaucoup plus importante quantitativement que la première vague et en précision spatiale (proposé par Nowak et Bullier 1997).

Ces trois premiers modèles proposent des mécanismes de reconnaissance très différents, et en présentent des solutions globales : reconnaissance par composants, par classification et par alignement. L'approche des trois modèles suivants est plus proche des neurosciences, dans le sens où ils proposent des mécanismes d'encodage qui tentent d'éviter par construction les hypothèses problématiques des modèles généraux - processus descendants et traitements de classe.

2.4 routage dynamique

principe

Pour pouvoir reconnaître des objets dans n'importe quelle partie du champ visuel, van Essen, Anderson et Olshausen (Olshausen et al. 1993) proposent un mécanisme de déplacement attentionnel. Ce principe est analogue à celui discuté à propos du modèle de Kosslyn, postulé pour éviter la multiplication de représentations redondantes dans le système visuel. Le but du modèle de est de fournir un mécanisme biologiquement plausible de déplacement et de changement d'échelle des zones saillantes de l'entrée visuelle (Figure 6). Ce mécanisme est supposé faire partie intégrante de l'attention, elle-même nécessaire à la reconnaissance des objets.

Les auteurs considèrent que « la capacité à reconnaître un grand nombre de patterns complexes est computationnellement trop lourde pour utiliser un mécanisme neuronal répliqué séparément pour chaque lieu dans le champ visuel. (…) L'attention visuelle est un mécanisme qui régule le flux d'information pour extraire l'information du champ visuel approprié, dans un format approprié, vers un centre de reconnaissance de haut niveau approprié ».

Ayant postulé que l'attention doit « formater » la représentation de l'objet avant toute reconnaissance, les auteurs extraient les caractéristiques de la fenêtre attentionnelle :

• L'attention doit pouvoir être dirigée à différents lieux et à différentes échelles spatiales du champ visuel. Ces déplacements peuvent être initiés par des traits saillants ou des influences descendantes. Lors de l'émergence de traits saillants, les déplacements s'effectuent en un délai de 50-100 ms. L'attention est alors focalisée sur un voisinage du trait ayant déclenché le déplacement (et non lui seul) - mécanisme de gain.

• L'attention visuelle agit comme un goulot qui réduit à un niveau gérable la quantité de données atteignant les centres de reconnaissance (d'après les auteurs, 0.1% de l'information transmise par le nerf optique dépasse ce goulot).

• Les relations spatiales doivent être préservées dans la fenêtre attentionnelle ; cependant sa résolution spatiale est limitée à une résolution équivalente à 30x30 pixels.

Le principe est donc de postuler la très grande précision d'un seul traitement de reconnaissance, prenant généralement ses informations de la zone fovéale ; lorsqu'une zone de saillance est détectée à un autre endroit du champ visuel, l'attention focalisée sur ce lieu permet un recentrage des informations provenant de ce lieu par rapport au traitement.

architecture

Le mécanisme proposé est le suivant :

1. Une aire de saillance est définie à partir des traits de bas niveaux et/ou de la luminance ; cette aire est le résultat d'un mécanisme de *pop-out*.

2. Une de ces aires est sélectionnée (sur la base de sa taille ou de sa luminance), et la fenêtre attentionnelle est dimensionnée à sa taille et orientée sur sa position. Cet ajustement provient d'influences ascendantes.

3. Le contenu de la fenêtre (à haute résolution) est soumis à un module de reconnaissance (mémoire associative).

4. L'objet contenu dans la fenêtre est analogue à une vue mémorisée, donc reconnu. Sinon, la fenêtre est réajustée en taille et position pour coïncider avec la taille de la meilleure vue sélectionnée ; ce second ajustement provient donc d'influences descendantes.

5. Cette partie de la scène est inhibée, et le mécanisme attentionnel dirige la fenêtre vers l'aire de saillance suivante.

Ce modèle fait explicitement appel à deux modules de contrôle : le premier, chargé d'ajuster une première fois la fenêtre à l'aire de saillance ; le second ajuste la fenêtre au contenu de la mémoire associative. Ces deux modules sont le cœur du mécanisme de routage.

capacités et prédictions

Ce modèle accorde à l'attention visuelle un rôle de premier plan pour la reconnaissance. Les traits saillants sont détectés pré-attentivement et définissent une aire de saillance, sur laquelle se focalise l'attention, afin de formater l'entrée visuelle pour qu'une vue mémorisée y corresponde. Ce formatage a l'avantage de réduire le problème combinatoire de la reconnaissance – nombre d'entrées possibles pour nombre de vues mémorisées. Il a l'avantage de préserver de façon implicite les relations spatiales à l'intérieur de la fenêtre attentionnelle. Pour reconnaître les objets de la fenêtre, les neurones des aires de haut niveau (V4 et IT) doivent pouvoir moduler dynamiquement leurs champs récepteurs.

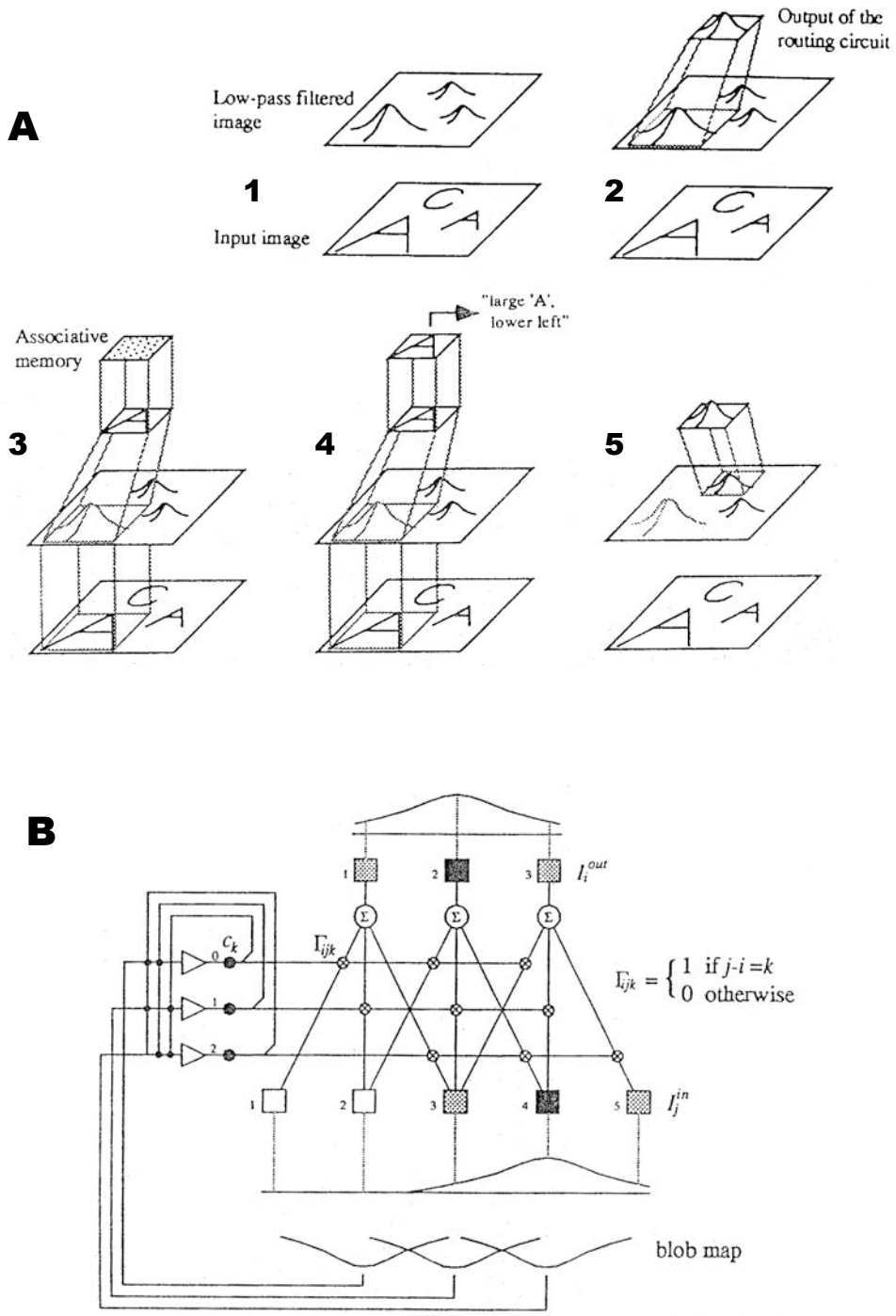


Figure 6 : Modèle de Olshausen, van Essen et Anderson (1993).

A. Stratégie attentionnelle de reconnaissance d'objets dans une scène. Après segmentation pré-attentive, les objets sont localisés puis recentrés pour être reconnus.

B. Circuit de recentrage à 1 dimension. La boucle de contrôle reçoit ses entrées pondérées par zones d'activités, et inhibe les autres zones.

Les neurones qui contrôlent les déplacements et recentrages font partie d'un module séparé et explicite, auquel le pulvinar peut être candidat (cet aspect du modèle est repris par LaBerge 1995). Le nombre de neurones nécessaire à ce mécanisme simple devrait être de l'ordre de 10^6, compatible avec le nombre total de neurones dans le pulvinar. Les aires pariétales postérieures peuvent implanter la carte des zones de saillance, comme autant de cibles de l'attention ; ce rôle peut être aussi joué par le colliculus supérieur, plus directement lié à la rétine. Par ailleurs, l'ensemble forme un code neuronal pour la position courante et la taille de la fenêtre attentionnelle, accessible pour et pouvant être influencé par d'autres modalités.

Ce mécanisme attentionnel est sériel. La ré-allocation de la fenêtre vers une nouvelle aire de saillance prend au minimum 50 ms, donc astreint à ce rythme la rapidité de reconnaissance. Ce modèle fait donc dépendre crucialement la reconnaissance de l'allocation d'attention visuelle.

plausibilité

Ce modèle est compatible avec les données psychologiques dont il est inspiré, à savoir l'allocation successive d'attention dans différents lieux du champ visuel, son caractère transitoire et d'inhibition au retour vers une zone déjà explorée (Posner et al. 1980; Klein 1988). Sur le plan neurophysiologique, le modèle propose l'hypothèse forte du déplacement et recentrage dynamique des champs récepteurs des cellules de V4 et IT, en fonction précisément du déplacement et de la taille de la fenêtre attentionnelle. Quelques expériences ont déjà montré la modulation des réponses neuronales à l'intérieur du champ récepteur en fonction de l'attention, mais aucune n'a fait état d'un déplacement ou d'un recentrage.

2.5 propagation asynchrone

principe

La rapidité du système visuel des primates dans les tâches d'identification amène S.Thorpe et J.Gautrais à proposer l'asynchronie des décharges neuronales comme base du codage dans le système visuel. Au lieu de considérer que seule la fréquence de décharge puisse transmettre de l'information, les auteurs font l'hypothèse que cette information est aussi présente dans le délai que met un neurone à intégrer les potentiels excitateurs qu'il reçoit (Figure 7) pour générer un potentiel d'action. Le but du modèle (Gautrais 1997) est de montrer que la reconnaissance d'objets est possible sur la seule base de ce délai : les neurones qui constituent leur réseau n'ont en effet, par principe, la possibilité de n'émettre qu'un seul potentiel d'action. L'intégration s'effectue donc à partir des délais relatifs des décharges des neurones afférents.

architecture et fonctionnement

Le réseau de neurones construit pour la détection des visages (Van Rullen et al. 1998) est constitué de trois couches organisées en hiérarchie ascendante (feed-forward), convergentes vers des neurones capables de répondre à la présence de visages, et de localiser l'endroit où ils se trouvent. La première couche possède des neurones sélectifs à l'orientation. Les neurones de la deuxième couche sont sélectifs à des traits constituant les visages (yeux, bouche) ; la troisième intègre ces traits lorsqu'ils présentent la configuration spatiale d'un visage. Le réseau ainsi constitué possède une architecture parallèle massive : les neurones de chaque couche sont organisés rétinotopiquement, et pour un même lieu les neurones de différents types (chaque orientation, ou composante, ou visage) déchargent d'autant plus rapidement que l'activité de leurs afférents coïncide avec leur sélectivité. Cette asynchronie fait que les neurones activés le plus tôt sont porteurs de l'information la plus représentative du stimulus, en comparaison des neurones « concurrents » de la même couche.

capacité

Les auteurs montrent que le réseau est capable de détecter la présence de visages à partir de photographies. Il est capable d'encoder plusieurs visages en même temps, les visages étant détectés d'autant plus rapidement qu'ils correspondent le mieux aux patterns mémorisés. L'encodage effectué suit la rétinotopie : chaque neurone terminal répond à un visage dans son champ récepteur, à la fréquence spatiale correspondant à la taille de ce champ récepteur. Ainsi le modèle résout les invariances par translation et par taille en assumant l'existence de neurones très sélectifs à la catégorie des visages à plusieurs fréquences spatiales et à plusieurs endroits.

La propriété par laquelle les informations les plus pertinentes sont transmises en premier a été étendue à un mécanisme d'attention spatiale (Van Rullen et Thorpe 1999) : les auteurs proposent d'accroître l'efficacité synaptique des neurones dont le champ récepteur est inclus dans le 'spot' attentionnel. Par les propriétés temporelles du réseau, un objet ainsi présenté dans l'aire spatiale privilégiée sera reconnu plus tôt que les autres objets : cette précédence fait que l'objet est sélectionné au détriment des autres constituants de l'image. Ces mécanismes peuvent donc rendre compte de manière simple de la principale propriété de l'attention spatiale sélective.

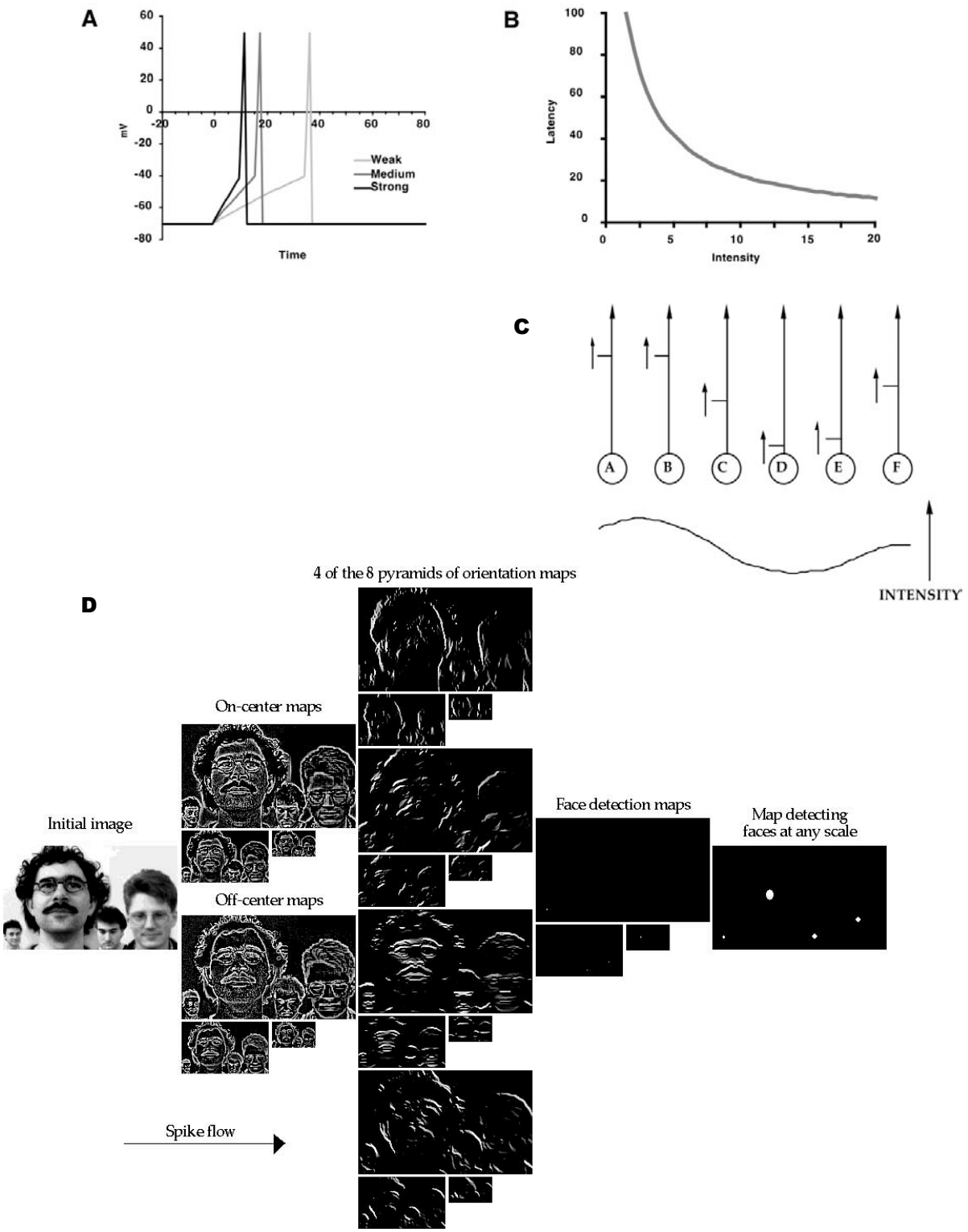


Figure 7 : Modèle de Thorpe, Gautrais, Delorme et van Rullen (1998).

A-C. Propriétés neuronales du réseau. La latence d'émission d'un potentiel d'action est fonction de l'ordre des entrées.
B. Architecture et fonctionnement du réseau FaceNet. Les visages sont détectés en parallèle à plusieurs endroits et à différentes échelles spatiales.

2.6 encodage invariant à la vue

principe de la modélisation

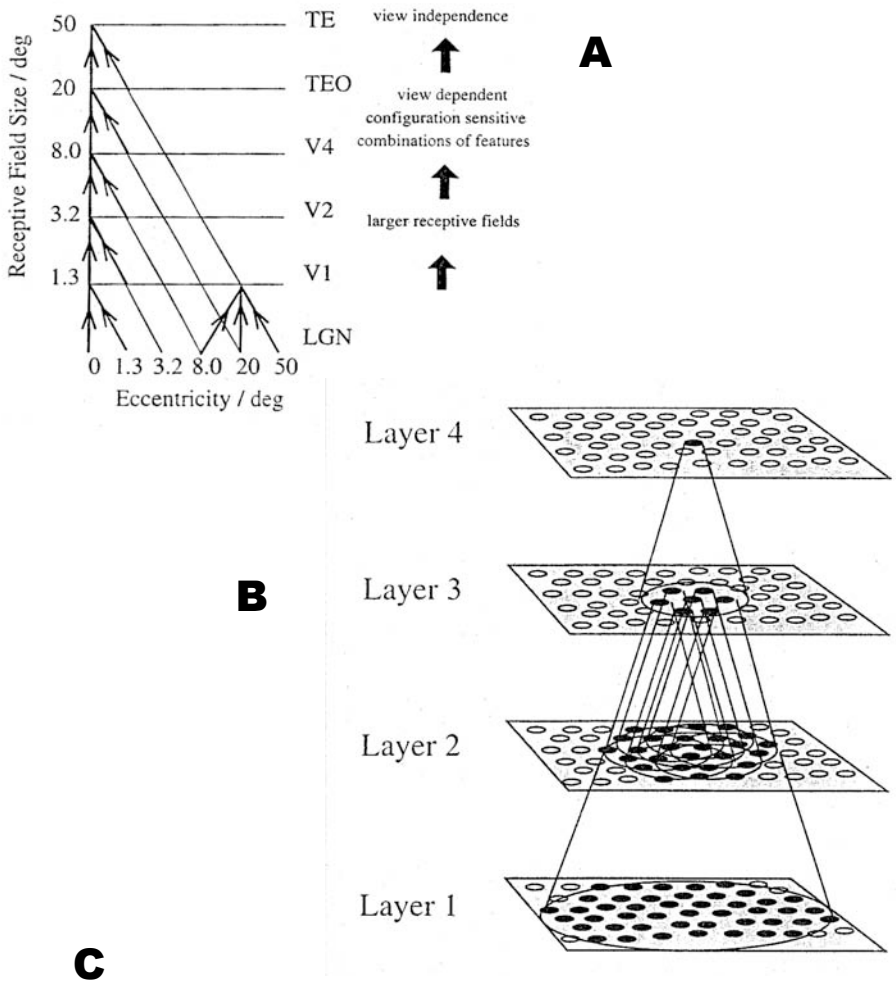
Pour rendre compte de la reconnaissance d'objets, le système doit pouvoir se montrer invariant à la translation, à la taille et à l'angle de vue, comme le codage observé dans le cortex inféro-temporal.. Pour rendre compte de ces invariances, Wallis et Rolls 1997 ont élaboré un modèle dans lequel ils appliquent quatre principes biologiquement plausibles : (1) une architecture neuronale en couches convergentes, (2) une propagation de l'information en cascade (feed-forward), (3) une représentation distribuée, et (4) une règle d'apprentissage proche du mécanisme de potentiation synaptique à long terme.

Réseau

Le modèle de Wallis et Rolls 1997 est composé de 4 couches de neurones connectées en cascade (Figure 8). La connectivité entre couche est convergente, c'est-à-dire que chaque neurone reçoit une centaine de connexions d'une région de la couche précédente, de telle sorte que chaque neurone de la quatrième couche est connecté à tous les neurones de la couche d'entrée. Chaque couche possède le même mécanisme d'intégration : un neurone effectue la somme pondérée de ses entrées, et reçoit de chacun de ses plus proches voisins une connexion inhibitrice : ces neurones voisins ont une sélectivité sensiblement différente après l'apprentissage. Cette inhibition locale intra-couche réalise une compétition « douce », notamment pour conduire à un compromis entre sélectivité et représentations distribuées (compétition douce plutôt que « winner-take-all »). Après compétition, le taux de décharge du neurone est normalisé pour maintenir constant le taux de décharge moyen d'une couche pour un même stimulus. Le réseau ainsi spécifié reçoit en entrée des images prétraitées par filtrage spatial orienté analogue au traitement effectué par les cellules simples de V1. Les quatre couches du réseau représentent donc V2, V4, et les aires temporales postérieure et antérieure.

apprentissage

La règle d'apprentissage implantée dans ce réseau n'est pas l'algorithme classique de rétro-propagation ; elle s'inspire au contraire de la modification synaptique à long-terme (potentiation ou dépression) observée sur les neurones corticaux, et utilise des présentations successives d'un même objet pour assurer son invariance dans la représentation du réseau. Chaque neurone ayant déchargé vers la couche supérieure maintient son taux de décharge (la « trace ») le temps qu'un autre stimulus puisse être présenté ; si le neurone répond aussi à ce nouveau stimulus, la synapse en lien avec la couche supérieure est renforcée.



Règle d'apprentissage par trace :

$$\delta w_{ij} = k m_i r'_j$$

$$m_i^{(t)} = (1-\eta) r_i^{(t)} + \eta m_i^{(t-1)}$$

où r'_j est l'entrée j du neurone i,

w_{ij} le poid j du neurone i,

η est l'influence de la trace,

$m_i^{(t)}$ la valeur de la trace de i au temps t

Figure 8 : Modèle de Rolls et Wallis (1997)

A. Convergence des projections feed-forward dans le système visuel. Les auteurs proposent que les propriétés d'invariance apparaissent naturellement à chaque étape.
B. Architecture hiérarchique du réseau. En couche 4, chaque neurone a pour champ récepteur le champ visuel complet. Le calcul est effectué par convolution comme dans le modèle précédent.
C. Règle d'apprentissage. L'activité d'un neurone se prolonge dans le temps, permettant aux neurones efférents d'être plus sensibles à une nouvelle entrée. Les propriétés d'invariance (translation, rotation, taille) dans les couches hautes apparaissent dès lors que la succession des entrées est cohérente : les mêmes neurones codent le même pattern malgré ses variations successives en position, vue ou taille. Les poids sont normalisés après chaque calcul.

Le réseau peut ainsi réaliser un code invariant à la translation si un objet a été placé successivement à plusieurs endroits lors de l'apprentissage. Coder un autre objet supposera de le présenter avec une séquence de déplacement analogue.

capacités du modèle

Les auteurs montrent que l'apprentissage converge pour réaliser un codage distribué, sélectif de chaque objet. Ce codage peut être invariant par translation si l'apprentissage est réalisé en différentes positions, invariant à l'angle de vue (pour des visages) si l'apprentissage est effectué par vues successives, ou invariant à la taille des objets. D'autre part, le système généralise correctement les lieux non appris : cette propriété émerge lorsque la séquence de déplacements lors de l'apprentissage comporte des déplacements adjacents successifs alternés avec des déplacement de plus longue portée (analogue à des saccades longues).

plausibilité

Le modèle s'inspire de la connectivité ascendante du système visuel, en couches successives, avec une architecture en feed-forward réalisant une compétition entre neurones par inhibition latérale. Les objets sont représentés par une population de neurones. L'hypothèse principale est donc que l'apprentissage des objets s'effectue grâce au maintien de l'activité des neurones pendant une durée plus longue que le stimulus. Le fait que les représentations distribuées et invariantes sont obtenues progressivement dans la hiérarchie des couches est un point fort du modèle : les représentations intermédiaires sont ici utilisées pour un codage d'exemplaires, mais pourrait l'être aussi bien pour un codage de catégories.

limitations

Il reste que le mécanisme de compétition réalisé par inhibition latérale suppose pour fonctionner de manière satisfaisante qu'au sein d'une même couche tous les neurones de sélectivités différentes aient déchargé. Le calcul ne peut donc s'effectuer par couche que de manière synchrone. D'après les auteurs, la normalisation des taux de décharge, nécessaire pour l'apprentissage, peut être réalisée par un effet de shunt de neurones feed-back inhibiteurs. Si le réseau fonctionne de manière très satisfaisante pour la reconnaissance d'un objet, les auteurs ne décrivent pas son comportement lorsque plusieurs objets lui sont présentés ; on peut supposer que le mécanisme de compétition « douce » permettant un encodage distribué se révélera peu efficace lorsque un stimulus concurrent ou une scène complexe vient modifier l'encodage.

3 les solutions proposées

3.1 construction des représentations

Pouvoir reconnaître un objet, c'est en avoir une forme de représentation invariante : si un consensus s'est dégagé ces dernières années en faveur de représentations analogues à des vues plutôt que des représentations en trois dimensions, il reste à déterminer comment et de quoi sont constituées ces vues. L'utilisation des vues ne semble poser de problèmes à aucun modèle : celui qui en est le plus éloigné par sa conception, celui de Hummel et Biederman, peut s'accommoder très bien d'un encodage dépendant de la vue dans ses couches hautes, à partir du « vocabulaire » descriptif réalisé dans ses couches 3 et 5. Quant au problème de la variabilité entre les vues d'un même objet, il peut être résolu par un mécanisme d'encodage distribué (à grande sélectivité grâce aux propriétés de dépression synaptique, Abbott et al. 1997), comme dans le modèle de Wallis et Rolls qui réalise aussi par *feature matching* une sorte d'interpolation des vues déjà encodées de manière analogue à l'interpolation prévue par les modèles du *template matching*. L'alternative se situe plutôt à propos de ce vocabulaire descriptif de vues : la construction des représentations visuelles se fait-elle à partir d'une description structurale, d'une construction progressive de traits réguliers, ou de transformations qui maintiennent la structure d'image ?

Contrairement aux évidences qui montrent que le mécanisme d'amorçage (priming) est davantage sensible aux composantes que partagent les images plutôt qu'aux traits d'angles et de contours (expériences de Biederman et Cooper 1991), les modèles du *template matching* conçoivent de transformer ou d'aligner des vues complètes au moyen de ces indices : modifier progressivement des distances ou des conditions d'éclairage pour accéder à une vue caractéristique de l'objet est considéré par leurs auteurs comme plus proche de ce qu'on connaît des neurones de la voie ventrale. On peut aussi émettre des doutes quant à l'utilisation de dessins dans les expériences de Biederman : la reconnaissance d'objets dessinés n'est-elle pas plus proche de l'acte du dessin plutôt que de l'identification de ces objets dans leurs milieux naturels ? Si c'était le cas, la catégorisation des relations spatiales décrite dans la RPC serait un mécanisme propre à la voie dorsale, mécanisme impliqué dans le phénomène de reconnaissance pour le cas particulier du dessin. Il est possible qu'il existe des neurones sélectifs aux catégories de position, mais l'utilisation de ces catégories pour reconnaître des images naturelles nous semble superflue.

La reconnaissance illusoire est le principal problème rencontré par les modèles de *feature matching :* si l'encodage d'objet est seulement caractérisé par une collection de traits ou composantes, sans posséder de repères spatiaux explicites (RPC) ou implicites (vues complètes), un même ensemble de traits peut correspondre à plusieurs objets ou même provoquer l'illusion d'un objet qui n'est pas dans l'image. Les modèles qui cherchent à rendre compte d'une reconnaissance invariante à la position se heurtent au

problème de « reconnaître » aussi les objets aux traits mélangés (modèle de Mel, Mel 1997). Certains auteurs insistent sur le rôle des processus descendants (top-down) pour pallier cet inconvénient et sur le rôle du contexte ; la place des processus descendants sera discutée dans un paragraphe ultérieur. Cependant, se pose de manière beaucoup plus cruciale pour ces modèles le problème du *binding*, de l'association de traits pour former une représentation cohérente.

3.2 liage et invariance à la position

Le problème du binding est de reconnaître individuellement des traits particuliers dont l'arrangement spatial doit être conservé à n'importe quel endroit du champ visuel. Poggio et Ullman résolvent ce problème en mettant en œuvre des normalisations sur la position par des traitements de haut niveau ; Kosslyn l'évite, en postulant un déplacement de la fenêtre attentionnelle pour chaque objet à reconnaître : l'objet est donc toujours centré, cette fois-ci par rapport au « spot » de l'attention. Des solutions plus plausibles sont avancées par von der Malsburg qui propose un liage dynamique par synchronisation des réponses neuronales, comme celui implanté dans le modèle de Biederman.

Comme il a été relevé plus haut, le problème est d'autant plus crucial pour les modèles du *feature matching* : ils doivent mettre en œuvre un liage qui à la fois associe les traits entre eux (de quoi est composé l'objet) et rende compte de leur configuration spatiale. Or réaliser le premier assure d'une reconnaissance invariante (par translation, rotation, taille), mais réaliser le second pénalise cette invariance (rend la représentation rigide). Le modèle de Thorpe et Gautrais contourne ce problème et réalise l'invariance en postulant des représentations multiples parallèles et rétinotopiques, mais le modèle n'a pas prétention de plausibilité sur ce plan. C'est à cet égard que la solution proposée par Wallis et Rolls est décisive pour ces modèles : en plus du parallélisme massif, l'invariance par translation est achevée progressivement dans la structure du réseau où les derniers neurones ont un champ récepteur de la taille de la couche d'entrée. La rétinotopie complète des premières couches se transmet partiellement à la couche suivante, dans un réseau convergent vers des représentations de plus en plus élaborées. La structure résout le problème du binding puisque ces représentations élaborées sont construites à l'aide de cette convergence. C'est somme toute ce qui est observé dans le système visuel, si l'on admet le parallélisme à tous les niveaux de traitements ; la solution proposée par Rolls d'un mécanisme de potentiation synaptique permettant de saisir les régularités de l'entrée par redondance dans le temps est particulièrement élégante.

3.3 accès aux catégories

Les modèles actuels de la reconnaissance des objets ne réalisent correctement la reconnaissance qu'au niveau des catégories de base, à moins qu'ils ne se spécialisent dans la reconnaissance de membres d'une catégorie particulière, la plus rencontrée étant la

classe des visages. L'absence de modèles plus complets capables de reconnaître aussi bien individus que classes n'est pas seulement un choix de simplicité pour les modèles, dont on pourrait penser qu'ils puissent réaliser un modèle complet par la multiplication des modules ou des neurones constituant les réseaux. L'illustration du problème est particulièrement claire pour les modèles de la reconnaissance par vues complètes (*template matching*), qui proposent que l'accès à la catégorie de base soit un préalable à la reconnaissance de l'objet individualisé : tous ces modèles suggèrent que l'accès à la catégorie permette de définir des transformations nécessaires pour individualiser l'objet (l'instancier). Le problème n'est pas ici dans le codage, puisque nous voyons le modèle de Poggio-Edelman (possédant des neurones répondant à la vue complète) et le modèle de Ullman (proposant un codage distribué) se heurtant à cette nécessité. Cette nécessité vient de l'idée que l'accès à la catégorie est réalisable par une reconnaissance grossière, ou la reconnaissance d'une vue « floue » de l'objet (Poggio), et que l'accès direct au niveau sous-ordonné est un cas particulier, restreint au cas rare où un indice particulièrement saillant ou le contexte permette un raccourci dans le processus de reconnaissance (voir le problème de la sélection initiale du modèle de Ullman) : l'inspiration de cette idée vient évidemment de l'expérience citée au début de ce chapitre, dans laquelle une moyenne des objets membres est perçue elle aussi comme membre de la classe.

Or les modèles psychologiques de formation des catégories (perceptives ou abstraites) et de classification reconnaissent depuis les travaux de Rosch 1975 (cf. Rosch 1978; Kleiber 1990), que « le prototype est simplement un moyen commode pour parler de typicalité », et que « ce sont les *effets* prototypiques qui sont important pour l'organisation catégorielle » (Kant 1996). Ces effets sont dus à une combinaison typique de traits, plutôt qu'à un prototype défini par une instance de la catégorie, et les traits typiques, « qui correspondent à l'intersection des sous-catégories (..), sont caractérisés par des similarités locales et non plus globales entre membres de la catégorie ». Les modèles du *feature matching* sont donc plus à même de conjuguer identification et catégorisation.

La réponse la plus complète à ce problème est donnée par Biederman, dont le modèle est conçu précisément pour rendre compte de cette combinaison de traits. Le réseau est capable d'effectuer une tâche d'identification d'objets dessinés quelle que soit leur catégorie, et une extension simple à ce réseau serait de mettre en œuvre un accès simultané à l'instance de l'objet et à sa catégorie de base. Les traits typiques sont ici les géons, liés dynamiquement au prix d'un accès explicite aux catégories de relations spatiales.

4 le temps : une contrainte essentielle

voir plusieurs objets

De nombreux modèles ne peuvent traiter qu'un seul objet à la fois ; en conséquence, le parallélisme indéniable des voies visuelles n'est souvent considéré que comme

mécanisme permettant de tester plusieurs hypothèses de reconnaissance d'un seul objet présenté (Ullman), ou un dédoublement des représentations sur chaque hémisphère ou quadrants. Afin de reconnaître plusieurs objets, les modèles font appel aux mécanismes sériels de l'attention visuelle alors conçue comme boucle de contrôle chargée de sélectionner les objets à reconnaître, à focaliser les traitements sur des zones cruciales de l'entrée, et à provoquer des déplacements attentionnels et saccades oculaires pour recentrer ces zones. Le schéma de Kosslyn illustre parfaitement ce point de vue, dont Olshausen propose un mécanisme plausible. Si c'était le cas, alors la reconnaissance d'objets présentés sur la rétine extrafovéale devrait prendre plus de temps qu'une reconnaissance fovéale. D'un autre côté, les modèles incluant des mécanismes de reconnaissance parallèles peuvent s'accommoder d'entrées multiples, comme les modèles de Mel, Thorpe-Gautrais, Rolls-Wallis, et dans une certaine mesure, le modèle de Ullman. Si ce parallélisme se vérifie, alors la reconnaissance de plusieurs entrées simultanées doit pouvoir au contraire être effectuée avec une latence similaire à une reconnaissance unique.

place des processus descendants

De manière cruciale dans le schéma de Kosslyn, le contexte joue un rôle d'amorce sur les représentations mémorisées pour initier la reconnaissance ; ce problème de 'sélection initiale' est aussi présent dans tous les modèles du *template matching*, sauf dans les modèles de traitement par raffinement ascendant. Ce problème est lié aux algorithmes utilisés, fonction dont il faut spécifier un état initial ou à défaut un contexte de reconnaissance. En plus de l'initiation, les processus descendants sont omniprésents chez Kosslyn, par exemple dans l'utilisation de l'imagerie mentale, ou chez Ullman par le test d'hypothèses. Il y a quelques années, cette position devenue généralisée a amené Henderson à qualifier les influences du contexte et les processus descendants de « position théorique par défaut », et à se demander si l'on « passera totalement à côté des processus traités de manière plus ascendante en choisissant de négocier tous les problèmes difficiles en vision en postulant des traitements top-down » (Henderson 1992). Le rôle du contexte peut en effet être contesté au moins dans la génération de saccades (Henderson et Hollingworth 1999) ; d'autre part les études en présentations séquentielles rapides d'images (RSVP, par exemple Intraub 1980) laissent à penser que ni le contexte ni les processus descendants ne jouent de rôle dans la reconnaissance rapide.

rôle de l'attention

De même que les influences contextuelles, l'utilisation des mécanismes attentionnels est omniprésente pour remédier aux problèmes rencontrés dans la reconnaissance visuelle : recentrage de l'entrée visuelle pour des présentations extrafovéales, liage des propriétés de l'objet afin d'unifier la représentation mentale, amorçage de traits visuels dans une seconde boucle de traitement. Si ces mécanismes existent, ils peuvent être utilisés par le traitement visuel dans des cas particulièrement difficiles ; leur utilisation systématique ne semble pas concevable si l'on considère que la reconnaissance est l'une des principales fonctions du

système visuel, qui utilise probablement des mécanismes simples et rapides. Les mécanismes attentionnels sont invoqués à toutes les étapes du traitement visuel dans le schéma de Kosslyn (mis à part le premier encodage pré-attentif dans un 'buffer' analogue à la première aire corticale visuelle V1). Si l'attention est en elle-même une ressource limitée pouvant être allouée à différentes tâches aux dépens de son efficacité, on voit mal comment la reconnaissance visuelle pourrait dépendre crucialement d'une ressource si précieuse à son fonctionnement.

vitesse de traitement

Tous les modèles postulent une reconnaissance plus rapide pour les vues familières d'objets ; par un processus analogue, la catégorie visuelle est utilisée par les mécanismes du *template matching* comme une première étape vers la reconnaissance des exemplaires. Si ces postulats sont en accord avec les données expérimentales, le problème vient du temps nécessaire pour reconnaître les exemplaires. Pour mener à terme la reconnaissance d'un objet après la reconnaissance de sa classe, Kosslyn propose l'utilisation additionnelle des couleurs et textures, Ullman postule une série d'alignements, rotations et transformations avec les nombreux modèles mémorisés, Poggio propose un mécanisme analogue à la classification d'un grand nombre de vues. Vu les traitements sophistiqués requis, le temps pris pour la reconnaissance complète devrait plus que doubler celui de la reconnaissance de classe. La classification d'un objet est en effet la plus immédiate dans ces modèles, grâce à l'utilisation d'une vue 'floue'. Or la plupart de ces modèles utilisent des dessins comme entrée visuelle : si ces objets étaient placés en situation dans une scène naturelle, on peut aisément supposer que ces traitements et transformations devraient prendre beaucoup plus de temps, ou être aidés, comme on l'a vu, par le contexte. La solution alternative est l'emploi de 'templates' généralisé à tous les objets, comme proposé par les modèles de Thorpe et Wallis-Rolls, dont les convolutions successives ascendantes peuvent être considérés comme autant de raffinements parallèles convergeant vers des représentations possibles. Le temps pris par la reconnaissance est à notre sens un caractère crucial pour considérer ces modèles comme plus proche de la réalité.

la contrainte du temps

La contrainte du temps pris par le traitement visuel est mise en évidence par le nombre d'objets qui peuvent être intégrés par le système visuel entre les saccades oculaires. Cette capacité est portée à quatre objets dans la revue récente de Gordon et Irwin 1996 ; si les yeux effectuent un minimum de quatre saccades par seconde, cela porte déjà à moins de 100 ms par reconnaissance d'objet ! Cette reconnaissance peut être plus longue si le traitement est réalisé par une architecture en pipeline, mais précisément dans ce cas l'utilisation de boucles d'itérations est très improbable. Ce genre de contrainte temporelle a été souligné par Thorpe et Imbert 1989, en prenant en compte les vitesses de conduction corticales et le temps d'intégration synaptique. L'argument a été repris et confirmé dans d'autres travaux (Nowak et Bullier 1997; Gautrais et Thorpe 1998). L'activation sélective

des neurones des aires inférotemporales du primate est observée vers 110-130 ms ; si on alloue un temps de transmission de 1 à 10 ms entre aires corticales, et une dizaine de synapses à traverser en partant de la rétine (5-10 ms par synapse), il reste très peu de temps pour que des boucles en retour puissent agir. Le traitement visuel ne peut donc reposer de façon aussi essentielle sur des boucles de feed-back pour la reconnaissance visuelle (voir aussi Rolls 1991; Maunsell et Gibson 1992). Il semble donc que les modèles postulant une reconnaissance 'lente', par étape après une reconnaissance de catégorie ou au moyen d'itérations successives soient incompatibles avec la vitesse que l'on peut supposer de la reconnaissance visuelle.

Cette vitesse du traitement visuel apparaît donc comme un critère crucial d'élaboration des modèles de la vision : elle donnerait des indications sur l'architecture fonctionnelle du système visuel et les mécanismes en jeu dans les tâches de reconnaissance d'objets. Cette vitesse peut être mesurée dans le cas de dessins ou de stimuli simples, mais elle donnerait des indications plus cruciales encore dans le cas de scènes naturelles - pouvant être extrêmement complexes d'un point de vue de modélisateur.

VITESSE DU TRAITEMENT VISUEL

Dans la partie précédente, nous avons montré la place cruciale qui doit être accordée au temps de traitement des stimuli naturels pour la compréhension et l'élaboration de modèles de la vision. Dans la première partie de ce chapitre, nous constaterons que les études chronométriques déjà menées ne permettent pas de répondre clairement à notre préoccupation. D'une part les études utilisant des stimuli naturels sont rares, d'autre part les méthodes électrophysiologiques chez l'homme ne nous semblent pas adaptées à notre question. Les études en potentiels évoqués procèdent par extraction et analyse de pics d'activités neuronales pour dégager un phénomène sous-jacent *per se* ; elles ne permettent pas de déterminer le moment où un traitement visuel complexe peut être considéré comme terminé. Pour répondre à cette question, nous serons amené à mener des expériences de caractère plus exploratoire et conduit à développer notre propre méthode d'analyse.

Nous nous attacherons donc à mettre en œuvre une tâche expérimentale visant à mesurer le temps nécessaire au système visuel pour analyser des scènes naturelles. Pour cela, nous choisirons une tâche de catégorisation de photographies contenant ou non un animal, avec réponse en go / no-go, dans une situation qui nous semble exigeante pour le système visuel. La mesure du « temps nécessaire pour analyser une image » sera réalisée par une méthode s'attachant à préciser les divergences des potentiels évoqués par les cibles et les distracteurs de la tâche. *L'activité différentielle* qui en résulte sera discutée en terme de latence afin de cerner le temps nécessaire aux traitements indispensables à l'exécution de la tâche.

Le chapitre suivant complétera cette première tentative par une série d'expériences visant à mieux comprendre les premiers résultats obtenus en terme de contrainte temporelle du traitement visuel, en variant les positions des images, leurs couleurs, et la catégorie cible de la tâche.

1 quelques indices de mesure

1.1 temps de réaction comportementaux

Dans le champ de la psychologie expérimentale, la technique la plus utilisée pour étudier les caractéristiques du fonctionnement cérébral a été longtemps la mesure des temps de réaction comportementaux. Cette technique a permis d'aborder un large éventail de questions ; elle a pu apporter nombre de réponses déterminantes, dont les résultats et les méthodes servent largement de références à des études utilisant des techniques plus récentes d'enregistrement (potentiels et champs magnétiques évoqués, neuroimagerie fonctionnelle). Dans le domaine de la reconnaissance visuelle, les premières études cherchant à mesurer le temps requis pour identifier des objets ont été menées il y a plus de cent ans : Cattell 1886 a cherché à mesurer ce temps en demandant à des sujets de nommer des images qu'il leur présentait. Ce temps, environ 750 ms, a été largement confirmé depuis (Oldfield et Wingfield 1965 ; Potter et Faulconer 1975 ; Intraub 1979 ; Carr et al. 1982 ; Fraisse 1984 ; Biederman et al. 1987 ; Henderson et al. 1987 ; pour une revue, voir Thorpe 1988 ; Thorpe et Imbert 1989). Ce temps se modifie en fonction de la familiarité aux images et en fonction de la prototypicalité de leurs vues, comme indiqué en première partie (de 600 ms à 1200 ms pour les objets les moins familiers). Cependant ces latences mesurent non seulement la reconnaissance visuelle des objets, mais aussi le temps nécessaire pour énoncer leurs noms.

Les expériences de Oldfield et Wingfield 1965, et Intraub 1979 ont testé l'approche consistant à présenter des images d'objets pendant des temps très brefs, et à augmenter ce temps jusqu'à ce qu'une reconnaissance puisse être possible. Les résultats ont cependant montré que la reconnaissance était correcte pour des stimulations aussi brèves que 10 ou 15 ms, laissant penser que les images était traitées de telle façon que les informations portées par elles étaient disponibles après que l'image ait disparu. L'introduction d'un masque après l'image porte le délai de présentation nécessaire à 100-120 ms, mais rien ne permet de dire que le traitement visuel ait été suffisant à ces latences pour permettre une reconnaissance.

Les expériences de Potter (Potter et Faulconer 1975 ; Potter et al. 1979) et Intraub (Intraub 1980 ; Intraub 1981) ont utilisé les techniques de présentations séquentielles rapides de séries de 16 images (RSVP), en variant les temps de présentation de 114 à 333 ms. Les sujets devaient signaler la détection de scènes cibles en appuyant sur un bouton-réponse, ces scènes pouvant apparaître n'importe où dans la séquence d'images. En fonction de la définition des scènes cibles, les résultats ont montré que même les présentations courtes permettaient encore aux sujets des détections correctes : 60% de cibles détectées dans le cas d'images particulières, 46% pour une catégorie d'objets, et

35% pour une catégorie définie négativement. Sur la base de ces résultats, Potter proposa 100 ms comme délai suffisant pour analyser une image, cette valeur exprimant un ordre de grandeur (voir Potter et Intraub 1999). Dans tous les cas, ces études montrent qu'une reconnaissance visuelle rapide est possible dans des cas difficiles, où le flux visuel est artificiel et non continu, et les informations de contexte absentes.

1.2 potentiels évoqués de surface

a) visages

Chez l'homme, l'existence d'activités neuronales liées au traitement des entrées sensorielles est facilement mise en évidence par les potentiels électriques enregistrés à la surface du scalp ; évoqués par des présentations d'images, ils sont appelés potentiels évoqués visuels. L'existence de potentiels visuels évoqués spécifiquement par des présentations de visages est bien connue (Jeffreys et al. 1992 ; Seeck et Grüsser 1992). Des études plus précises utilisant des électrodes implantées ont permis de préciser leurs bases cérébrales (essentiellement le gyrus fusiforme) (Allison et al. 1994 ; Halgren et al. 1994 ; Kanwisher et al. 1997). Ces potentiels 'visages' peuvent couvrir un domaine temporel très large.

Dans une expérience visant à déterminer quels paramètres de ces stimuli peuvent être encodés par ces potentiels évoqués visuels, Bobes et al. 1994 rapportent trois expériences dans lesquelles les sujets doivent déclarer si des visages connus sont corrects ou non. Les photographies utilisées sont manipulées de telle sorte que les traits composant les visages (1) n'appartiennent pas tous à la même personne, (2) sont mélangés, (3) sont légèrement déplacés. Les effets de ces distorsions par rapport aux visages de référence sont visibles sur les potentiels évoqués à partir de 250 ms (N374) pour les traits incongrus, 370 ms (P435) pour les traits mélangés et 300 ms (P486) pour les traits déplacés. Les auteurs suggèrent qu'un processus de reconnaissance de visages familiers (la description des traits en mémoire, leur contenu) provoque le premier effet à 250 ms, tandis que le second serait lié à un encodage et une analyse de la structure du visage (la configuration, d'après la théorie d'encodage structurel de Bruce et al. 1992).

D'autres études ont montré qu'il existait des potentiels 'visages' à des latences bien plus courtes : 130-170 ms (Botzel et Grusser 1989; George et al. 1996 ; Jeffreys 1989). Mais il n'est pas encore établi que ces potentiels 'visages' se différencient de ceux répondant à la présentation de mots, de pseudo-mots ou pseudo-fontes, en comparaison à d'autres formes d'objets (Shimoyama et al. 1992 ; Schendan et al. 1998). Pour ces derniers auteurs, un effet de familiarité à ces formes spécifiques peut expliquer ces traitements rapides. Même si ces potentiels sont particuliers aux visages, il est possible que ces traitements puissent recruter des voies spécialisées dans des reconnaissances de formes aux

configurations particulières et typiques ; il n'est pas sûr que ces latences puissent se généraliser à la reconnaissance d'autres objets.

b) autres catégories

Les tâches dans lesquelles des sujets doivent décider si l'image affichée représente ou non un objet (object decision task) ont été utilisées dans des contextes d'études de la mémoire (évoquant des potentiels aux latences de l'ordre de 400 ms), et plus généralement au cours d'études sur les effets d'amorçage. A notre connaissance, aucune étude de décision d'objets n'a été menée pour caractériser des potentiels évoqués visuels à d'autres catégories que les visages. Dans une étude de potentiels évoqués par les visages comparés à d'autres catégories comme des personnages, outils et fleurs, Seeck et Grüsser 1992 ont montré des différences entre 130 et 160 ms pour les visages et personnes comparées aux autres catégories, ces autres catégories étant indissociables. D'après Jeffreys et al. 1992 ou Shimoyama et al. 1992, de nombreuses catégories d'objets évoquent une onde positive au vertex entre 130-170 ms, similaire à celle évoquée par les visages mais moins ample. Une autre étude (Jeffreys 1996) montre que la latence de ce potentiel varie en fonction de la pertinence de la forme (structure possible ou impossible). On peut inférer de ces résultats que les processus cérébraux à la base de la reconnaissance des visages se déroulent dans des aires proches de celles activées dans la reconnaissance des objets ; cependant ces activités peuvent refléter un encodage visuel intermédiaire avant une perception et une reconnaissance complète.

Une étude plus parlante a été menée par Rudell 1992 mettant en évidence ce qu'il appelle le 'potentiel de reconnaissance' : une onde positive occipitale s'initiant à 170 ms jusqu'à 350 ms observable pour les mots écrits et les images signifiantes pour les sujets. Les stimuli contrôles étant des caractères chinois ou arabes, ou des amas de traits, il est malheureusement impossible sur ces expériences de savoir si cette activité est aussi évoquée par des non-mots ou objet inconnu : si c'était le cas, ce potentiel pourrait aussi bien refléter une analyse visuelle intermédiaire précédant une reconnaissance complète. Les études dans lesquelles les sujets décident si les lettres affichées représentent ou non un mot de leur langue (décision lexicale) montrent des activités corrélées à leur réponse vers 400 et 600 ms, sans que cette activité soit motrice (Bentin et Peled 1990 ; Karayanidis et al. 1991 ; Holcomb 1993 ; Kounios et Holcomb 1994). Par exemple, Rugg 1983 attribue cet effet à un processus d'évaluation du stimulus ; ses résultats montrent aussi que les potentiels évoqués à 100 et 180 ms peuvent refléter une reconnaissance des lettres, aussi bien dans le cas des mots que des non-mots. Comme pour les objets et les visages, ces potentiels évoqués par des lettres correspondent aux données d'enregistrements corticaux chez l'homme dans les gyri fusiforme et temporal inférieur vers 200 ms (Allison et al. 1994 ; Nobre et al. 1994).

c) effets d'amorce

Un champ de recherche a été exploré dans le cadre de l'étude des processus et des représentations impliqués dans la compréhension des scènes visuelles et du langage. Ces études, en mettant en jeu le mécanisme d'amorçage (priming), cherchent à savoir si les processus verbaux et visuels partagent les mêmes représentations (un « système conceptuel commun »). Si ces études ne sont pas d'un intérêt direct pour notre propos, elles utilisent des tâches visuelles dans lesquelles le sujet doit choisir entre deux réponses ; d'autre part, depuis Barrett et al. 1988 l'ensemble des recherches du domaine utilisent les potentiels évoqués visuels comme variables d'études – la neuroimagerie fonctionnelle très récemment.

Les effets mesurés dans de tels processus visuels ont des latences tardives : Barrett et Rugg 1990 montrent des effets à partir de 400 ms, mais une tâche phonologique vient s'ajouter à la tâche visuelle. S'affranchissant de cette contrainte, les mêmes auteurs Barrett et Rugg 1990 montrent que des différences sont visibles entre potentiels évoqués par des dessins reliés à l'amorce et des dessins non reliés, à partir de 300 ms lors d'un amorçage sémantique (i.e. dessins reliés par une association sémantique de type couteau-fourchette ou clef-serrure). Une expérience similaire est relatée par Pratarelli 1994 dans laquelle les sujets doivent décider si les deux stimuli sont appariés (deux boutons-réponses). L'amorce est constituée de mots entendus par le sujet environ 1.5 s auparavant ; le deuxième stimulus est constitué par un dessin de l'objet nommé lui correspondant, s'il est apparié. Dans ce type de tâche, les auteurs mettent en évidence un effet de l'amorçage du premier sur le second stimulus à partir de 250 ms. Cependant, comme dans les tâches précédentes, cette différence de traitement est due à l'effet d'amorce et non uniquement au traitement visuel du second stimulus.

Dans le même contexte, Holcomb et McPherson 1994 rapportent une tâche dans laquelle le sujet doit décider si le dessin présenté est ou non un objet (deux boutons-réponses). Comme les autres études du domaine, ils mettent en évidence un effet d'amorçage causé par une autre vue de l'objet, présentée environ 1s avant. Plus précisément, il y a trois types de stimuli dans cette tâche : objet apparié à l'amorce, objet non-apparié, non-objet (non-reliés à l'amorce). Du point de vue de la rapidité de traitement, il est intéressant de comparer ici les potentiels évoqués par les objets non-reliés à l'amorce et par les non-objets. Cette comparaison montre une différence significative entre ces deux potentiels visible à deux moments, à 200 et 550 ms. Cette différence de potentiels est caractérisée comme étant plus négative sur les électrodes frontales pour les non-objets. Ce résultat n'est malheureusement pas discuté ; il est cependant un indice de reconnaissance à 200 ms de dessins d'objets en comparaison de dessins aléatoires (possédant les mêmes propriétés visuelles que les dessins d'objets en termes de contrastes, fréquences spatiales etc.), cette reconnaissance pouvant refléter une configuration visuelle connue des sujets, comme observé dans le cas des décisions lexicales.

d) attention visuelle

Les théories psychologiques de l'attention ont très tôt caractérisé ses effets dans l'espace, c'est-à-dire décrit le statut spécifique de la zone du champ visuel privilégiée, comme nous l'avons évoqué en première partie. Les expériences en potentiels évoqués reflètent cette prédominance des études sur l'attention spatiale ; de nombreux exemples peuvent être cités, mais nous nous limiterons ici à énoncer les principaux résultats pouvant être utiles à nos préoccupations (pour une revue, voir par exemple Heinze et Mangun 1995 ou Hillyard et al. 1998).

Les effets de l'attention spatiale peuvent être très précoces : dès 75 ms, latence correspondant aux premiers potentiels corticaux enregistrés chez l'homme (ex. Halgren et al. 1994), de nombreuses expériences montrent une modification des réponses neuronales lorsqu'un stimulus visuel est détecté dans une zone du champ visuel privilégiée. Cet effet ne peut être attribué qu'à la correspondance entre l'état d'attention préparatoire des sujets et la présence d'un stimulus visuel, dont l'effet supprime probablement les réponses neuronales aux autres stimuli présentés dans le reste du champ (Posner et Dehaene 1994 ; Anllo Vento 1995 ; Luck et Hillyard 1995). Le second effet de l'attention spatiale est mis en évidence à partir de 140 ms ; si son interprétation est encore controversée, les études récentes s'accordent à l'interpréter en termes de gain et de facilitation de l'entrée visuelle (Luck et al. 1994 ; Heinze et Mangun 1995; Hillyard et AnlloVento 1998 ; Hillyard et al. 1998). Ces deux effets attentionnels ne semblent donc avoir aucun rapport avec une reconnaissance du contenu des entrées visuelles.

Les études expérimentales visant à caractériser les processus de l'attention non-spatiale sont plus récentes, ou difficilement dissociables de l'attention spatiale : de nombreuses expériences ont cherché à spécifier le rôle de l'attention dans le problème du liage, ou dans l'émergence de représentations en mémoire à court terme. Elles sont généralement de peu de recours pour la question posée ici ; cependant certains résultats dans des protocoles de recherche visuelle (visual search) permettent d'obtenir quelques indices sur la rapidité du traitement visuel (voir Sperling et Melchner 1978; Treisman et Gelade 1980; Kahneman et al. 1983; Treisman et al. 1983; Klein 1988; Tallon-Baudry et al. 1997).

En potentiels évoqués, une expérience de Sugita (Sugita 1995) met ainsi en œuvre des stimuli composés de cercles barrés ou non barrés, ainsi que de cercles incomplets et d'ellipses, pour tenter de caractériser les recherches visuelles parallèles et sérielles. Lors de la tâche à double choix, les résultats montrent que les stimuli faisant l'objet d'une recherche parallèle (entraînant un 'pop-out') évoquent une différence entre cibles et distracteurs à partir de 150 ms, plus négative pour les cibles sur le site occipital d'enregistrement (4 électrodes au total). Dans le cas des recherches sérielles, le même phénomène se produit plus tard (débute vers 200 ms, pic vers 250 ms). Cet effet possède la caractéristique d'accroître sa latence en fonction directe du nombre de distracteurs : les

auteurs proposent de l'interpréter comme un corrélat de la tâche de recherche. Si cet effet se vérifie, il est probable que ces latences soient caractéristiques de la durée du traitement visuel nécessaire dans ces tâches.

Les études associant potentiels évoqués visuels et comportement, portant spécifiquement sur l'attention non-spatiale, peuvent se grouper en trois sous-ensembles : les plus nombreuses sont les études sur l'attention à la couleur (Hillyard et Munte 1984 ; Wijers et al. 1989; Anllo Vento et Hillyard 1996 ; Karayanidis et Michie 1996 ; Smid et al. 1997 ; Anllo-Vento et al. 1998) ; puis viennent les études sur l'attention à la fréquence spatiale (Harter et Previc 1978 ; Previc et Harter 1982 ; Kenemans et al. 1993 ; Zani et Proverbio 1995), enfin à l'orientation (Harter et Guido 1980 ; Rugg et al. 1987; Kenemans et al. 1993). Les résultats de ces études montrent tous que l'effet de sélection attentionnelle de traits non spatiaux est caractérisé par une première onde négative postérieure suivie d'une onde frontale positive (SN et SP), observables à partir de 150 ms jusqu'à 350 ms. La SN est probablement à mettre en correspondance avec la déflexion décrite par Sugita 1995. Pour l'ensemble de ces attributs visuels simples, ce reflet de la sélection attentionnelle indique que le traitement visuel a déjà pu être suffisant pour la reconnaissance.

1.3 latences de réponse des neurones

A côté des méthodes d'exploration non invasives utilisées chez l'homme, les techniques d'enregistrement chez l'animal permettent de s'intéresser directement aux latences de réponse des neurones. L'analogie avec l'homme doit cependant être menée avec précautions car les primates non humains se révèlent toujours plus rapides que l'homme à déclencher une réponse comportementale. Dans diverses tâches visuo-motrices, les temps de réaction de l'homme sont supérieurs d'environ 50% à ceux du singe. Cette différence pourrait être attribuée à la taille plus réduite de leur cerveau qui aurait pour résultats d'abaisser substantiellement les temps de conductions intercorticaux. Nous présentons ici quelques études montrant la sélectivité de certains neurones à des catégories d'entrées visuelles, ces quelques exemples étant suffisants pour indiquer les latences typiquement observées dans les aires cérébrales visuelles.

Rolls et ses collaborateurs ont mené de nombreuses études d'enregistrement unitaire chez le singe vigile à partir des années 1970. Notamment, cette équipe a mis en évidence des réponses sélectives de neurones à des entrées visuelles particulières, réponses neuronales dont les latences permettent d'établir la rapidité avec laquelle certains traitements visuels peuvent être effectués. Par exemple, Rolls et al. (Rolls et al. 1976; Rolls et al. 1979) ont montré que dans l'hypothalamus latéral des primates se trouvent des neurones ne répondant que dans le cas où l'animal regarde un aliment ; ces neurones ont la propriété de répondre à tout stimulus visuel associé à la nourriture, y compris des stimuli neutres comme des formes simples à condition que l'animal y ait associé la disponibilité

d'aliments. Ces neurones répondent à des latences d'environ 150 ms après la présentation d'un stimulus, laissant penser que le traitement d'analyse visuelle pour déterminer l'identité de l'entrée est déjà terminée à ce moment.

D'autres données concernant la catégorie de neurones répondant spécifiquement aux visages chez le primate forment depuis quelques années un ensemble cohérent : localisés essentiellement dans l'aire visuelle supérieure du sulcus collatéral (STS), ces neurones commencent à répondre sélectivement à des latences de 100 à 140 ms (Perrett et al. 1982 ; Desimone 1991 ; Oram et Perrett 1992 ; Young et Yamane 1993; Rolls et al. 1994). Le comportement de ces neurones montre le résultat d'un traitement visuel élaboré, puisque quelques-uns de ces neurones ne répondent qu'à des directions précises du regard, abaissant brusquement leur activité dès que le regard dévie de seulement 15 à 20 degrés (Perrett et al. 1987). D'autres neurones enregistrés possèdent des propriétés remarquables, comme ceux observés par Rolls et al. 1982 dans le thalamus antérieur ne répondant qu'à partir de la seconde présentation d'un objet. Cette réponse se manifeste étonnamment quelle que soit l'identité de l'objet, à partir de 130 ms dès la deuxième présentation. La reconnaissance de cet objet est donc clairement établie dès 130 ms, montrant que le traitement visuel est suffisant à cette latence pour déterminer la familiarité de l'objet.

D'autres études apportent des informations plus complètes sur le décours des activations des aires visuelles (par exemple Celebrini et al. 1993; Mikami et al. 1994; Nakamura et al. 1994; Vogels 1999). L'intérêt porté par les neurophysiologistes aux latences des réponses neuronales est un phénomène assez récent ; à notre connaissance, trois revues assemblent des données collectées dans de nombreuses études (Maunsell et Gibson 1992 ; Nowak et Bullier 1997 ; Schmolesky et al. 1998). Ces études montrent les latences avec lesquelles répondent les principales aires corticales impliquées dans l'analyse visuelle des images statiques (voie ventrale). En V1, les premières décharges neuronales apparaissent à partir de 40 ms, généralement 50 ms ; en V2, 60 ms ; V4, 70 ms ; TE, 70-80 ms ; IT, 90 ms ; STS 90-100 ms.

Ces latences sont inférieures à celles observées chez l'homme lors de reconnaissances d'objets, de visages et de mots affichés (Halgren et al. 1994) : des pics d'activité (non les latences des premières décharges) sont observables en V1 vers 75 ms, et la plupart des aires visuelles répondent à des stimulations pendant les 110 ms suivantes : gyrus lingual postérieur 75-105 ms (aires 17-18 correspondant à V1-V2 chez le primate), gyrus fusiforme 130 ms (bases des aires 19 et 37 , corr. V4 ?), occipito-temporal 180 ms, gyrus angulaire 190-250 ms, gyrus supramarginal 240-310 ms, sulcus temporal supérieur 430 ms, hippocampe 630 ms). Par contre, dès 150-220 ms, des activations diffusent vers les cortex pariétaux, temporaux latéraux et limbiques. Dans le cas du gyrus fusiforme, des activations ont pu être observées à 130 ms dans le cas des visages, suivies à 240 ms par une activation du gyrus temporal médian (corr. IT ?). L'étude de Allison et al (Allison et al. 1994) a mis en évidence un ensemble de zones du cortex temporal inférieur répondant à partir de 170-200 ms à des visages, véhicules, insectes, lettres et chiffres, et à la couleur.

Bien que la disposition respective de ces aires varie entre individus, les auteurs concluent à 4 voies temporales inférieures : visages, lettres, nombres et couleur, portées par les gyri fusiforme, para-hippocampique et temporal inférieur, qui répondent à ces latences.

2 nécessité d'une expérience dédiée à la mesure

Nous venons de voir que les données disponibles laissent à penser que le traitement visuel nécessaire à la reconnaissance des objets pouvait être très court. Mais ce parcours rapide de la littérature nous a permis de constater que la perception de scènes naturelles était rarement étudiée (Potter et Faulconer 1975; Potter et al. 1979 ; Intraub et Hoffman 1992) ; la plupart des études étaient menées avec des stimuli simples ou familiers, pouvant abréger de façon non négligeable les traitements nécessaires à leur reconnaissance.

Cette absence est doublée d'un problème de méthode : l'approche classique étudie un processus cérébral, au moyen des variations des composantes évoquées que la méthode cherche à isoler. La réponse à notre question passe par la recherche d'un signe électrophysiologique – quel qu'il soit - indiquant que les traitements perceptifs suffisants pour une tâche à accomplir ont été effectués. Nous allons voir pourquoi la méthode classiquement utilisée se révèle inadaptée pour déterminer les contraintes temporelles qui nous intéressent, puis nous présenterons une tâche et une méthode d'analyse.

2.1 la chronométrie et les potentiels évoqués

a) les questions posées en électrophysiologie

L'EEG est le reflet en surface d'activités électriques profondes généralement corticales, provenant des décharges simultanées de groupes de neurones (principalement la sommation des potentiels synaptiques vers les cellules pyramidales, Vaughan et al. 1988). On enregistre les différences de potentiels entre chaque électrode et une électrode de référence ; l'expérimentateur s'intéresse généralement aux variations de potentiels en termes d'amplitude ou de fréquence, mais aussi au degré de synchronisation entre électrodes. La moyenne de ces mesures lors d'essais successifs permet d'extraire la réponse évoquée à une stimulation en s'affranchissant de l'activité électrique spontanée indépendante de la stimulation. Une réponse évoquée est donc constitué d'une série de composantes – ou événements – associées à la stimulation ou à la réponse comportementale des sujets, en fonction de la synchronisation choisie. Chaque événement est caractérisé par un pic de mesure (amplitude, latence), séparé du suivant d'au moins 30 millisecondes.

La première préoccupation des études d'électrophysiologie humaine de surface est de faire correspondre ces événements aux différentes étapes des modèles de la psychologie,

en termes de comportement, de traitement de l'information, ou de propriété émergente. Citons par exemple la correspondance entre le potentiel VPP (vertex positive potential, 150-200 ms) et le traitement des objets et visages (Botzel et Grusser 1989 ; Jeffreys et al. 1992), la synchronie de phase d'oscillations gamma et la perception de scènes visuelles cohérentes (Tallon-Baudry et al. 1997; Tallon-Baudry et al. 1998; Rodriguez et al. 1999). Une préoccupation attenante est d'identifier à chaque événement isolé et spatialisé une source profonde, associée à une aire cérébrale – utile pour les applications cliniques, et pour les analogies avec la physiologie des primates non-humains.

Une fois acquise cette correspondance entre les modèles et les événements évoqués, le second type de préoccupation est d'affiner en retour les modèles, en étudiant la variation des indices électrophysiologiques en fonction des conditions expérimentales (cf. Renault et al. 1982; Renault et al. 1988). Nous pouvons citer les études sur l'attention, sur la chronométrie mentale (traitements sériels ou parallèles, échanges inter-hémisphériques, compatibilité stimulus-réponse), sur les types des représentations perceptives, sur le développement cérébral et la récupération fonctionnelle.

b) des traitements de données incompatibles avec notre problème

Pour mettre en évidence un événement particulier extrait de la réponse évoquée, l'analyse consiste à isoler l'onde dans sa plage de fréquences et distribution spatiale, et à en étudier la latence, l'amplitude et la distribution en fonction des stimuli, de la tâche et des temps de réaction. La variable principale de quasiment toutes les études en potentiels évoqués est le pic d'activité : l'extraction de ce pic est grandement facilité par un lissage temporel des données dans la plage de fréquence des potentiels cognitifs. Or nous avons cherché dans les données publiées à caractériser une latence, propre aux tâches effectuées, permettant de soutenir que le traitement visuel de reconnaissance devait être suffisamment avancé pour permettre à la tâche d'être réalisée : par exemple, un effet de sélection attentionnelle d'un attribut cherché ne peut être réalisé que si le système visuel a analysé suffisamment l'entrée pour pouvoir décider de l'existence de cet attribut. Un effet de l'attention ou d'amorçage à une catégorie visuelle ne peut avoir lieu que si une activité spécifique à cette catégorie est déjà présente. En conséquence, nous avons cherché à extraire des potentiels évoqués non pas la latence du pic de l'onde associée à la tâche, mais la latence à laquelle cette onde commence à apparaître. L'extraction des pics d'activité par lissage rend plus flou le début de ces activités ; d'autre part les études publiées permettent rarement une comparaison directe des potentiels entre eux. Dans tous les cas, aucune donnée n'est disponible sur les latences de séparation entre potentiels évoqués par les diverses conditions expérimentales. Or l'initiation de ces différences est très variable par rapport aux latences des pics d'activités (les références citées montrent des variations de 15 à 80 ms) ; cette latence est la principale variable de notre étude.

2.2 méthodes

Au moyen de la technique des potentiels évoqués, nous voulons estimer le temps nécessaire pour analyser une scène naturelle. Il s'agit de disposer de photographies, d'une tâche à effectuer par les sujets, et d'une méthode d'analyse.

a) le choix de la tâche

Les potentiels évoqués par des stimuli visuels sont le reflet d'un ensemble de processus associés à l'entrée visuelle. D'un point de vue computationel, ces processus peuvent être liés à l'encodage des attributs des images, à celui de leur contenu, à la mémorisation, ou à d'autres processus dépendant plus crucialement de la tâche accomplie par le sujet, comme une verbalisation ou un geste moteur. Nous avons vu au chapitre précédent que l'estimation au moyen des potentiels évoqués du « temps nécessaire pour analyser une image » se heurte à la question de savoir si cette analyse est terminée, c'est-à-dire suffisante pour la tâche à accomplir, ou si le potentiel observé est le reflet d'un encodage visuel intermédiaire indiquant l'analyse en cours. Disposer d'éléments de comparaison implique donc d'enregistrer d'autres potentiels servant de références, obtenus au cours de la même tâche, évoqués par des stimuli possédant les mêmes caractéristiques physiques. Il s'agit donc de mettre en œuvre une tâche de choix. Ce choix opéré par le sujet sera une manière de spécifier une « sortie » au système visuel, et nous nous focaliserons sur les indices pouvant montrer que cette sortie a eu lieu.

Cette tâche doit être suffisamment élaborée pour nécessiter un traitement visuel le plus complet possible, dans lequel tous les stimuli seront analysés de manière identique. Une tâche de reconnaissance basée sur un jugement de contenu, en fonction de ce que les images représentent, s'impose si les indices visuels cherchés sont suffisamment variées pour ne pas réduire la tâche à une détection de traits simples encodés dans le système visuel primaire.

Cette tâche de reconnaissance se doit par contre d'être suffisamment simple pour ne pas faire appel à d'autres processus (langage, jugement de grandeur ou de position, mémoire) qui pourraient interférer avec les mécanismes visuels étudiés et parasiter l'analyse des résultats.

Nous avons choisi de demander au sujet si l'image présentée contenait ou non un animal. Un animal peut appartenir aux différentes classes – insectes, reptiles, mammifères, poissons, oiseaux - dans de nombreuses postures, dans des environnements variés, seul ou en groupe, sous plusieurs angles de vue, pouvant être partiellement occulté ou photographié en gros plan. Cette diversité nous assure d'un jugement perceptif basé sur des indices qui peuvent être propres à chaque image, sans permettre au sujet d'élaborer une stratégie reposant sur des indices simples et répétés, ce qui distingue cette tâche d'une

détection simple. Les images distracteurs doivent comporter la même diversité en terme de couleurs, fréquences spatiales et types de scènes.

Afin d'être effectuée rapidement, cette tâche est réalisée en go / no-go, c'est-à-dire que le sujet a pour consigne de « répondre si un animal est présent dans l'image », sous-entendu « ne rien faire dans le cas contraire », qui est la manière la plus simple d'effectuer une tâche de choix. Dans notre cas, la réponse go est associée à la présence de la cible plutôt qu'à son absence, ce qui permet d'associer un temps de réaction aux images de type animaux et aux fausses détections.

b) l'analyse des potentiels évoqués

Nous disposons donc de deux potentiels évoqués moyens, l'un associé à l'ensemble des images contenant un animal et pour lesquelles le sujet a répondu, l'autre évoqué par l'ensemble des images distracteurs pour lesquelles le sujet n'a pas répondu. Le choix des stimuli permet de poser l'hypothèse que les traitements des aspects bas-niveau ou physiques des images sont les mêmes pour les deux types de stimuli. Les potentiels évoqués reflétant ces traitements devraient par conséquent être les mêmes : de visu par superposition, ou par test statistique d'égalité, aucune différence ne devrait apparaître à ce niveau.

Les divergences entre les deux signaux seront analysées en terme de différences de traitement : soit à cause du contenu des images (catégorie visuelle), soit de la tâche de choix accomplie, soit en terme de geste moteur. Nous pouvons noter dès à présent que le fait de disposer de potentiel de référence ne lève pas la possibilité d'un processus *visuel* propre à la catégorie animal, de manière analogue à l'encodage structurel des visages ; mais contrairement aux visages pour lesquels cet encodage semble nécessaire à la reconnaissance, un tel processus serait le reflet d'un traitement catégoriel au niveau super-ordonné, précisément à cause de la diversité des stimuli, et à ce titre, au delà peut-être de l'analyse visuelle de l'image en cours. Quant aux autres causes de divergences, tâche de choix et geste moteur, elle sont par elles-mêmes une conséquence du traitement visuel, en soulignant son caractère suffisant. Les divergences entre potentiels évoqués par les cibles et les distracteurs reflétera donc un traitement visuel - nécessaire pour la tâche - achevé.

Un potentiel évoqué étant par lui même une moyenne, nous appliquerons le test d'égalité de Student entre potentiel cible et distracteur sur chaque échantillon, au cours du temps. Le problème des égalités de séries temporelles a été abordé par Rugg et Coles 1995 ; ces auteurs ont proposé de considérer comme significativement différents des processus mettant en défaut ($p>0.05$) un test d'égalité sur une quinzaine de mesures successives. Dans notre cas, nous considérerons la latence de la première différence significative comme résultat de notre mesure si cette différence se poursuit sur au moins une quinzaine d'échantillons (soit une quinzaine de millisecondes). Il est permis de considérer que les causes de variance entre chaque essai sont les mêmes, qu'il s'agisse

d'un essai cible ou d'un distracteur ; cependant en l'absence de test de Fisher (d'égalité des variances) le test de Student est robuste à partir d'une trentaine d'essais, que la répartition soit normale ou non. Dans la pratique, nous serons bien au delà de ce nombre.

c) l'analyse des réponses comportementales

Cette mesure de latence montrant une différence de traitement peut être étendue à l'analyse des temps de réaction comportementaux. Les mesures de moyennes et médianes de ces temps sont utiles pour montrer des différences entre processus mentaux. Dans nos études, nous disposerons principalement des temps de réaction aux essais 'go' réussis par les sujets, et les temps de réaction aux distracteurs sur lesquels les sujets ont fait une erreur. Ces faux positifs sont en nombres plus réduits si la tâche est réussie correctement par les sujets ; ils sont le reflet de processus probablement plus complexes, en tout cas mettant en échec les processus dont nous cherchons à mesurer la latence. Une comparaison directe des moyennes ou médianes entre essais réussis et faux positifs permet de donner quelques indications sur la manière dont est effectuée la tâche, mais ne permet pas une analyse plus poussée.

Par contre, la présentation en histogrammes de ces deux distributions peut permettre de disposer de plus d'informations ; comme en potentiels évoqués, ces deux distributions sont indifférenciées aux latences courtes, avec des valeurs nulles ou petites. Lorsque ces distributions sont égales en nombre d'essais par intervalle, les processus ayant amené à ces réponses comportementales peuvent être considérés comme indépendants de la tâche à accomplir, puisqu'entraînant un taux de réussite égal à la chance. Si le nombre d'essais par intervalle devient différent pour les distributions des essais réussis et des faux positifs, cela indique que la tâche comportementale biaise significativement le nombre d'essais dans les distributions : ce biais reflète que la tâche est réussie au-delà du seuil de chance sur l'intervalle considéré (à condition qu'il y ait autant de cibles et de distracteurs dans les stimuli). Ainsi les représentations en histogrammes permettent de disposer de la latence à laquelle les temps de réaction ne peuvent plus être considérés comme des anticipations, mais commencent à refléter la réussite - rapide - de la tâche.

Dans la pratique, le test du chi2 entre deux séries de chiffres mesure si la différence de ces deux séries suit une loi normale, c'est-à-dire une somme de tirage aléatoires. Sur chaque intervalle de temps, le nombre d'essais réussis et le nombre de faux positifs peut être comparé à leur moyenne : une égalité montre que l'écart entre les deux peut être attribué à la chance. Nous relèverons pour chaque expérience le premier intervalle de temps sur lequel le test du chi2 entre le nombre d'essais réussis et le nombre de faux positifs d'une part, et leur moyenne d'autre part, montrera une différence significative [1]- si

[1] exemple : sur l'intervalle 230-240 ms, il y a 3 faux positifs et 6 essais réussis. Leur moyenne est 4.5, de partie entière 5. Nous testons l'égalité de deux processus stochastiques : le premier a donné 3 lors d'une première réalisation, puis 6 à la seconde ; le second a réalisé 5, puis de nouveau 5. Le test du chi2 entre les deux réalisations (3 ;6) et (5 ;5) montre que les deux processus sont égaux à $p>0.3$

cette différence persiste sur les intervalles suivants, une quinzaine… Cette mesure de temps de réaction indiquera la latence avec laquelle la tâche commence à être réussie.

d) les sources dipolaires

Depuis une dizaine d'années, les techniques de localisation de sources des potentiels évoqués de surface se sont suffisamment développées pour permettre une localisation approximative des aires cérébrales ayant généré ces activités. Ces localisations reposent sur l'hypothèse que toute distribution de potentiel de surface possédant une topographie simple peut être expliquée par une source dipolaire unique profonde (cérébrale). La topographie simple est un gradient circulaire à la surface, si cette surface est une sphère. Dans le cas du scalp, cette sphéricité est loin d'être évidente ; elle peut être obtenue par approximation en modifiant les emplacements réels des électrodes du bonnet EEG (par projection et correction de sphéricité). Des techniques plus récentes permettent d'affiner ces localisations en considérant une ellipsoïde, ou un modèle réel obtenu par maillage (Yvert et al. 1997).

L'hypothèse de source unique profonde est valide dans les cas connus pour lesquels un groupe de neurones organisés en colonnes corticales adjacentes peut être le seul lieu d'où provient le potentiel enregistré. Cette hypothèse est cohérente avec la principale explication des potentiels évoqués de surface, c'est-à-dire celle faisant intervenir un ensemble de neurones, pour lesquels l'activité simultanée des dendrites produit un ensemble de différences de potentiels locales, dont la somme est enregistrée à la surface si leur topographie est favorable (résultante locale non nulle et orientation orthogonale à la surface). Par contre, il est clair que des phénomènes neuronaux comme l'encodage distribué sont moins compatibles avec ce modèle ; mais il est peu probable qu'un encodage distribué puisse être efficace sur de grandes distances. Un tel encodage peut impliquer de nombreux neurones d'une même aire, rendant pas pertinente une localisation unique. Nous pouvons supposer aussi que des phénomènes électriques très distribués ne puissent pas être enregistrés à la surface comme potentiels évoqués.

Les techniques mises en œuvre pour la modélisation dipolaire des activités de surface reposent, comme de nombreuses solutions approchées aux problèmes inverses, sur les algorithmes de recuit simulé dont la fonction d'énergie est ici l'erreur commise entre la modélisation et les données. Une fois définie une situation d'origine à un ou plusieurs dipôles dont les emplacements et les orientations sont connus, l'amplitude de ces dipôles est ajustée pour rendre compte au mieux des données (ce calcul est analytique dans le cas sphérique) ; l'erreur commise en chaque électrode est calculée, puis les dipôles déplacés aléatoirement jusqu'à obtenir une direction minimisant l'erreur ; cette direction empruntée définit un nouvel emplacement et de nouvelles orientations à partir desquelles les amplitudes sont de nouveau ajustées. Le processus continue jusqu'à converger vers un minimum d'erreur. Ce minimum peut être local (non robuste), ou plus global s'il est le résultat de plusieurs modélisations à conditions initiales différentes. Plusieurs solutions

peuvent être trouvées. On voit donc que la recherche des sources dipolaires à des potentiels évoqués procède de méthodes d'approximation supposant de l'utilisateur un certain nombre de choix : le choix du nombre de dipôles, de leur positions et orientations initiales, et des contraintes entre dipôles – contraintes de symétries ou d'orientations, pouvant être fixées pour abaisser le nombre de libertés sur les déplacements. D'autre part, c'est à l'utilisateur de définir correctement l'intervalle de temps sur lequel sera ajusté le modèle, en fonction d'une topographie de surface stable, et d'un nombre suffisamment grand d'échantillons permettant un ajustement fiable. Enfin, c'est à l'utilisateur d'effectuer le tri des solutions multiples pour n'en retenir que les plus probables.

Les stratégies de recherche que nous avons utilisées s'appuient sur les méthodes habituellement employées : n'accroître le nombre de sources que lorsque l'algorithme ne converge pas vers des solutions stables, restreindre la plage d'action d'une source à une seule activité. Cette plage d'action est définie prioritairement par le front ascendant d'une activité, le meilleur ajustement étant trouvé aux pics (où les données possèdent le plus de variance). L'ajustement des sources est effectué en deux étapes, la première fixant la localisation par l'utilisation de sources insensibles aux variations d'orientations (un triplet orthogonal de dipôles ou 'source régionale'), puis en ajustant l'orientation sur un seul dipôle si les variations d'amplitudes sont équivalentes sur le triplet. Enfin, l'implication des deux hémisphères cérébraux pouvant être considéré comme identique dans nombre de cas expérimentaux, nous avons utilisé de nombreuses fois la symétrie de localisation pour contraindre des paires de dipôles.

La localisation des sources de l'activité différentielle de l'expérience principale 'Animal / non-animal' a été effectuée par B. Doyon. C'est son assistance qui a permis de mener à bien la recherche des sources de l'expérience 'présentations alternées', et ses conseils qui ont permis l'initiation à la théorie et à la délicate pratique du logiciel BESA (Scherg et Berg 1991) v.2.0 puis v.99 que nous avons utilisés pour les études de sources suivantes.

3 l'expérience animal / non-animal

3.1 protocole et hypothèses

sujets, tâche et stimuli

Les 15 sujets (7 hommes et 8 femmes) avaient un âge compris entre 22 et 45 ans. La consigne était de relâcher le plus rapidement possible le bouton d'une souris placée sous leur main dominante, dès que le sujet apercevrait un animal dans la scène présentée. Les stimuli étaient affichés sur un écran vidéo placé à 1m, centralement par rapport au point de fixation (une croix blanche sur fond noir). Les sujets étaient avertis de la rapidité

d'apparition des stimuli : pendant 20 ms (2 trames non-entrelacées à 100 Hz), et de la durée interstimulus aléatoire autour de 2 s (entre 1.5 s et 2.5 s).

Les scènes présentées comme stimuli étaient choisies parmi un ensemble de 4300 photographies en couleur disponible dans le commerce (Corel). Les photographies utilisées comme cibles contenaient des animaux des différentes classes – insectes, reptiles, mammifères, poissons, oiseaux - dans de nombreuses stations, dans des environnements variés, seul ou en groupe, sous plusieurs angles de vue. Les distracteurs étaient composés de scènes naturelles (paysages divers, bâtiments, plans de fleurs, fruits, légumes). Chaque scène (de 12.6x8.6 cm, 384x256 pixels, 6,6x4,5°) n'était vue qu'une fois par un sujet. Le type de scènes présentées (cibles de la tâche ou distracteurs) était tiré au hasard.

Il était demandé au sujet de catégoriser un minimum de 700 photographies (un sujet allant jusqu'à 2000), groupées en séries de 100. Chaque série expérimentale durait approximativement 4 mn ; entre chaque série le sujet se reposait à sa convenance pendant un minimum de 3 mn. Les sujets avaient la possibilité d'interrompre le défilement des images au cours de chaque série, en ne rappuyant pas immédiatement sur le bouton après la détection d'une cible : de cette manière, les sujets pouvaient se détendre (dos, clignement des yeux, étirements) étant donné la fatigue occasionnée par l'expérience durant au total approximativement une heure et demie.

poste expérimental et traitement des données

Le poste expérimental est composé d'une têtière relié au bonnet d'enregistrement, de l'écran haute définition pour l'affichage des stimuli et de la souris dans la pièce expérimentale derrière une cloison ; de l'autre côté, se trouvent le système d'amplification d'EEG Synamps et deux micro-ordinateurs type PC. L'un des PC contrôle la stimulation (carte graphique VSG 2.2 Cambridge Vision Research, et carte horloge DCC5p Industrial Computer Source) et affiche les numéros, types d'images et résultats comportementaux au cours de la séance. L'autre PC contrôle l'enregistrement synchronisé à la stimulation et les traitements de l'électroencéphalogramme au moyen du logiciel NeuroScan, et permet de corriger en direct les dérives de tension pouvant saturer l'amplification.

L'électroencéphalogramme était capté au moyen d'un bonnet 20 voies (norme 10-20, electrocap). Chaque voie, dont le contact au moyen d'une électrode Ag/AgCl était d'une résistance inférieure à 20 k ohms, était amplifiée (gain 1000) séparément, filtrée (passe-bas à 100Hz, 12dB/octave), échantillonnée à 1000 Hz (résolution 0.084 μV/bit), et enregistrée en continu au cours de chaque série. Les traitements en fin de séance comprenaient le découpage de l'EEG en période de 500 ms ([-100 ;400] ms post-stimulus), l'association de chaque période avec les résultats comportementaux et le type de l'essai, le rejet d'artefacts liés aux mouvements oculaires (critère [-50 ;50] μV sur les électrodes FP1 et FP2) et aux rythmes de perte de vigilance (critère généralement de [-20 ;20] μV sur [-100 ;0] ms sur les électrodes FZ, CZ et PZ). Les essais étaient ensuite triés et moyennés en fonction de leur

type (cible ou distracteur) et des réponses fournies. Les potentiels évoqués par les scènes étaient moyennés sur environ 340 essais par condition et par sujet.

Les deux potentiels évoqués étaient superposés et comparés au moyen du test d'égalité de moyennes (Student) pour chaque échantillon (1 ms). La moyenne interindividuelle des différences cibles-distracteurs de chaque sujet permettait une mesure de significativité de groupe, moyenne testée contre l'hypothèse nulle (Student apparié, ddl 14). Ces différences entre cibles et distracteurs étaient considérées comme significatives si 15 valeurs consécutives de t-test dépassaient le niveau 0.05 de significativité (Rugg et Coles 1995). Les cartographies de surface des potentiels évoqués étaient réalisées au moyen des outils fournis par le logiciel Neuroscan, c'est-à-dire une interpolation simple entre électrodes dans le plan de projection défini à l'estime par l'utilisateur.

hypothèses

• Nous cherchons à mettre en évidence un signe électrophysiologique indiquant un traitement différent pour les scènes cibles et les scènes distracteurs dans cette tâche de catégorisation en go / no-go. Lors de la réalisation de cette étude en 1995, l'hypothèse principale était qu'il serait possible de révéler le processus moteur en réponse aux cibles réussies ; cette latence plus courte que les réponses comportementales pouvant déjà fournir une limite supérieure du temps pris par le traitement visuel pour décider de la présence d'animaux dans les images - les réactions comportementales étant obtenues dans une tâche où il était demandé aux sujets de répondre le plus vite possible.

• Les études électrophysiologiques disponibles en go / no-go (Simson et al. 1977 ; Pfefferbaum et al. 1985 ; Gemba et Sasaki 1989 ; Jodo et Kayama 1992 ; Sasaki et al. 1993) ont montré un accroissement d'amplitude de la N2 pour les réponses no-go (pic à 280 ms, visible à partir de 220 ms) de distribution fronto-centrale. Par mise en évidence de la divergence des potentiels évoqués associés aux cibles et aux distracteurs, cette activité aux environs de 220-240 ms pourrait révéler un traitement différencié indiquant que la reconnaissance de la catégorie cible était effectuée. Ici aussi cependant, les stimuli utilisés dans ces études sont des diodes électroluminescentes (LED), ou plus rarement des dessins (Nasman et Rosenfeld 1990) : l'incidence de l'utilisation de scènes naturelles ne pouvait pas être prédit.

3.2 article publié

Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

Centre de Recherche Cerveau & Cognition, UMR 5549, 31062 Toulouse, France

How long does it take for the human visual system to process a complex natural image? Subjectively, recognition of familiar objects and scenes appears to be virtually instantaneous, but measuring this processing time experimentally has proved difficult. Behavioural measures such as reaction times can be used[1], but these include not only visual processing but also the time required for response execution. However, event-related potentials (ERPs) can sometimes reveal signs of neural processing well before the motor output[2]. Here we use a go/no-go categorization task in which subjects have to decide whether a previously unseen photograph, flashed on for just 20 ms, contains an animal. ERP analysis revealed a frontal negativity specific to no-go trials that develops roughly 150 ms after stimulus onset. We conclude that the visual processing needed to perform this highly demanding task can be achieved in under 150 ms.

Neurophysiological measurements of the latencies of selective visual responses can be used to provide estimates of visual processing time[3]. For example, it is known that higher-order visual areas such as the primate superior temporal sulcus contain neurons that can respond selectively to faces with latencies of ~ 100 ms[4–6]. In humans, face-selective evoked potentials have been demonstrated using both surface ERP recordings[7,8] and implanted intracerebral electrodes[9–11]. Such potentials typically peak at ~ 200 ms after stimulus onset, but may start as early as 140 ms. It is unclear, however, whether such latencies are typical of visual processing in general. One problem is that face processing may involve highly specialized and optimized neural pathways, and although there have been a few reports of early differential responses to other stimuli, including words[10,12,13] and line drawings[12], no previous ERP studies have attempted to measure processing times for more natural scenes. A second problem is that the existence of a short latency differential response does not imply that visual processing has been completed—responses to faces, for example, could correspond to an early processing stage such as 'structural encoding'. One can tackle this problem by using a task that requires the subject to make some sort of categorical judgment about the stimulus, an approach that has been used in a variety of studies using not only photographs of faces[14–17], but also both line drawings[18,19] and photographs[20] of everyday objects. However, differential effects in such tasks occur at considerably longer latencies, typically involving the N400 component of the ERP.

The present study used a task that provides a very serious challenge to the processing capacities of the human visual system. Subjects performed a go/no-go categorization in which they had to decide on the basis of a 20-ms presentation whether an image contained an animal or not. Earlier studies using rapid sequential visual presentation (RSVP) had shown that subjects can detect photographs of animals in a string of images at high presentation rates[21], but this is the first time such a scene categorization task has been performed using ERPs. We used a set of over 4,000 commercially available colour photographs, of which roughly half were used as targets and included a wide range of animals in their natural environments (mammals, birds, reptiles, fish); the remainder were distractors that included pictures of forests, mountains and lakes, as well as buildings, flowers and fruit. As in a number of other studies[8,9,11,16,19], each stimulus was only ever seen once, thus eliminating the possibility of stimulus-specific learning effects.

Despite the very high demands made on the visual system by such a task (the subjects had no *a priori* information about the type of animal to look for, its position or size, or even the number of animals present), performance was remarkably good. The average proportion of correct responses was 94%, with one of the fifteen subjects achieving 98% correct responses. The median reaction times on 'go' trials was 445 ms, although this value varied considerably between subjects, from a minimum of 382 ms to as much as 567 ms (Fig. 1). This remarkable level of performance was possible despite the very brief presentations, which effectively rule out the use of eye movements during image processing.

Whereas the behavioural reaction times put an upper limit on the time required for visual processing, the analysis of event-related potentials provided a much stronger constraint. By comparing average brain potentials generated on correct 'go' trials with those generated on correct 'no-go' trials, we were able to demonstrate that the two potentials diverge very sharply at ~ 150 ms after stimulus onset. The effect was particularly clear at frontal recording sites, and was characterized by a nearly linear increase in the voltage difference over the following 50 ms or so, the potential being more negative on no-go trials (Fig. 2). All 15 subjects showed the effect (Fig. 3), and although the onset latency varied somewhat between subjects, the differences were very minor compared with the very large differences in behavioural reaction times. Furthermore, there was no correlation whatsoever between behavioural reaction time and the onset latency for the differential response. This makes it unlikely that the differential

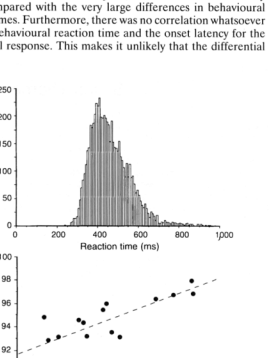
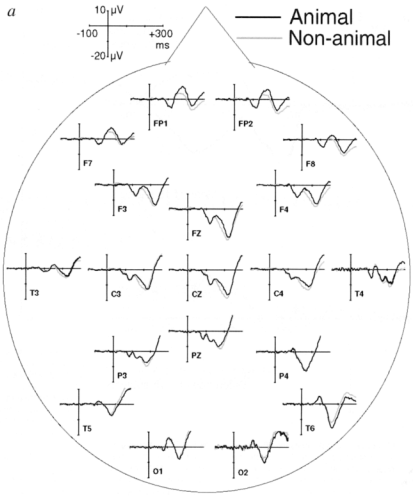
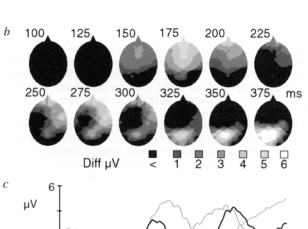


FIG. 1 Behavioural performance measures. *a*, Distribution of reaction times on target trials for the 15 subjects participating in the study (7 males and 8 females, aged between 22 and 45 years of age), each of whom performed at least 700 trials (range, 700–2,000). Subjects pressed a button to start a sequence of trials and, after an interval of 1 to 2 seconds, an image was presented at the centre of the screen. A small fixation cross was present before and after stimulus presentation. Subjects were instructed to release the button if they saw an animal ('go' trials), and to keep their finger on the button otherwise ('no-go' trials). Target and distractor trials were presented at random, with roughly equal probability in blocks of 100 trials. All the images were 384 by 256 pixels in size and were presented for 20 ms (two frames at 100 Hz) using a Cambridge Vision Research VSG 2/2 graphics board mounted in a PC compatible computer. *b*, Accuracy as a function of median reaction time for each of the 15 subjects. The dashed line plots a linear regression ($r = 0.623$) and indicates the presence of a speed–accuracy trade-off.





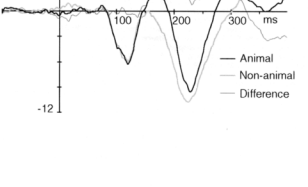
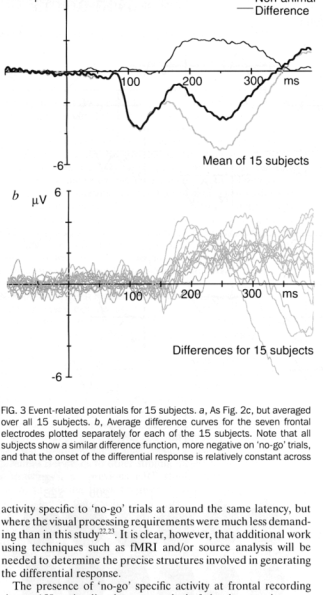


FIG. 2 Evoked potential data for one subject. *a*, Event-related potentials plotted for a 400-ms period starting 100 ms before stimulus onset. Solid lines plot the average response on correct target trials; grey lines plot averages for correct distractor trials. Data were obtained from one subject during 4 separate sessions (a total of 1,580 trials). The two averages overlap until ~150 ms, at which point a clear difference emerges between the potentials on animal and non-animal trials which was clearest at frontal recording sites. Recordings were made using a 20-electrode Electrocap bonnet in the 10–20 configuration connected to a Neuroscan SynAmps system sampling at 1,000 Hz. Signals were digitally filtered using a low pass filter with a cut-off frequency of 100 Hz and a notch filter to remove 50 Hz mains interference. Potentials on individual trails were baseline corrected on the basis of the 100 ms preceding stimulus onset, and any trials contaminated by eye movement artefacts were excluded from the analysis. Data analysis was performed using the SCAN software suite. *b*, Two-dimensional plots of the difference between correct 'go' and 'no-go' trials averaged over 25-ms time slices, starting at 100 ms. No difference can be seen in the first two slices, but from 150 ms onwards a clear difference appears which affects all the frontal electrodes. At around 325 ms, there are clear signs of a lateralized difference at more posterior sites which probably reflects motor activity in this right-handed subject. *c*, Event-related potentials on target (thick black lines) and distractor (grey) trials calculated by averaging the responses for all seven frontal electrodes (FP1, FP2, F3, F4, F7, F8 and FZ) for this one subject. The difference curve (thin black line) demonstrates the sharp onset of the differential response which is more negative on no-go trials. Using the statistical procedure proposed by Rugg *et al.* for determining the onset of the differential effect[28] (at least 15 consecutive *t*-test values exceeding the 0.05 level of significance) the first significant effect was seen for electrode F8 at 152 ms (d.f. = 1,578). All seven electrodes reached significance by 157 ms ($2.20 < t < 3.74$), and the level of significance increased monotonically to reach a peak at 186 ms (mean *t*-score for the seven electrodes, 6.72).

effect seen at frontal sites is related to motor activation. We did see lateralized differential activity that could be related to motor preparation (Fig. 2*b*) but this was localized more posteriorly and did not occur until considerably later (300 to 400 ms after stimulus onset).

What could be producing the early differential activity? We can rule out some simple systematic difference between animal and non-animal images because the same differential effects were seen for a very wide range of images. Presumably, the difference must be related to some sort of decision-related activation that occurs only once the necessary visual processing has been completed. One possibility is that the difference results from activity that is specific to 'go' trials, perhaps related to target detection. If target detection was indeed critical, then the differential response should start earlier on trials where the subjects responded earlier—that is, on trials where the target animal was easier to detect. However, Fig. 3*c* shows that there was no difference whatsoever between the latency of the differential activity evoked on 'fast' trials and 'slow' trials. The most plausible explanation of this result is that the difference is not generated by 'go'-related neural activity, but rather by neural activity that is specifically generated on 'no-go' trials. This could reflect a role for frontal areas in inhibiting inappropriate behavioural responses, an interpretation supported by a number of earlier studies that reported frontal



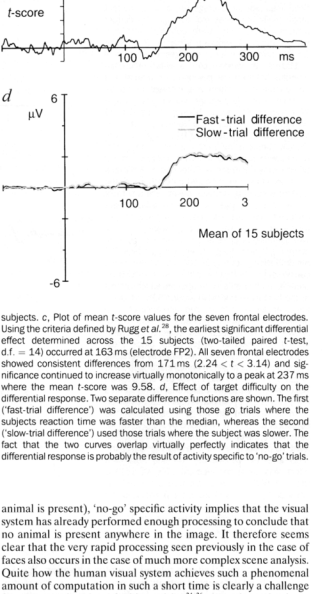


FIG. 3 Event-related potentials for 15 subjects. *a*, As Fig. 2*c*, but averaged over all 15 subjects. *b*, Average difference curves for the seven frontal electrodes plotted separately for each of the 15 subjects. Note that all subjects show a similar difference function, more negative on 'no-go' trials, and that the onset of the differential response is relatively constant across subjects. *c*, Plot of mean *t*-score values for the seven frontal electrodes. Using the criteria defined by Rugg *et al.*[28], the earliest significant differential effect determined across the 15 subjects (two-tailed paired *t*-test, $d.f. = 14$) occurred at 163 ms (electrode FP2). All seven frontal electrodes showed consistent differences from 171 ms ($2.24 < t < 3.14$) and significance continued to increase virtually monotonically to a peak at 237 ms where the mean *t*-score was 9.58. *d*, Effect of target difficulty on the differential response. Two separate difference functions are shown. The first ('fast-trial difference') was calculated using those go trials where the subjects reaction time was faster than the median, whereas the second ('slow-trial difference') used those trials where the subject was slower. The fact that the two curves overlap virtually perfectly indicates that the differential response is probably the result of activity specific to 'no-go' trials.

activity specific to 'no-go' trials at around the same latency, but where the visual processing requirements were much less demanding than in this study[22,23]. It is clear, however, that additional work using techniques such as fMRI and/or source analysis will be needed to determine the precise structures involved in generating the differential response.

The presence of 'no-go' specific activity at frontal recording sites at 150 ms implies that a great deal of visual processing must have been completed before this time. Indeed, although activity related to target detection could be explained relatively easily (the presence of an eye or feathers would be enough to decide that an animal is present), 'no-go' specific activity implies that the visual system has already performed enough processing to conclude that no animal is present anywhere in the image. It therefore seems clear that the very rapid processing seen previously in the case of faces also occurs in the case of much more complex scene analysis. Quite how the human visual system achieves such a phenomenal amount of computation in such a short time is clearly a challenge for current theories of object vision[24–26], but given the large number of processing stages involved in primate visual system, it seems likely that much of this processing must be based on essentially feed-forward mechanisms[3,4,27]. □

Received 21 November 1995; accepted 22 March 1996.

1. Potter, M. C. & Faulconer, B. A. *Nature* **253,** 437–438 (1975).
2. Rugg, M. D. & Coles, M. G. H. *Electrophysiology of Mind: Event-related Brain Potentials and Cognition* (Oxford University Press, Oxford, 1995).
3. Thorpe, S. J. & Imbert, M. *in Connectionism in Perspective* (eds Pfeifer, R., Schreter, Z., Fogelman-Soulié, F. & Steels, L.) 63–92 (Elsevier, Amsterdam, 1989).
4. Oram, M. W. & Perrett, D. I. *J. Neurophysiol.* **68,** 70–84 (1992).
5. Perrett, D. I., Rolls, E. T. & Caan, W. *Exp. Brain Res.* **47,** 329–342 (1982).
6. Rolls, E. T. & Tovee, M. J. *Proc. R. Soc. Lond. B* **257,** 9–15 (1994).
7. Jeffreys, D. A. & Tukmachi, E. S. *Exp. Brain Res.* **91,** 340–350 (1992).
8. Seeck, M. & Grüsser, O. J. *Exp. Brain Res.* **92,** 338–349 (1992).
9. Halgren, E., Baudena, P., Heit, G., Clarke, M., Marinkovic, K. & Clark, M. *J. Physiol., Paris* **88,** 1–50 (1994)
10. Allison, T., McCarthy, G., Nobre, A., Puce, A. & Belger, A. *Cereb. Cortex* **4,** 544–554 (1994).
11. Allison, T. *et al. J. Neurophysiol.* **71,** 821–825 (1994).
12. Nobre, A. C., Allison, T. & McCarthy, G. *Nature* **372,** 260–263 (1994).
13. Rudell, A. P. *EEG. Clin. Neurophysiol.* **83,** 77–82 (1992).
14. Barrett, S. E., Rugg, M. D. & Perrett, D. I. *Neuropsychologia* **26,** 105–117 (1988).
15. Barrett, S. E. & Rugg, M. D. *Neuropsychologia* **27,** 913–922 (1989).
16. Holcomb, P. J. & MacPherson, W. B. *Brain Cogn* **24,** 259–276 (1994).
17. Bobes, M. A., Valdes Sosa, M. & Olivares, E. *Brain Cogn.* **26,** 1–22 (1994).
18. Barrett, S. E. & Rugg, M. D. *Brain Lang.* **38,** 424–437 (1990).
19. Barrett, S. E. & Rugg, M. D. *Brain Cogn* **14,** 201–212 (1990).
20. Pratarelli, M. E. *Brain Cogn* **24,** 137–157 (1994).
21. Intraub, H. *J. Exp. Psychol. Hum. Perc. Perf.* **7,** 604–610 (1981).
22. Gemba, H. & Sasaki, K. *Neurosci. Lett.* **101,** 263–268 (1989).
23. Sasaki, K., Gemba, H., Nambu, A. & Matsuzaki, R. *Neurosci. Res.* **18,** 249–252 (1993).
24. Vetter, T., Hurlbert, A. & Poggio, T. *Cereb. Cortex* **5,** 261–269 (1995).
25. Bülthoff, H. H., Edelman, S. Y. & Tarr, M. *Cereb. Cortex* **5,** 247–260 (1995).
26. Hummel, J. E. & Biederman, I. *Psychol. Rev.* **99,** 480–517 (1992).
27. Celebrini, S., Thorpe, S., Trotter, Y. & Imbert, M. *Vis. Neurosci.* **10,** 811–825 (1993).
28. Rugg, M. D., Doyle, M. C. & Wells, T. *J. cog. Neurosci.* **7,** 209–227 (1995).

ACKNOWLEDGEMENTS. This work was supported by the Human Capital and Mobility program, the CNRS and the Conseil Régional Midi-Pyrenees.

CORRESPONDENCE and requests for materials should be addressed to S.T. (e-mail: thorpe@cerco.ups-tlse.fr).

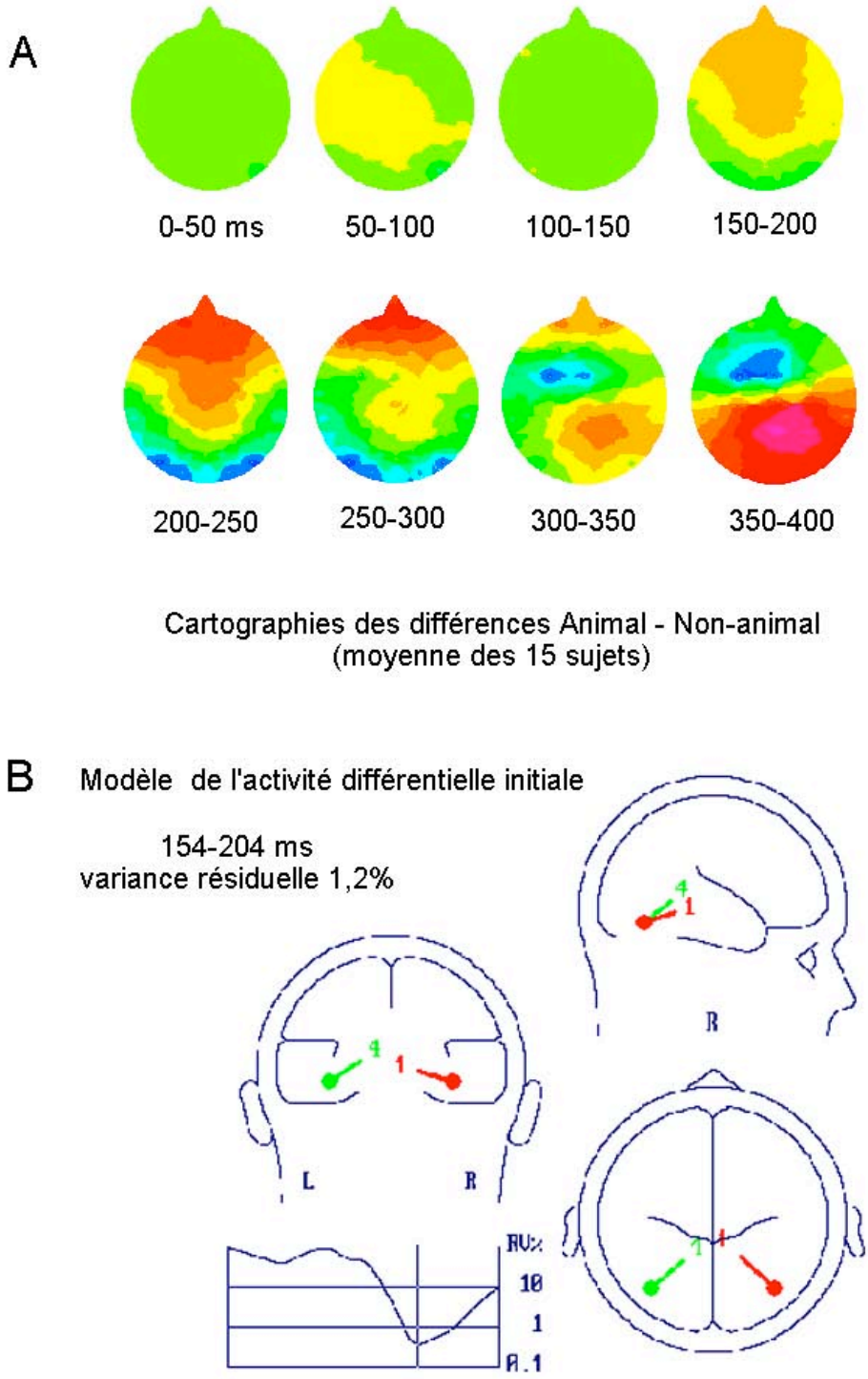


Figure 9 : Cartographies et modèle de l'activité différentielle initiale

A. Cartographies de l'activité différentielle moyenne des 15 sujets entre 0 et 400 ms. L'activité possède une topographie uniforme entre 150 et 230 ms (pic moyen à 205 ms). L'activité tardive vers 370 ms est probablement liée à la préparation motrice.

B. Modèle dipolaire de l'activité différentielle entre 154 et 204 ms. Deux sources occipitales symétriques (aires visuelles extrastriées) suffisent à expliquer plus de 98% de la variance.

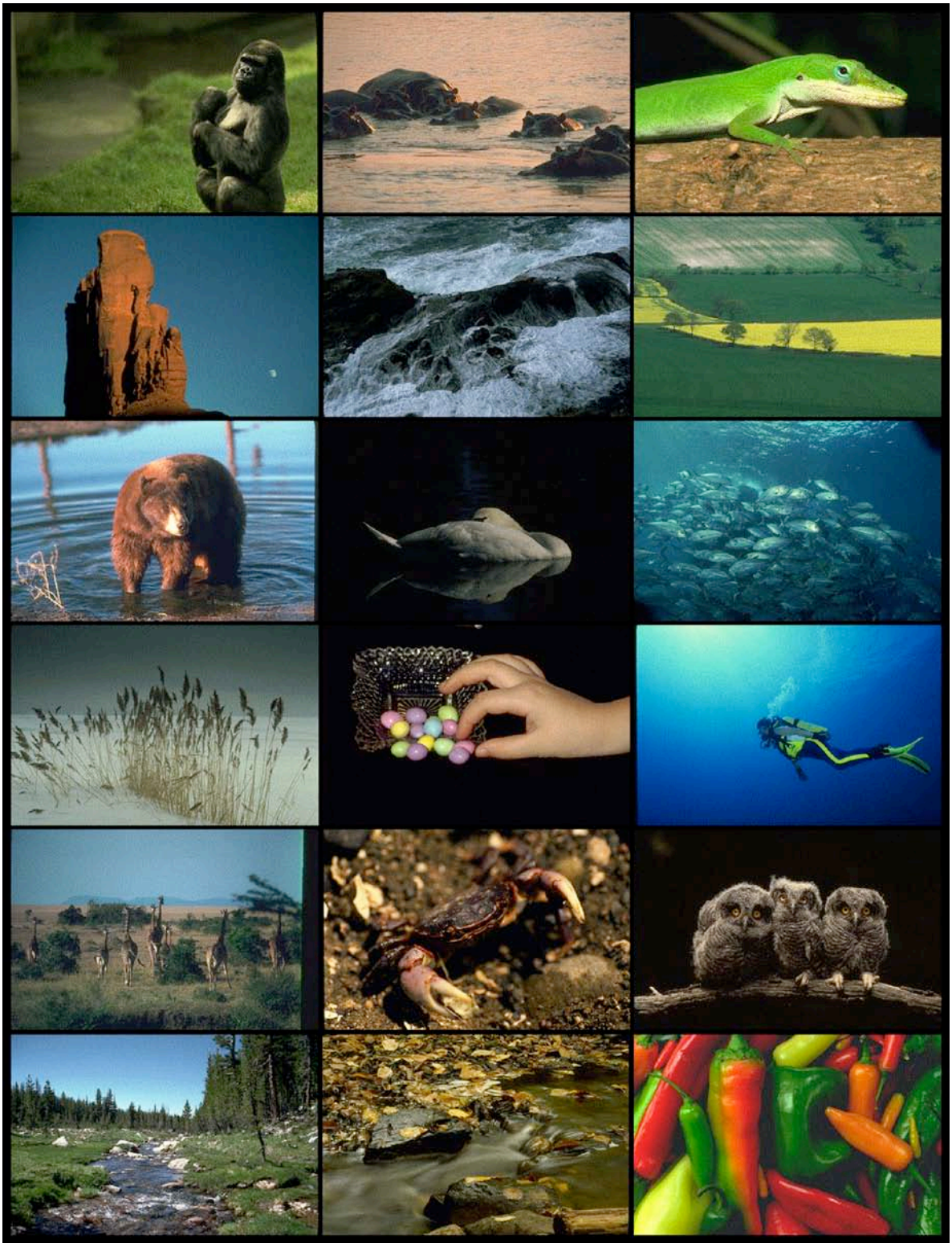


Figure 10. Exemples de stimuli.

Les photographies étaient choisies pour inclure tout type d'animaux et de vues. Les scènes étaient aussi variées que possible en luminance, fréquences spatiales, couleurs, positions des cibles. Les humains ne sont pas considérés comme appartenant à la catégorie animal.

3.3 sources principales de l'activité différentielle

Nous avons cherché à mettre en évidence les sources de l'activité différentielle visible à partir de 150 ms. L'intérêt de rechercher les bases cérébrales d'une *différence* de potentiels évoqués va de soi dans l'approche que nous avons définie ici : la différence permettant d'exprimer la latence avec laquelle deux processus commencent à diverger, la modélisation de cette différence permet d'obtenir des indices sur le lieu dont elle provient.

Ce genre de modélisation n'est pas courant en électrophysiologie de surface. Les raisons en sont principalement la difficulté d'interprétation des potentiels évoqués, et la méfiance portée aux méthodes soustractives qui ont pu être utilisées systématiquement en psychologie expérimentale. L'étude de différences n'a en effet de sens que si les deux conditions comparées sont très proches du point de vue des bases cérébrales impliquées : comment justifier l'étude d'une différence de N400 obtenue dans le cas d'amorçage sémantique et perceptif ? Dans le cas du protocole mis en œuvre ici, les résultats montrent que les signes électrophysiologiques des traitements effectués sont identiques jusqu'à 150 ms, puis divergent de manière continue jusqu'à environ 210 ms. Ces deux propriétés (il s'agit de la première différence rencontrée, elle est continue sur environ 60 ms) nous permettent de postuler que les bases cérébrales sont identiques pour les deux processus cibles et distracteurs jusqu'à 150 ms et qu'un seul phénomène vient les différencier pendant 60 ms. La modélisation de ce seul phénomène respecte donc les hypothèses des modélisations dipolaires, à savoir l'isolement de topographies de potentiels stables et l'utilisation d'un nombre minimal de sources pouvant les expliquer. Ceci nous conduit à attribuer quelque justesse aux modélisations de différences, à condition que le protocole n'amène que des hypothèses simples et séparées dans le temps[1]. Dans cette première expérience et celles du chapitre suivant, nous ne modéliserons *que* ces 'premières différences', afin de travailler avec des hypothèses simples et de confirmer la robustesse des résultats sur plusieurs populations et dans différentes conditions expérimentales.

Nous attendons de ces modélisations qu'elles fournissent avant tout des indications de provenance globale de cette activité différentielle. En effet, cette activité observée dans cette première expérience semblait provenir d'aires frontales, étant donné sa forte amplitude sur les sites frontaux. Les résultats de la modélisation effectuée sur 160-210 ms indique au contraire une forte implication bilatérale des lobes occipitaux, dont la représentation réduite à deux dipôles symétriques suffit à expliquer plus de 98 % de la variance du signal. Le relatif éloignement de ces deux sources du pôle occipital permet d'exclure une activité réduite à l'aire visuelle primaire : ces sources expriment une forte activité bilatérale des aires visuelles extrastriées (Figure 9). Des précisions supplémentaires sur les aires impliquées ne seront cherchées qu'au moyen de l'IRM fonctionnelle (chapitre 4).

[1] A notre connaissance, la modélisation systématique de différences n'a été menée que dans une seule étude, celle de Anllo-Vento et al. 1998.

4 discussions et conséquences

résumons

• La tâche de catégorisation d'images 'Animal vs non-animal' est effectuée rapidement et très bien réussie par les sujets, malgré le temps de présentation court (20 ms), la complexité des scènes (photographies), l'absence d'aide du contexte et l'impossibilité de faire appel à l'exploration oculaire.

• Les potentiels évoqués par les scènes contenant un animal et celles n'en contenant pas sont identiques jusqu'à 150 ms post-stimulus.

• Ces potentiels divergent de manière continue de 150 ms jusqu'à environ 210 ms. Cette activité provient d'une différence d'activation des aires visuelles bilatérales extrastriées.

4.1 la rapidité du traitement visuel

D'où peut provenir cette différence d'activité observable à partir de 150 ms ? La grande diversité des images utilisées permet d'exclure toute différence systématique entre les scènes d'animaux et non-animaux, du point de vue des fréquences spatiales, couleurs ou traits saillants. Cette différence peut être reliée à un processus de décision en lien avec la tâche à accomplir, ne pouvant apparaître que lorsque le traitement visuel nécessaire à l'analyse de la scène a été achevé. La présence d'une activité différentielle plus tardive (300-400 ms) au dessus des aires motrices, ainsi que la localisation de l'activité à 150 ms dans des aires visuelles – et sa latence - semble exclure tout processus directement moteur dans ce phénomène. Comme il a été discuté dans l'article présenté, l'observation du fait que la latence ne se modifie pas en fonction du temps de réaction comportemental laisse penser à une activité davantage liée aux distracteurs qu'aux cibles de la tâche (l'hypothèse étant qu'à une scène difficile à analyser seront associés des temps de réaction plus longs). Une telle activité liée aux réponses 'no-go' est à faire correspondre avec les hypothèses d'inhibition déjà observées dans des protocoles go / no-go, mais à des latences plus grandes et avec des stimuli moins complexes (Gemba et Sasaki 1989 ; Jodo et Kayama 1992). L'inhibition no-go liée à la couleur d'une LED serait-elle plus longue à s'initier que celle liée à l'absence d'un animal dans une scène naturelle ? Il semble plutôt qu'un tel phénomène d'inhibition motrice (décrit comme de sources frontales dans les études citées) soit probablement inclus dans les activités plus tardives, entre 210 et 350 ms dans notre tâche. Plutôt qu'à une inhibition motrice, l'activité à 150 ms pourrait être associée à une activité dans les aires visuelles (qui peut être inhibitrice elle aussi, comme suggéré par le potentiel évoqué plus ample aux distracteurs qu'aux cibles), comme indiqué par la localisation extrastriée des sources.

Une possibilité est que cette activité différentielle reflète une activation des neurones répondant à des traits visuels spécifiques aux animaux (pattes, plumes, yeux, etc.) comparés aux autres scènes naturelles. Dans ce cas, de nombreux neurones peuvent participer à l'émergence de cette activité (un ensemble de neurones dans des aires analogues à V4 ou IT du primate non-humain), résultant alors de la compilation d'activations spécifiques à chaque image d'animaux. L'émergence d'un potentiel signifierait que ces neurones sont groupés en colonnes corticales de même orientation, probablement proches donc incluses dans une zone corticale délimitée ; de tels potentiels évoqués n'ont jamais été décrits dans les publications que nous avons consultés (en IRMf, certaines études relatent l'activation du gyrus occipital médian dans le cas de stimuli animaux, par exemple Martin et al. 1996). Si c'était le cas, l'absence de variation de la latence de l'activité différentielle en fonction du temps de réaction comportemental ne peut être expliqué que par un parallélisme massif des voies visuelles portant l'information vers ces aires en temps constant. D'autre part, la variabilité observée pourrait provenir de la décision liée à la tâche (effet d'accumulation et de seuil, boucles feed-back, accès périodique à l'information sensorielle Dehaene 1993) ou dans un délai d'exécution de la réponse motrice. Une telle activation n'est pas incompatible avec l'inhibition proposée dans le paragraphe précédent, si cet ensemble de neurones forme un réseau de connexions inhibitrices avec les autres neurones activés par la présentation des scènes de type distracteur ; cette éventualité sera approfondie dans l'étude menée en IRMf. Si l'activité différentielle observée ici est le résultat d'activations spécifiques aux traits visuels définissant un ensemble d'animaux, il est tentant de comparer cette activation à celle liée à la présentation de visages, observée à des latences comparables ou plus courtes. Cependant, l'activité liée aux visages est très réduite lors de présentations de visages aux traits mélangés, donc très sensible à la qualité des stimuli visuels présentés permettant ou non un encodage de structure (e.g. George et al. 1996). Dans le cas d'une activité 'Animal', nous observons que cette activité se montre insensible à la configuration visuelle puisqu'évoquée par de nombreuses vues variées d'espèces différentes, laissant penser davantage à un encodage de type catégoriel plutôt que de structure visuelle. Dans ce cas aussi, 150 ms indiquent la latence à laquelle le traitement visuel est suffisant pour décider de la présence ou de l'absence d'une catégorie aussi variée que celle des animaux.

Dans tous les cas, cette étude montre qu'une tâche de catégorisation de scènes complexes peut être effectuée sans aide contextuelle comme le postulent certains modèles computationnels de la reconnaissance visuelle. Elle peut être effectuée sans saccades oculaires aidant à centrer une zone d'intérêt de l'image. La latence de la première différence observée en potentiels évoqués est insensible à la complexité de la scène, si celle-ci se reflète dans le temps de réaction comportemental. Cette rapidité fournit des contraintes fortes quant aux modèles de la reconnaissance visuelle : étant donné les vitesses de conduction des voies neuronales et le nombre d'étapes que l'influx nerveux doit traverser pour que des neurones sélectifs à des réponses visuelles élaborées puissent répondre, de nombreuses informations doivent être traitées par le système visuel en un seul passage dans une architecture parallèle ascendante (Thorpe et Imbert 1989 ; Celebrini et al.

1993 ; Thorpe et al. 1996 ; Fabre-Thorpe et al. 1997 ; Nowak et Bullier 1997; Gautrais et Thorpe 1998).

4.2 l'activité à 150 ms et autres indices de traitement

Dans cette partie, nous passons en une revue sommaire une sélection des différents indices de traitements étudiés dans la littérature pouvant être impliqués dans l'activité différentielle « animal ». Si les latences de ces indices ne correspondent pas toujours à la plage sur laquelle cette activité différentielle s'étend, ils peuvent en partager des mécanismes et des aires sous-jacentes. Nous discuterons donc des liens possibles entre les activités enregistrées dans notre tâche et les composantes des potentiels évoqués lors de différentes tâches en go/no-go, ou impliquées dans des mécanismes attentionnels et visuels.

a) la détection de cible et la P3

La détection d'une cible évoque une composante potentiel évoqué spécifique, de distribution centro-pariétale, entre 300 et 600 ms (P3 pariétale ou P3b, Sutton et al. 1965). Cette P3b est connue pour refléter un traitement requérant de l'attention portée à la cible (Banquet et al. 1981 ; Ragot et Renault 1981; Knight 1996) ; elle n'est pas affectée par les attributs physiques du stimulus, mais modulée par la difficulté de la tâche (plus ample). Des mesures d'EEG, d'IRMf et de lésions convergent pour en situer la source dans le gyrus supra-marginal (aire 40 de Brodman, Menon et al. 1997). La P3b est précédée par une P3a fronto-centrale (250-500 ms) qui semble refléter une réponse d'orientation ou d'inhibition de réponse ; elle est amplifiée lors de réponses no-go (Kiefer et al. 1998), et semble provenir de sources proches de la P3b. Les différences d'activités montrées dans notre tâche n'excluent pas la présence d'une composante fronto-centrale vers 250 ms. Cette composante ne semble cependant pas correspondre à l'activité à partir de 150 ms observée dans les aires visuelles.

b) l'inhibition de réponse et la N2

La N2 est une composante négative et fronto-centrale, qui atteint un pic entre 250 et 300 ms ; elle est visible à partir de 200 ms. Dans des tâches en go / no-go, la N2 est plus large lorsque la réponse doit être inhibée (Simson et al. 1977). Toutes les études associent cet effet à un processus d'inhibition dans un système de contrôle d'exécution d'actions (Gemba et Sasaki 1989). Par exemple, Jodo et Kayama 1992 ont montré que l'amplitude de la N2 était liée à « l'effort » nécessaire pour inhiber la réponse : lors d'une expérience dans laquelle il est demandé aux sujets d'être le plus rapide possible, rendant plus difficile l'inhibition, la N2 est clairement plus ample. La N2 a été aussi observée lorsqu'une réponse orale doit être inhibée (comptage de cibles, Pfefferbaum et al. 1985). Dans la modalité auditive, cet effet est visible à partir de 200 ms (pic à 240 ms) ; elle semble

prendre sa source dans les cortex préfrontaux inférieurs (Kiefer et al. 1998) pour les deux modalités. Comme nous l'avons vu précédemment, le début de l'activité différentielle que nous avons observée ne correspond pas aux latences auxquelles la N2 s'amplifie pour l'inhibition de réponse - sans que la présence de cette composante soit à exclure.

c) la sélection de traits (SN)

La négativité de sélection (selection negativity, SN) a été mise en évidence par Harter et Aine, en 1984 (Harter et al. 1984 ; Aine et Harter 1986) ; il s'agit d'une réponse évoquée spécifique au traitement d'un stimulus possédant un trait attendu (i.e. privilégié par l'attention). Cette réponse a été mise en évidence dans l'attention sélective à la couleur, la fréquence spatiale, l'orientation de barres, et à la forme. La SN est de polarité négative, maximale sur les sites occipitaux ; elle se caractérise comme un décalage lent et endogène qui recouvre les composantes PE reliées à un traitement passif de l'information (Ritter et al. 1984; Hillyard et Mangun 1986 ; Hillyard et Mangun 1987 ; Rugg et al. 1987 ; Wijers et al. 1989 ; Naatanen et al. 1992 ; Plendl et al. 1993 ; Mangun 1995 ; Smid et al. 1996 ; Czigler et al. 1998; Michie et al. 1999 ; Smid et al. 1999; Wang et al. 1999). Avant d'évoquer la possibilité d'identifier l'activité différentielle dans la tâche animal / non-animal à une SN « animal » ou « forme », nous allons en énumérer quelques caractéristiques.

La variabilité de la SN en latence et en amplitude est une propriété caractéristique : par exemple, la SN spécifique à la forme s'étale de 120 ms à 320 ms (Smid et al. 1996) en fonction de la tâche à accomplir et des stimuli. Dans l'étude de Smid, Jakob et Heinze, 1997, la SN est utilisée comme « index de traitement sélectif », dont il est étudié la latence et l'amplitude pour déterminer le traitement parallèle et la conjonction de la couleur et des formes globales et locales. Les latences des SN « couleur » et « forme globale », à 200 ms, sont retardées d'une cinquantaine de millisecondes (250 et 260 ms respectivement) lorsque la discriminabilité de ces traits est plus faible ; la SN « forme locale » ne semble pas soumise à ces variations, et reste de l'ordre de 280 ms. Les interactions couleur et forme globale évoquent une SN à 250 ms, retardée à 270 ou 280 ms lorsqu'un des facteurs est moins discriminable. Cette même variabilité est observée par Proverbio et al., 1998, en plus d'une asymétrie d'amplitude de la SN, en fonction de la forme globale et locale (grandes lettres composées de petites) : elle est située en occipital gauche pour la forme locale, comparé à l'occipital droit pour la configuration globale. D'autre part, la latence de la SN est plus rapide pour la forme globale (pic à 230 ms pour la configuration globale, à 300 ms pour la forme locale).

Anllo-Vento et al. 1998 montrent une SN de sélection à la couleur, qui débute à 160 ms, atteint un pic à 220-240 ms ; elle est accompagnée par une positivité (*selection positivity*) SP qui débute au même moment et pique à 210 ms, fronto-centrale. Ces deux ondes de différences (différence couleur attendue / non-attendue) se montrent plus amples lorsque les essais sont des cibles mais leur latence n'est pas affectée. La modélisation de

source de ces ondes nécessite 2 paires de dipôles symétriques (il s'agit de dipôles de différences, comme dans notre étude); la première paire rend compte de la déflexion initiale 160-250 ms, localisée par leurs auteurs dans le sulcus collatéral BA 19, ou le gyrus fusiforme postérieur pouvant correspondre à IT chez l'homme. La deuxième paire est active de 190 ms à 240 ms, et semble correspondre à une activation de l'aire 6 prémotrice, beaucoup plus ample pour les cibles que pour les non-cibles.

Cette localisation occipitale et ces propriétés semblent correspondre à l'activité différentielle que nous avons observée, dans le cas Animal. Cependant, si l'activité différentielle 'Animal vs non-animal' à 150 ms est une négativité de sélection, alors elle possède des propriétés remarquables : (1) elle n'est pas modulée en latence par l'attention, comme si les traitements spécifiques dont elle est le reflet avaient atteint leur rapidité optimale (nous le mettrons en évidence au chapitre 3), (2) elle est très précoce, comparée aux SN spécifiques à la formes observées dans les études récentes (Michie et al. 1999), même chez des sujets surentraînés (Czigler et al. 1998). Il est cependant évident que la topologie de l'activité différentielle 'Animal' ressemble à celle de la SN, notamment dans l'étude de Anllo-Vento et al. 1998 sur l'attention à la couleur[1].

d) la focalisation de l'attention sur une cible (N2pc)

La N2pc (N2 postérieure controlatérale) est une composante mise en évidence par la différence entre les potentiels évoqués par une cible ipsilatérale et une autre cible controlatérale. La cible controlatérale est plus négative sur les sites occipitaux et temporaux (cette propriété pouvant être inversée pour l'hémichamp visuel supérieur chez quelques sujets). Cet effet peut débuter à 170 ou 200 ms, et s'étend sur une cinquantaine de millisecondes ; son amplitude est rarement au delà de 3 µV (souvent de 0.5 µV dans Luck et al. 1997). Elle est visible dans l'intervalle compris entre la fin de N2 et le début de P3 ; la N2pc est à peine visible sur le flan descendant de la N2, contrairement à l'activité différentielle 'Animal' observée ici. La N2pc est plus ample pour les stimuli présentés dans la partie basse du champ visuel, plus ample aussi lorsque des distracteurs viennent enrichir le stimulus ; enfin la N2pc dure plus longtemps et est plus ample lorsque la discrimination est plus complexe (conjonction de traits par rapport à un seul). Elle est plus ample aussi lorsque les sujets doivent faire une saccade vers la cible, plutôt que de répondre manuellement.

C'est pour cet ensemble de raisons que Luck et al. 1997 associent la N2pc à la focalisation de l'attention, en analogie étroite avec les neurones de V4 reflétant la sélection d'une cible en même temps que la suppression de distracteurs pendant une tâche de recherche visuelle (Chelazzi et al. 1993 ; Tassinari et al. 1994 ; Luck et al. 1997. Ces deux effets apparaissent vers 175 ms, controlatéralement au stimulus. Les auteurs associent la N2pc avec l'activation de V4, qui inclut des neurones possédant les mêmes propriétés (V4

[1] Nous verrons au chapitre 3 que cette activité ne se réduit pas à une sélection de couleur, puisqu'elle ne se modifie ni en latence ni en topographie quand elle est évoquée par des stimuli noir et blanc 'Animal'.

et TE ne s'activent en effet pratiquement qu'à la stimulation des hémichamps controlatéraux ; V4 possède des cartes séparées des champs visuels haut et bas, TE n'en possède pas). Dans une étude sur l'attention, Wijers et al. 1997 voient les effets de l'attention retardée par l'isoluminance des stimuli lors de tâches de détection de cibles : la N2Pc localisée comme inféro-temporale est retardée de 40-50 ms pour les cibles de la tâche de recherche, alors que les non-cibles ne voient pas le même retard. D'après les auteurs, la N2pc reflète un processus automatique de *template matching*. Ces propriétés ne semblent pas correspondre à l'activité différentielle 'Animal', bien qu'une action de l'attention sur les traits visuels de la catégorie animal soit très probable.

e) la saillance des catégories visuelles connues (les P150)

les P150 rassemblent des ondes évoquées par deux ensembles de stimuli : les visages, et les mots. Ces ondes commencent à 125 ms, atteignent un pic à 150-170 ms et possèdent une large distribution centrale. La P2 « visage » ou VPP (*vertex positive potential*) apparaît à la suite de présentations schématiques de visages, ou visages distordus, modérément mélangés, ou illusoires, en présentation dégradée ou inversée (Botzel et Grusser 1989 ; Jeffreys 1989 ; Seeck et Grüsser 1992 ; Allison et al. 1994 ; Jeffreys 1996 ; Jemel et al. 1999). Ce potentiel est dans ces cas d'amplitude moins grande, et apparaît comme nécessitant des traitements supplémentaires et latéralisés pour l'encodage structurel des visages[1] (George et al. 1996). La P150 « mot » apparaît à la suite de présentation de mots, de pseudo-mots, non-mots, pseudo-fontes, et non aux autres formes d'objets (Shimoyama et al. 1992 ; Schendan et al. 1998). Shendan et al. rapportent que la P150 est indissociable dans leur expérience entre visage et mots. Ils en déduisent qu'elle reflète davantage un traitement spécifique aux catégories visuelles bien connues des sujets (en accord avec Jeffreys 1996), plutôt que spécifiques aux visages en tant que tels, en même temps qu'aux mots. Les générateurs de la P150 semblent se trouver très en avant dans la voie ventrale, bien que leur localisation précise ne soit pas un sujet de consensus (gyrus fusiforme postérieur, jonction occipito-temporale, amygdale, hippocampe ou gyrus cingulaire).

Quelle que soit la nature des similitudes entre les potentiels évoqués par les visages et les mots, le fait que les P150 reflètent des traitements automatiques spécifiques à certaines catégories visuelles peut faire penser à un mécanisme analogue aux traitements induits par la tâche « animal ». Cependant, la P150 apparaît avant le début de l'activité différentielle que nous avons mesurée, et ne possède pas la même distribution spatiale. De même, les générateurs que nous avons mis en évidence semblent moins antérieurs que les générateurs de la P150. Il est donc probable que les traitements en jeu dans l'activité différentielle 'Animal' soient très différents des traitements liés aux visages et aux mots : les stimuli « animaux » possèdent de plus nombreux traits visuels pouvant être indices de

[1] Une autre composante mise en évidence par Bentin et al. 1996 montre une spécificité aux visage ; la N170 est évoquée de la même façon par tous les traits composant un visage, mais non par des visages d'animaux, de mains ou autres objets.

la catégorie, et leur ensemble est moins homogène. Il est possible dans ce cas que les traitements nécessaires dans notre tâche soient plus élaborés que lors des tâches évoquant les P150 ; ainsi, de la même façon que l'hypothèse d'une détection simple peut être écartée dans la tâche « animal », un mécanisme très direct et spécialisé semblable à la reconnaissance des visages et des mots ne semble pas plausible. Au contraire, la catégorisation « animal » semble mettre à contribution plusieurs mécanismes de la voie ventrale analogues à la sélection de traits (reflétée par les SN formes ou couleur) ou à la focalisation de l'attention sur une cible (reflétée par les N2pc), sans requérir cependant une attention préparatoire. Dans notre tâche, la rapidité de ces traitements non spécialisés n'en est que d'autant plus étonnante.

L'activité différentielle 'Animal vs Non-Animal' observée à partir de 150 ms ne peut donc directement s'assimiler à des activités déjà étudiées en potentiels évoqués. Il est cependant probable qu'elles soit uneforme de négativité de sélection : elle en possède tous les attributs, hormis une latence étonnamment précoce pour des formes aussi variées que celles de la catégorie Animal, sélectionnées dans des images de scènes naturelles. Les expériences qui vont suivre vont nous permettre d'en explorer les caractéristiques.

VARIATIONS AUTOUR DE LA CATEGORISATION VISUELLE RAPIDE

1 présentations alternées par hémichamps

Les résultats présentés au chapitre précédent ont montré que le système visuel humain possède la capacité de traiter une scène visuelle en moins de 150 millisecondes, sans utiliser d'information contextuelle – les scènes très diverses n'étant présentées qu'une seule fois - et sans mouvement oculaire d'exploration. Cependant, les images étaient présentées centralement par rapport au point de fixation, et la prise de vue professionnelle assure au sujet de la photographie - à l'animal à détecter - une place généralement centrale. Ces saccades d'exploration n'étaient donc pas nécessaires pour que les images soient traitées par la fovéa, zone privilégiée dans laquelle la densité des photorécepteurs assure un échantillonnage précis et une large représentation corticale.

Cette première variation de l'expérience standard vise à mesurer les performances des sujets et la vitesse du traitement en vision extrafovéale. La tâche accomplie par les sujets sera la catégorisation animal / non-animal de photographies présentées aléatoirement à gauche, au centre ou à droite ; cette nouvelle tâche - plus difficile- sera alternée avec la tâche en vision centrale seule, en tant que contrôle. Les résultats seront discutés en termes d'acuité nécessaire à la reconnaissance d'objets, de la possibilité d'un recentrage des entrées latéralisées, en terme de supériorité hémisphérique de la catégorisation 'Animal', et en coût temporel du partage de l'attention sur une portion plus large du champ visuel.

1.1 expérience et résultats

a) protocole et hypothèses

stimulation et sujets

Cette variante propose aux sujets d'effectuer la catégorisation Animal / non-Animal dans deux situations distinctes : la première consiste à présenter les photographies en vision centrale seule, les stimuli centrés sur le point de fixation ; cette tâche à une position est similaire à l'expérience du chapitre précédent, la situation standard. La seconde situation consiste à présenter les photographies de manière aléatoire en trois positions : centralement, ou avec une excentricité de 3,5° à gauche ou à droite du point de fixation (voir Figure 11). En chaque position, le nombre de scènes cibles était égal au nombre de scènes distracteurs.

15 sujets (13 droitiers), d'un âge variant de 22 à 55 ans, ont été testés dans les deux situations par blocs de 100 essais. L'ordre des tâches à une et trois positions était aléatoire ; les sujets étaient informés au début de chaque bloc de la tâche qui leur était proposée. Comme dans l'expérience standard, la taille des images (verticales ou horizontales) était de 384x256 pixels, soit 6,6°x4,5° vues à 1.10 m, et l'affichage de 20 ms. Le bonnet EEG utilisé dans cette expérience différait de celui de l'expérience précédente (standard) en ce qu'il lui était ajouté 11 électrodes supplémentaires (voir Annexe).

hypothèses du protocole

• La tâche à une position étant similaire à l'expérience standard, nous attendons confirmation des résultats précédents.

• La comparaison des essais présentés au centre dans la tâche à une position et celle à trois positions permet d'estimer, en termes de temps de réaction et de vitesse du traitement visuel, le coût du partage de l'attention sur une zone du champ visuel ayant doublé de surface.

• La brève présentation des stimuli dans les hémichamps gauche et droit assure la latéralisation de l'entrée visuelle dans l'hémisphère controlatéral, et permet de comparer la performance des deux hémisphères lors de la tâche, en temps de réaction comportemental et potentiels évoqués. Les études mettant en œuvre une perception latéralisée ont cherché à mesurer l'effet de facilitation pouvant résulter des traitements de la perception et de l'action dans le même hémisphère cérébral. Lorsqu'un stimulus est présenté en situation extrafovéale et que l'action doit être produite par l'hémisphère stimulé, le temps de réaction simple est plus rapide de 4 à 6 ms par rapport à une réponse controlatérale (depuis Poffenberger 1912). Cet effet a été interprété comme le délai de transmission par le corps calleux. Dans les tâches de choix, cet effet est amplifié (de 20 à 50 ms) du fait du conflit à résoudre entre position du stimulus et position du geste moteur reflétant la décision - les tâches de choix étudiées mettant en œuvre deux boutons réponses. Cet effet n'est pas dû à un biais favorisant l'hémisphère stimulé, puisqu'un même conflit peut être occasionné de façon intrahémisphérique (Ragot et Lesevre 1986). La tâche go / no-go n'entraînant pas de tel conflit spatial, nous pouvons donc postuler qu'une différence de temps de réaction en réponse aux présentations gauches et droites ne sera due qu'au délai de transmission du corps calleux (les temps de réaction aux présentations gauches étant alors plus longs pour les droitiers, de 4 à 6 ms), ou à une supériorité hémisphérique de la tâche de catégorisation que nous pourrons mettre en évidence.

b) résultats comportementaux

La tâche à une seule position a été réussie à 95.1% ; ce fort taux de réussite est comparable à celui observé lors de l'expérience standard. Le temps de réaction médian est de 414 ms. Les présentations centrales de la tâche à trois positions aléatoires présentent des

résultats identiques : 94.9% et 417 ms (écarts entre les deux tâches non significatifs pour les présentations centrales, voir Figure 12). Ces premiers résultats mettent en évidence l'absence de coût comportemental lié au partage attentionnel sur un champ visuel agrandi.

Les résultats des présentations latérales de la tâche à trois positions sont par contre significativement différents de ceux des présentations centrales. Considérés ensemble, la réussite a été de 91.9% et la médiane des temps de réaction de 430 ms (écarts 3% et 12 ms, $p<0.0001$). Aucune différence significative entre les résultats aux présentations gauches et droites n'a pu être observé dans le groupe. La même analyse effectuée séparément pour le sous-groupe des 13 sujets droitiers montre une différence entre gauche et droite de 3 ms, qui n'est cependant pas significative. Ces données ne vont donc pas dans le sens de l'existence d'une supériorité hémisphérique ; elles ne montrent pas non plus un délai significatif reflétant une transmission callosale – à moins que celle-ci ne compense exactement celle-là, au sens d'une supériorité hémisphérique droite. Une telle compensation chez les sujets droitiers devrait se solder par un effet cumulatif chez les sujets gauchers, effet qui n'est clairement pas observé chez nos deux sujets gauchers.

c) résultats électrophysiologiques

Les résultats présentés sont ceux de la moyenne de groupe des 13 sujets droitiers. Ces résultats (obtenus avec un bonnet à 32 voies dans lesquelles les sites occipitaux sont très représentés) montrent tout d'abord que l'activité différentielle de polarité positive entre cibles et distracteurs au dessous des sites frontaux se complète par les deux activités occipitales de polarité inverse bien plus visibles ici, et compatibles avec le modèle dipolaire de l'expérience standard. Ces activités différentielles occipitales sont clairement significatives dans l'étude de groupe. Par contre, bien que tous les sujets présentent cette configuration d'activité, l'activité occipitale n'atteint un seuil de significativité exploitable que chez certain d'entr'eux. Ceci devient problématique pour tester la significativité d'effets occipitaux en tests appariés ; nous ne pourrons donc fournir pour ces sites que les résultats des potentiels évoqués moyens du groupe.

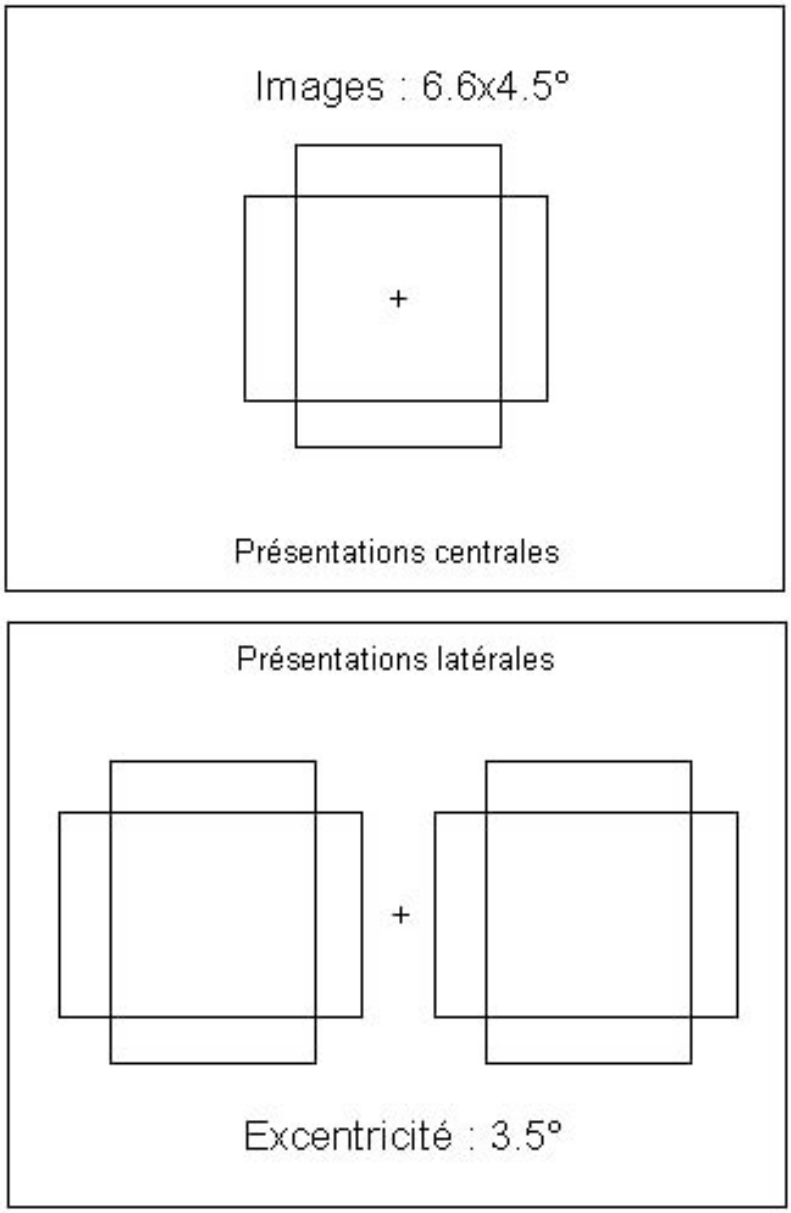


Figure 11 : Positions des stimuli sur l'écran.

Les stimuli de toutes les expériences présentées ont une taille de 384x256 pixels, de 6.6x4.5 degrés d'angle visuel (12.6x8.6 cm à 1.10 m). Dans l'expérience standard et à trois positions, les images pouvaient apparaître verticalement ou horizontalement ; dans les expériences suivantes, les stimuli étaient présentés horizontalement. Dans la tâche à trois positions, les bords latéraux des photographies étaient à 6.8° pour les stimuli horizontaux, et 5.8° pour les stimuli verticaux.

A

sujets	Pourcentages de réussite					Temps de réaction médian (ms)				
	1 Position	3 Positions				1 Position	3 Positions			
	Centre	Centre	Gauche	Droite	Latérales	Centre	Centre	Gauche	Droite	Latérales
12	97.0	98.7	95.4	93.4	94.4	425	434	448	444	445
13	95.1	95.8	92.7	93.2	92.9	400	397	412	401	408
14	94.1	93.5	87.9	87.5	87.7	446	438	466	462	463
·2 (G)	95.2	93.9	90.8	91.8	91.3	362	385	391	397	392
23	96.3	95.8	94.2	94.7	94.4	388	392	396	393	395
26	94.6	93.5	91.2	89.4	90.3	467	465	480	479	479
27	95.2	95.8	92.2	90.2	91.2	410	426	448	438	442
28	96.8	97.1	94.1	94.3	94.2	390	394	399	396	398
30	97.1	97.0	93.1	94.0	93.5	366	369	405	394	400
35	92.6	90.1	89.8	87.7	88.7	402	415	448	430	437
37	91.8	91.4	90.9	91.3	91.1	433	433	433	442	436
38	92.0	91.6	87.5	87.3	87.4	482	477	480	497	490
39	96.8	96.4	93.7	93.9	93.8	411	410	414	409	412
50 (G)	97.0	97.7	94.0	95.2	94.6	422	429	449	437	441
57	94.9	95.7	93.5	93.6	93.6	399	398	408	409	408
moyenne	**95.1**	**94.9**	**92.1**	**91.8**	**91.9**	**414**	**417**	**432**	**428**	**430**

tests Student appariés (ddl=14)	écart	t	p
% réussite au centre, 1P vs 3P	0.167	0.628	0.5401
% réussite 3P, **centre vs latéral**	2.993	8.288	<,0001
% réussite 3P, gauche vs droite	0.233	0.775	0.4512
médianes au centre, 1P vs 3P	-3.767	-1.722	0.1071
médianes 3P, **centre vs latéral**	-12.433	-5.682	<,0001
médianes 3P, gauche vs droite	3.233	1.404	0.1821
droitiers (ddl=12), % réussite 3P, gauche vs droite	0.438	1.418	0.1816
droitiers (ddl=12), médianes 3P, gauche vs droite	3.231	1.302	0.2173

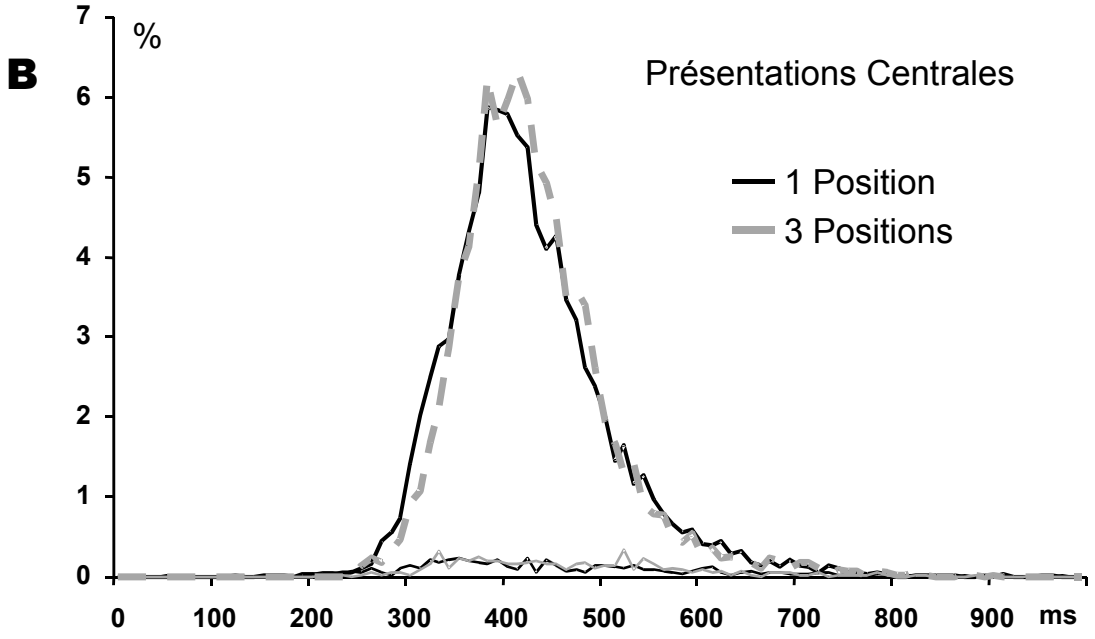


Figure 12 : Résultats de l'expérience 'présentations alternées'

A. Résultats comportementaux des tâches à une position (1P) et trois positions (3P). Pour chaque sujet sont indiqués les pourcentages de réussite et les temps de réaction médians dans les différents cas. En dessous ont été testées les différences entre les positions centrales dans les deux tâches, les différences entre la position centrale 3P et la moyenne des positions latérales, et enfin les différences entre les positions latérales gauche et droite. Les tests ont été reproduits à part pour les 13 sujets droitiers. (G dans la colonne 'sujets' indique les sujets gauchers).

B. Histogrammes des temps de réaction lors des présentations centrales des deux tâches. Les lignes épaisses et fines différencient les cibles réussis des distracteurs ratés (faux positifs). Dans la tâche à 1 position, le nombre de cibles réussis est significativement plus grand que le nombre de faux positifs à partir de l'intervalle 270-280 ms. Dans la tâche à trois positions, ce test est significatif sur l'intervalle 280-290 ms ($p<0.005$).

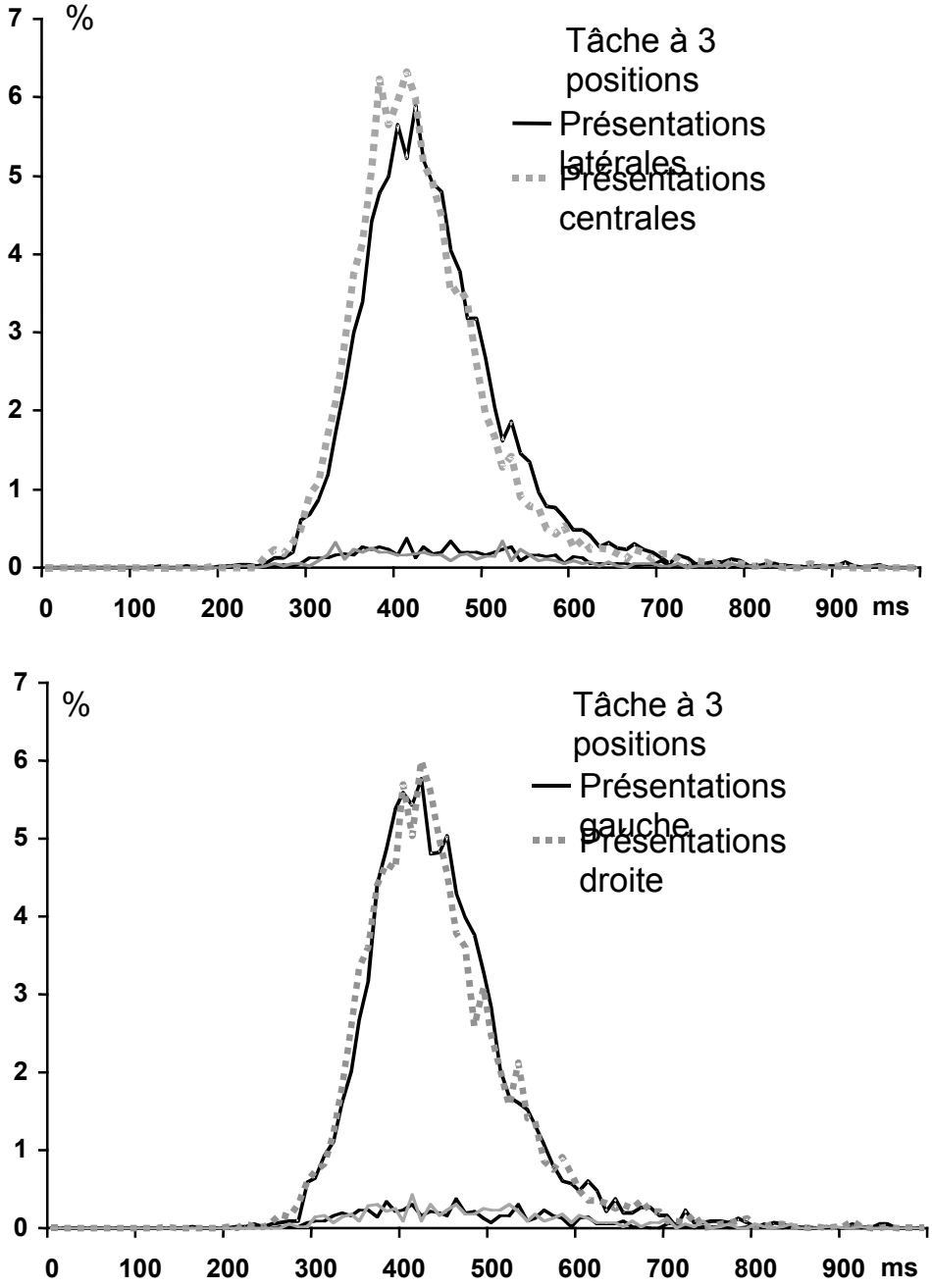


Figure 13 : Histogrammes des temps de réaction lors de la tâche à 3 positions.

Les lignes épaisses et fines différencient les cibles réussies des distracteurs ratés (faux positifs). Le nombre d'essais réussis est significativement plus grand que le nombre d'essais ratés à partir de l'intervalle 280-290 ms en présentation centrale et pour la moyenne des présentations latérales. A Gauche, ce test est significatif sur l'intervalle 290-300 ms et à droite sur 280-290 ms($p<0.005$).

effet du partage de l'attention sur un champ visuel élargi

La comparaison des potentiels évoqués dans les conditions centrales montre clairement l'effet du partage de l'attention : que ce soit sur les cibles ou les distracteurs, la négativité à 170 ms est plus ample de 1.5 µV dans la tâche à trois positions ; cet effet est clairement visible à partir de 140 ms (Figure 14).

• La P1 [75-130 ms] est connue pour être la première composante des potentiels évoqués sensible à l'attention visuo-spatiale (pour une revue, Hillyard et al. 1998) dans un contexte lieu privilégié / non-privilégié. Cet effet a été interprété par un effet de suppression des entrées non-attendues (Posner et Dehaene 1994 ; Anllo Vento 1995 ; Luck et Hillyard 1995). Le résultat présenté ici montre que la présentation centrale est autant « attendue » dans la condition à une position qu'à trois positions aléatoires. Ce résultat va dans le sens d'une compensation du partage attentionnel par une plus grande allocation globale.

• Cette plus grande allocation est démontrée par l'effet d'amplification de la N1 [150-190 ms]. Malgré la complexité de cette composante (frontale 140 ms, pariétale 15-160 ms, occipitale 170-190 ms), les études récentes s'accordent à l'interpréter en termes de gain et de facilitation de l'entrée visuelle - à la différence de la P1 (Luck et al. 1994 ; Heinze et Mangun 1995; Hillyard et AnlloVento 1998 ; Hillyard et al. 1998). Cette onde est donc plus à même de refléter la quantité d'attention allouée pour traiter l'entrée visuelle.

activités différentielles en présentations centrales

Les différences d'activités entre cibles et distracteurs de la tâche dans les deux conditions se superposent de la même manière sur les sites frontaux et sur les sites occipitaux (Figure 15). Les séparations que l'on peut remarquer entre 100 et 170 ms sur les électrodes frontales entre les deux courbes ne sont pas statistiquement significatives (test de Student apparié entre les deux conditions pour les 13 sujets). Ces courbes répliquent le principal résultat de l'expérience standard.

Ce résultat est en lui-même surprenant, si l'on considère l'effet d'amplification observé dans la tâche à trois positions sur les potentiels évoqués cibles et distracteurs en présentation centrale. Nous voyons que même si le recrutement attentionnel est accru par la tâche à trois positions, ce recrutement ne modifie pas le traitement requis pour décider de la présence d'un animal – les latences et amplitudes des différences sont similaires dans les deux conditions.

présentations latéralisées

A la latéralisation des entrées visuelles correspondent des potentiels évoqués controlatéraux occipitaux amples, ainsi que des activités différentielles latéralisées. Lors

des présentations gauches et droites, les différences cibles-distracteurs possèdent une topographie en miroir (voir les données et les modèles dipolaires correspondants, Figures 17-19) : sur les sites frontaux, l'activité différentielle évoquée par les présentations gauches possède un pôle négatif sur les sites occipitaux droits et un pôle positif sur les sites frontaux gauches ; la configuration est inversée pour les activités différentielles évoquées par les présentations droites. La Figure 16A montre que les activités différentielles frontales sont très comparables entre présentations gauche, droite et centrale lorsqu'elles sont mesurées sur les sites frontaux en regard des maxima de ces activités. Les latences des différences cibles-distracteurs ne sont pas significativement différentes sur les sites frontaux entre droite et gauche ; elles approchent la significativité entre présentations centrale et latérales (Figure 16B).

Les reflets occipitaux de ces différences cibles–distracteurs sont clairement significatifs pour le groupe. Si leur amplitude est plus grande sur les sites controlatéraux, les latences semblent cependant montrer une asymétrie en faveur des sites occipitaux gauches (présentations gauches : O1' 170 ms, O2' 183 ms ; présentations droites, O1' 145 ms, O2' 165 ms, voir Figures 17 et 18). Ces résultats de groupe semblent aller dans le sens d'un avantage de l'hémisphère gauche sur l'hémisphère droit. La latence plus courte aux présentations droites exprime le même avantage. Il est remarquable que la réponse aux présentions gauches soient plus précoces sur les sites ipsilatéraux aux présentations, que sur les sites controlatéraux. Cependant, ces résultats de groupe ne peuvent pas être confirmés par des tests statistiques appariés, les activités différentielles cibles-distracteurs occipitales individuelles ne faisant qu'approcher la significativité chez un trop grand nombre de sujets. D'autre part, ces différences de latéralisation sont probablement liées aux activités évoquées entre 110 et 140 ms, qui voient elles aussi une asymétrie (double pic de N1 controlatérale pour les présentations gauches, dont le second pic marque une différence cibles-distracteurs approchant la significativité). Nous ne pouvons donc pas conclure à une supériorité hémisphérique claire. Cette différence Animal vs non-Animal précoce, entre 110 et 140 ms, sera étudiée plus profondément dans les expériences suivantes.

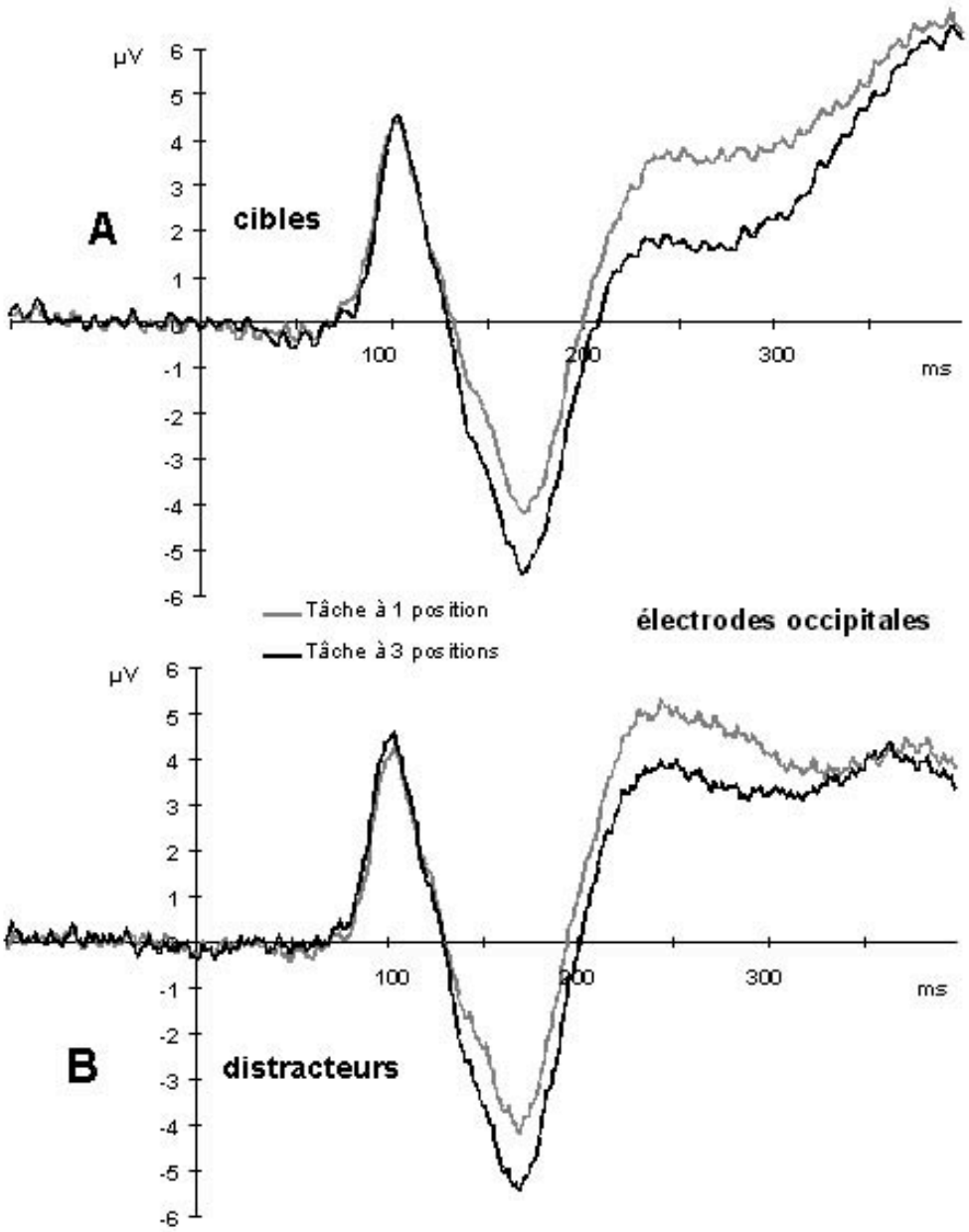


Figure 14 : Potentiels évoqués par les cibles et distracteurs des tâches à une et trois positions.

L'effet du partage de l'attention sur un champ visuel triple (tâche à trois positions) est visible à partir de 140 ms sur les électrodes occipitales, de manière identique sur les potentiels évoqués cibles (**A**) et distracteurs (**B**). Il n'influe pas sur la latence de la N1, et augmente significativement son amplitude ; cet effet est décrit dans les tâches requérant une plus grande attention.

Présentations Centrales

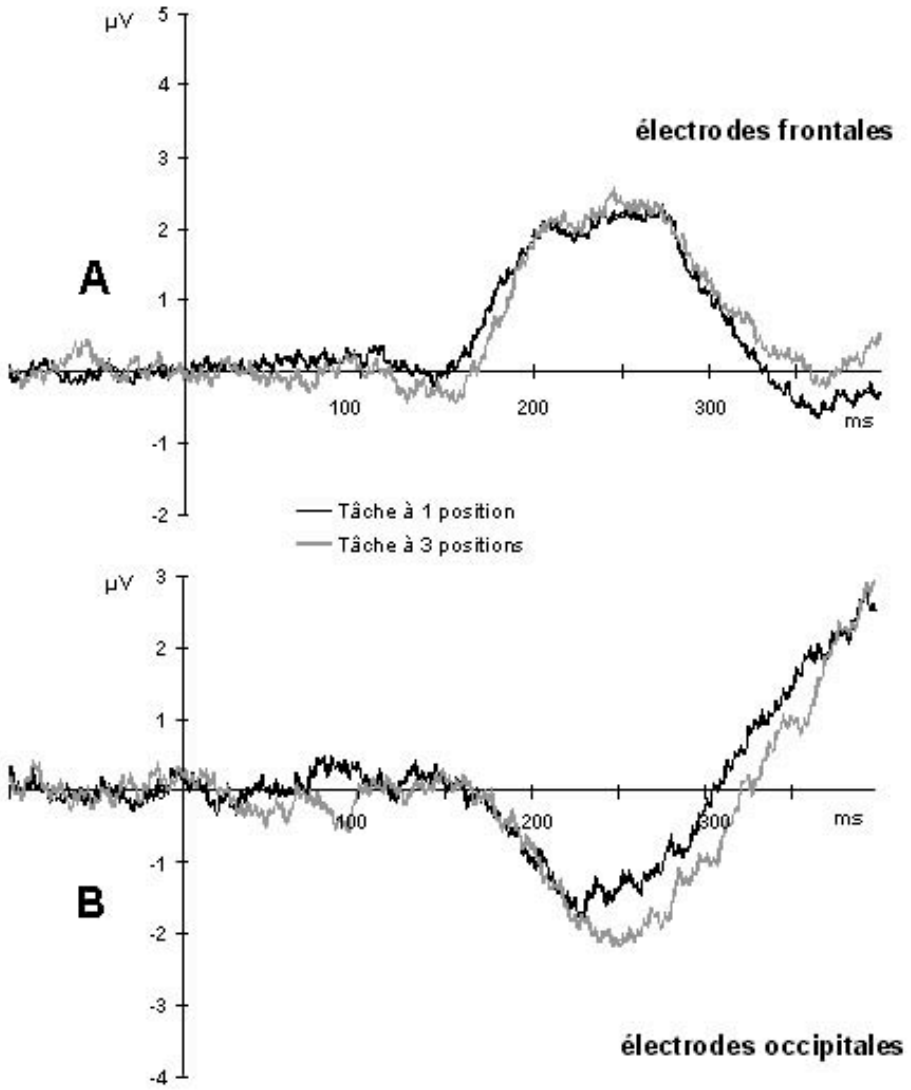
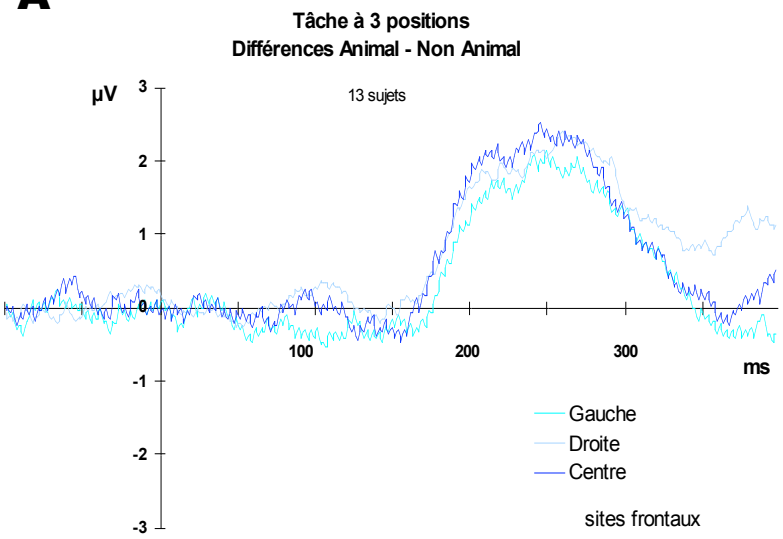


Figure 15 : Activités différentielles dans les tâches à une et trois positions.

La figure précédente a montré un plus grand recrutement attentionnel sur les cibles et les distracteurs de la tâche. On peut observer ici que ce recrutement attentionnel ne modifie ni en amplitude ni en latence l'activité différentielle initiale ; entre 150 et 230 ms, les deux courbes sont remarquablement superposées, sur les électrodes frontales (moyenne des 7 électrodes frontales, **A**) comme sur les électrodes occipitales (moyenne, **B**). Après 230 ms, un effet attentionnel est visible sur les électrodes occipitales, mais ne modifie pas la topographie de l'activité (cartographies).

A



B

	1 Position	3 Positions			
sujets	Centre	Centre	Gauche	Droite	Latérales
12	154	167	178	187	183
13	180	188	188	178	183
14	171	183	184	180	182
23	166	150	185	143	164
26	157	172	ns	ns	ns
27	ns	ns	ns	ns	ns
28	196	214	207	210	209
30	204	157	173	160	167
35	166	177	ns	ns	ns
37	171	200	212	186	199
38	188	175	190	211	201
39	196	176	177	191	184
57	166	178	186	186	186
Groupe	176	175	182	184	183
moyenne	176	178	188	183	186

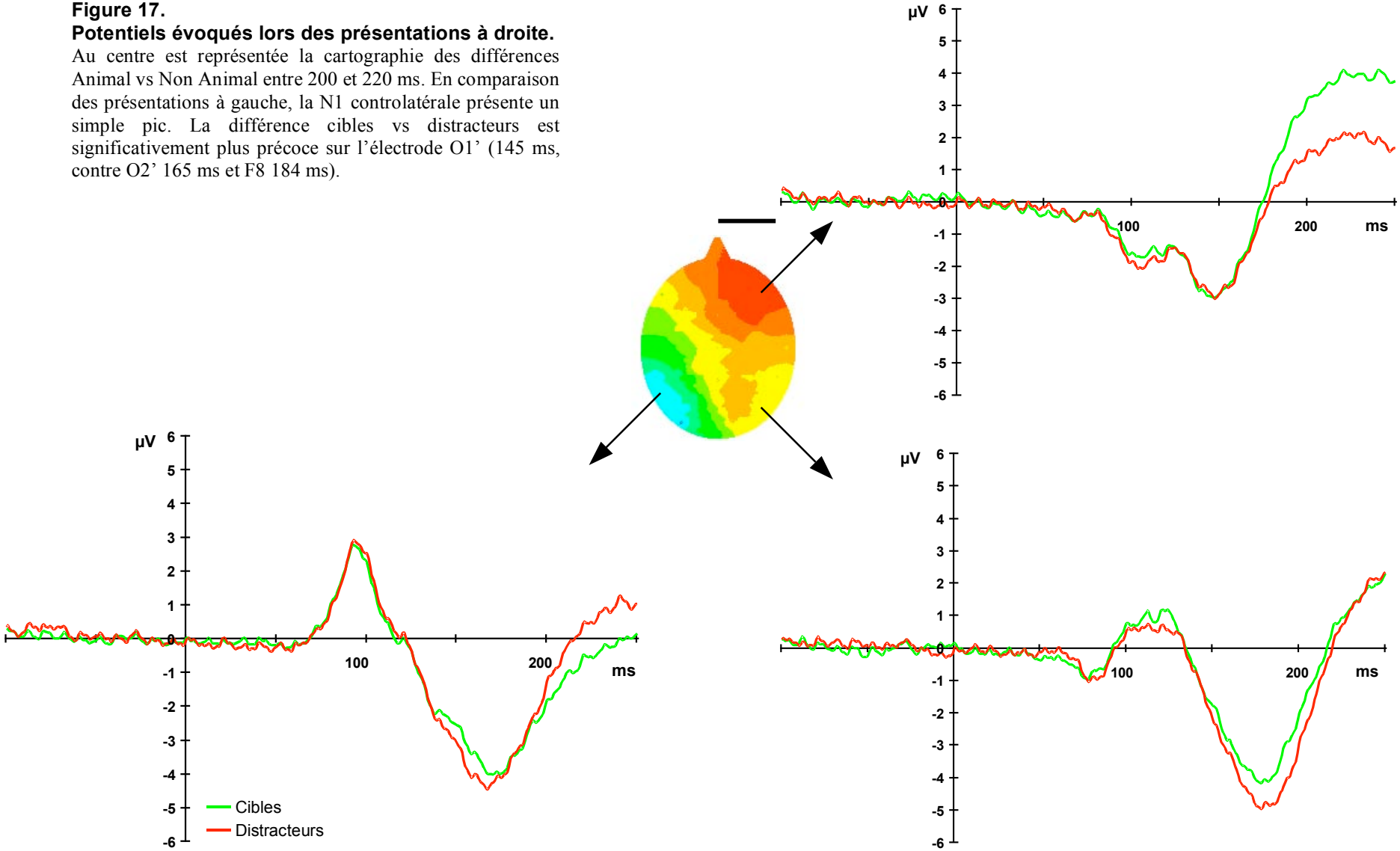
	ddl	écart	t	p
Centre, 1P vs 3P	11	-1.8	-0.3	0.77
3P, Centre vs Lat.	9	-7.0	-2.2	0.06
3P, Gauche vs Droite	9	4.8	0.8	0.44

Figure 16 : Potentiels évoqués lors des présentations latérales

A. Différences Animal – Non Animal lors des présentations latérales. Les activités différentielles représentées sont les moyennes des sites frontaux sur lesquelles ces activités sont maximales dans le groupe des 13 sujets droitiers (Centre : FP1, FP2, F3, F4, F7, F8, FZ ; Gauche : FP1, F3, F7, FZ ; Droite : FP2, F4, F8, FZ).

B. Résultats électrophysiologiques individuels et significativité de groupe. Les colonnes indiquent pour chaque sujet, en millisecondes, les instants à partir desquels la différence Animal – Non Animal devient significative sur les électrodes frontales (Centre : FZ ; Gauche : F7 ou FP1 ; Droite : F8 ou FP2). La ligne 'Groupe' indique les résultats provenant des potentiels évoqués moyens pour le groupe des 13 sujets. La ligne 'Moyenne' indique la moyenne des 13 sujets. En dessous sont indiqués les résultats des tests de Student appariés pour les 13 sujets entre les présentations centrales dans les tâches à une et trois positions, entre la présentation centrale et la moyenne des présentations latérales pour la tâche à trois positions, et enfin entre les présentations gauches et droites. Aucune différence n'apparaît significative ; seules les présentations latérales ont tendance à être plus lentes que la présentation centrale.

Figure 17.
Potentiels évoqués lors des présentations à droite.
Au centre est représentée la cartographie des différences Animal vs Non Animal entre 200 et 220 ms. En comparaison des présentations à gauche, la N1 controlatérale présente un simple pic. La différence cibles vs distracteurs est significativement plus précoce sur l'électrode O1' (145 ms, contre O2' 165 ms et F8 184 ms).



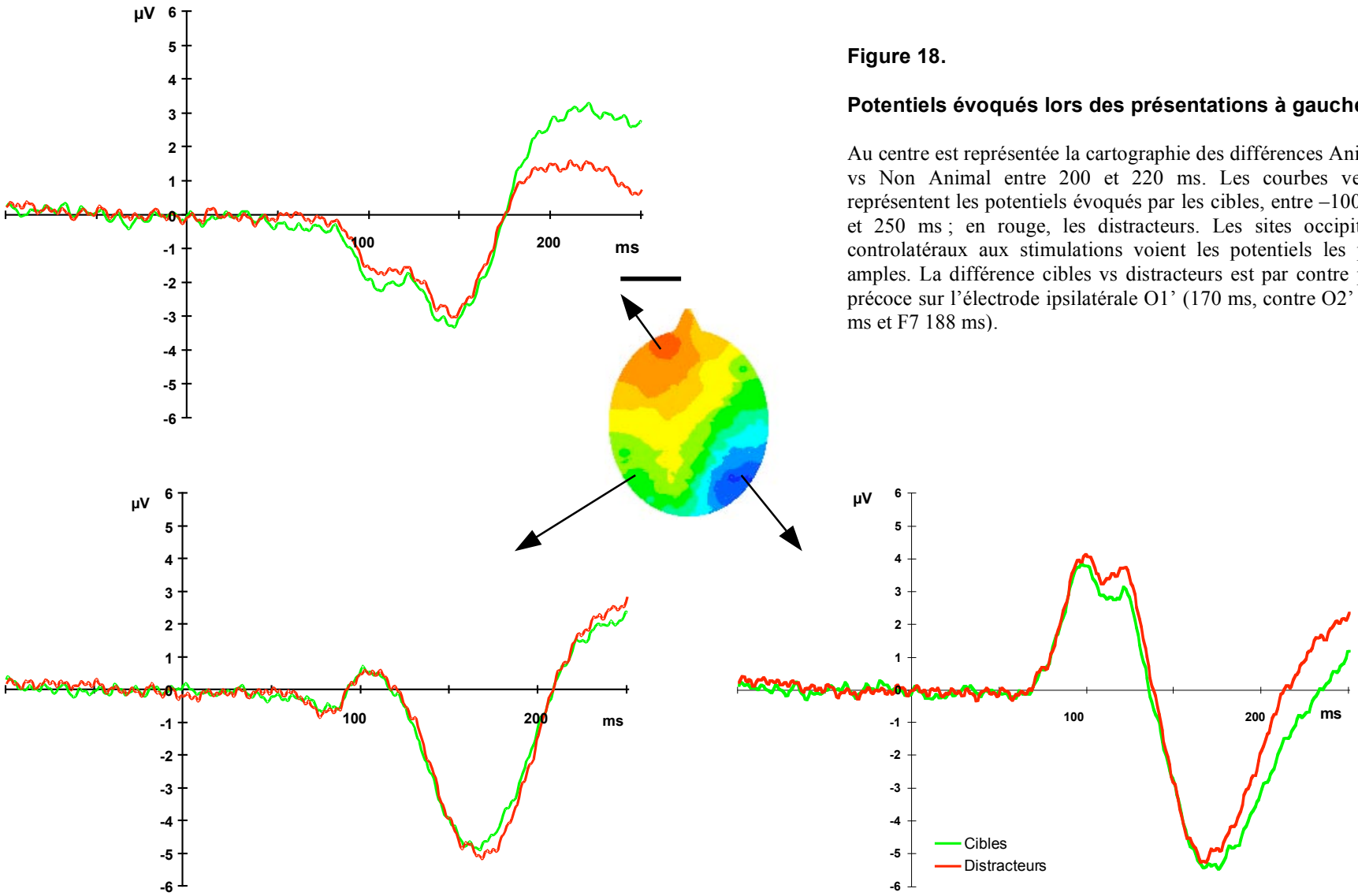


Figure 18.

Potentiels évoqués lors des présentations à gauche.

Au centre est représentée la cartographie des différences Animal vs Non Animal entre 200 et 220 ms. Les courbes vertes représentent les potentiels évoqués par les cibles, entre –100 ms et 250 ms ; en rouge, les distracteurs. Les sites occipitaux controlatéraux aux stimulations voient les potentiels les plus amples. La différence cibles vs distracteurs est par contre plus précoce sur l'électrode ipsilatérale O1' (170 ms, contre O2' 183 ms et F7 188 ms).

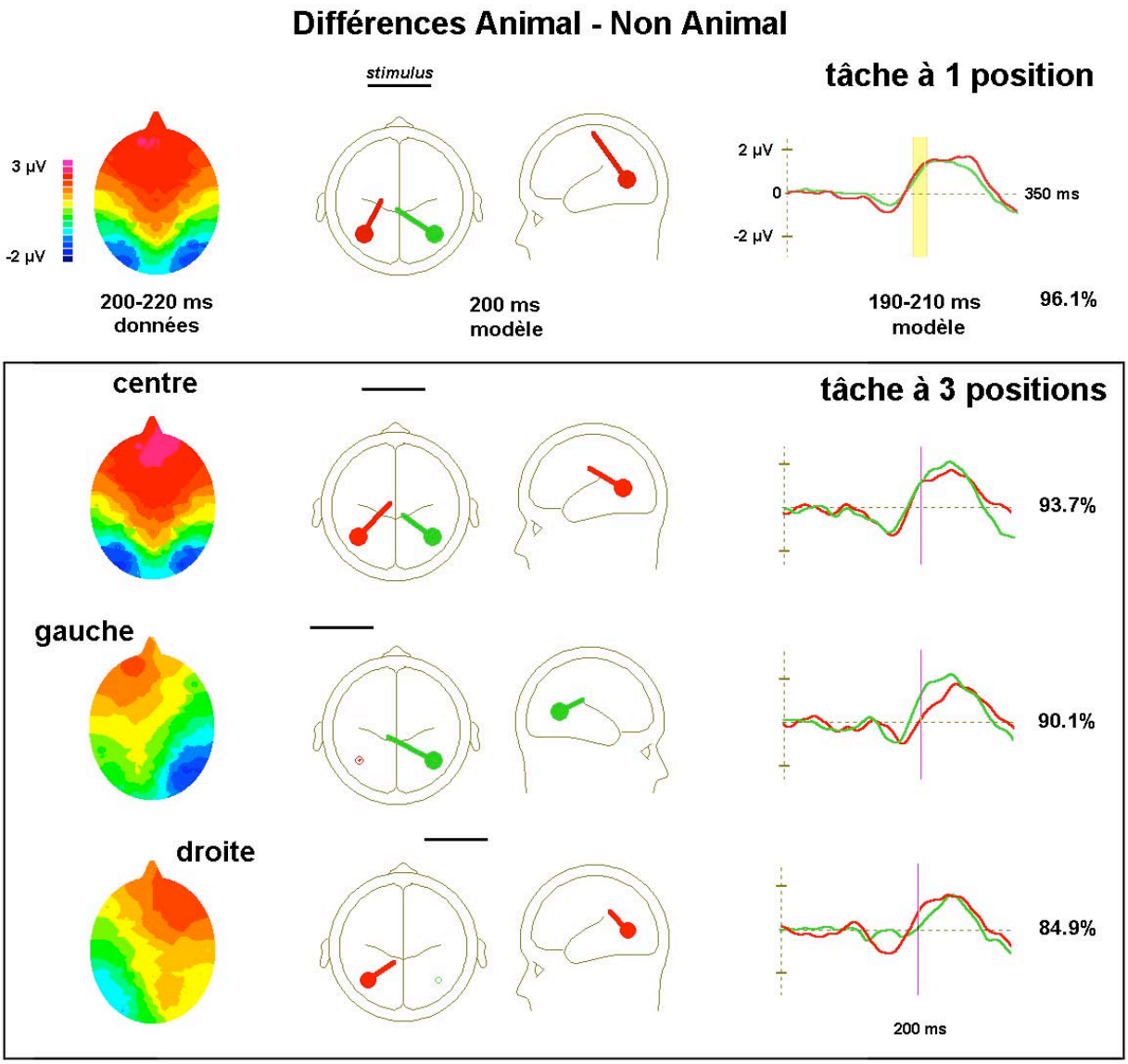


Figure 19 : Modèles des différences Animal - Non Animal de l'expérience 'présentations alternées'

Les modélisations ont été effectuées sur [190-210ms], intervalle commun à toutes les conditions expérimentales. L'activité de chaque hémisphère a été réduite à 1 source, hypothèse validée dans le cas de l'expérience standard. Dans cette expérience, les modèles des présentations centrales 1P et 3P ont été trouvés équivalents au modèle standard (Figure 9). Ce modèle ajusté en orientation suffit à expliquer correctement les présentations latéralisées. Les différences cibles - distracteurs sont principalement dues à une activation différenciée de l'hémisphère controlatéral, droit ou gauche.

d) modèles dipolaires

choix, méthode et hypothèses

La recherche des sources dipolaires vise ici à vérifier l'origine visuelle des activités différentielles et à spécifier l'effet de la latéralisation des entrées sur ces sources. Comme souligné lors de l'expérience précédente, nous ne chercherons pas ici comme dans les expériences suivantes à préciser les multiples sources pouvant expliquer l'ensemble du phénomène dans sa dynamique ni à en inférer les différentes aires impliquées. Afin de conserver des hypothèses simples, nous ne modéliserons que les *premières* différences observables entre les potentiels évoqués par les cibles et distracteurs.

Lors de l'expérience standard, les sources de l'activité différentielle ont été trouvées dans les cortex visuels extrastriés, symétriquement pour les deux hémisphères. Les variations d'amplitudes similaires pour les deux sources nous ont permis de conclure à l'égale implication des cortex visuels dans la tâche. Nous observerons ici si cette propriété de symétrie est conservée pour les présentations par hémichamps, et si l'amplitude et la latence des sources permet de conclure à une spécialisation hémisphérique, ou du moins à un effet de supériorité. L'accroissement du nombre de sites d'enregistrement nous permet d'espérer une meilleure résolution. La méthode employée s'appuie donc sur le modèle effectué à partir de l'expérience standard. Nous vérifierons qu'il s'applique correctement aux données de la tâche à une position, et nous l'ajusterons à ces données.

Chercher les sources des activités différentielles à partir des données des présentations latéralisées est techniquement une entreprise délicate. L'hypothèse de symétrie entre hémisphères est un facteur majeur de stabilité et de fiabilité des solutions ; notre expérience sur les présentations latéralisées nous a amené à abandonner toutes les solutions trouvées sans hypothèse de symétrie, ces solutions étant très instables, trop diverses, et à notre point de vue incohérentes. Il se peut que nous soyons passé à côté d'une vérité surprenante par simple incrédulité. Quoi qu'il en soit, la stratégie retenue a été de mettre en œuvre les modèles des présentations centrales en ajustant les orientations des dipôles, et de retenir les solutions expliquant plus de 80% de variance et dont le décours temporel semblait cohérent (amplitudes inférieures à 3 µV, décours similaires aux dipôles des présentations centrales).

En conséquence, les hypothèses de la modélisation sont les suivantes :

• Une seule source par hémisphère. Le comportement principal d'un hémisphère peut être résumé par un seul dipôle (modélisation du front ascendant des activités différentielles, entre 190 et 210 ms, intervalle utilisable pour toutes les présentations).

• Symétrie de position. L'activité d'une structure dans un hémisphère ne rend possible que l'activation d'une structure voisine dans l'autre hémisphère. Cette contrainte

n'empêche pas le traitement d'être latéralisé ; mais *s'il existe* une activité dans l'autre hémisphère, *alors* elle sera sensiblement au même endroit. Cette hypothèse est donc aussi valable pour les présentations non centrales.

• les sources des présentations non centrales doivent pouvoir être comparées à celles des présentations centrales.

résultats

Comme le montre la Figure 19, tous les modèles sont cohérents avec celui mis en œuvre pour l'expérience standard (cf. Figure 9). Les données des présentations centrales amènent aux modèles les plus explicatifs. Ces bons résultats confirment que la dynamique des activités différentielles peut être simplifiée par une source dans chaque hémisphère : ils valident l'approche choisie pour les présentations latéralisées. Les modèles des présentations latéralisées expliquent correctement les données, bien que moins bons que les modèles précédents. Le choix d'un seul dipôle par hémisphère montre ici sa limite : un seul dipôle se trouvant actif sur 190-210 ms, l'explication (même suffisante) est nécessairement moins bonne que lorsque deux dipôles rendent compte des données (variance 85-90%).

Présentations centrales :

• Sur la plage 150-200 ms, les modèles des présentations centrales ne montrent pas de différence majeure entre la tâche à une position et à trois positions, notamment sur leurs latences. Ce résultat est identique à celui des potentiels de surface.

• Les deux sources des modèles ont une amplitude égale sur la plage d'ajustement. Elles confirment l'égale implication des hémisphères dans le traitement visuel lié aux animaux lors de présentations centrées sur la fovéa, même en situation d'attention partagée.

Présentations latéralisées :

• la situation « en miroir » des présentations gauche et droite ne laisse place à aucun indice de supériorité hémisphérique. Miroir du lieu de l'activité : entre 150 et 210 ms, la source principale est controlatérale. Miroir en décours temporel : ces sources controlatérales ont la même latence.

• la source controlatérale est la plus active entre 150 et 210 ms : l'hémisphère stimulé reflète en premier la différence Animal – non Animal ; comme suggéré par les modèles, l'hémisphère ipsilatéral pourrait refléter cette même différence environ 10 ms plus tard. La différence observée dans les potentiels évoqués occipitaux entre présentations gauche et droite n'est pas expliquée par ce modèle [190-210 ms].

1.2 discussions et conséquences

résumons

• La catégorisation Animal / non-Animal sur des photographies présentées pendant 20 ms en dehors de la fovéa sur les hémichamps gauche ou droit, dans une situation où le sujet doit partager son attention sur un champ d'environ 25°, est réalisée facilement, avec une faible altération de performance en termes de temps de réaction et de taux de réussite par rapport à la présentation centrale.

• Le partage de l'attention sur un champ agrandi n'affecte ni les performances, ni la latence, ni l'amplitude des activités différentielles.

• Aucune différence dans l'implication des hémisphères gauche et droit n'a été remarquée en taux de réussite et temps de réaction. Les latences ou amplitudes des différences cibles–distracteurs (potentiels évoqués ou sources supposées) n'expriment pas de manière claire un avantage hémisphérique à partir de 150 ms. Lors d'une présentation par hémichamp, l'activité différentielle principale prend son origine dans l'hémisphère controlatéral.

a) pas de supériorité hémisphérique « animal »

Nous avons vu en première partie que certains traitements de la perception visuelle se trouvent être effectués plus rapidement par un hémisphère que par un autre : notamment le traitement analytique et détaillé (détection de traits, facilitation), l'encodage de relations spatiales, l'accès aux vues prototypiques pour l'hémisphère gauche ; relations plus globales (entre parties d'un objet, contexte), jugements des distances, perspectives et postures, reconnaissances de vues atypiques, soit un ensemble de capacités nécessitant de traiter des informations partielles et incomplètes, pour l'hémisphère droit.

Du point de vue de la reconnaissance des objets, quelques expériences concluent à une supériorité hémisphérique gauche (McKeever et Jackson 1979 ; Vitkovitch et Underwood 1992). Cependant d'après Biederman et Cooper 1991, cet effet peut provenir de plusieurs supériorités connexes : demander au sujet de nommer l'objet reconnu amène clairement un avantage à l'hémisphère gauche ; le même avantage peut provenir de l'utilisation de dessins d'objets requérant une discrimination des hautes fréquences spatiales, ou d'une tâche de classification au niveau de base pouvant être effectuée par des contrastes de catégories visuelles. L'expérience menée par Biederman tente de s'affranchir de ces biais : le temps de réaction est donné par le nom de l'objet énoncé par le sujet, mais la mesure cherche à déterminer si l'effet d'un amorçage préalable est plus important sur un hémisphère que sur l'autre. Les auteurs utilisent des stimuli symétriques (en miroir) pour éviter des biais dus à une interférence motrice (une tasse dont l'anse est à droite) ou tout biais d'orientation des vues latérales, balancés entre les sujets ; les stimuli incluaient quelques animaux. Les résultats ne montrent sur l'effet d'amorce aucune différence entre

hémisphères, sur 52 sujets droitiers, l'effet d'amorce étant par lui-même très significatif. De même, les expériences de Durnford et Kimura 1971; Levine et Banich 1982 , et Sergent 1987 concluent à l'absence de supériorité hémisphérique.

Nos résultats à partir de photographies d'animaux vont dans le sens d'une absence de supériorité hémisphérique pour cette catégorie. Les modélisations des activités différentielles entre les scènes avec animaux et scènes sans animaux dans chacune des expériences montre une égale implication des hémisphères gauche et droit dans la tâche.

Une confirmation beaucoup plus claire vient des résultats de l'expérience dans laquelle les stimuli sont présentés de façon alternée à gauche, au centre et à droite du point de fixation : aucune différence significative n'est apparue en termes de temps de réaction, de taux d'erreur, ou de latence des activités différentielles résultant des présentations latérales.

Deux conséquences peuvent être tirées de ces résultats :

• Le codage des représentations « animal » semble distribué de façon équivalente dans les deux hémisphères, rendant moins probable la modularité de la représentation d'objets.

• Il est probable que les traitements pour lesquels un hémisphère se trouve supérieur à l'autre ne soient pas impliqués dans notre tâche de catégorisation visuelle rapide. Comme nous l'avons vu au premier chapitre, il peut s'agir de l'encodage des configurations spatiales pour la reconnaissance (modèles de Kosslyn, de Biederman), de l'accès privilégié aux vues prototypiques comme préalable à des traitements spécifiques à la classe (modèle de Poggio et Edelman), ou de manipulation « géométrique » de l'entrée visuelle pour accéder aux représentations mémorisées (modèle de Ullman).

b) peu de mobilisation attentionnelle

Nos résultats montrent que le partage attentionnel sur un champ double de celui de la tâche à présentation centrale seule n'induit pas de baisse de performance. La psychologie de l'attention (notamment Eriksen et Yeh 1985) souligne que les ressources attentionnelles peuvent être allouées sur une petite partie du champ visuel avec une très haute résolution perceptive, ou être allouées sur un champ plus large au détriment de la résolution : certains auteurs attribuent à cette fenêtre perceptive un échantillonnage constant d'environ 30 « pixels » de diamètre (Olshausen et al. 1993). Nous aurions donc pu nous attendre à une baisse de performance pour la présentation centrale dans la tâche à trois positions, si la tâche requérait une forte résolution perceptive. Le fait qu'une résolution relativement faible permet d'accomplir la tâche sans perte dramatique de performance est montré par la réussite des sujets lors des présentations latéralisées, où pour lesquelles la résolution est limitée par la baisse d'acuité en dehors de la fovéa. Les implications de ce constat seront discutées au point suivant.

L'absence de baisse de performance laisse deux alternatives : l'agrandissement de la fenêtre attentionnelle a été compensé par un accroissement des ressources allouées à la tâche ; ou bien la tâche ne requiert pas d'attention visuelle particulière – elle peut être réalisée de façon pré-attentive. L'augmentation d'amplitude des composantes N1 (constatée pour les potentiels évoqués par les cibles et les distracteurs de la tâche à présentations alternées, par rapport à la présentation centrale seule) va dans le sens d'une compensation. Il est remarquable cependant que cette compensation ait pu être complète, c'est-à-dire que les sujets aient pu atteindre une performance comparable. Cette disponibilité des ressources attentionnelles souligne donc le fait que la tâche en présentation centrale puisse être réalisée sans nécessiter une grande mobilisation de ces ressources.

Un effet fondamental de l'allocation d'attention sur un lieu du champ visuel est d'accroître la quantité d'information extraite de ce lieu (Downing 1988 ; Lavie et Tsal 1994). En conséquence, les traitements qui sous-tendent la catégorisation visuelle rapide des scènes naturelles ne s'effectuent pas sur une grande quantité d'information extraite du stimulus. Ces informations peuvent être de forme, de couleur ou de texture, de leur conjonction et de leur agencement dans l'espace.

Qu'apportent les différentes théories de l'attention sur le traitement de ces informations ? Selon Treisman et Gelade 1980, l'attention est nécessaire pour intégrer dans un même traitement des informations encodées séparément lors de l'entrée visuelle. Ce point de vue est corroboré par des expériences de détections, montrant qu'en dehors du foyer attentionnel, une intégration couleur-forme est difficile – dans les cas ambigus. Le peu d'attention requise dans notre tâche implique alors dans ce cadre que le traitement ne repose que sur certains de ces traits de façon relativement distincte. Cette conséquence est contraire à la forte intégration nécessaire à la reconnaissance postulée dans certains modèles (notamment ceux de Biederman ou Kosslyn). La théorie alternative de Desimone et Duncan 1995 (biaised competition model) conçoit l'attention comme un mécanisme neuronal qui favorise le traitement des informations pertinentes pour la tâche. Ce point de vue apporte un autre éclairage aux résultats présentés ici : le peu d'attention pour réaliser efficacement la tâche souligne le peu d'informations concurrentes à celles nécessaires à la reconnaissance, donc l'aspect automatique des traitements effectués. Autrement dit pour le système visuel, cette reconnaissance semble aussi naturelle que le sont les scènes. La facilité avec laquelle le système visuel effectue cette catégorisation de scènes est en accord avec la proposition de Luck et al. 1997, pour qui l'attention est un mécanisme rendu nécessaire par l'ambiguïté de l'encodage des informations visuelles, uniquement lorsqu'une caractéristique pertinente pour la tâche est partagée par plusieurs stimuli dans un même champ récepteur (*ambiguity resolution theory*). Dans notre tâche, la présence de deux traits caractéristiques aux animaux renforce au contraire la probabilité de la cible, rendant la résolution d'ambiguïté superflue.

Dans tous les cas, les résultats indiquent que cette catégorisation rapide de scènes naturelles peut être accomplie avec des ressources attentionnelles largement distribuées, de résolution perceptive relativement faible. Ce résultat va à l'encontre d'un point de vue largement rencontré dans la modélisation de la vision, qui s'appuie sur les propriétés de l'attention pour résoudre les problèmes difficiles de la reconnaissance visuelle. Ce point de vue va dans le sens des expériences récentes de Braun et Julesz 1998 cherchant à estimer le coût attentionnel de tâches de détection et de discrimination de traits simples. Bien que ce fait demeure controversé, ces auteurs montrent que les tâches de discrimination de couleurs et de formes peuvent s'effectuer en dehors du foyer attentionnel.

c) la reconnaissance ne nécessite pas une forte acuité visuelle

Lors de tâches de reconnaissance d'objets, une saccade oculaire joue le rôle d'amener la zone fovéale de la rétine vers une région particulière d'une scène. La fovéa est caractérisée par une forte densité de photorécepteurs permettant une grande acuité, qui semble requise pour l'analyse des objets : les expériences de Nelson et Loftus 1980 ont montré que la détection d'un changement de forme est réduite pratiquement au niveau de la chance dès que l'objet se trouve au delà de 1.8° du point de fixation. De même, lors de la lecture, les expériences de Rayner et Bertera 1979 ont montré que les sujets ne peuvent reporter plus de 10% des mots d'une phrase en l'absence de leur vision fovéale sur une étendue de 2°.

Malgré cette importance évidente de la vision fovéale, certains indices montrent que des informations utiles à l'identification d'objets peuvent être extraites de la vision extrafovéale : Pollatsek et al., 84, ont notamment observé un bénéfice important sur la rapidité d'identification lorsque l'objet est d'abord aperçu à 10° d'excentricité avant saccade, même dans un contexte où plusieurs objets sont présents dans le champ visuel (Henderson et al. 1987 ; Henderson et al. 1989). De façon plus indirecte, les expériences de Wolfe et al. 1998 vont dans le même sens en montrant que dans les tâches de recherche visuelle, la difficulté à traiter les stimuli excentrés n'est due qu'à un biais résultant de l'attention focalisée, et non à une moindre acuité incompatible avec la tâche. Ces indices d'une reconnaissance extrafovéale de traits et d'objets dessinés sont corroborés par les expériences d'exploration de scènes, où les premières saccades se portent davantage sur les objets incongrus (informatifs) que sur les objets en accord avec le sens général de la scène (Antes 1974 ; Loftus et Mackworth 1978) ; cependant, les expériences de De Graef et al. 1990, et Henderson et Hollingworth 1999 précisent que les informations vers lesquelles se portent ces saccades sont davantage visuelles (contrastes et contenus fréquentiels) que sémantiques. Henderson et al. 1997, dans une série d'expériences d'exploration libre d'un tableau de dessins où une partie du champ visuel peut être masquée (scotome artificiel), montre que l'absence de vision fovéale peut être exactement compensée par un plus grand nombre de saccades extrafovéales. Cette compensation est quasi-complète en terme de taux de réussite pour les tâches d'identification, mais limitée par l'acuité extrafovéale lorsque des détails (hautes fréquences) sont nécessaires à l'exécution de la tâche. Lorsqu'un objet

est artificiellement absent d'une partie du champ extrafovéal, les fixations fovéales sur l'objet ont une durée plus longue. Ces résultats montrent que l'information fovéale n'est pas nécessaire à l'encodage des objets dessinés, et qu'un traitement de l'information extrafovéale facilite la reconnaissance.

Les résultats de notre expérience vont plus avant dans le sens d'une reconnaissance extrafovéale : l'impossibilité d'effectuer des saccades limite l'exploration de la scène, constituée ici non plus de dessins (4 dans l'expérience d'Henderson) mais d'une photographie. Nous montrons que la reconnaissance de la catégorie animal n'est que très peu affectée (en taux de réussite et temps de réaction) par des présentations excentrées de 3.5°, les bords latéraux des photographies allant jusqu'à presque 7°.

Citons enfin une expérience récente menée par Thorpe et al. 1999 qui accentue les résultats discutés ici. Les auteurs ont étendu la tâche 'animal' en présentations alternées jusqu'à 9 positions : une centrale, et 4 positions symétriques jusqu'à 75° d'excentricité, les images étant présentées pendant 28 ms. Les résultats des 10 sujets de l'expérience montrent que la réussite baisse linéairement en fonction de l'excentricité, mais se maintient au dessus du niveau de la chance même pour les présentations les plus excentrées. Les performances étant égales à droite comme à gauche pour une excentricité donnée, les auteurs concluent à l'absence de supériorité hémisphérique. De fait, les auteurs montrent que les baisses de performances observées en présentations latérales s'expliquent le mieux par la baisse de précision d'échantillonnage due à l'accroissement des champs récepteurs des cellules ganglionnaires de la rétine. La baisse observée ici (3% à 3.5° d'excentricité) est très proche de la valeur estimée par interpolation linéaire dans leur étude.

d) vision latérale : traitement parallèle plutôt que recentrage

Le modèle de Olshausen et al. 1995 en accord avec les choix proposés par Kosslyn, propose un mécanisme neurobiologique de recentrage dynamique de l'entrée visuelle grâce à l'attention focalisée en un lieu de l'espace. Comme nous l'avons analysé en première partie, ce routage dynamique permet de faciliter la reconnaissance de représentations invariantes en taille et en position à partir d'une entrée « pré-formatée ». Les auteurs confient donc à un mécanisme de déplacement attentionnel la capacité de reconnaître des stimuli à plusieurs endroits du champ visuel. En application de ce principe, les auteurs proposent un délai de ré-allocation de 50-100 ms en accord avec les résultats de Nakayama et Mackeben 1989, Krose et Julesz 1989. Comme nous l'avons vu dans le premier chapitre, ces chiffres correspondent au délai de réorientation d'une cible vers une autre ; dans notre expérience où l'attention est partagée au préalable dans un large champ, le recentrage et le déplacement de la fenêtre attentionnelle sur le stimulus peuvent être assimilés au délai d'allocation simple mesuré par Eriksen et Hoffman 1972 et Posner 1978, soit 30-50 ms.

Comment s'applique le modèle d'Olshausen dans notre expérience ? Dans un premier temps, la fenêtre attentionnelle est agrandie en préparation à la tâche jusqu'à une

taille incluant les trois positions possibles pour le stimulus ; cette attention partagée sur un large champ conduit à une résolution très faible à l'intérieur de la fenêtre. L'affichage du stimulus (de forte luminance comparé au fond noir) conduit dans un deuxième temps à la ré-allocation de l'attention vers la position de l'image et à sa taille, accroissant ainsi la résolution perceptive à l'intérieur de la fenêtre par le mécanisme de routage. Cette succession simple est le mécanisme minimum requis pour qu'une image de la tâche à trois positions se trouve au centre de la fenêtre attentionnelle, de façon identique à la tâche à une seule position. Il est possible qu'à ce minimum s'ajoute un biais lié à la position précédente, accroissant ainsi le délai de ré-allocation.

Or les résultats présentés ici indiquent que ce mécanisme attentionnel devrait prendre au plus 10-20 ms, données inférieures au délai minimum d'allocation simple mesuré jusqu'à présent. Ces données nous semblent donc exclure un tel mécanisme de routage dans notre tâche, pour expliquer la capacité des sujets à reconnaître la présence d'un animal en vision latérale dans un contexte d'attention partagée. D'autre part, un mécanisme quelconque de contrôle et de routage (ne reposant pas sur l'attention) qui devrait recentrer dynamiquement l'information en ce laps de temps nous semble peu compatible avec la relative lenteur des circuits neuronaux. L'alternative pour expliquer la rapidité de reconnaissance en vision latérale est de postuler une large redondance des représentations (multiples et rétinotopiques), se résolvant progressivement par convergence (modèle de Rolls et Wallis) ou restant parallèle dans les couches les plus hautes (Thorpe et Gautrais).

Un tel parallélisme de traitement a déjà été montré dans des tâches de recherche visuelle, dans les cas de traits visuels simples et de textures pouvant être traités en parallèle en présentation parafovéale, bien que nécessitant une recherche sérielle en vision fovéale Fiorentini 1989; ici cependant, la cible de la recherche est beaucoup moins bien définie par ses attributs visuels, la catégorie animal pouvant être représentée sous de nombreuses formes prototypiques.

2 l'absence de couleur dans les scènes naturelles

L'expérience précédente a montré que l'attention focalisée n'est pas requise pour réaliser la catégorisation rapide de scènes naturelles, au moins quand il s'agit de détecter la présence d'un animal. Une conséquence directe de ce constat est que peu d'informations sont nécessaires pour réaliser la tâche ; parmi ces informations, les couleurs de la scène aident-elles à la reconnaissance ? Intuitivement, on peut supposer que cette information joue un rôle important dans la performance des sujets, et que son absence pourrait soit allonger le temps de réaction, soit abaisser notablement la réussite.

Un ensemble d'expérience montre en effet que la couleur joue un rôle dans tous les aspects de la vision : dans l'encodage des textures, lors des tâches de reconnaissance et de détection d'objets, mais aussi dans la perception de la profondeur et du mouvement. Ces derniers cas sont les plus surprenants, quand on constate l'importante projection de la voie magnocellulaire vers les aires pariétales. Par exemple, Troscianko et al. 1991 concluent à l'utilisation des gradients de couleur pour l'encodage de la profondeur. L'utilisation de la couleur est démontrée dans la détection du mouvement apparent (Ramachandran et Gregory 1978 ; Papathomas et al. 1991), chez le singe (Dobkins et Albright 1994) et lors de pathologies (Cavanagh et Anstis 1991). Cette implication de la couleur a cependant des difficultés à être expliquée par une information d'origine parvocellulaire, puisque la physiologie tend à montrer le parallélisme des voies du traitement de la couleur, et du mouvement ou de la profondeur. Récemment, des résultats laissent à penser que les informations de couleurs pour les aires pariétales pourraient provenir d'une autre source : Baker et al. 1998 montrent que deux mécanismes différents sont à l'œuvre dans la détection du mouvement par la couleur. D'après l'expérience récente de Cavanagh et al. 1998, l'existence d'une voie de la couleur autre que parvocellulaire s'impose. Les auteurs montrent que la détection du mouvement à partir de la couleur est indépendant des centres occipitaux-temporaux, et même de MT : seule une voie alternative peut expliquer les performances de patients atteints d'une achromatopsie centrale dans une tâche de détection du mouvement à partir de la couleur. Il est ainsi probable que ces informations de couleur aient une origine différente, non parvocellulaire, de celles atteignant la voie ventrale. La question même de l'encodage des informations de couleur a été récemment mise en question par Calkins et Sterling 1999, pour lesquels cette information pourrait transiter via les voies koniocellulaires du corps genouillé[1].

D'autres expériences ont démontré l'importance de la couleur dans les tâches de détection et de reconnaissance d'objets, notamment pour les tâches lors desquelles le sujet doit nommer l'objet présenté (Wurm et al. 1993 ; Boucart et Humphreys 1997; Cohen

[1] D'après Benevento et Port 1995 le pulvinar contient une classe importante de neurones sélectifs à des combinaisons de forme et couleur, et répondent à des latences moyennes de 80 ms (entre 31 et 186 ms), soit bien avant les réponses inférotemporales aux mêmes indices.

1997 ; Grossman et al. 1997 ; Syrkin et Gur 1997). Cependant, l'implication de la couleur comme indice à la reconnaissance des formes demeure sujet à controverses : Biederman et Ju 1988 ont trouvé que l'absence de couleur ne réduit pas la performance d'une tâche de catégorisation de cibles. Dans l'expérience présentée ici, nous nous attacherons à mesurer l'effet de l'absence de couleur dans les scènes naturelles sur la tâche de catégorisation rapide.

2.1 expérience et résultats

a) protocole et hypothèses

17 sujets ont effectué la catégorisation animal / non animal d'au moins 1000 photographies chacun (il leur été demandé au moins 10 séries de 100 stimuli, soit environ 1h d'expérience) ; ces stimuli, pris dans un nouvel ensemble de photographies par rapport aux expériences précédentes, n'étaient vus qu'une seule fois, soit en couleur soit en niveaux de gris. Les stimuli en couleur était choisis aléatoirement au début de chaque série expérimentale.

Les autres caractéristiques du protocole restaient inchangées par rapport à la tâche de référence : présentations centrées sur le point de fixation, durée d'affichage d'un stimulus 20 ms, intervalle interstimulus aléatoire entre 1.8 et 2.5 s, stimuli de 384 x 256 pixels soit 6.6° x 4.5° d'angle de vue à 1.10 m. L'enregistrement de l'EEG était effectué au moyen d'un bonnet 32 voies ; seuls les essais réussis et non-artefactés étaient utilisés pour le calcul des moyennes, au moyen du logiciel Neuroscan.

b) résultats comportementaux

Globalement, la réussite aux stimuli couleur est plus élevée de 1.5% par rapport aux stimuli N&B (Figures 20 et 21). Cette faible différence apparaît comme significative dans les résultats de groupe. Le temps de réaction médian aux cibles réussies est de 441 ms aux photographies couleur, et 447 ms en niveau de gris. Cette baisse de rapidité est très faible, cependant significative à $p<0.005$ ($t = -3.3$ apparié) pour le groupe. Ce résultat indique la tendance de la plupart des sujets à être plus lents en réponse aux stimuli non colorés, approchant à peine la significativité individuelle $p<0.08$ pour 3 sujets parmi les 14 ayant cette tendance ; la tendance inverse n'atteint jamais la significativité (différences médianes de 1.5 ms, 1.5 et 3 ms pour les 3 sujets). Comme indiqué sur la Figure 21, la différence entre les médianes des faux positifs en couleur et N&B n'est pas significative.

Ces résultats comportementaux montrent donc que la performance avec les stimuli en niveaux de gris est plus faible, mais que cette baisse est extrêmement ténue.

c) résultats électrophysiologiques

Le première observation des résultats électrophysiologiques montre la différence extrêmement faible entre les activités différentielles entre les images couleur et noir et blanc. Aucune différence de latence n'est visible sur les sites occipitaux ; les sites frontaux semblent refléter un faible avantage pour les images couleur. Ces résultats montrent que ces potentiels sont évoqués essentiellement par des informations de luminance.

activité principale 'Animal vs non-Animal'

Les comparaisons 'animal vs non-animal' pour les photographies en couleur et en niveau de gris semblent montrer un léger avantage des stimuli en couleur de 5 ms sur les sites frontaux (Tableau B de la Figure 21) ; cependant, cet avantage ténu ne se retrouve pas sur les sites occipitaux. L'activité différentielle à 150 ms n'est donc pas vraiment affectée par la présence ou l'absence d'indices chromatiques dans les stimuli. En revanche il semble exister, avant 150 ms, une première activité différentielle précoce entre essais 'Animal' et 'Non-Animal' qui n'est pas similaire dans les deux conditions de stimulations (Figures 22 et 23), et que allons décrire maintenant.

activités avant 150 ms

Afin de caractériser cette différence 'Animal vs non-animal' avant 150 ms, nous avons représenté sur la Figure 23 les sites sur lesquels cette différence est présente, ainsi que les cartographies de significativité pour les stimuli en niveau de gris et en couleur. Cette activité émerge sur les sites centraux et pariétaux à partir de 100 ms et devient clairement significative entre 110 et 140 ms. Cette activité est davantage prononcée pour les stimuli en couleur qu'en niveau de gris, pour lesquels les pôles positifs des pics d'activités sont plus marqués.

Cette activité 'Animal précoce' se différencie nettement par sa topographie de l'activité liée à la coloration des stimuli (activité 'couleur vs N&B' représentée en Figure 24, indépendamment de 'Animal' et 'non-Animal'), malgré sa proximité temporelle. Il est donc possible de distinguer clairement ces deux phénomènes, l'un lié à la catégorie 'Animal' entre 110 et 140 ms de distribution centro-pariétale, et l'autre liée à la coloration des stimuli entre 130 et 160 ms de distribution pariéto-temporale et de polarité inverse.

d) modèles dipolaires

La Figure 25 indique les modèles dipolaires obtenus pour les deux activités 'Animal vs Non-Animal' que nous venons de distinguer. L'activité précoce observée sur les sites centro-pariétaux semble prendre son origine dans des aires du plancher occipito-temporal extrastriées, tandis que le modèle de l'activité différentielle principale indique une source plus antérieure et plus haute. Ces résultats ont été obtenus indépendamment l'un de l'autre. Nous pouvons noter que le modèle de l'activité différentielle s'initiant à 150 ms est

cohérent avec ceux obtenus lors des modélisations précédentes. Ces résultats indiquent que les deux processus à l'origine de ces différences 'animal vs non-animal' ne partagent pas les mêmes bases cérébrales.

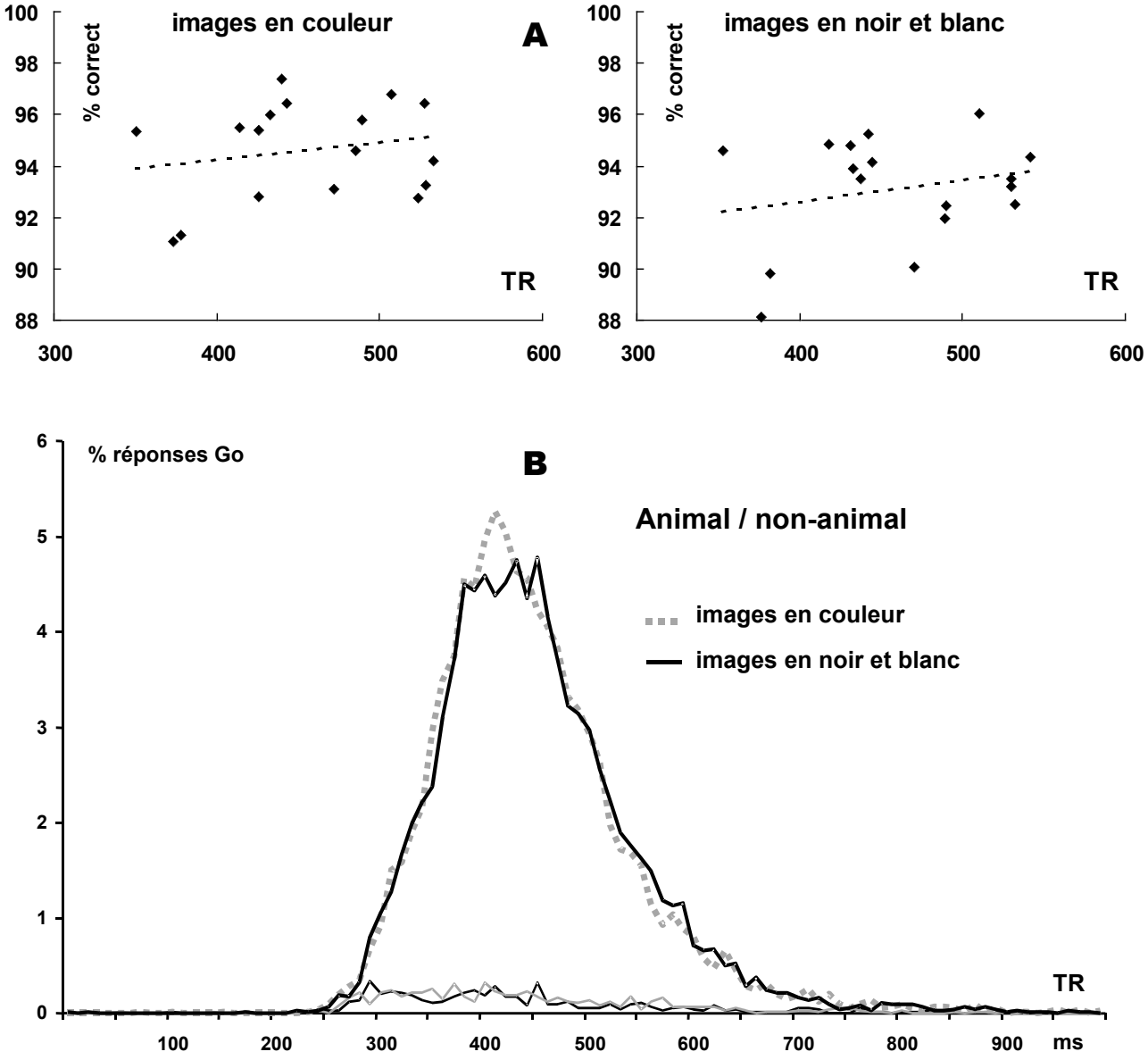


Figure 20 : Résultats de l'expérience 'la couleur dans les scènes naturelles'.

A. Réussite des sujets en fonction du temps de réaction. On observe une baisse globale de la réussite des sujets pour les images en noir et blanc, indiquée par le décalage de 1.5% des droites de corrélation (corrélation faible, r=0.2 dans les deux cas).

B. Histogrammes des temps de réaction. La distribution des réponses aux cibles réussies se sépare de celle des faux positifs sur l'intervalle 290-300 ms, pour les deux types de stimuli ($p<0.001$).

A

	Temps de Réaction médian (ms)				% réussite	
	Cibles réussies		Distracteurs ratés			
sujets	Couleur	N&B	Couleur	N&B	Couleur	N&B
aa	420	423	391	439	96.0	94.8
ad	465	472	439	443	94.6	92.0
br	515	516	445	535	94.2	93.2
cb	411	421	450	497	92.8	93.5
cdt	507	532	450	543	92.8	94.4
cm	336	337	349	336	95.4	94.6
co	517	521	581	560	96.5	93.5
df	431	433	470	481	96.5	94.2
fs	467	463	455	407	95.8	92.5
gr	411	418	411	399	95.4	93.9
mma	362	369	319	327	91.1	88.2
mp	489	494	810	460	96.8	96.1
rvr	406	411	405	415	95.5	94.9
sg	428	430	424	406	97.4	95.3
sjt	455	456	487	499	93.1	90.1
sm	373	372	321	347	91.3	89.8
vr	510	524	513	514	93.3	92.5
moyenne	**441**	**447**	454	447	94.6	93.2
écart C-N&B		**-5.2**	**ms**	6.5	ms	**1.5**	**%**
t apparié (ddl=16)		-3.25		0.283		4.954	
p		0.005		0.781		0.0004	

B

		images **Couleur**		images **N&B**	
Sujets	sites	FP2		FP2	
aa	FP2	169	169	177	177
ad	FP1	159	163	166	166
br	FP1	188	198	210	217
cb	C3	181	182	184	185
cdt	FP1	177	179	189	186
cm	F7	147	188	156	200
co	F4	208	177	210	215
df	PZ	159	162	156	162
fs	FP2	186	203	183	183
gr	F4	160	160	161	182
mma	FP1	177	211	184	213
mp	CB1	160	187	161	173
rvr	CB1	165	195	160	203
sg	FZ	160	162	168	168
sjt	CZ	182	230	196	225
sm	FP2	183	183	178	178
vr	CB1	165	192	192	228
Groupe	FP2	**173 ms**		**178 ms**	
					$p<0.08$

Figure 21 : Résultats comportementaux et électrophysiologiques.

A. Résultats comportementaux : temps de réaction et réussite aux stimuli couleur et noir et blanc.
B. Résultats électrophysiologiques : latence de la première différence cibles – distracteurs observable sur les sites frontaux. Les mesures électrophysiologiques et comportementales donnent le même résultat : un avantage très faible et à peine significatif pour les stimuli colorés.

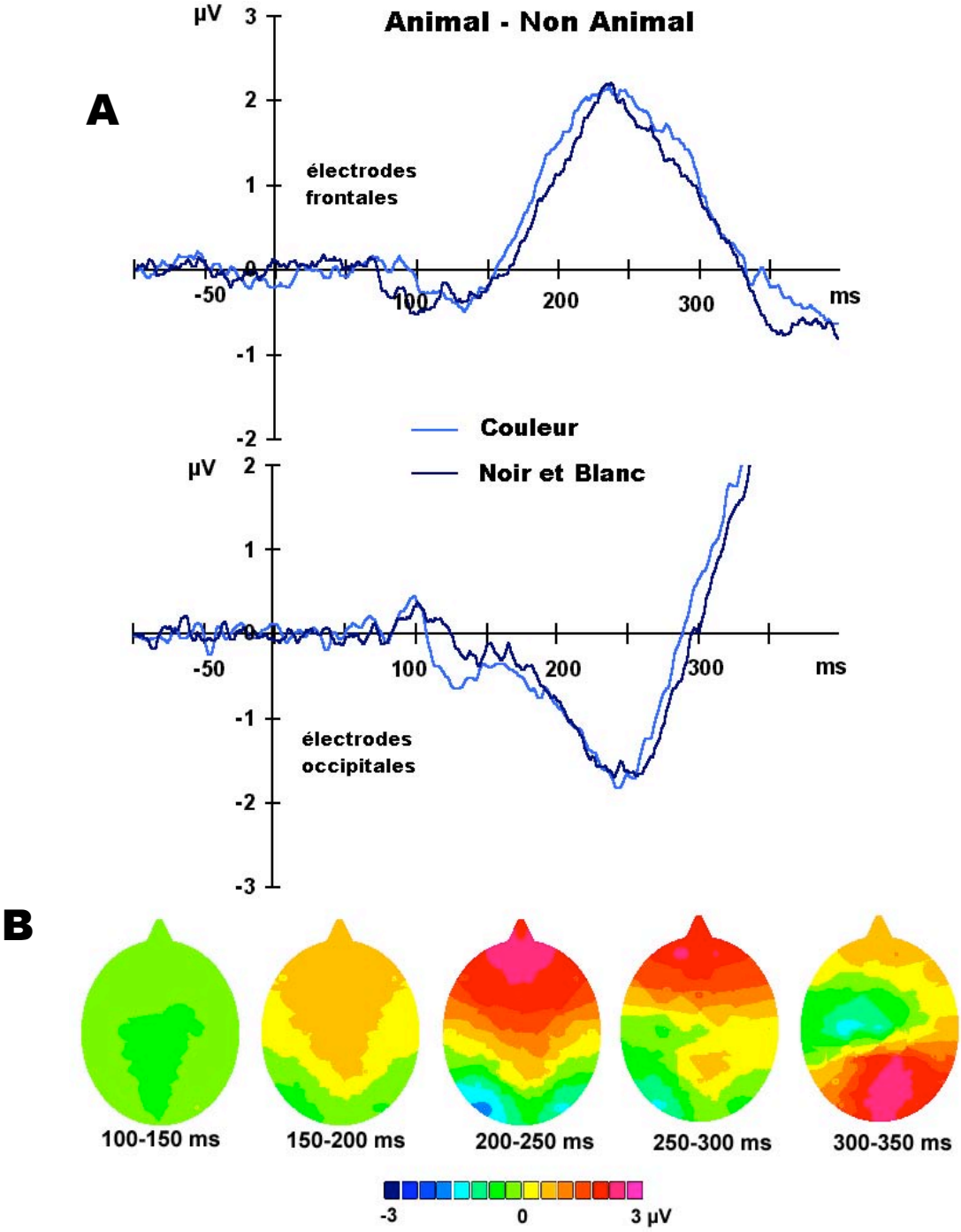


Figure 22 : Résultats électrophysiologiques.

A. Activités Animal vs Non-animal sur les sites frontaux et occipitaux : stimuli couleur et N&B. L'avantage ténu pour les stimuli colorés observable sur les électrodes frontales ne se retrouve pas sur les sites occipitaux. Une autre activité semble émerger avant 150 ms, différente pour les stimuli en couleur et en niveau de gris (cf. Figure suivante).

B. Cartes des activités différentielles au cours du temps. La topographie de l'activité 'Animal vs Non-Animal' à partir de 150 ms jusqu'à 250 ms est identique à celles des expériences précédentes ; selon toute vraisemblance, l'activité suivante est motrice. Les stimuli couleur et N&B sont mélangés (les cartographies sont identiques dans les deux conditions, à partir de 150 ms).

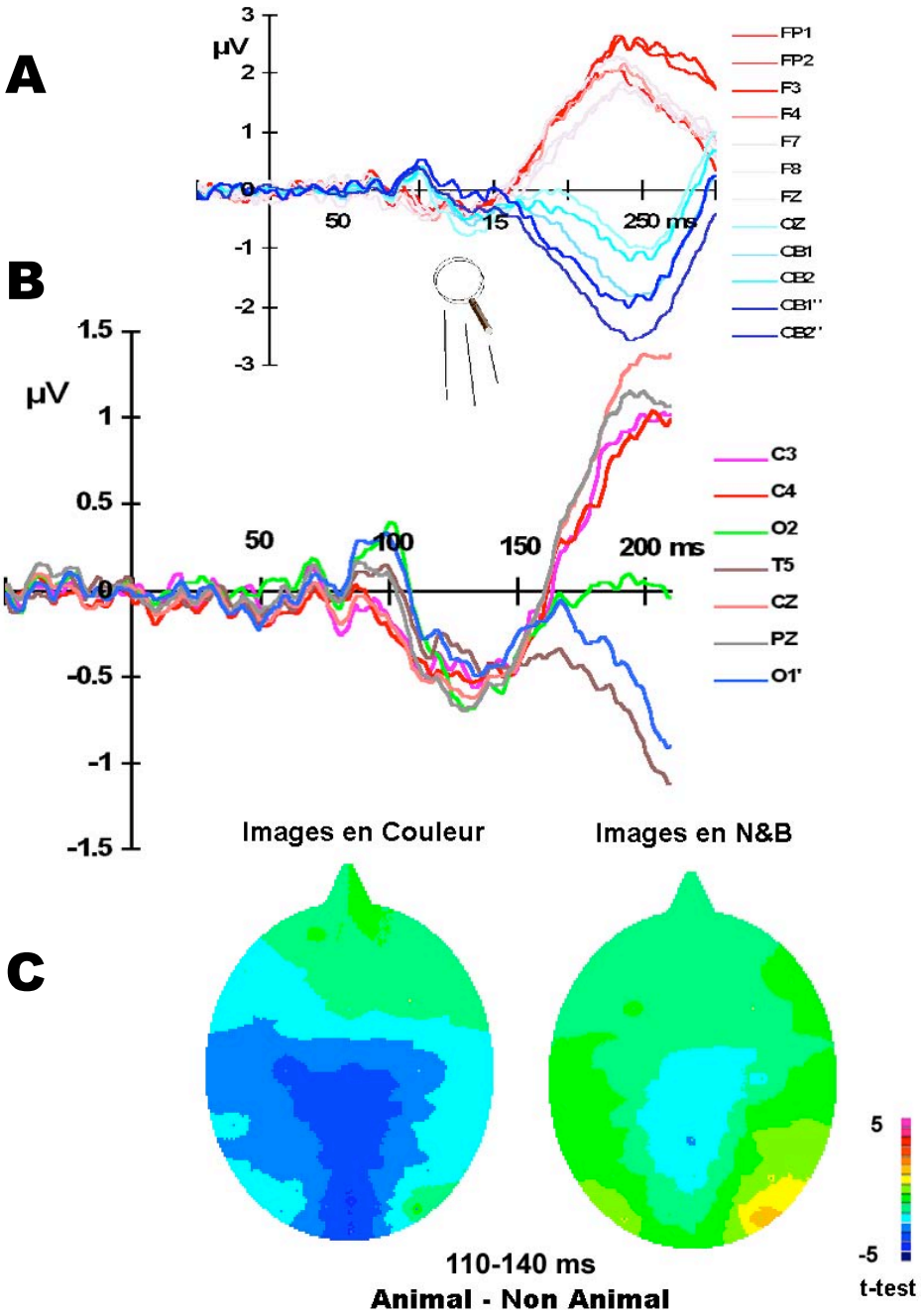


Figure 23 : Activités Animal - Non Animal avant 150 ms

A et B. Activités différentielles Animal vs Non-animal. Les stimuli couleur et noir et blanc sont mélangés, pour mettre en évidence la robustesse des activités différentielles. **(A)** L'activité principale 150-250 ms possède un décours temporel globalement identique sur toutes les électrodes (d'où les modélisations simplifiées). **(B)** Avant 150 ms, l'activité différentielle 'Animal vs Non-Animal précoce' est plus centrale et de polarité négative. Sont représentés les électrodes sur lesquelles cette activité est la plus ample.

C. Cartographies de t-test 'Animal vs non-Animal précoce' : stimuli couleur et N&B. Les résultats sur chaque électrode des t-test entre potentiels évoqués par les scènes cibles et ceux évoqués par les scènes distracteurs sur 110-140 ms sont présentés sous forme de carte. Il s'agit de t-test appariés par sujet, effectués séparément pour les scènes colorées et noir et blanc. Ces résultats montrent une différence Animal vs Non-animal significative avant 150 ms. D'autre part, cette activité est dépendante de la coloration des stimuli, donc davantage liée à l'entrée sensorielle que l'activité suivante à 150 ms.

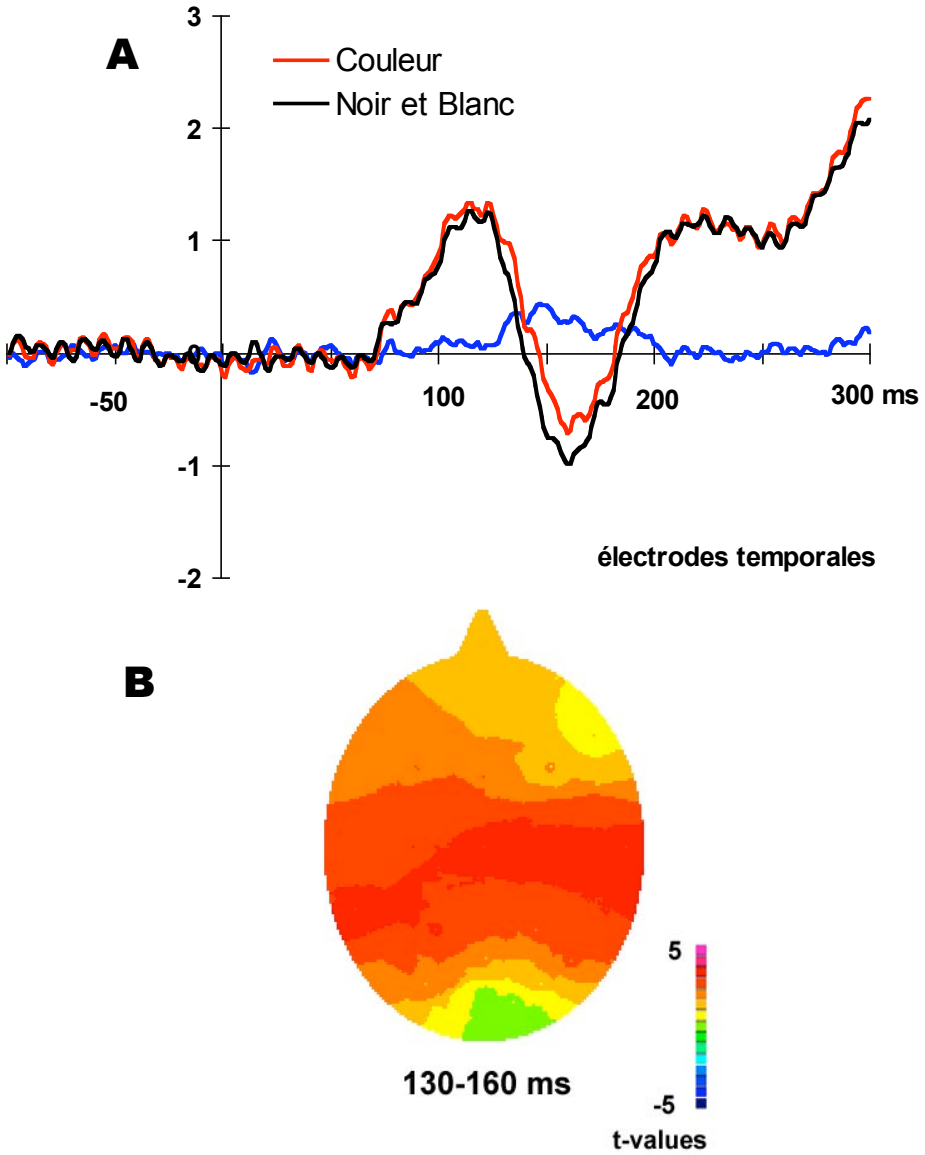


Figure 24 : Activités Couleur vs Noir et Blanc

A. Potentiels évoqués par les stimuli couleur et noir et blanc. Quel que soit le statut de cible ou distracteur des images, la figure représente les potentiels évoqués par les images colorées et non colorées, sur les électrodes temporales (moyenne de 6 électrodes : T7, T8, P7, P8, TP7, TP8). La courbe bleue indique la différence Couleur – N&B, dont la significativité est représentée en (B).

B. Cartes de t-test des différences Couleur vs N&B sur 130-160 ms. La différence des potentiels évoqués par les stimuli couleur vs noir et blanc (courbe bleue en **A**) a été effectuée pour chaque sujet ; la significativité de cette différence par t-test apparié est indiquée ici sur 130-160 ms. Cette activité se différencie nettement par sa cartographie de l'activité Animal vs Non-animal observable sur la période 110-140 ms.

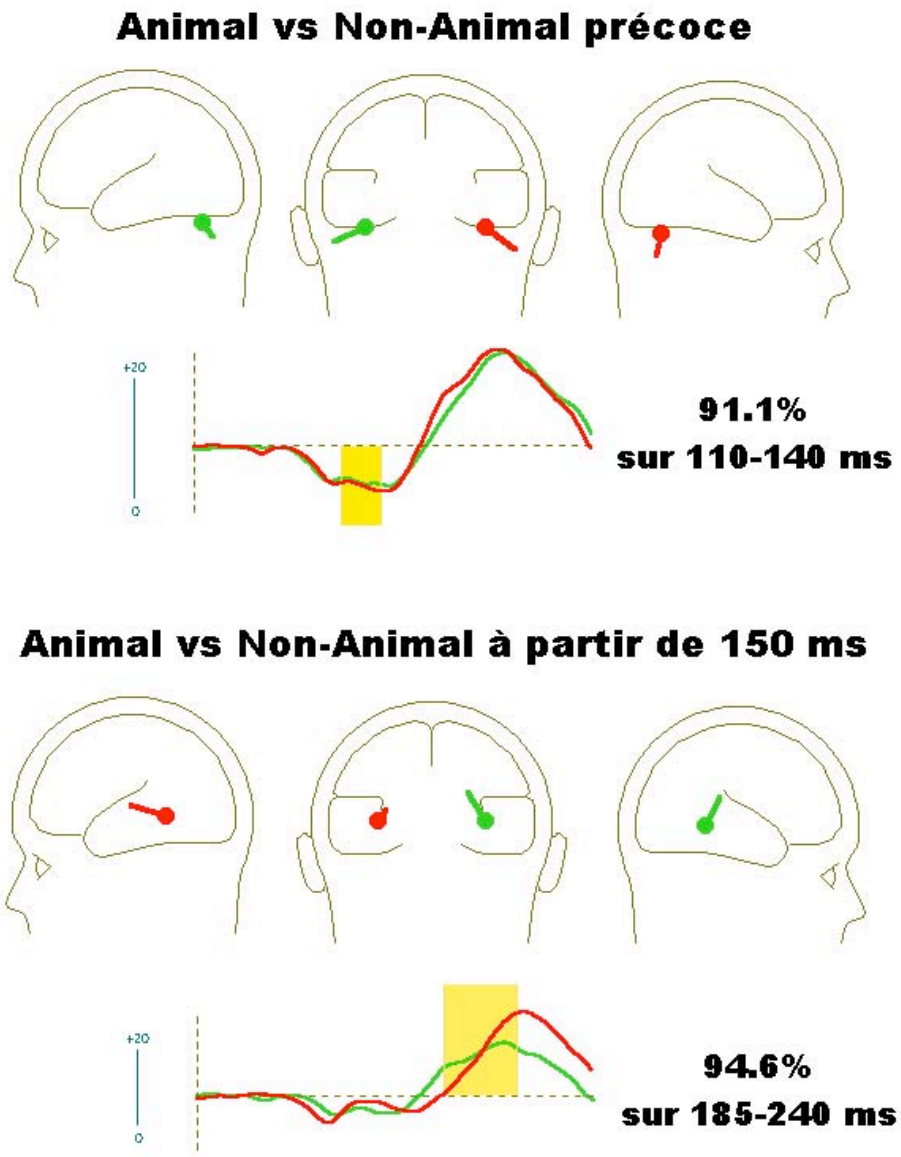


Figure 25 : Modèles des activités Animal vs Non-Animal (expérience 'absence de couleur')

Le pourcentage indiqué est celui de la variance du signal expliquée par le modèle sur la plage d'ajustement. Les courbes représentent les variations d'intensité des sources ajustées : ces variations ne sont compatibles avec le modèle que sur la plage d'ajustement (indiqué en jaune). Les deux activités différentielles ('précoce' et 'à partir de 150ms') ont été modélisées indépendamment, et amènent à des solutions stables. Pour tous les modèles, la stratégie repose sur l'ajustement de la localisation d'une source régionale, puis de l'ajustement de l'orientation de la source réduite à un dipôle si les variations d'intensité sont cohérentes sur les 3 axes de la source. Les plages d'ajustement ont été choisies en fonction de la stabilité de la topographie des activités. Comme pour tous les modèles présentés ici, ces modélisations ne visent pas à expliquer un phénomène complet mais à donner une indication de provenance des différences observées.

2.2 discussions et conséquences

résumons

• L'information couleur n'est pas requise pour la reconnaissance visuelle rapide. Son absence ne baisse que très peu les performances comportementales, et n'a pas d'effet clair sur l'indice électrophysiologique du traitement effectué.

• Une activité différentielle plus précoce liée à la catégorie Animal est observée entre 100 et 140 ms. Cette activité 'Animal' se différencie clairement de l'activité liée à la coloration des stimuli observée entre 130 et 160 ms.

a) le peu d'information nécessaires pour réussir la tâche

Ces résultats confirment les conséquences de l'expérience précédente, où nous avions proposé que le faible recrutement attentionnel observé est la conséquence du peu d'information nécessaire aux traitements. Le fait que la couleur ne soit pas nécessaire pour réaliser correctement la tâche souligne donc l'aspect pré-attentif des traitements visuels effectués lors de la catégorisation rapide des scènes naturelles.

D'autre part le peu d'attention requis est cohérent avec la position d'A. Treisman qui propose que la disponibilité simultanée de la couleur et de la forme pour une tâche conjuguant ces deux facteurs demande un investissement attentionnel (Treisman et Gelade 1980). En accord avec cette théorie, l'expérience de Lavie 1997 montre que les processus de discriminations des formes et des couleurs sont séparés en dehors du foyer de l'attention. Dans notre expérience, le faible recrutement attentionnel peut être expliqué dans ce cadre par le fait que couleur ne soit pas traitée de manière conjointe à la forme. Il serait alors possible que cette séparation des traitements soit le cas de l'ensemble des tâches de reconnaissance.

Le fait que la couleur intervienne peu est-il particulier au protocole expérimental ? Les stimuli colorés sont mélangés aux stimuli en niveau de gris de façon égale et aléatoire, pour que le sujets n'établissent pas une stratégie adaptée aux deux conditions. Mais certains pourraient argumenter que l'indice de couleur ne vaut que pour la moitié des stimuli : une stratégie de reconnaissance faisant intervenir la couleur abaisserait donc de façon notable la performance globale des sujets, qui se trouvent donc réduits à n'utiliser que les indices de formes basés sur les contrastes de luminance. Une première réponse à ce point de vue est que la couleur est encodée pré-attentivement par le système visuel, et se trouve donc en théorie disponible pour renforcer les traitements – lorsqu'elle est présente ; son peu de disponibilité dans les stimuli renforcerait d'ailleurs l'effet de saillance. Si cette information pouvait faciliter notablement la tâche, il est probable que la différence de performance serait bien au-delà des 5 ms et des 1.5% observés. Une seconde réponse plus complète vient des comparaisons avec les expériences incluant des stimuli colorés seuls.

Les performances des sujets lors de ces expériences sont très comparables aux performances observées ici (tâche standard 94% sur 15 sujets, ici 93.8% sur 17 sujets). Soit les traitements effectués sont les mêmes, et cela montre que la couleur n'est pas prise en compte dans la tâche incluant les stimuli colorés, soit la stratégie provoquée par l'absence de couleur se révèle aussi efficace dans la réalisation de la tâche. Dans tous les cas, la couleur n'est pas une information nécessaire pour la catégorisation rapide.

Cependant, il est possible que ce point soit spécifique à notre tâche. L'examen des photographies utilisées montre que les indices de couleur ne sont que de peu d'intérêt pour la détection d'une catégorie aussi vaste que celle des animaux, et aussi diversement représentée dans les stimuli utilisés : une couleur saillante du stimulus ne concorde pas souvent avec l'endroit où un animal ou l'une de ses parties peut être détecté - malgré la présence parmi les stimuli de quelques papillons ou poissons de couleurs éclatantes. La plupart des couleurs ne sont pas typiques de la catégorie animal. L'observation de la couleur n'est donc pas une stratégie efficace pour effectuer la tâche : la couleur n'est pas 'diagnostique' pour la catégorie Animal, au sens de Oliva et Schyns 1997. D'autre part, les sujets effectuent un très grand nombre d'essais, leur permettant d'affiner leur expertise. Ainsi un indice saillant comme la couleur peut capter l'attention du sujet (pop out), mais ne correspond généralement pas aux représentations pertinentes pour réaliser la tâche. Le fait que l'activité différentielle 'Animal précoce' soit différente pour les stimuli couleur et N&B suggère que l'information de couleur est présente. Mais ni les potentiels évoqués ni le comportement n'indiquent que la décision puisse être prise en utilisant cet indice. Dans ces conditions, il est possible que cette information ne soit pas favorisée par le système visuel « préparé » pour la détection d'animaux. Ce point de vue rentre dans le cadre des théories considérant le traitement visuel comme entièrement parallèle, et l'attention comme un biais favorisant le traitement des indices pertinents pour la tâche (Desimone et Duncan 1995).

Il est donc probable que le peu d'information de couleur utilisé par les traitements provienne des aspects particuliers de la tâche : la diversité des représentations de la catégorie animal laisse peu de place à d'autres indices que la forme. Cependant, la rapidité avec laquelle s'effectue la catégorisation des scènes laisse penser que ce sont les mécanismes des traitements qui n'impliquent pas la couleur, comme nous allons en discuter maintenant.

b) la rapidité de la reconnaissance minimise le rôle de la couleur

Quelques études montrent que l'information de couleur aide à la reconnaissance visuelle : l'étude de patients atteints d'une agnosie aperceptive, mais ayant conservé intactes leur perception des couleurs et la capacité à reconnaître des formes géométriques régulières, montre qu'une reconnaissance d'objets dessinés peut être effectuée sur la base de la couleur (Grossman et al. 1997).

Ostergaard et Davidoff 1985 ont développé une série expérimentale pour clarifier l'impact de la couleur sur la reconnaissance visuelle. La tâche consistait à nommer les objets reconnus. Les résultats montrent que nommer des objets colorés est plus rapide que des objets en noir et blanc. Cependant une dissociation est montrée entre la reconnaissance visuelle des objets, sur laquelle la couleur n'avait pas d'effet, et la nomination des objets, facilitée par la présence de couleur. Dans une série d'expériences visant à vérifier si la couleur renforce la reconnaissance, Wurm démontre clairement que les temps de réaction sont plus rapides pour les stimuli colorés, dans une tâche où les sujets doivent nommer les objets reconnus (Wurm et al. 1993). Les auteurs montrent que cet effet est indépendant du contenu fréquentiel des stimuli et de l'acuité des sujets. Cependant, les réponses les plus rapides le sont aux objets prototypiques, indépendamment de la couleur. Comme pour Ostergaard et Davidoff, l'avantage de la présence de couleurs n'intervient qu'une fois l'objet reconnu.

Quelques cas cliniques confirment ce point de vue : Luzzatti et Davidoff, 1994 ont examiné deux patients montrant un déficit à trouver la couleur d'objets qu'ils avaient des difficultés à nommer, alors qu'ils étaient capables de nommer les couleurs seules, comme si la connaissance de la couleur d'un objet était subordonnée à la reconnaissance préalable de cet objet (Luzzatti et Davidoff 1994. Voir aussi Humphreys et Boucart 1997).

Nos résultats sont en accord avec le point de vue des études précédentes : la reconnaissance visuelle ne requiert pas la couleur pour être rapide. La rapidité de la reconnaissance peut être due à la prototypicalité des objets, ou, à l'opposé, à cause de leur spécificité les rendant facilement reconnaissables. Ainsi les dessins utilisés par Davidoff sont-ils peut-être plus aisément reconnaissables que des photographies, laissant la possibilité à la couleur d'intervenir dans les traitements visuels de reconnaissance plus longs.

Dans une analyse plus détaillée des résultats comportementaux de notre expérience, nous avons cherché si les sujets les plus lents présentaient plus que les sujets rapides un avantage lié à la présence de couleur dans les scènes. Aucune relation n'a été trouvée entre un avantage à la couleur et le pourcentage de réussite des sujets (corrélation entre la différence [TR médian couleur – TR médian N&B] et réussite : essais réussis r=0.08, faux positifs r=0.04). Cependant une autre étude menée chez des sujets moins entraînés par A. Delorme et al. (Delorme et al. 1998 ; Delorme et al. 1999) montre un avantage pour les stimuli colorés chez les sujets lents, alors que chez les sujets rapides, la couleur n'aide pas à la reconnaissance.

Ces résultats ne vont donc pas à l'encontre d'un point de vue pour lequel la couleur est une information supplémentaire à la forme et aide à accélérer la reconnaissance. Par exemple, Syrkin et Gur suggèrent que la couleur est utilisée dans la reconnaissance visuelle pour renforcer les traitements basés sur la luminance. Cette hypothèse s'appuie sur leurs expériences (Syrkin et Gur 1997) qui montrent que la discrimination entre cercle et ellipse

est grandement facilitée par la conjonction d'un contraste de luminance et d'un contraste de couleur, par rapport à un contraste lumineux isochromatique ou un contraste chromatique isolumineux. Les auteurs mesurent que le contraste de couleur aide d'autant plus à la discrimination que le contraste de luminance est grand. En ce sens, la reconnaissance est facilitée par la présence de couleur, mais il est probable que leur tâche soit jugée difficile par les sujets, et leurs temps de réaction longs.

c) le rôle prépondérant de la voie magnocellulaire

Quels mécanismes neuronaux et quelles voies physiologiques peuvent-être impliqués dans cette tâche ? La tâche de catégorisation visuelle recrute probablement de nombreuses étapes de la voie ventrale, impliquées dans la reconnaissance des objets. Parmi ces étapes, chez le singe, l'aire inférotemporale contient des neurones répondant préférentiellement à des stimuli précis (visages ou formes simples, parties ou objets entiers), et d'autres répondant à des catégories visuelles (Vogels 1999). Chez l'homme, de nombreuses études montrent une activité dans le gyrus fusiforme (voir l'étude en IRMf). Cette étape est considérée comme un passage obligé lors des tâches de reconnaissance, voire comme l'étape ultime du traitement particulier à l'entrée visuelle (avant des traitements plus intégratifs impliquant mémoire ou réponse comportementale). Dans cette aire, l'information de couleur semble peu affecter les réponses individuelles des neurones : seul un petit pourcentage parmi ceux-ci ont leur activité renforcée par la présence de couleur dans les stimuli (Tanaka et al. 1991 ; Nakamura et Kubota 1995; Booth et Rolls 1998).

Or parmi le flux d'informations visuelles atteignant le cortex (trois voies d'origines distinctes : les voies sous-corticales magnocellulaire, parvocellulaire et koniocellulaire se prolongeant dans les voies corticales magno- , blob- et interblob- dominantes dans V1), la voie parvocellulaire P semble seule véhiculer l'information de couleur[1] (Tootell et al. 1988), disponible à une forte résolution permettant une analyse fine d'une image statique. La voie magnocellulaire M est traditionnellement impliquée dans l'extraction du mouvement ou de la structure des objets à partir de ce mouvement, et possède une résolution spatiale plus grossière. Chez l'homme non plus, cette voie ne semble pas contenir d'information de couleur : Milner et al. montrent que des dommages à la voie magnocellulaire n'affectent pas la vision des couleurs (Milner et al. 1991) - un patient atteint d'agnosie des formes préserve une perception des couleurs consciente, allant dans le sens d'une origine parvocellulaire de cette information par rapport à la voie alternative proposée par Cavanagh (Cavanagh et al. 1998). Les études concernant la vitesse avec laquelle cette information transite indiquent que, chez le singe, la voie M est plus rapide d'environ 20 ms que la voie P (Nowak et al. 1995; Nowak et Bullier 1997) - l'information de mouvement et de forme grossière étant disponible avant les détails et les couleurs.

[1] voir l'article récent de Calkins et Sterling 1999 pour une implication de la voie koniocellulaire sur le traitement de la couleur. Cette voie est probablement plus lente que la voie parvocellulaire.

D'après ce qui précède, l'idée selon laquelle la réussite de la tâche de catégorisation visuelle rapide est basée sur l'activité précoce de chaque aire corticale traversée (les premiers potentiels d'action) amène a supposer que (1) la couleur n'étant pas cruciale et la voie M rapide, la catégorisation rapide recrute des traitements basés sur les informations amenées par la voie M, et (2) cette information, bien que moins détaillée, est suffisante pour décider de la présence d'un animal. Cette proposition est étayée par les résultats de l'expérience citée plus haut (Delorme et al. 1998) montrant que l'absence de couleur n'a pas d'incidence sur les réponses les plus rapides. Cette proposition est aussi en accord avec les résultats de l'expérience de Thorpe et al. 1999, qui montrent que la catégorisation 'animal' est possible et réussie au dessus du niveau de la chance en présentant les stimuli en périphérie rétinienne extrême (jusqu'à 75° d'excentricité). A cette excentricité, la reconnaissance des objets repose essentiellement sur des analyses achromatiques à basse fréquence spatiale.

d) une activité 'Animal' avant 150 ms

Cette expérience a permis de mettre en évidence avant 150 ms une différence entre les potentiels évoqués par les animaux et ceux évoqués par les autres scènes. Cette activité est ici clairement significative sur 110-140 ms ; lors des expériences précédentes, cette activité n'avait pu être mise en évidence : présente bien que non-significative, elle peut être observée par exemple sur la Figure 18. Cette activité différentielle entre les scènes d'animaux et les autres scènes est-elle de même nature que l'activité à partir de 150 ms ? Etant plus précoce, est-elle davantage liée à l'encodage visuel des stimuli ? Nous observons ici que cette activité précoce varie notablement avec la coloration des stimuli, alors que l'activité à 150 ms ne varie pas (en amplitude ou topographie) ou peu (en latence). Ceci semble aller dans le sens d'un encodage visuel lié aux attributs physiques des stimuli. En conséquence, l'activité à 150 ms est peut-être liée plus directement à la tâche. Cependant, ces quelques observations sont insuffisantes pour argumenter de façon plus décisive, et nous tâcherons au cours des deux expériences suivantes de répondre de façon plus directe.

3 détecter la couleur dans les scènes naturelles

Jusqu'à présent, la tâche demandée aux sujets a été de répondre à la présence d'un animal dans l'image : la catégorie animal est cherchée, et nous pouvons penser que dans une certaine mesure, s'il existe des représentations cérébrales d'animaux ou d'indices signalant leur présence, alors elles sont amorcées dans le système visuel du sujet prêt à effectuer la tâche, son attention préparatoire est « focalisée sur eux ». La rapidité avec laquelle s'effectue cette catégorisation d'image est peut-être un corollaire de cet état du sujet, tendu, en quelque sorte, vers la catégorie cible. Notons que ceci n'enlève rien aux résultats précédents : le système visuel est simplement capable d'effectuer les traitements aussi rapidement, avec les implications dont nous avons essayé de discuter. Il peut s'agir justement d'une situation limite : une forte proportion de sujets ont participé aux séries expérimentales précédentes, et vu le nombre d'essais requis par séance, nous pouvons nous attendre à ce qu'une certaine expertise à détecter les animaux puisse rendre à terme cette tâche quasi-automatique.

Or l'existence d'un automatisme semble se dégager des résultats : nous avons observé qu'un plus grand recrutement attentionnel a pour effet de compenser un champ visuel élargi, mais n'affecte pas la rapidité de traitement. La situation de base ne semble pas *a-posteriori* exiger une attention soutenue. L'expérience de Thorpe et al. 1999 présentant ces mêmes photographies en périphérie lointaine de la rétine montre des taux de réussite différents du hasard quand bien même les sujets n'ont pas conscience d'avoir perçu un animal. Ces indices laissent à penser que les résultats rapportés ici sont des mesures d'un mécanisme automatique de la vision.

Pour séparer les effets d'une expertise de ceux d'une reconnaissance automatique, nous proposons d'observer ici s'il existe un indice de traitement associé à la présence d'un animal dans l'image lors d'une tâche indépendante de cette catégorie. Nous avons demandé à de nouveaux sujets d'effectuer une tâche de détection de couleur dans les scènes de l'expérience précédente, scènes contenant des animaux ou non. Afin d'éviter qu'une attention particulière soit portée sur la catégorie animal, il nous a semblé important de sélectionner des sujets n'ayant jamais entendu parler des expériences ci-dessus ; de même, le mot « animal » n'était pas prononcé par l'expérimentateur jusqu'à la fin de la séance. La moitié des sujets avaient pour consigne de détecter les images contenant de la couleur, l'autre moitié de détecter les images en noir et blanc.

3.1 expérience et résultats

a) protocole et hypothèses

14 sujets droitiers (9 femmes) ont effectué la catégorisation de photographies en fonction de la présence de couleur. 7 sujets avaient pour consigne de relever le bouton d'une souris le plus rapidement possible dès que l'image affichée était en couleur ; 7 autres sujets devaient répondre aux images en noir et blanc (niveaux de gris). Les photographies n'étaient vues qu'une seule fois (soit en N&B, soit en couleur) présentées pendant 20 ms. Ces photographies étaient les mêmes que les stimuli de l'expérience précédente : la moitié contenait donc un animal, et les stimuli couleur était choisis aléatoirement au début de chaque série expérimentale. Il était demandé aux sujets au moins 10 séries de 100 stimuli, soit environ 1h d'expérience. Les autres caractéristiques du protocole restaient inchangées : présentations centrées sur le point de fixation, intervalle interstimulus aléatoire entre 1.8 et 2.5 s, stimuli de 384 x 256 pixels, 6.6° x 4.5° d'angle de vue à 1.10 m. L'enregistrement de l'EEG était effectué au moyen d'un bonnet 32 voies ; seuls les essais réussis et non-artefactés étaient utilisés pour le calcul des potentiels évoqués moyens.

hypothèses du protocole

• Nous cherchons préalablement si la présence d'animaux dans certains stimuli biaise les temps de réaction ou la réussite aux tâches de détection de présence/absence de couleur. L'existence d'un tel biais montrerait l'interférence d'une perception catégorielle sur une tâche de 'bas niveau' visuel ; son absence indiquerait que la tâche a été accomplie indépendamment des objets présents dans l'image.

• Les potentiels évoqués aux images contenant des animaux comparés aux images n'en contenant pas permettent de déterminer si un encodage de la catégorie animal existe malgré une tâche dirigée sur un autre attribut des images. L'absence d'un tel signe ne permettra pas de conclure.

• Si de tels signes d'encodage existent, la comparaison (en latence et en topographie de surface) avec les signes de reconnaissance observés dans les expériences précédentes permet de déterminer s'il s'agit des mêmes phénomènes. La robustesse de la latence et de la topographie des activités différentielles 'animal – non animal' déjà observée rend cette comparaison possible, même s'il ne s'agit pas du même groupe de sujets.

• Des indices supplémentaires sur la rapidité du système visuel peuvent être obtenus en comparant les potentiels évoqués aux cibles et aux distracteurs des deux tâches. (1) La comparaison cibles vs distracteurs, indépendamment de la tâche, permettra de mettre en évidence des indices liés à l'adéquation du stimulus à la tâche (l'ensemble de stimuli étant identique pour les deux conditions). (2) La comparaison stimuli couleur vs stimuli N&B permettra de révéler un encodage commun de la couleur indépendamment de la cible de la

tâche. Dans les comparaisons ci-dessus, la validation de la méthode sera permise par l'examen détaillé des potentiels évoqués seuls.

b) résultats comportementaux

Notons que les sujets ont rapporté être surpris de la difficulté de la tâche. La plupart ont souligné la nécessité de maintenir une attention vigilante pour ne pas se tromper. Cette vigilance s'explique par la présence de quelques photographies dans lesquelles les indices de couleur ne sont pas saillants (la difficulté d'une image représentant un ours blanc sur la banquise a été mentionnée fréquemment). Cette impression subjective laisse penser que les sujets étaient suffisamment absorbés dans la tâche pour ne pas porter une attention particulière aux images contenant un animal.

Prise globalement sur l'ensemble des deux tâches (14 sujets, Figure 26), la réussite a été de 93%, et la moyenne des temps de réaction individuels médians de 502 ms (moyenne des TR moyens 517 ms). La tâche détection de couleur a été moins bien réussie que la détection de N&B (92% contre 94%), mais effectuée plus rapidement (481 ms contre 524 ms) bien que nous ne puissions établir la significativité de ces comparaisons (deux populations différentes de 7 sujets). Les temps de réaction médians, pour les tâches prises ensemble, sont comparables entre les images contenant un animal et les autres scènes (animaux 503 ms, non-animaux : 501 ms, $p>0.1$), bien que les scènes contenant un animal aient tendance à être moins réussies (animaux : 92.6%, non-animaux 94.1, $p>0.01$).

Cette tendance se retrouve dans les deux tâches prises séparément : les images contenant un animal sont à peine moins réussies (Couleur 91% vs 93%, N&B 93% vs 95%), mais cet effet n'est significatif dans aucune des tâches (Figure 27). Les temps de réaction médians ne sont pas significativement différents (Couleur 484 ms vs 478 ms, N&B 524 ms pour animaux et non-animaux). De même, les analyses des erreurs ne laissent apparaître aucune différences entre les deux types de scènes (Figures 27-29). Au vu de ces résultats, nous pouvons affirmer que les tâches ont été accomplies indépendamment du contenu des images, la présence de la catégorie animal dans les stimuli n'intervenant pas dans le comportement des sujets.

c) résultats électrophysiologiques

les couleurs et les tâches

Les potentiels évoqués par les stimuli colorés diffèrent de ceux évoqués par les stimuli noir et blanc entre 115 et 145 ms (Figure 30), de façon indépendante de leur statut de cibles ou de distracteurs : la différence 'Couleur vs N&B' est de même signe dans les deux tâches, et d'une topographie remarquablement similaire. Cet effet est observable bilatéralement sur les sites temporaux et occipitaux latéraux. Cette activité liée à la couleur est identique (signe, amplitude, latence, topographie) à celle observée dans l'expérience

précédente, dans laquelle d'autres sujets effectuaient la catégorisation 'Animal' sur des photographies soit colorées, soit N&B (voir Figure 24).

Les potentiels évoqués par les cibles diffèrent de ceux évoqués par les distracteurs à partir de 150 ms indépendamment de la coloration des stimuli (tâches mélangées). Cet effet de la tâche est lui aussi similaire aux différences cibles – distracteurs dans la tâche 'Animal'. La Figure 31 montre que la topographie de cette activité est comparable dans les deux tâches, bien que provenant de groupes de sujets différents, effectuant des tâches où les cibles de l'une sont les distracteurs de l'autre, et vice-versa.

Ces résultats montrent une dissociation claire entre les effets de la coloration des stimuli indépendamment des tâches, et les effets des tâches indépendamment de la coloration des stimuli.

animal vs non-animal

Les potentiels évoqués par les scènes contenant des animaux et celles n'en contenant pas divergent nettement sur les électrodes occipito-temporales gauches à partir de 170 ms (l'effet est clairement significatif entre 180 et 230 ms, Figure 32). Il est remarquable que cet effet soit similaire chez des sujets différents engagés dans des tâches différentes. La latence, le signe et la topographie de cette différence occipito-temporale « Animal vs Non-animal implicite » sont comparables aux activités « Animal vs Non-animal explicites » observées sur les mêmes sites lors des expériences précédentes. Le reflet fronto-central Animal vs Non-animal est différent entre les tâches détection de couleur et détection de N&B ; il n'est pas possible ici de décider si cet effet est lié aux tâches ou aux sujets. Cependant, la comparaison de ces effets frontaux (voir Figure 45) met en évidence la similarité de leur décours temporels, chez les deux groupes de sujets, ainsi que la présence de plusieurs composantes dans cette activité.

Comme lors de l'expérience précédente, il est possible de distinguer une première différence 'Animal vs non-Animal précoce' à partir de 110 ms, visible surtout dans la tâche de détection d'images noir et blanc sur les électrodes frontales, d'une amplitude comparable à celle observée précédemment. Une deuxième différence est visible sur les électrodes frontales et occipitales gauches à partir de 150 ms, d'une topographie analogue à l'activité différentielle 'Animal vs non-animal' de toutes les expérience précédentes. Son amplitude est ici clairement très amoindrie, les animaux ne sont pas catégorie cible de la tâche.

d) modèles dipolaires

Les modélisations dipolaires des activités ont été effectuées sur les données de groupe des 14 sujets. Le modèle de la Figure 33 présente l'activité différentielle 'cibles vs distracteurs' pour des stimuli couleur et N&B représentés de manière identique dans les catégories cibles et distracteurs (mélange des deux tâches). Nous pouvons observer que les

activités de ces sources sont plus tardives, et localisées davantage à la jonction pariéto-temporale que les activités liées aux animaux.

Par contre, l'activité 'Animal vs Non-Animal' mise en évidence entre 150 et 300 ms alors que les sujets sont impliqués dans une autre tâche, est ici clairement occipito-temporale. Cette localisation est très proche de celle observée dans le cas des différences précoces. Ce résultat est cohérent avec l'hypothèse d'un processus d'encodage automatique dans les aires temporales, dont l'activité devient significative à partir de 110 ms lorsqu'un animal est recherché, et observable à partir de 150 ms lorsque cette catégorie visuelle n'intervient pas dans la tâche à accomplir.

A

	Moyenne	Médiane	Dév. Std	% réussite
amc*	597	577	120	88
bd*	534	513	98	95
cm*	550	535	97	94
fs*	523	512	87	95
ic*	485	462	105	96
md*	575	569	86	95
sv*	520	500	113	93
br	541	524	105	95
cg	448	439	87	91
cha	568	557	95	92
eg	441	427	77	94
ii	499	481	95	90
lp	439	433	85	94
vr	523	508	92	91
détection de la couleur	494	481	91	92
*détection du N&B	540	524	101	94
tâches mélangées	517	502	96	93

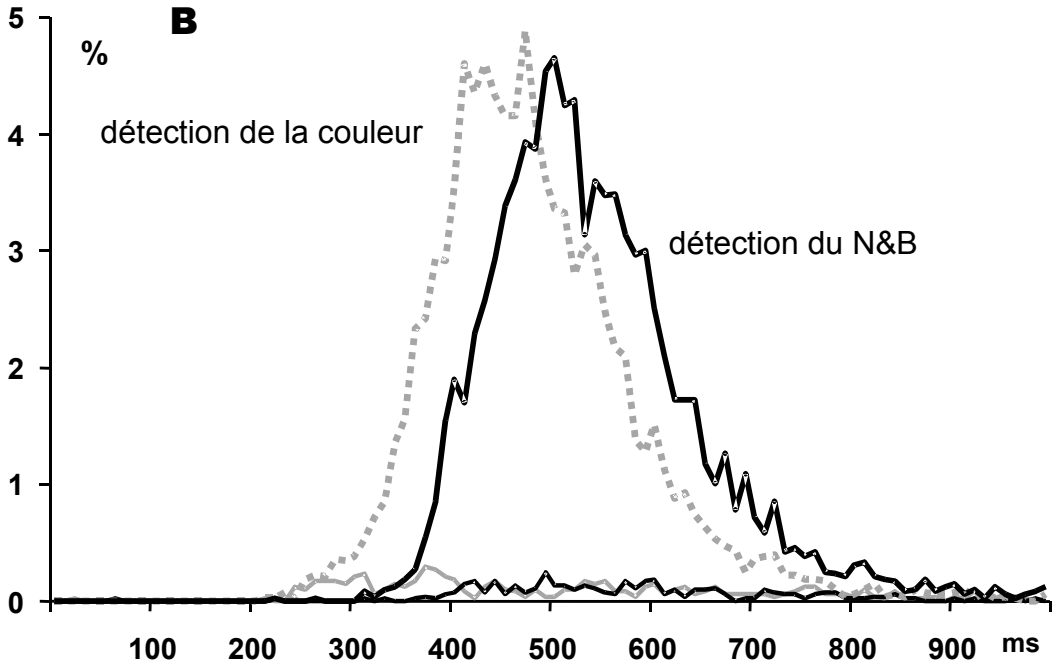


Figure 26 : Résultats de l'expérience 'détecter la couleur dans les scènes'

A. **Tableau des résultats individuels.** La tâche détection de couleur a été moins bien réussie que la détection de N&B, mais effectuée plus rapidement ; il s'agit cependant de deux populations différentes de 7 sujets. Les moyennes, médianes et variances sont exprimées en millisecondes.

B. **Histogrammes des temps de réaction des deux groupes.** Les distributions des essais réussis se séparent des faux positifs à partir de 310 ms pour la tâche de détection de couleur, et à partir de 350 ms pour le noir et blanc.

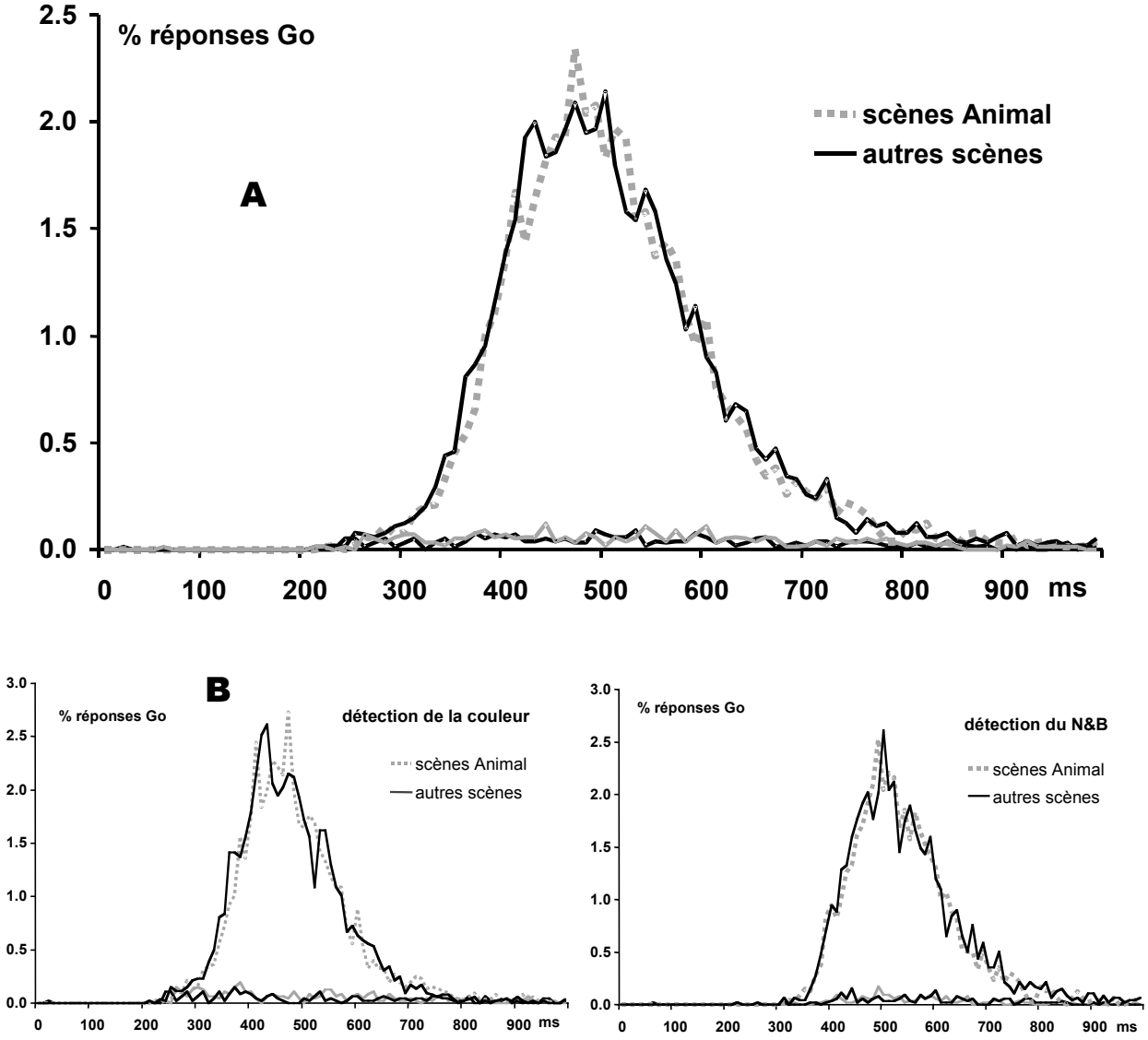


Figure 27 : Influence des scènes Animal dans les tâches de détection

A. Histogrammes des réponses aux deux types de scènes, tâches mélangées. La distribution des essais réussis se différencie de celle des essais ratés sur l'intervalle 320-330 ms, indépendamment du type de scène (Animal ou autres scènes). Il n'y a d'autre part que cinq intervalles de 10 s sur lesquels les tests d'égalité (chi2) entre les distributions Animaux et autres scènes sont inférieurs à p=0.05 : les deux types de scènes ne sont pas traitées différemment.

B. Histogrammes des réponses aux deux types de scènes, tâches séparées. Séparer les analyses sur les deux tâches donnent essentiellement les mêmes résultats.

A

Essais réussis

	Non-Animal				Animal			
	Moyenne	Médiane	Dév. Std	% réussite	Moyenne	Médiane	Dév. Std	% réussite
amc*	597	577	122	90,8	597	577	119	86,0
bd*	533	512	100	95,1	534	515	96	94,4
cm*	550	536	96	95,0	550	535	98	93,8
fs*	526	514	98	95,5	519	510	77	95,3
ic*	482	458	103	96,2	488	466	106	96,6
md*	581	574	91	95,0	569	563	80	94,5
sv*	524	498	120	94,4	516	501	106	91,5
br	534	518	100	94,5	548	529	110	96,5
cg	442	433	83	93,4	453	445	92	89,3
cha	570	564	89	92,0	566	551	101	91,8
eg	444	427	80	95,0	439	427	75	92,7
ii	492	477	93	92,1	507	486	98	88,1
lp	436	427	89	94,2	442	438	80	93,9
vr	514	503	79	93,3	532	512	105	88,3
détection de la couleur	490	**478**	88	**93**	498	**484**	94	**91**
*détection du N&B	542	**524**	104	**95**	539	**524**	97	**93**
tâches mélangées	516	501	97	94,1	517	503	95	92,6

B

tests appariés entre essais Animal et Non-animal réussis

	Moyennes	Médianes	Ecarts-types	% réussite
détection de la couleur (ddl=6)				
écart Animal - Non-animal	8	5	7	-2
t	2,276	1,550	1,560	-2,042
p	0,063	0,172	0,170	0,087
*détection du noir et blanc (ddl=6)				
écart Animal - Non-animal	-3	0	-7	-1
t	-1,245	-0,187	-2,131	-2,016
p	0,259	0,858	0,077	0,090
tâche mélangées (ddl=13)				
écart Animal - Non-animal	2	3	0	-2
t	0,944	1,153	-0,037	-2,920
p	0,362	0,270	0,971	0,012

Figure 28 : Comparaisons des scènes Animal et des autres scènes : essais réussis

Erreurs

A

	Non-Animal			Animal		
	Moyenne	Médiane	Dév. Std	Moyenne	Médiane	Dév. Std
amc*	612	605	140	547	564	228
bd*	607	532	139	593	576	119
cm*	530	515	115	579	587	110
fs*	528	496	136	493	487	82
ic*	431	413	47	542	507	173
md*	593	579	116	636	603	133
sv*	580	564	126	576	536	174
br	759	776	167	747	760	159
cg	358	316	91	390	353	130
cha	671	746	218	621	654	188
eg	575	573	169	497	519	109
ii	576	592	108	611	605	175
lp	362	345	113	360	318	105
vr	405	405				
détection de la couleur	529	536	144	538	535	144
*détection du N&B	554	529	117	567	551	146
tâche mélangées	542	533	129	553	544	145

B **tests appariés entre essais Animal et Non-animal ratés**

	Moyennes	Médianes	Ecarts-types
détection de la couleur (ddl=5)			
écart Animal - Non-animal	-13	-23	0
t	-0,688	-1,228	-0,003
p	0,522	0,274	0,998
*détection du noir et blanc (ddl=6)			
écart Animal - Non-animal	12	22	29
t	0,542	1,152	1,202
p	0,607	0,293	0,275
tâche mélangées (ddl=12)			
écart Animal - Non-animal	1	1	15
t	0,051	0,090	0,998
p	0,960	0,930	0,338

Figure 29 : Comparaisons des scènes Animal et des autres scènes : faux positifs

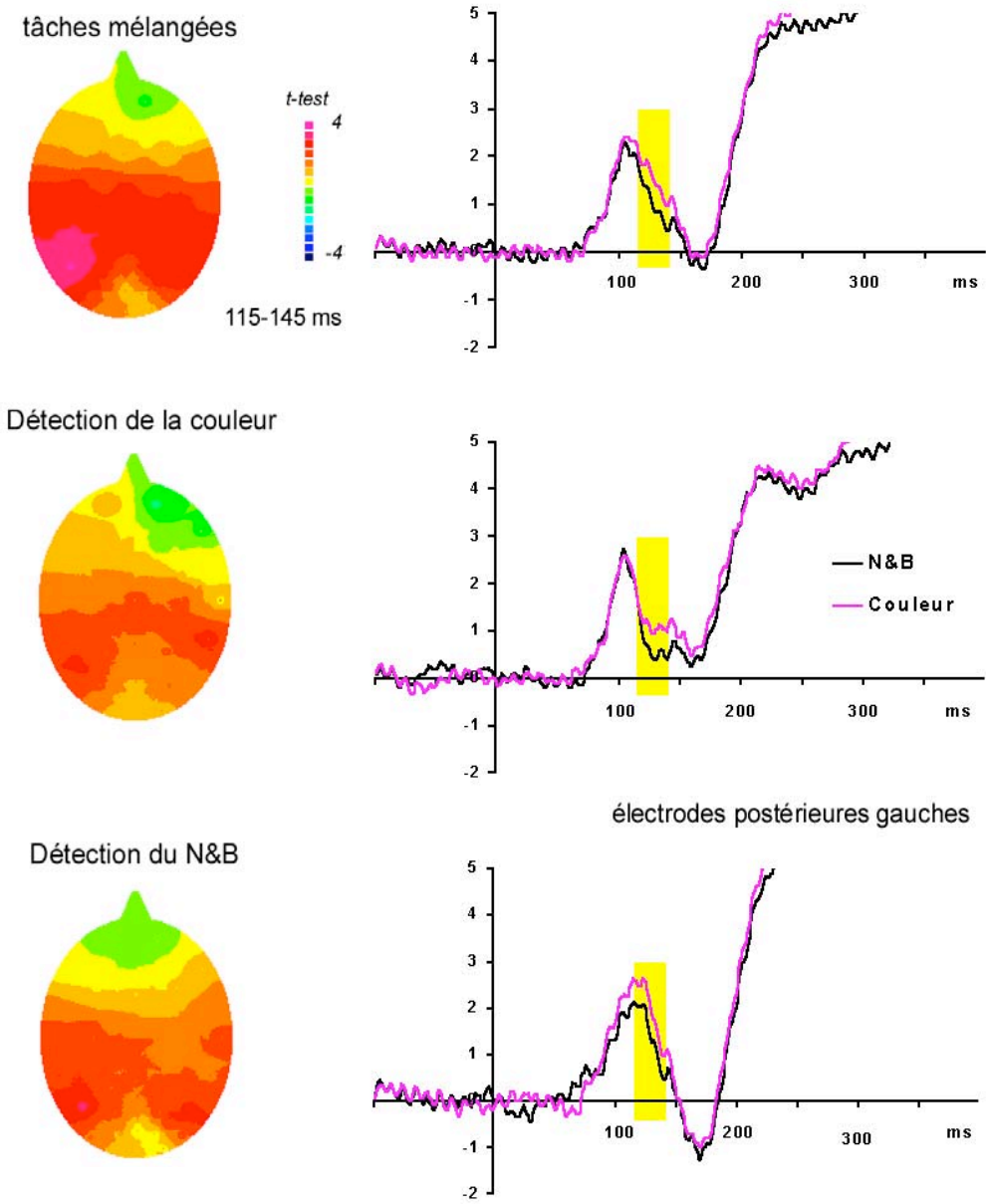


Figure 30 : Effets de la coloration des stimuli dans les deux tâches

La colonne de gauche présente les cartes de t-test différenciant les potentiels évoqués aux stimuli couleur vs les potentiels évoqués aux stimuli noir et blanc quelle que soit la tâche : dans la tâche de détection de couleur, les scènes colorées sont cibles tandis qu'elles sont distracteurs dans l'autre tâche. La colonne de droite montre ces potentiels (en µV) au cours du temps ; la plage colorée en jaune indique l'intervalle sur lequel les cartes de t-test ont été représentées. Il est remarquable de constater que la présence de la couleur dans les scènes intervient de la même façon, qu'elle soit cible ou non. Cet effet est d'autant plus robuste que les sujets sont différents d'une tâche à l'autre. La topographie de la différence couleur vs N&B observée ici sur 115-145 ms est identique à celle observée chez d'autre, sujets sur 130-160 ms, lorsque la couleur n'intervient pas dans la tâche (tâche Animal, différences scènes couleur vs scènes N&B, Figure 24).

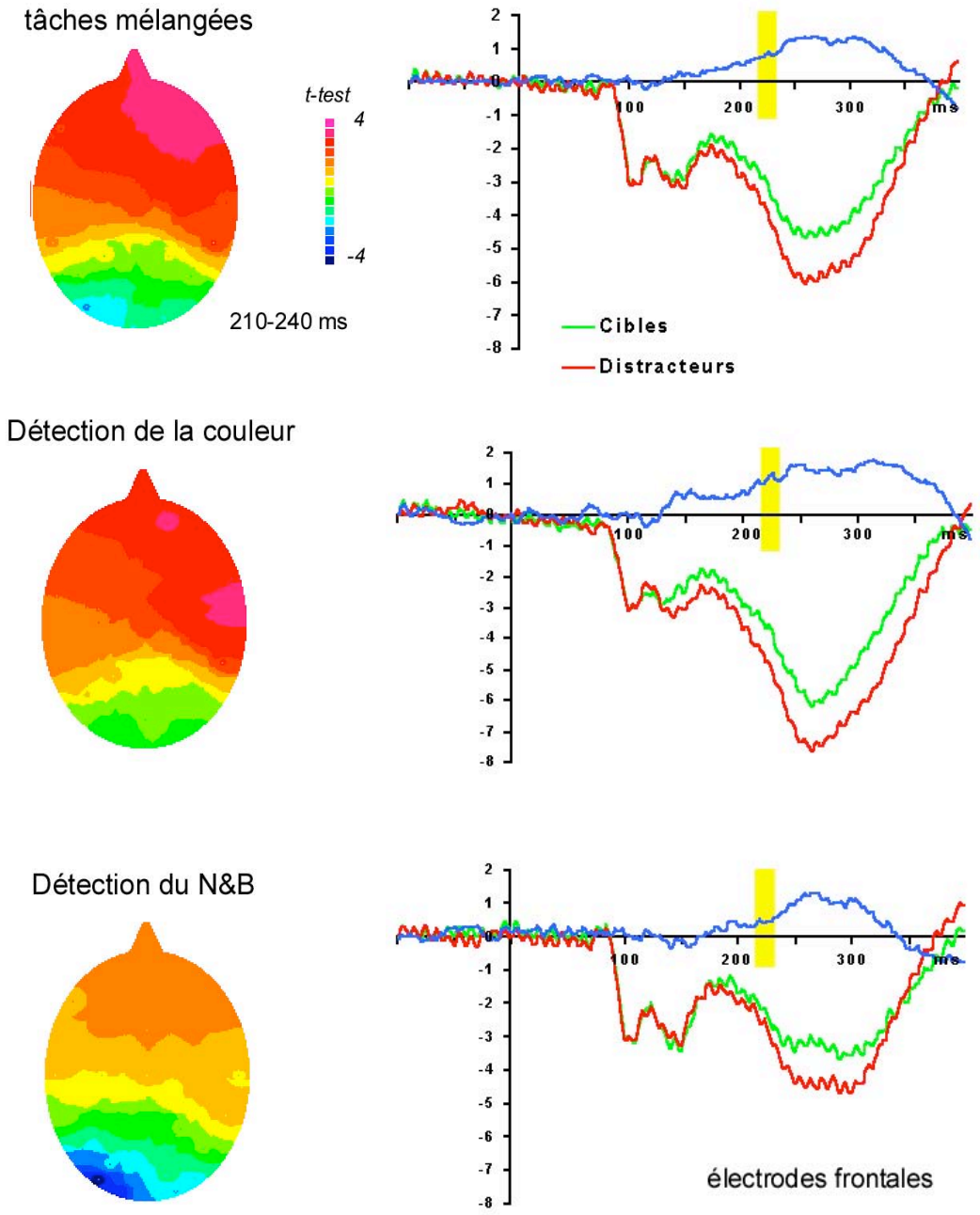


Figure 31 : Cibles et Distracteurs dans les deux tâches

Même présentation que la figure précédente. Sont présentés ici les potentiels évoqués par les cibles et les distracteurs des tâches de détection de couleur et de détection des scènes en noir et blanc. Les stimuli cibles sont les scènes en couleur dans la tâche de détection de couleur, tandis que les cibles sont les autres scènes dans l'autre tâche. Les différences entre cibles et distracteurs sont indiquées en bleu sur les graphiques (µV). Les sujets sont différents dans les deux tâches. **Détection de la couleur :** les potentiels cibles et distracteurs se séparent à 120 ms, mais cette séparation est clairement provoquée par la différence couleur vs N&B montrée dans la figure précédente. **Détection du Noir et blanc :** l'effet de la couleur est inverse, et le potentiel de différence ne devient positif que vers 170 ms. **Tâches mélangées :** l'effet de la couleur se compense dans les potentiels cibles et distracteurs, et leur différence débute à 150 ms pour s'accroître linéairement. Cette différence est donc liée à un traitement fonction de l'adéquation du stimulus à la tâche.

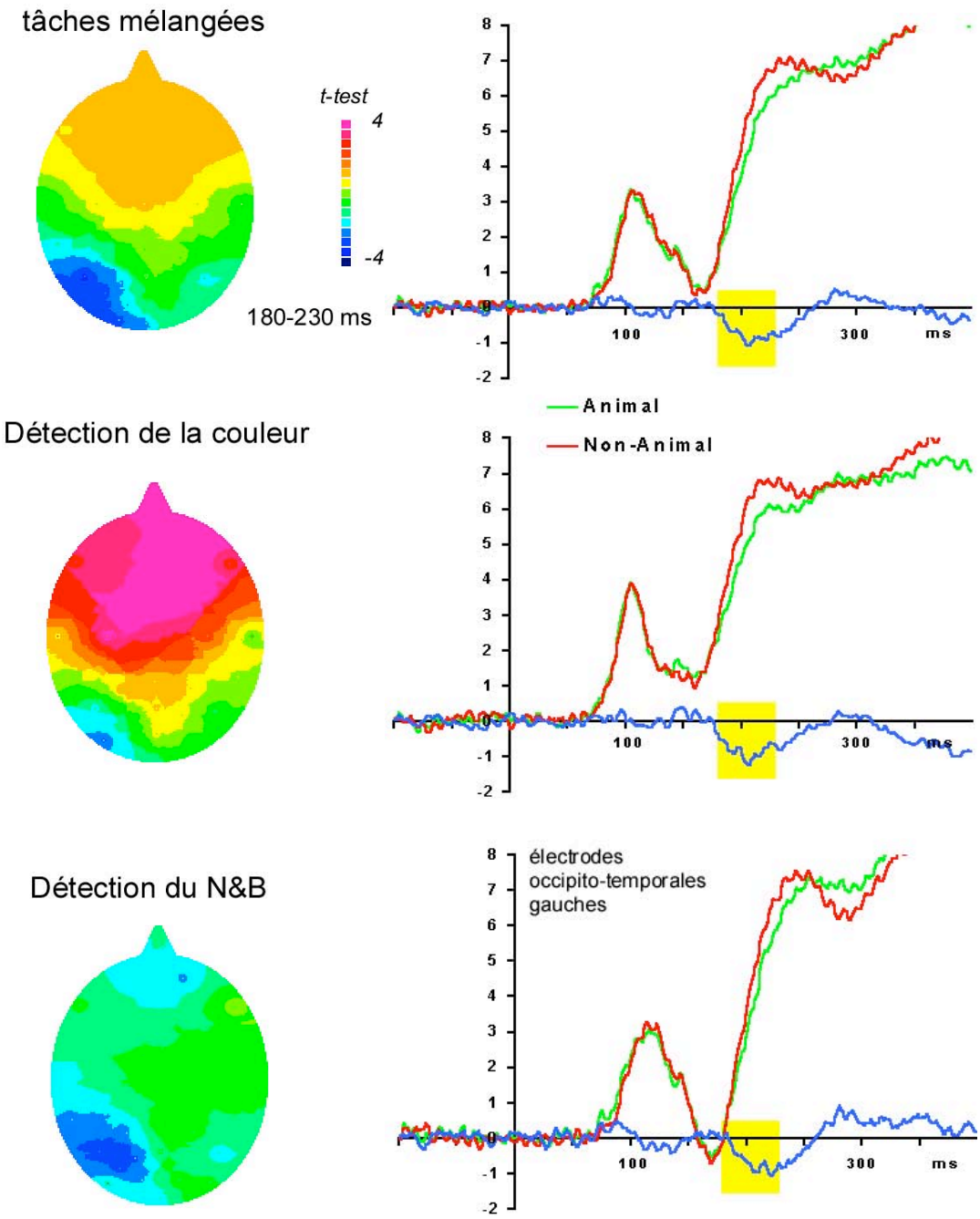


Figure 32 : Différences Animal – non Animal dans les deux tâches

Les sujets de cette expérience n'étaient pas avertis des expériences précédentes 'Animal / Non-animal'. Pour ces nouveaux sujets, les résultats comportementaux montrent que les tâches de détection de couleur et de noir et blanc n'étaient pas affectées par la présence d'animaux dans les scènes (Figures 27-29). Cependant, les potentiels évoqués par les scènes Animal et les autres scènes divergent nettement à partir de 180 ms, comme montré ici sur les électrodes occipito-temporales gauches. Cet effet est identique pour les deux populations de sept sujets engagés dans des tâches différentes, et semble indiquer un traitement automatique lié à la catégorie Animal. L'activité différentielle observée possède une topographie que l'on peut comparer avec l'activité différentielle standard ; elle est ici bien moins ample (1µV) et plus tardive (retard d'environ 30 ms). La catégorie Animal n'est pas pertinente dans cette tâche ; une réduction et un retard pour une cause similaire ont été mis en évidence dans le cas de la couleur (Figures 24 et 30).

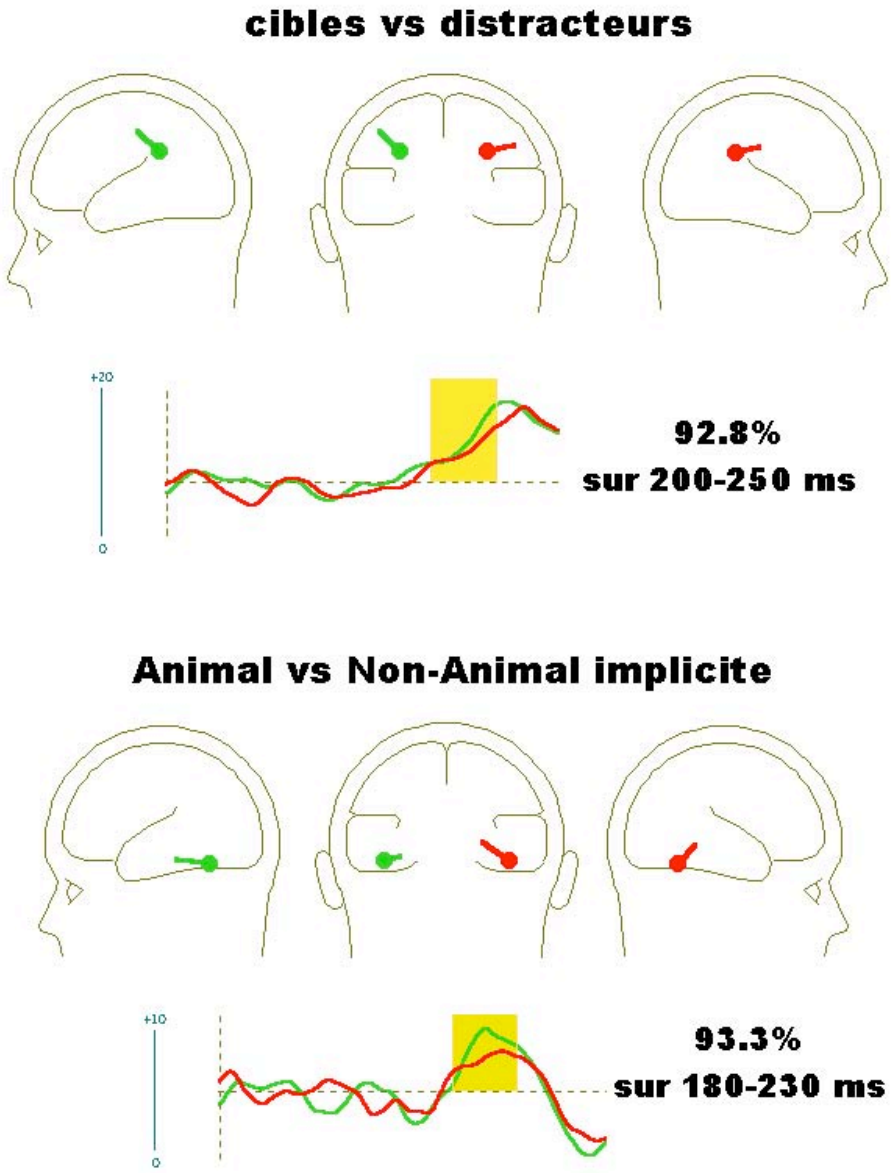


Figure 33 : Modèles des activités lors de l'expérience 'détection de la couleur'

Le pourcentage indiqué est celui de la variance du signal expliquée par le modèle sur la plage d'ajustement (en jaune). Les deux types d'activités ('cible – distracteur' et 'animal vs non animal implicite') ont été modélisés indépendamment, et amènent à des solutions stables. L'activité cible – distracteur de détection de la présence de couleur a une localisation plus pariétale que les activités cible – distracteur de détection de la présence d'animaux. L'activité différentielle 'animal implicite' se trouve être au même endroit (plancher temporo-occipital) que l'activité 'animal précoce' observée entre 110 et 140 ms dans l'expérience précédente ; elle est de polarité inverse.

3.2 discussions et conséquences

résumons

• Nous avons observé que des sujets impliqués dans des tâches de détection de couleur ne manifestent pas dans leur comportement une dépendance au contenu des images : les réponses aux stimuli contenant des animaux ne sont pas différentes des autres.

• Malgré cela, la présence d'animaux dans les stimuli induit une différence de potentiel significative, visible à partir de 150-170 ms, d'une topographie similaire à l'activité différentielle rapportée dans les tâches de reconnaissance d'animaux (très robuste sur les sites occipito-temporaux gauches). Cette activité est moins ample que dans les tâches où la catégorie animal est activement recherchée ; elle est davantage prononcée en réponse à des stimuli colorés. Elle semble être précédée d'une activité visible sur [110-140 ms] de distribution fronto-centrale, analogue à l'activité précoce observée dans l'expérience précédente.

• La présence de la couleur évoque une activité propre sur [115-145 ms], indépendamment du statut cible ou distracteur de cet attribut. Cette activité est identique dans les deux tâches (signe, latence, topographie). Elle est d'une topographie similaire à celle observée dans l'expérience précédente (chez d'autres sujets qui ne cherchaient pas la couleur), d'une latence avancée d'environ 20 ms.

• La différence cibles vs distracteurs indépendamment de la tâche est visible à partir de 150-170 ms. Cette activité possède des topographies analogues dans les deux tâches, indépendamment de la nature de la cible (couleur, N&B). Elle semble similaire à l'activité observée dans le cas de la détection de la catégorie animal dans les expériences précédentes.

a) l'encodage visuel et l'adéquation à la tâche

Cette expérience nous a permis de dissocier les effets de la coloration des stimuli et les effets des tâches. Les résultats montrent que l'effet de la coloration des stimuli est d'abord observable entre 115 et 145 ms, indépendamment du fait que la couleur soit cible ou non. Cet effet est donc lié à un encodage dans les aires visuelles des attributs de couleur, et compatible avec d'autres études (Tobimatsu et al. 1996; Anllo-Vento et al. 1998 ; Manolas et al. 1999 . Ce reflet électrophysiologique d'encodage de la couleur est d'amplitude identique, lorsque les sujets portent attention à cet attribut, à l'effet de la couleur observé dans l'expérience précédente dans laquelle la couleur n'était pas un indice visuel pertinent. Ce premier encodage semble donc indépendant, du point de vue de l'amplitude, du recrutement attentionnel lié à la tâche. Sa latence, par contre, semble être plus précoce lorsque la présence de cet attribut intervient dans la décision à prendre (Nobre et al. 1998). D'autres expériences sont cependant requises pour établir si la pertinence d'un

attribut modifie la latence à laquelle il est encodé par le système visuel ; nous pouvons noter toutefois qu'un même décalage est observé dans le recrutement des aires temporales lorsque la catégorie cible intervient dans la tâche à accomplir et lorsqu'elle en est indépendante.

La différence d'activité s'initiant à 150 ms est clairement liée au statut cible et distracteur de la tâche. Dans cette expérience, malgré l'opposition des statuts couleur et N&B chez les deux groupes de sujets, les potentiels frontaux évoqués par les distracteurs sont plus négatifs que pour les cibles. Cette propriété est identique à celle observée dans les expériences précédentes où la cible est la catégorie Animal, et dont l'activité possède une topographie proche. De même, les dernières expériences ont permis de localiser avec davantage de précision les sources des activités différentielles liées aux différences cibles-distracteurs : ces processus semblent provenir davantage d'aires situées à la jonction pariéto-temporale plutôt que temporales seules. Ces résultats montrent donc que ces activités sont probablement le reflet électrophysiologique de processus engageant des décisions comparables, lié à l'adéquation du stimulus présenté à la tâche à accomplir.

Les expériences récentes de Rufin van Rullen au laboratoire confirment cette dissociation entre encodage visuel et pertinence pour la tâche. Dans cette série expérimentale, les sujets doivent alternativement décider de la présence d'animaux dans l'image, ou de la présence de moyens de transport ; les images de la catégorie cible deviennent stimuli distracteurs lors de la série suivante. Les potentiels évoqués lors des deux tâches de catégorisation mélangées montrent des différences cible vs distracteur à partir de 150 ms, comme les activités présentées ici, bien que les stimuli soient identiques dans les deux conditions. Le statut commun cible ou distracteur peut donc seul expliquer les différences observées. Par contre, des différences d'activités clairement significatives entre 110 et 140 ms séparent les catégories 'animal' et 'moyens de transport', montrant que l'encodage visuel de ces catégories s'effectue à ces latences par des populations neuronales différentes. Cet effet d'encodage des catégories visuelles avant 150 ms confirme les observations que nous avons effectuées précédemment, sur l'existence d'une différence 'Animal – non Animal' entre 110 et 140 ms, davantage liée, d'après ce qui précède, à un encodage différent entre les scènes animaux et les autres scènes qu'à une décision liée à la tâche.

Ainsi la nature de l'activité différentielle déclinée au long de cette étude est un phénomène lié à la pertinence des stimuli pour la tâche, et pas *seulement* un encodage visuel des stimuli. Cet aspect met d'autant plus en relief la rapidité du système visuel dans des traitements complexes, puisqu'une décision sur le statut du stimulus en rapport avec la tâche ne peut être pris qu'une fois la catégorie cible reconnue. Comme nous allons le discuter maintenant, les processus de décision s'appuient probablement sur les bases cérébrales de cette reconnaissance de haut-niveau.

b) une reconnaissance implicite de la catégorie animal

Nous avons vu que la présence d'animaux dans les stimuli induit une différence de potentiel visible à partir de 150-170 ms. Cet effet peut-il provenir d'un mécanisme d'amorçage lié à la répétition de la catégorie animal lors de l'expérience ? En effet, le nombre de photographies contenant un animal est égal au nombre d'autres scènes, et des effets d'amorçage ont été montrés dans de nombreuses études en potentiels évoqués (Boddy et Weinberg 1981 ; Leiphart et al. 1993 ; George et al. 1997; Paller et Gross 1998). Les études d'imagerie sur l'amorçage perceptif lié à la répétition (*repetition priming*) montrent l'activation des mêmes aires cérébrales lors de l'amorçage que lors de tâches de reconnaissance (Buckner et Koutstaal 1998 ; Schacter et Buckner 1998; Badgaiyan et al. 1999). Ainsi dans notre tâche, si la répétition de la catégorie amorce les représentations d'animaux, cet amorçage est indissociable d'un phénomène de reconnaissance implicite, et lui est probablement identique (voir Musen et Treisman 1990; Treisman 1992; DeSchepper et Treisman 1996). La latence de cet effet est comparable aux latences observées lors des expériences précédentes où d'autres sujets entraînés cherchent explicitement la catégorie animal dans les images ; ceci semble souligner le caractère automatique de cette reconnaissance. Dans tous les cas, il démontre que l'attention visuelle n'est pas indispensable aux mécanismes de reconnaissance, confirmant les indices accumulés au long de cette étude.

Le reflet électrophysiologique de cette reconnaissance implicite possède une topographie similaire à l'activité différentielle rapportée dans les tâches de reconnaissance d'animaux ; il est donc probable que les bases cérébrales de ces deux phénomènes soient en parties communes. Cependant, contrairement à l'encodage visuel précoce observé pour la couleur et les animaux entre 110 et 140 ms, l'activité liée à la reconnaissance est beaucoup plus ample lorsque la catégorie animal est activement recherchée. Cette plus grande amplitude peut être liée à l'attention visuelle portée sur la catégorie cible, ou à la décision comme mentionné précédemment. Nous pouvons penser que l'activité liée à la reconnaissance visuelle de la catégorie est fortement renforcée lorsque cette catégorie est pertinente pour la tâche. Dans tous les cas, la conséquence est que la 'décision neuronale', indiquant l'adéquation du stimulus à la tâche, partage les mêmes bases cérébrales que la reconnaissance visuelle de la catégorie (Schyns et Oliva 1999).

Ces résultats sont compatibles avec des études récentes sur la catégorisation perceptuelle ou reconnaissance de 'type', lorsqu'une information perceptive entre en correspondance avec une représentation d'objet en mémoire, à partir de 150 ms (Bentin et al. 1996 ; Schendan et al. 1997).

4 images contenant des formes simples

Nous avons vu que l'activité différentielle 'Animal – Non Animal' à 150 ms provenait de deux phénomènes liés : d'une part, une réponse des structures cérébrales à la présence de la catégorie 'Animal' (mise en évidence par l'encodage automatique de cette catégorie quand les sujets sont engagés dans une autre tâche), et d'autre part une réponse des structures cérébrales au *statut* cible et distracteur (visible lorsque les stimuli sont contrebalancés dans deux tâches où les cibles de l'une sont distracteurs de l'autre). Ainsi, la rapidité de cette tâche de catégorisation est d'autant plus marquante, puisqu'à 150 ms un indice de la décision à prendre est déjà visible.

Nous allons compléter ici cette série d'études en nous demandant si la même tâche effectuée sur des formes géométriques simples permet de conclure à la même vitesse de *décision.* Les formes à discriminer seront des carrés et des ronds. Pour ces formes, des éléments de décision peuvent être disponibles dès les aires visuelles primaires, étant donné la présence de cellules répondant préférentiellement à des angles dans leur champ récepteur, pouvant fournir des informations fiables sur la présence de carrés.

4.1 expérience et résultats

a) protocole et hypothèses

17 sujets (9 femmes) ont effectué au cours des mêmes séances la catégorisation de photographies 'Animal vs Non Animal' en tâche contrôle, et une tâche de catégorisation d'images 'Carrés vs Ronds'. Les deux tâches étaient effectuées alternativement, par séries de cent images chacune. Il était demandé aux sujets un minimum de sept séries de chaque. La tâche 'Animal' était inchangée par rapport aux autres expériences. Sur les 17 sujets, 15 participaient à l'expérience pour la première fois.

Lors de la tâche 'Carrés vs Ronds', les sujets avaient pour consigne de relâcher le bouton d'une souris le plus rapidement possible dès que l'image affichée contenait des carrés. Les stimuli étaient des images achromatiques sur fond gris moyen, contenant aléatoirement soit des carrés soit des cercles. Pour éviter que la tâche se réduise à une détection simple, la disposition aléatoire des formes dans l'image changeait à chaque essai et les contours étaient remplis aléatoirement d'un niveau de gris (voir Figure 34). Les dix formes affichées étaient de même surface, mais pouvaient être masquées les unes par les autres. Les autres caractéristiques du protocole restaient inchangées : présentations centrées sur le point de fixation, affichage 20 ms, intervalle interstimulus aléatoire entre 1.8 et 2.5 s, stimuli de 384 x 256 pixels soit 6.7° x 4.5° d'angle de vue à 1 m. L'enregistrement de l'EEG était effectué au moyen d'un bonnet 32 voies ; seuls les essais réussis et non-artefactés étaient utilisés pour le calcul des potentiels évoqués moyens.

hypothèses du protocole

• S'agissant d'images contenant des formes simples au lieu de scènes naturelles complexes contenant des animaux, nous cherchons préalablement à vérifier si la tâche 'Carrés vs Ronds' est effectuée plus rapidement que la tâche 'Animal vs non-Animal'.

• La complexité de cette tâche est moindre, au regard des informations à traiter pour décider de la présence d'un carré. Cette information peut être disponible dès le cortex visuel primaire, ce qui n'est pas le cas pour la tâche 'Animal' : l'analyse des potentiels évoqués dans les deux tâches, enregistrés chez les mêmes sujets, permettra d'observer si un indice électrophysiologique de décision « cible ou distracteur » est décelable plus tôt que dans la tâche 'Animal'.

• De manière similaire, nous tâcherons d'observer si un encodage visuel des stimuli peut être décelé plus tôt, montrant alors l'implication d'une aire visuelle plus précoce de la voie ventrale.

b) résultats comportementaux

La performance des sujets est bien meilleure lors de la tâche 'Carrés vs Ronds' que dans la tâche 'Animal' : les taux de réussite sont de 96,6% et 91,6% respectivement, et les temps de réaction médians de 444 ms et 483 ms (moyennes 454 et 496 ms, voir Figure 36). Ces différences entre les résultats comportementaux des deux tâches sont significatives à $p<0.0001$ (*t* apparié). L'histogramme des temps de réaction (Figure 35) illustre le biais des réponses 'go' correctes par rapport aux réponses 'no-go' incorrectes ($p<0.001$) à partir de 300-310 ms pour la tâche 'Carrés vs Ronds', et à 330-340 ms pour la tâche 'Animal', pour l'ensemble des 17 sujets.

Comparée aux expériences précédentes, la tâche 'Animal' a cependant été moins réussie et plus lente[1] : les résultats précédents étaient de 445 ms, 417 ms et 441 ms en médiane, et 270 ms, 280 ms et 290 ms pour les temps de réaction les plus courts. Les sujets de cette expérience ont donc été bien moins performants. D'autre part, un examen plus précis des histogrammes de temps de réaction pour les essais les plus rapides (Figure 35) montre que le décalage des histogrammes observé dans la tâche précédente par exemple (voir Figure 26) est bien moins clair ici. Les temps de réaction sont donc globalement plus rapides de 40 ms dans la tâche 'Carrés vs Ronds' mais cette différence n'est pas aussi clairement visible pour les essais les plus rapides. Ces résultats demandent donc à être confirmés, par des sujets dont les résultats comportementaux seraient plus proche de ceux observés précédemment.

[1] Tous les sujets sauf un participaient à ce type d'expérience pour la première fois.

c) résultats électrophysiologiques

potentiels évoqués

Les potentiels évoqués par les stimuli simples ont une allure identique à ceux évoqués par les scènes jusqu'au premier pic négatif à 100 ms (Figure 37 A) visible sur les sites occipitaux. Les formes simples évoquent ensuite un deuxième pic d'activité à 130 ms, absent dans le cas des scènes. Ce premier résultat indique qu'un traitement supplémentaire est effectué pour les stimuli simples entre 100 et 150 ms ; cette activité ne permet pas de conclure à un traitement différents pour carrés et ronds.

différences cibles – distracteurs

Les cibles et distracteurs des deux tâches divergent nettement et de manière identique à partir de 150 ms, comme représenté sur la Figure 37 B. Ces activités différentielles restent identiques jusqu'à 220 ms, latence à partir de laquelle les stimuli simples évoquent des différences de plus grande amplitude tant sur les sites occipitaux que frontaux. Si la latence et l'amplitude de ces deux activités différentielles sont identiques, leur topographie mesurée sur 170-220 ms, bien que proche, n'est pas similaire sur tous les sites (Figure 38). Si, comme il a été suggéré lors de l'expérience précédente, ces activités différentielles reflètent un processus de décision, ce dernier résultat indique que les décisions 'cibles vs distracteurs' sont l'œuvre de processus de même latence et similaires sur 150-220 ms, qui ne partagent toutefois pas des bases cérébrales identiques pour 'carrés vs ronds' et 'animal vs non animal'.

activités précoces

Nous avons vu qu'aucune différence 'carrés vs ronds' n'était observable avant 150 ms. Par contre, comme indiqué sur la Figure 38 B et C, une différence 'Animal – non animal' est clairement significative sur 110-140 ms. Chez ces nouveaux sujets, cette activité possède un signe, une latence et une amplitude comparable à celles observées dans les expériences précédentes pour 'Animal vs Non-Animal précoce' ; sa topographie est déplacée légèrement en frontal. La présence de cette activité confirme qu'un encodage différent pour deux types de scènes peut être observé à ces latences, visible chez de nouveaux sujets.

d) modèles dipolaires

Nous avons ici cherché à vérifier les localisations des sources pouvant refléter les activités successives liées à la catégorie animal. La Figure 39 montre que les résultats obtenus chez ces sujets sont consistants avec les localisations précédemment obtenues : l'activité précoce modélisée sur 110-140 ms semble une nouvelle fois résulter de processus clairement temporaux, alors que l'activité différentielle à partir de 150 ms est plus antérieure et plus pariétale.

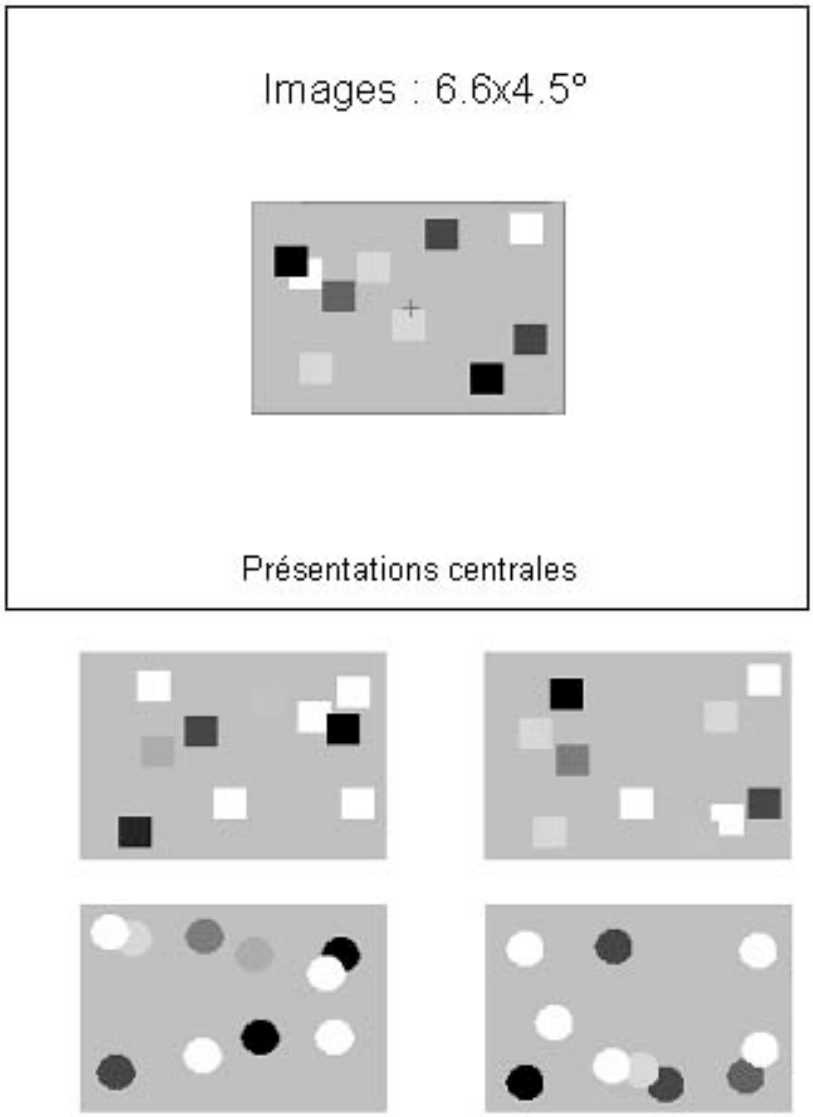


Figure 34 : Stimuli de la tâche 'Carrés vs Ronds'

Les stimuli sont constitués d'images achromatiques sur fond gris moyen, contenant aléatoirement soit des carrés soit des cercles au nombre de dix, disposés et colorés aléatoirement d'un niveau de gris à chaque essai. Les formes sont de même surface, mais pouvaient être masquées les unes par les autres. Les présentations sont centrées sur le point de fixation, et sont de même taille que les stimuli constitués de scènes naturelles. L'affichage dure 20 ms ; les cibles sont les carrés.

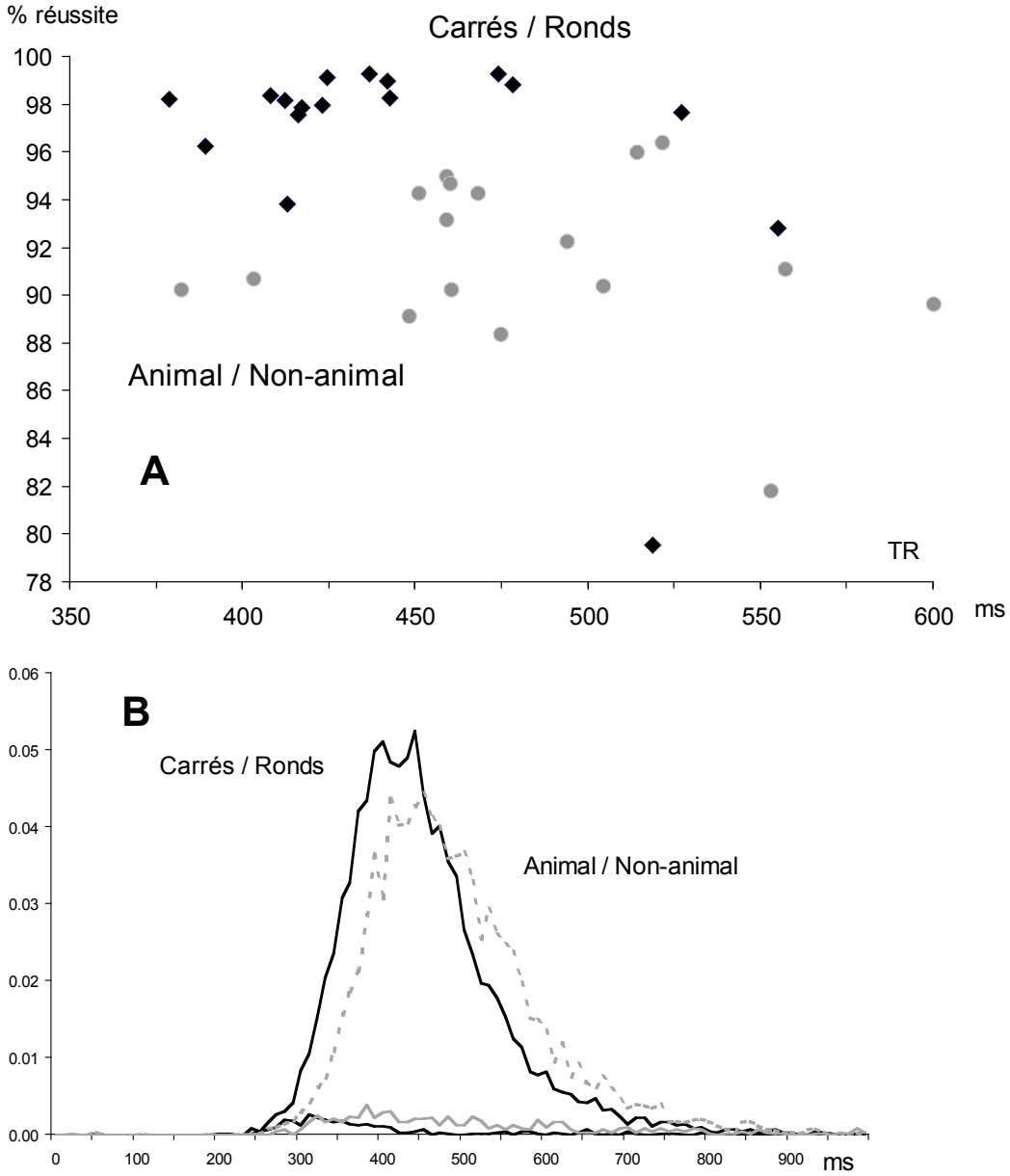


Figure 35 : Résultats comportementaux de l'expérience 'formes simples'

A. Résultats individuels dans les deux tâches. La tâche Carrés / Ronds est toujours mieux réussie et plus rapide que la tâche Animal (en moyenne, le taux de réussite est de 96,6% et 91,7%, temps de réaction médian de 444 et 482 ms, $p<0.001$ dans les deux cas). Les temps de réaction ne montrent pas de tendance linéaire avec le taux de réussite ($r=0.2$, et $r=0.04$ pour la tâche Animal).

B. Histogrammes des temps de réactions. En plus du décalage des temps de réaction de 40 ms (38 ms en médiane, 42 ms en moyenne), l'histogramme de la tâche Animal est plus étalé vers les temps de réaction longs (en moyenne, les écarts types sont de 81 et 89 ms, $p=0.0015$). Les différences entre les essais 'go' corrects et incorrects sont significatives à $p<0.001$ à partir de 300 ms dans la tâche Carrés / Ronds et 330 ms dans la tâche Animal pour cet ensemble de 17 sujets.

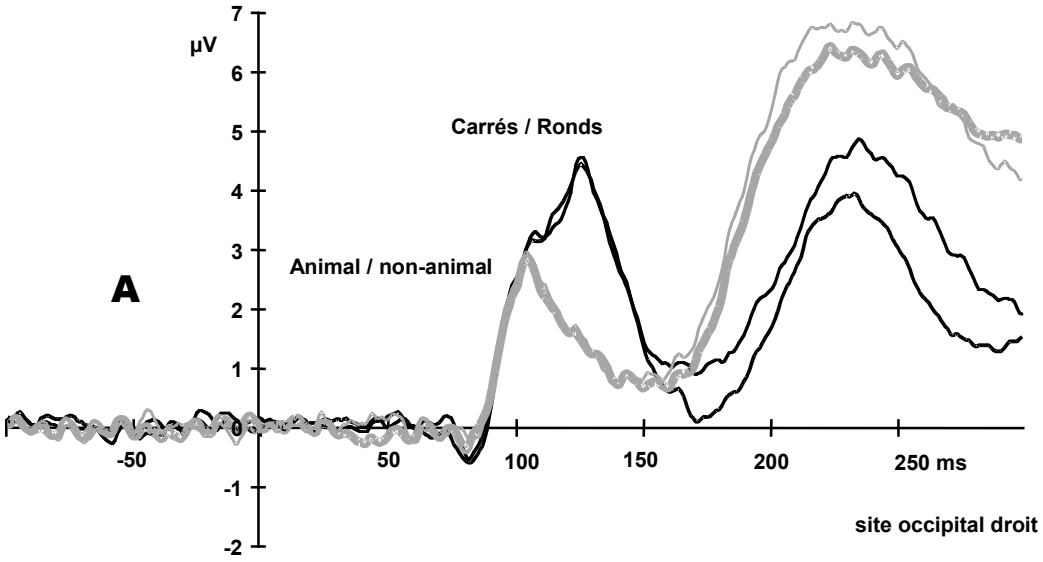
	Carrés / Ronds				Animal / Non-Animal			
	Moyenne	Médiane	% Correct	écartT	Moyenne	Médiane	% Correct	écartT
Sujet 02	385	379	98,3	58	387	382	90,3	57
Sujet 06	537	527	97,7	87	568	557	91,1	101
Sujet 08	418	408	98,4	84	475	461	90,3	105
Sujet 09	430	425	99,1	64	482	475	88,4	85
Sujet 10	430	423	98,0	68	481	459	93,2	87
Sujet 12	452	443	98,3	73	514	505	90,4	87
Sujet 14	451	442	99,0	68	484	468	94,3	79
Sujet 15	400	389	96,3	64	415	403	90,7	74
Sujet 17	420	413	93,9	80	459	448	89,1	84
Sujet 18	438	437	99,3	65	472	459	95,0	71
Sujet 19	549	519	79,6	148	570	553	81,8	138
Sujet 20	432	417	97,9	84	502	494	92,3	86
Sujet 21	485	474	99,3	66	535	522	96,4	77
Sujet 22	415	412	98,2	63	465	451	94,3	80
Sujet 23	565	555	92,8	124	617	600	89,7	119
Sujet 24	422	416	97,6	80	475	460	94,7	89
Sujet 25	495	478	98,9	96	533	514	96,0	96
moyenne	**454**	**444**	**96,6**	**81**	**496**	**483**	**91,6**	**89**

t-test appariés

	écart	t	p
Moyennes	-42	-9,766	<,0001
Médianes	-39	-9,102	<,0001
Ecart-types	-8	-3,818	0,0015
% réussite	5,0	7,051	<,0001

Figure 36 : Résultats individuels des tâches Carrés / Ronds et Animal / Non-Animal

La tâche Carrés / Ronds est plus rapide et mieux réussie ; elle se différencie aussi de la tâche Animal par une distribution plus resserrée autour de la moyenne (écart-type plus faible), visible sur les histogrammes. Les différences entre les deux tâches vont dans le même sens pour l'ensemble des sujets, sauf pour le sujet 19 pour lequel les temps de réaction sont lents et la réussite faible. Les moyennes, médianes et écart-types sont mesurés en millisecondes post-stimulus.



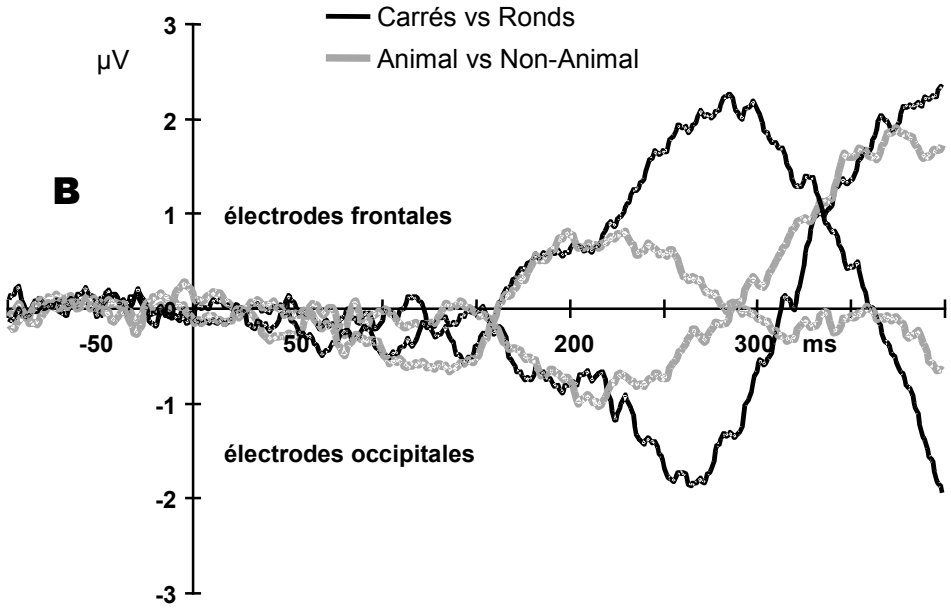


Figure 37 : Potentiels évoqués de l'expérience 'formes simples'

A. Potentiels évoqués occipitaux. Les lignes épaisses représentes les potentiels évoqués par les distracteurs, les lignes fines les potentiels évoqués par les cibles. Les potentiels évoqués par les stimuli simples (Carrés et Ronds) ont une allure identique à ceux évoqués par les scènes jusqu'au premier pic négatif à 100 ms. Les formes simples évoquent ensuite un deuxième pic d'activité à 130 ms, absent dans le cas des scènes. Ce premier résultat indique qu'un traitement supplémentaire est effectué pour les stimuli simples entre 100 et 150 ms. Aucune différence de traitement n'est visible entre carrés et ronds dans cette période.

B. Différences cibles- distracteurs. Les cibles et distracteurs des deux tâches divergent nettement, à la même latence (à partir de 150 ms). Ces activités différentielles restent identiques jusqu'à 220 ms, latence à partir de laquelle les stimuli simples évoquent des différences de plus grande amplitude, tant sur les sites occipitaux que frontaux.

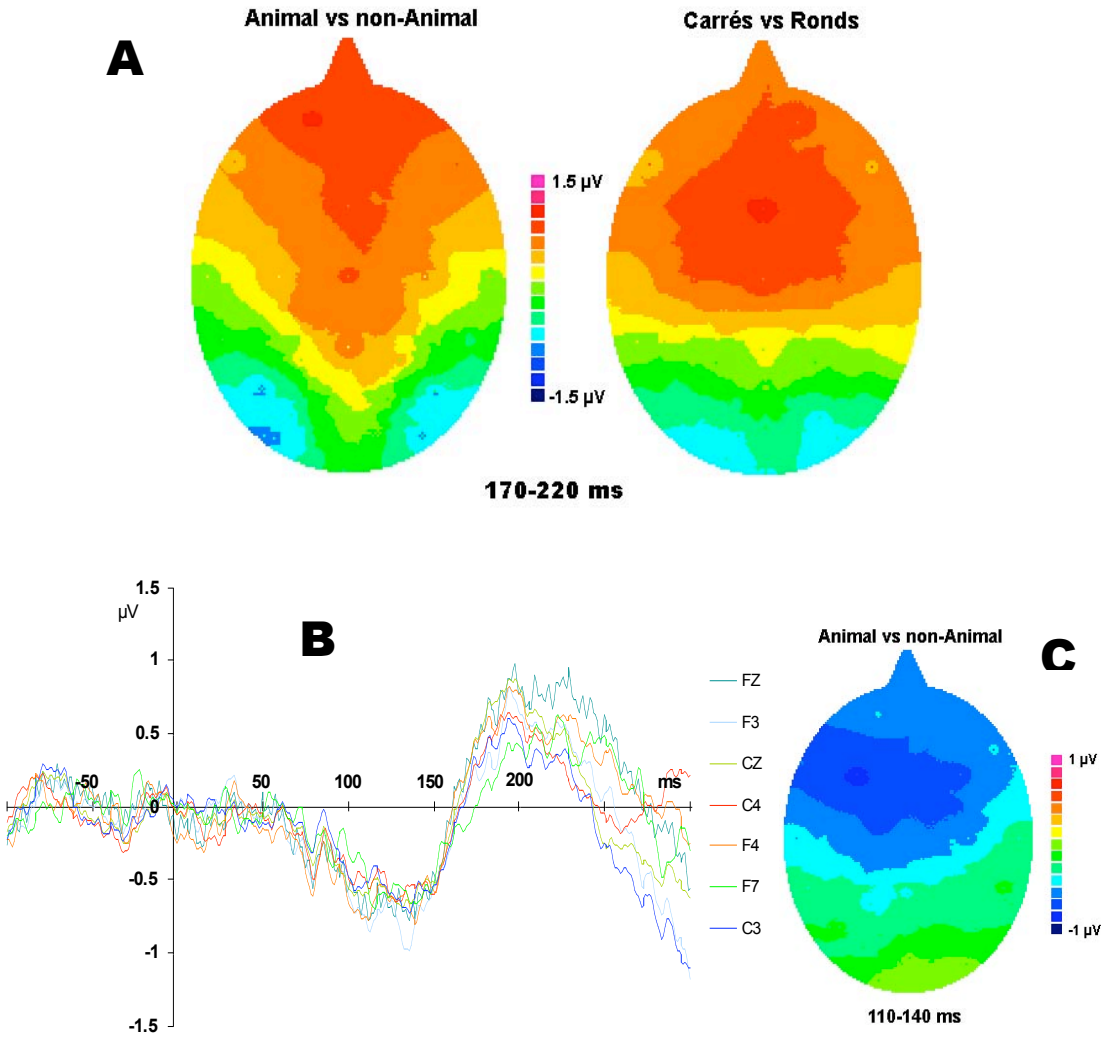


Figure 38 : Topographie des activités observées dans l'expérience 'formes simples'

A. Topographie des différences cibles – distracteurs dans les deux tâches. Si la latence et l'amplitude des activités différentielle 'Animal vs Non-Animal' et 'Carrés vs Ronds' sont identiques, leur topographie est proche mais distinctes. Les processus différenciant les cibles des distracteurs ne semblent donc pas partager les mêmes bases cérébrales dans les deux tâches.

B et C. Activité 'Animal vs Non Animal précoce'. En (**B**) sont représentées les différences 'Animal vs non-Animal' sur les électrodes où cette activité est clairement visible avant 150 ms. **C :** cartes du t-test apparié de 'Animal vs non-Animal' sur 110-140 ms pour chaque sujet. Cette activité est clairement significative, de topographie et d'amplitude comparable à celles observées dans les expériences précédentes. Elle est cependant plus frontale que pariétale, mais la localisation des sources de cette activité chez ces nouveaux sujets est cependant identique à l'activité précoce déjà observée (figure suivante). Aucune différence 'carrés vs ronds' n'est observable avant 150 ms.

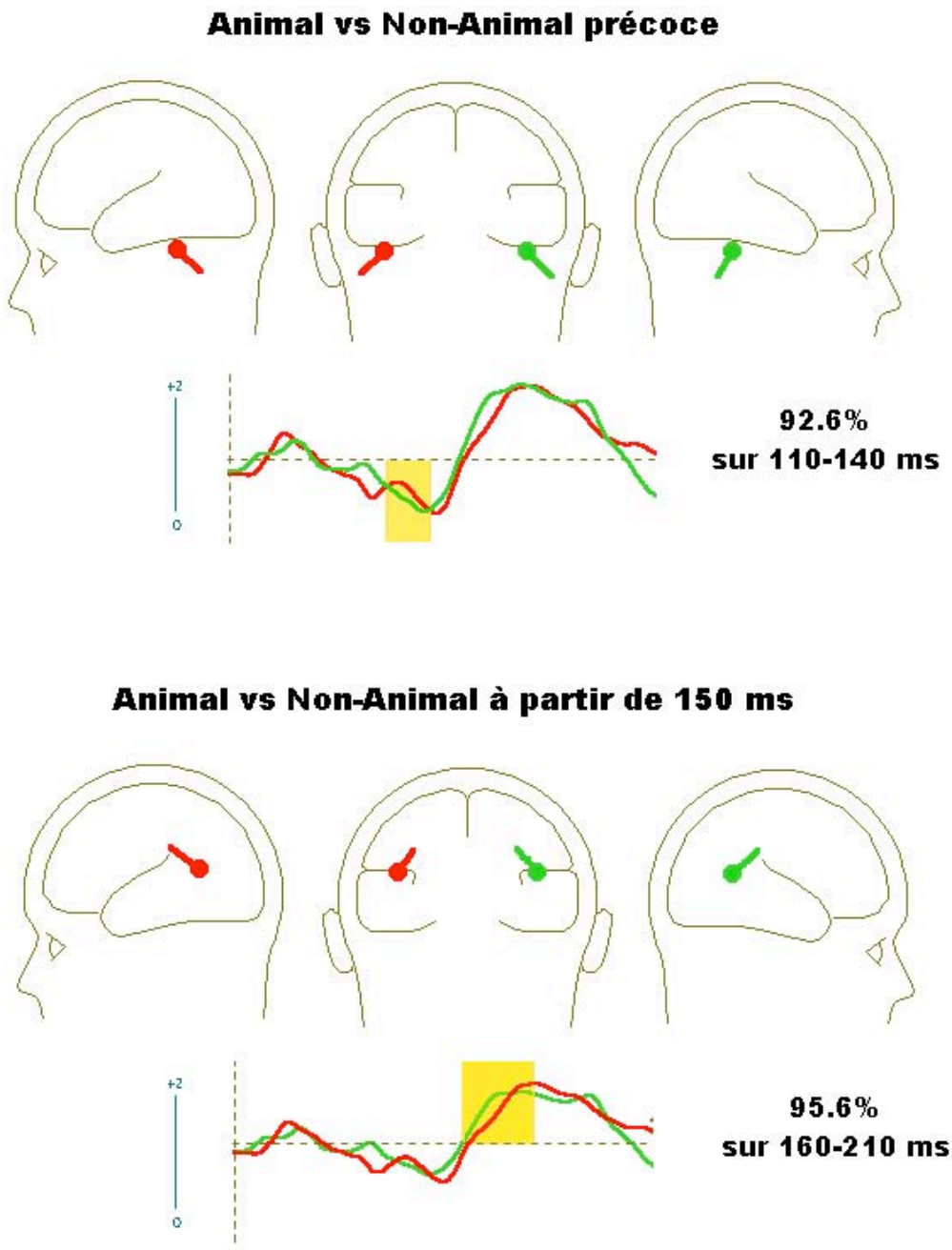


Figure 39 : Modèles des activités lors de l'expérience 'formes simples'

Sources de l'activité 'Animal précoce'. Malgré la topographie de l'activité 'Animal vs Non-Animal' à 110-140 ms déplacée en frontal, la localisation des sources est compatible avec les localisations précédemment obtenues : l'activité précoce modélisée sur 110-140 ms résulte de processus temporaux.

Sources de l'activité différentielle à partir de 150 ms. La localisation des sources de l'activité différentielle principale est ici légèrement plus haute, mais compatible avec celles observées dans les expériences précédentes.

4.2 discussions et conséquences

résumons

• La tâche 'Carrés vs Ronds' est mieux réussie (5%) que la tâche 'Animal' et effectuée plus rapidement (40ms), sans que cette rapidité n'apparaisse aussi clairement sur les essais les plus rapides.

• Les traitements visuels des stimuli simples et des scènes complexes sont clairement différentiés à partir de 100 ms.

• Malgré cela, les activités liées aux cibles et aux distracteurs n'apparaissent qu'à partir de 150 ms, et ce de manière identique dans les deux tâches jusqu'à 210 ms, malgré des stimuli fondamentalement différents. Ces activités semblent résulter des mêmes processus de décision cibles vs distracteurs, quelle que soit la complexité des stimuli.

• Une activité différentielle précoce [110-140 ms] pour 'animal vs non-animal' est observée comme lors des expériences précédentes ; elle en possède toutes les caractéristiques (amplitude, signe, latence, topographie, sources).

a) l'encodage précoce des formes simples et des scènes

La présence d'une activité électrophysiologique précoce entre 110 et 140 ms, différente pour les scènes contenant des animaux et celles n'en contenant pas, se confirme dans cette expérience. Les résultats des études précédentes ont suggéré que cette activité pouvait refléter un encodage visuel différent pour les deux types de scènes, sans lien direct avec la tâche à accomplir. Ce dernier point a été démontré par une étude de R. van Rullen précédemment cité. Cet encodage visuel différent pour des catégories visuelles ne s'observe cependant pas dans le cas d'images contenant des formes simples.

Cette absence d'encodage différent pour carrés et ronds reflète-t-elle une identité de catégorie visuelle entre les deux types de stimuli ? Nous croyons que la question se pose plutôt en termes de bases neuronales sous-tendant la représentation de ces formes : en effet, étant donné la physiologie du système visuel, la présence de formes simples peut être simplement encodée par des activités de neurones de toutes les aires visuelles (de V1 à IT et V4). Dans chacune de ces aires, de nombreux neurones peuvent répondre préférentiellement à ces formes, et les données disponibles (Fujita et al. 1992; Vogels et Orban 1993 ; Tanaka 1996) ne montrent pas que les colonnes corticales répondant à des formes différentes soient disposées en groupes différenciés. Ceci rend peu probable l'émergence à la surface du scalp de potentiels évoqués différents pour des formes telles que carrés et ronds.

Un second aspect distingue les deux tâches : autant la tâche 'carrés vs ronds' fait intervenir deux types de formes bien différenciées et des stimuli binaires (contenant soit des carrés, soit des ronds), la tâche 'animal vs non-animal' oppose une catégorie (visuelle ou sémantique) à un ensemble d'autres scènes dont le seul point commun est l'absence de cette catégorie. Ainsi, même si l'on peut douter d'une aire « spécialisée » dans l'encodage visuel d'animaux, il est certain que la présence d'un ensemble de traits visuels communs ou très probables dans cette catégorie (yeux, pattes, plumes etc.) permet d'impliquer des neurones et des aires corticales différentes en fonction des deux types de scènes. Une différence entre potentiels évoqués par les deux types de scènes est donc beaucoup plus probable dans la tâche 'animal' que dans le cas de la tâche 'formes simples'.

D'autre part une même dispersion est possible pour les neurones répondant à l'ensemble des traits animaux. Ceci nous engage à douter des bases cérébrales obtenues par modélisation dipolaire, pour les activités différentielles précoces 'animal'. Les colonnes corticales impliquées peuvent être en très grand nombre et dispersées dans des lieux éloignés (l'aire 31 pouvant être l'analogue de V4 chez l'homme est très grande (Sereno et al. 1995 ; Tootell et al. 1997), et d'une distance au cortex inféro-temporal très variable.

b) décision et comportement

Nous avons vu que le début des activités différentielles des deux tâches ne différaient pas en latence. Pourtant, les résultats comportementaux montrent que les images contenant des formes simples sont catégorisées plus rapidement en médiane. Comment concilier ces deux résultats apparemment contradictoires ?

Une possibilité est d'interpréter la décision de répondre cible contre distracteur comme un phénomène périodique, comme l'a développé S. Dehaene dans une étude comportementale (Dehaene 1993)**.** L'auteur propose que le traitement d'information ne soit pas un phénomène continu, et que l'accès à ces informations soit le fait d'un phénomène périodique. Cette proposition s'appuie sur un très grand nombre de données comportementales obtenues chez plusieurs sujets et dans différentes tâches. Les résultats de cette étude montrent que l'aspect indenté des histogrammes de temps de réaction est la juxtaposition d'un histogramme de type gaussien et d'une oscillation périodique ; cette oscillation possède une période propre pour chaque sujet d'environ 40 ms. Le point marquant de cette étude est que la fréquence de ces oscillations est pour chaque sujet une fonction des conditions expérimentales (tâche à accomplir et stimuli). L'auteur propose un modèle dans lequel l'accès de l'information perceptive pour décider de l'action à accomplir soit soumis à ces oscillations.

Nous proposons que ce phénomène soit considéré en regard du type d'information perceptive porté par le système visuel. Dans la tâche 'carrés vs ronds', les stimuli sont caractérisés par une grande cohérence de contenu : les formes sont clairement délimitées, et l'information de forme est redondante dans l'ensemble de l'image. Ainsi l'information

extraite par le système visuel est très probablement dénuée de toute ambiguïté sur le contenu des stimuli : la décision 'cible vs distracteur' de la tâche peut-être prise de manière certaine dès que le contenu de l'image est disponible pour ce traitement, c'est-à-dire dès le « premier accès » d'un traitement de type décisionnel à l'information visuelle redondante. A l'inverse, les scènes de la tâche 'animal vs non-animal' sont plus complexes et bien moins redondante, et peuvent porter des informations ambiguës sur la présence d'un animal dans l'image. La décision 'cible vs distracteur' de la tâche pourrait alors être prise rapidement dans des cas particulièrement aisés, mais généralement soumis à un effet de seuil ne pouvant être atteint que par « accumulation d'évidence » (Perrett et al. 1998) de neurones encodant l'animal présent dans l'image[1]. Cette hypothèse permet ainsi d'expliquer à la fois le décalage de 30 ms du temps de réaction médian, et l'ambiguïté observée sur les temps de réactions les plus rapides pour lesquels il n'est pas sûr qu'un tel décalage soit présent. Cette proposition est aussi compatible avec les résultats électrophysiologiques, qui montrent une latence identique pour le début des activités différentielles des deux tâches.

[1] Dans cette hypothèse, les temps de réactions les plus longs pourraient être la conséquence d'accès répétés aux informations visuelles, c'est-à-dire après un ou plusieurs cycles pouvant être la conséquence de boucles de retour (*feed-back*) dans les aires visuelles.

ETUDE EN IRMf

Dans une tâche de catégorisation, les essais cibles et distracteurs induisent une activité neuronale différente dès 150 ms. Cette activité est précédée par un encodage visuel de la catégorie recherchée. Afin d'avoir une idée plus précise des aires cérébrales impliquées, et peut-être d'en dégager d'autres mécanismes que ceux observés en électrophysiologie de surface, nous avons tenté d'étudier l'activité cérébrale au cours de l'exécution de cette tâche en effectuant l'acquisition d'images cérébrales par résonance magnétique nucléaire (RMN), grâce à l'outil médical IRM de l'hôpital Purpan. Cette technique permet d'acquérir des images volumiques du cerveau très précises par résonance du proton. Lors de tâches comportementales, le signal modifié par la présence de la déoxyhémoglobine permet d'obtenir un indice local du débit sanguin, donc indirectement de la consommation d'oxygène par les neurones (effet BOLD : *blood oxygen level dependent*). Il s'agit donc d'une mesure très indirecte des neurones impliqués dans la tâche, mais qui offre l'avantage d'une grande résolution spatiale. Cette étude a été menée avec la collaboration de l'unité 455 de l'INSERM, principalement avec K. Boulanouar et Y. Chatel pour le traitement et l'analyse des données avec le logiciel SPM, et J.P. Ranjeva pour le paramétrage des acquisitions et la mise au point de la synchronie avec les stimulations. Nous avons commencé cette étude en 1996 ; à cette époque, les études événementielles en IRMf étaient rares, et la mise au point de ce protocole nous a demandé beaucoup de temps. Nous présenterons, au lieu d'une description détaillée des résultats, un article sous presse relatant ces travaux.

1.1 à la recherche d'un protocole adapté

Une réalisation difficile par la méthode classique

De nombreuses études effectuées en IRM fonctionnelle utilisent une méthode de recueil des données « par blocs », où un seul paramètre change à chaque bloc de stimulation. Cette méthode, issue des études en TEP, a permis dans un premier temps de valider les première études en IRM fonctionnelle, puis elle a été conservée dans de nombreuses études comme paradigme de base. Si cette méthode éprouvée a montré son intérêt dans beaucoup de protocoles mettant en œuvre des tâches passives, ou révélant un contraste entre états attentionnels ou différentes conditions de stimulations visuelles, elle se trouve être pratiquement inutilisable lorsque le sujet doit accomplir des tâches actives de choix entre deux stimuli équiprobables, et que l'on cherche à contraster les réponses à ces deux stimulations (Figure 40).

Même si d'un point de vue statistique il est possible de varier le nombre de cibles et de distracteurs dans chaque bloc, et de mesurer une gradation de signal permettant de

conclure que telles aires corticales semblent suivre ces variations, le problème est rejeté en amont puisque introduire de tels biais modifie la tâche qu'on ne peut plus considérer comme identique dans chaque bloc. Cette méthodologie pose donc de grandes difficultés pour toute tâche de catégorisation, à tel point qu'elle a pu être considérée par Buckner et al. 1996 comme « la limitation majeure dans la conduite d'études cognitives de haut-niveau avec l'IRMf ».

Une méthode événementielle...

Directement inspirées des études en potentiels évoquées, les méthodes événementielles en IRMf (*event-related fMRI*) développées récemment permettent de concevoir des protocoles dans lesquels la stimulation est aléatoire (Schacter et al. 1997 ; Zarahn et al. 1997 ; Friston et al. 1998 ; D'Esposito et al. 1999). Ces méthodes ont suscité beaucoup d'intérêt ; leur principe est de synchroniser la présentation des stimuli avec une succession d'acquisitions IRM s'étalant sur une quinzaine de secondes, temps minimal nécessaire pour que l'hémodynamique d'une aire cérébrale activée retourne à un niveau basal (Figure 40). Outre la possibilité de mélanger les types de stimuli au long d'une même séquence de stimulation (*mixed trial designs*), c'est grâce à cette méthodologie que Savoy et al. 1995 ont détecté des changements de signal dans les aires visuelles à partir de stimulations d'une durée de seulement 34 ms.

Deux aspects essentiels de notre protocole peuvent donc être réalisés par cette technique : des stimuli présentés de façon équiprobables, et un temps de stimulation très bref.

...associée à un délai interstimulus court.

Le choix de mettre en œuvre une tâche de catégorisation identique à celle réalisée en potentiels évoqués impose un délai interstimulus court. Or nous avons vu que la réponse locale du flux sanguin à une stimulation est connue pour être à dynamique lente. En conséquence, des stimuli rapprochés verront leurs réponses hémodynamiques associées mélangée à celles des stimuli antérieurs et postérieurs, de telle manière que le signal mesuré ne permette plus de distinguer les réponses à différents essais. La variance du signal mesuré peut en effet être rendue très faible à cause d'un tel recouvrement, dès que le délai interstimulus est inférieur à 3s (Burock et al. 1998).

Il est probable cependant que ce recouvrement soit linéaire. En effet, quelques études ont montré l'additivité des réponses hémodynamiques sur plusieurs essais de même type (Boynton et al. 1996 ; Dale et Buckner 1997; Rosen et al. 1998). Il est à noter que cette linéarité est postulée dans les méthodologies par blocs, même si ces méthodologies s'avèrent plus robustes - parce que plus grossières - à des non-linéarités.

Cependant, quelle que soit la loi de recouvrement des réponses hémodynamiques successives, si une même aire est impliquée dans le traitement des deux types de stimuli,

mais avec une dynamique ou un niveau d'activité différents, ce recouvrement conduit à un niveau de signal dépendant de *l'ordre* de présentation des types de stimuli. Nous voyons donc que dans ce cas le plus défavorable, l'ordre des stimuli permet d'obtenir des indices sur l'implication d'une aire dans le traitement de chacun des types. Dans le cas plus favorable d'une aire impliquée dans le traitement d'un seul type de stimulus, le niveau de signal permettra aisément de la mettre en évidence, toujours à la condition d'une variance suffisante du niveau de recouvrement (Figure 41).

Conséquences

Lors de l'élaboration du protocole, nous avons donc dû modérer l'exigence d'un délai interstimulus de 2s pour permettre une variance suffisante du niveau de recouvrement. Nous l'avons fixé arbitrairement à 5s, considérant que ce ralentissement du flux d'image ne modifiait pas la nature de la tâche de catégorisation.

D'autre part, le recouvrement des hémodynamiques impliquant de maîtriser l'ordre d'apparition des stimuli, il importait que la séquence de stimulation contienne toutes les successions possibles de cibles et de distracteurs, ces successions étant représentées en nombre égal. Cette propriété est bien la définition du tirage aléatoire, mais cette propriété n'est atteinte qu'asymptotiquement ; s'il est classique d'utiliser une trentaine de tirages pour égaliser la probabilité d'occurrence de deux événements (pile et face, ou cible et distracteur), le nombre de tirages doit être beaucoup plus élevé pour égaliser la probabilité d'occurrence de quatre événements (par exemple TT, TD, DT et DD, dans lesquels cibles T et distracteurs D sont successifs). Pour cinq stimuli successifs, les événements de type TTDTD (deux cibles puis un distracteur suivis d'une cible et d'un distracteur) sont au nombre de 32. On voit qu'un tirage aléatoire seul ne pourrait égaliser la probabilité d'occurrence de ces 32 événements dans une séquence de stimulation qui resterait raisonnable pour la patience des sujets. Nous avons donc construit une séquence de stimulation possédant cette propriété jusqu'à 5 stimuli successifs (le nombre de TTDTD est égal au nombre de TTDDD etc.), c'est-à-dire contrôlée sur des étendues de 20s. Cette séquence de stimulation possède aussi les mêmes propriétés pour les ordres inférieurs 1,2,3 et 4 (c'est-à-dire que le nombre de T est égal au nombre de D ; mais aussi le nombre de TT est égal au nombre de DT etc. ; le nombre de TTD égale le nombre de DTD etc. ; idem pour TTDT etc. voir Figure 42).

Protocole

Cet ensemble de contraintes nous amène à mettre en œuvre le protocole suivant, très similaire à celui utilisé en potentiels évoqués (Figure 43).

- cibles et distracteurs sont présentés de manière imprévisible pour le sujet, en suivant cependant une séquence qui ne doit rien au hasard. Il nous semble impossible qu'un sujet humain puisse deviner la loi d'une telle séquence[1].

- le délai interstimulus est fixé à 5s. Il est invariant pour permettre un séquençage régulier du signal.

- les photographies couleur sont présentées en flash.

Il est demandé au sujet d'appuyer sur un bouton de souris le plus rapidement possible dès qu'un animal est présent dans l'image. L'oxygénation du sang est mesurée dans le cerveau de chaque sujet lors de 3 séries consécutives constituées chacune de 68 stimuli (env. 5 mn par série). Les photographies couleur sont projetées toutes les 5 s pendant 33 ms à partir de la sortie vidéo d'un Macintosh sur des lunettes a-magnétiques reliées par fibres optiques (montage de la société Resonance Technology Inc). Le champ visuel stimulé est de 30°.

Les temps de réaction sont enregistrés au travers de la souris a-magnétique liée au même Macintosh. L'imagerie cérébrale est réalisée par un système IRM à 1.5 Tesla (Siemens Magnetom Vision). Chaque volume cérébral acquis selon la technique échoplanar est composé de 12 coupes non-contiguës d'épaisseur 6mm (intervalle entre les coupes de 1.2mm ; TE 64ms, TR 2.5s, taille des images 128x128 interpolées, champ de vue 200mm). L'acquisition de chaque volume de 12 coupes prend environ 1.3s. Lors d'une série, les données sont régulièrement collectées chaque 2.5s, soit 1.5s et 4s après la présentation de chaque stimulus.

1.2 analyses et traitements des données

a) modélisation par système linéaire et statistique de corrélation

Modèle linéaire

Dans le domaine de l'imagerie fonctionnelle cérébrale, il est courant de définir un modèle *a-priori* de la réponse hémodynamique aux stimuli présentés, dont on recherche la corrélation avec les données expérimentales. Cette analyse par corrélation est largement éprouvée et utilisée dans les protocoles par blocs. Nous l'utiliserons dans le cas de notre protocole pour une première analyse de données, ce qui permettra de valider notre protocole avec un outil connu. C'est grâce à Kader Boulanouar et Yvan Chatel de l'unité INSERM 455 que nous avons manipulé le logiciel SPM développé par K. Friston (Friston et al. 1995), logiciel majoritairement utilisé pour ce type d'analyses.

[1] même si l'indice le plus visible de la construction de la séquence est qu'il n'existe pas de succession de plus de 5 stimuli de même type.

Le modèle linéaire de la réponse hémodynamique aux stimulations à été simplement défini en fonction du temps t par :

$$x(t) = \alpha\, c(t) + \beta\, d(t) \quad \text{pour un pixel donné}$$

où x(t) est la fonction à comparer avec les données mesurées, c(t) la réponse des aires activées par les stimulations cible au cours de la série d'acquisitions et d(t) les pixels répondant aux distracteurs.

Les fonctions c(t) et d(t) ont été définies comme la convolution d'une fonction de Poisson $P(\lambda)$ (représentant la réponse hémodynamique à une seule stimulation) à la séquence de stimulation modélisée par une suite de distributions de Dirac δ représentant un stimulus à un instant donné. Le choix d'une poissonnienne est habituel dans ce contexte ; λ choisi arbitrairement pour que la réponse hémodynamique atteigne son maximum à 6 s post stimulus (λ=6) dans un premier temps.

hypothèses du modèle

Comme on le voit, la mise en place de ce modèle suppose trois hypothèses :

1. la réponse hémodynamique à un stimulus est une Poisson de paramètre λ=6

2. le recouvrement des réponses à des stimuli de même type est linéaire (exprimé par la convolution)

3. le recouvrement des réponses à des stimuli de types différents est aussi linéaire (équation en t).

traitements et statistiques

Une fois défini ce modèle, l'analyse va simplement consister à déterminer les paramètres de corrélation α et β pour chaque pixel, établissant une carte statistique pour chacune des composantes (SPM). La carte des paramètres α est appelée contraste T-D, et représente les corrélations positives avec la séquence des cibles sur les pixels pour lesquels $\alpha>\beta$. La carte des paramètres β est appelée contraste D-T, et indique les pixels corrélés positivement avec la séquence des distracteurs, pixels pour lesquels $\beta>\alpha$.

Les données sont traitées au préalable en corrigeant les mouvement de tête des sujets pendant l'acquisition, par un algorithme de ré-alignement (estimation des moindres carrés dans plusieurs dimensions du déplacement par rapport à un volume de référence, et correction). Les volumes sont lissés spatialement (10 mm FWHM soit 4 mm d'écart type) pour atténuer l'importance des faux-positifs et homogénéiser les activations. Dans cette méthode, les valeurs d'activation sont ensuite sur-échantillonnées au taux du modèle, et les valeurs du modèle sont ajustées à basse fréquence (>120s) pour suivre les dérives lentes du signal, par hypothèse non reliées à la tâche. Les valeurs d'activation sont ensuite

normalisées linéairement à une valeur moyenne fixe (1000) ; les valeurs d'activation faibles, considérées comme du bruit, sont supprimées avant la corrélation. En outre, les données individuelles ont été préalablement normalisées spatialement (atlas stéréotaxique de Talairach et Turnoux) pour permettre des études de groupe. Le lissage temporel et spatial permet d'utiliser la théorie des champs gaussien, qui fournit des outils de correction aux comparaisons multiples.

b) comparaisons de moyennes

Du point de vue des données

Nous avons cherché à mettre en œuvre une autre méthode qui puisse analyser les données avec un minimum d'*a-priori.* En effet, malgré l'implantation particulièrement ingénieuse de la méthode par corrélation dans le logiciel SPM, cette méthode modélise ce que l'on connaît de l'hémodynamique, et la façon dont elle est couplée à la séquence expérimentale. Cette méthode nous permet donc d'observer *quelles aires cérébrales réagissent de la manière que nous attendons*. Notre expérience étant exploratoire, et néanmoins très proche de celle menée en potentiels évoqués, nous avons jugé opportun d'analyser les volumes IRM de manière analogue aux données électrophysiologiques.

L'idée est la suivante : au lieu de considérer les valeurs d'activation des volumes IRM comme des séries chronologiques, nous comparerons directement ces valeurs en fonction des stimuli qui les ont précédés. Chaque volume acquis est donc typé en fonction des stimuli précédents, et l'effet différentiel des stimulations est mesuré sur l'ensemble des volumes sélectionnés. L'analogie avec la méthode d'analyse mise en œuvre en potentiels évoqués est directe puisque nous pouvons comparer des effets cibles vs distracteurs sur plusieurs volumes successifs, et nous restreindre aux essais réussis par les sujets.

comparaisons de deux populations

Le mécanisme de base est le suivant : nous disposons de deux mesures d'activation (1 et 2) après chaque stimulus (de type cible **T** ou distracteur **D**) ; nous pouvons mesurer la différence d'activation entre cibles et distracteurs en sélectionnant par exemple toutes les acquisitions 2 qui ont été immédiatement précédées par une cible réussie, et les comparer aux acquisitions 2 précédées d'un distracteur réussi. Cette comparaison (notée **T2 vs D2**) pourra être réalisée par une statistique d'estimation comme le test t de Student sur les moyennes, ou par la statistique de rang comme le test de Mann-Witney. La Figure 43 montre que cette méthode permet un large éventail de comparaisons.

hypothèse de la méthode

Si la comparaison T2 vs D2 ci-dessus mesure l'effet du type du dernier stimulus sur l'acquisition qui lui a succédé 4s après, cette même acquisition est encore influencée par la réponse hémodynamique aux stimuli précédents. Or nous avons sélectionné les volumes de

type **....T2** vs **....D2** où « **.** » est de type quelconque ; sachant que la séquence de stimulation a été conçue pour qu'il y ait dans ces « **.** » autant de cibles que de distracteurs, l'hypothèse de la comparaison est que l'effet de ces stimuli sur l'acquisition 2 est du même ordre dans les deux ensembles, c'est-à-dire qu'ils ne sont pas significativement différents. Il en est de même dans toute comparaison effectuée par cette approche, par exemple dans la comparaison **TT1 vs DT1**, où l'on cherche à mesurer l'effet sur les acquisitions **1**, du type de stimulus présenté 9s avant et en contraignant le dernier stimulus à être une cible. L'hypothèse est ici que ce dernier stimulus cible influera de la même façon l'acquisition **1** dans les deux cas.

En résumé, avec cette approche nous postulons que les réponses hémodynamiques aux stimuli non pertinents pour la comparaison ne sont pas significativement différentes dans les deux ensembles sélectionnés. Cette hypothèse est permise par la séquence de stimulation que nous avons construite, qui nous permet d'affirmer que ces stimuli non-pertinents sont en nombre égal dans les deux conditions. Nous voyons que l'hypothèse de linéarité postulée par la méthode précédente exprime la même contrainte tout en étant plus restrictive sur son type (additif).

traitement des données et statistiques

Nous avons utilisé cette méthode d'analyse indépendamment de la méthode précédente. Nous nous sommes procurés le logiciel d'imagerie AFNI (Analyses of Functional NeuroImages) développé par R. Ward et R. Cox pour le Medical College of Wisconsin, qui permet d'utiliser les tests de Student et de Mann-Whitney sur des volumes cérébraux, en plus des méthodes classiques (Cox 1996). Partant des données extraites du système IRM, le pré-traitement des données a consisté ici en un ré-alignement des volumes par rapport à un volume de référence moyen, à la suppression des valeurs d'activation faibles considérées comme du bruit, puis sur chaque coupe à un lissage gaussien d'écart type 4 mm.

Le test non-paramétrique de Mann-Whitney ne pouvant être réalisé sur un grand nombre de volumes, nous avons au préalable vérifié que ce test et le t de Student donnaient les même résultats sur quelques séries expérimentales. Nous avons ensuite opté pour le t-test sur l'ensemble des 3 séries expérimentales effectuées par chaque sujet. Il est à noter que le résultat des t-tests effectués n'est pas corrigé pour les comparaisons multiples ; à notre connaissance, aucune étude statistique n'a été effectuée en ce sens.

1.3 article sous presse

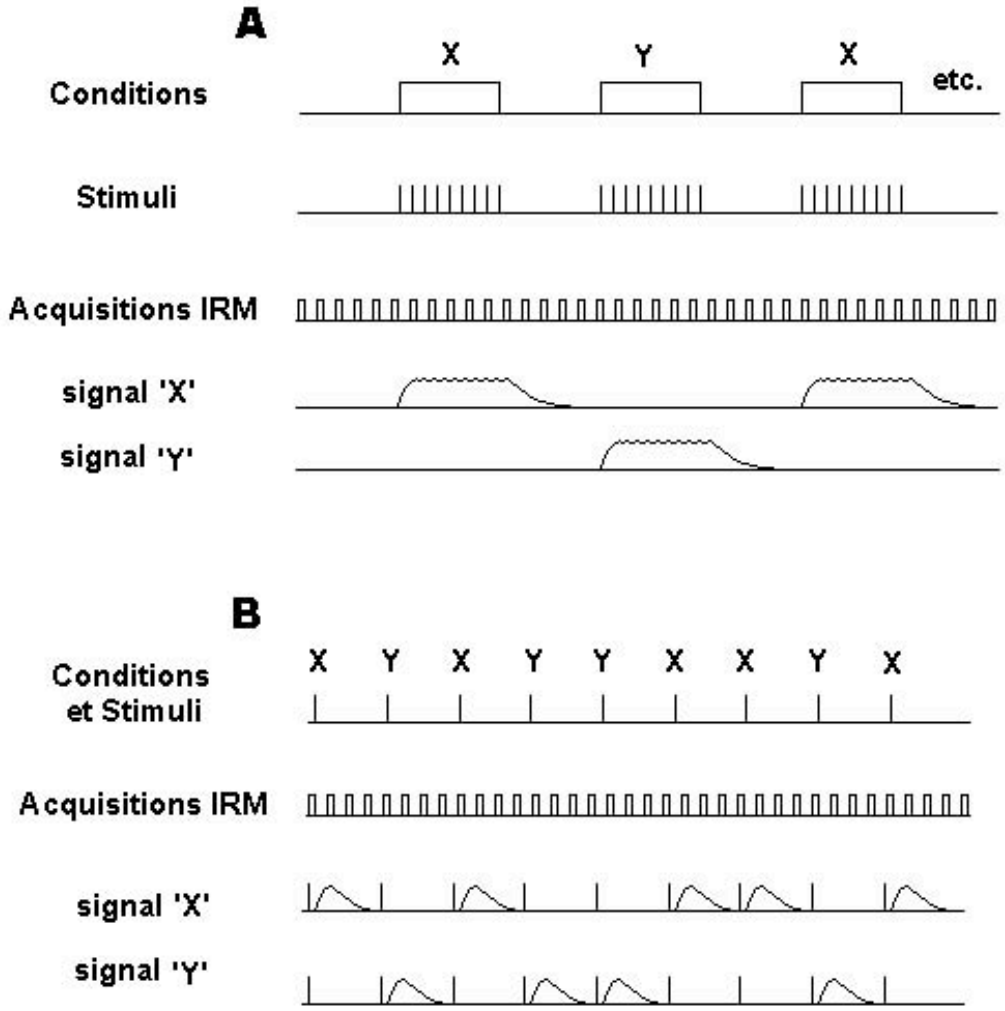


Figure 40 : protocoles d'IRM fonctionnelle

A. Protocole par bloc. Deux conditions expérimentales X et Y sont à comparer. Les stimuli et tâches relatifs à chaque condition sont groupés en blocs, séparés par des intervalles de repos 'neutre' vis-à-vis des tâches à accomplir. L'acquisition d'images IRM est continue, ce qui permet de contraster bloc par bloc les acquisitions obtenues en condition X et repos, Y et repos, X et Y, pour mettre en évidence les aires activées dans les différentes conditions. Les comparaisons s'effectuent soit par la méthode soustractive, soit par corrélation entre un signal idéal attendu (signal 'X' propre à une aire cérébrale répondant spécifiquement à la condition X) et le signal enregistré. Cette méthode issue des techniques TEP permet d'extraire les zones significativement corrélées à n'importe quel signal attendu.

B. Protocole événementiel. Le principe est ici d'espacer suffisamment les stimulations pour laisser l'hémodynamique évoquée retourner à un niveau basal (ex signal 'X'). Le protocole événementiel permet d'associer à chaque essai une réponse propre ; une réponse évoquée moyenne peut être extraite pour chaque condition expérimentale, ou la série temporelle être analysée par corrélation comme pour le protocole en blocs. L'avantage est l'ordre aléatoire des conditions expérimentales, et l'absence d'un recours systématique à une condition de repos au statut incertain. L'inconvénient réside essentiellement dans les longues périodes d'attente pour les sujets, les intervalles entre stimuli étant de l'ordre d'une quinzaine de secondes.

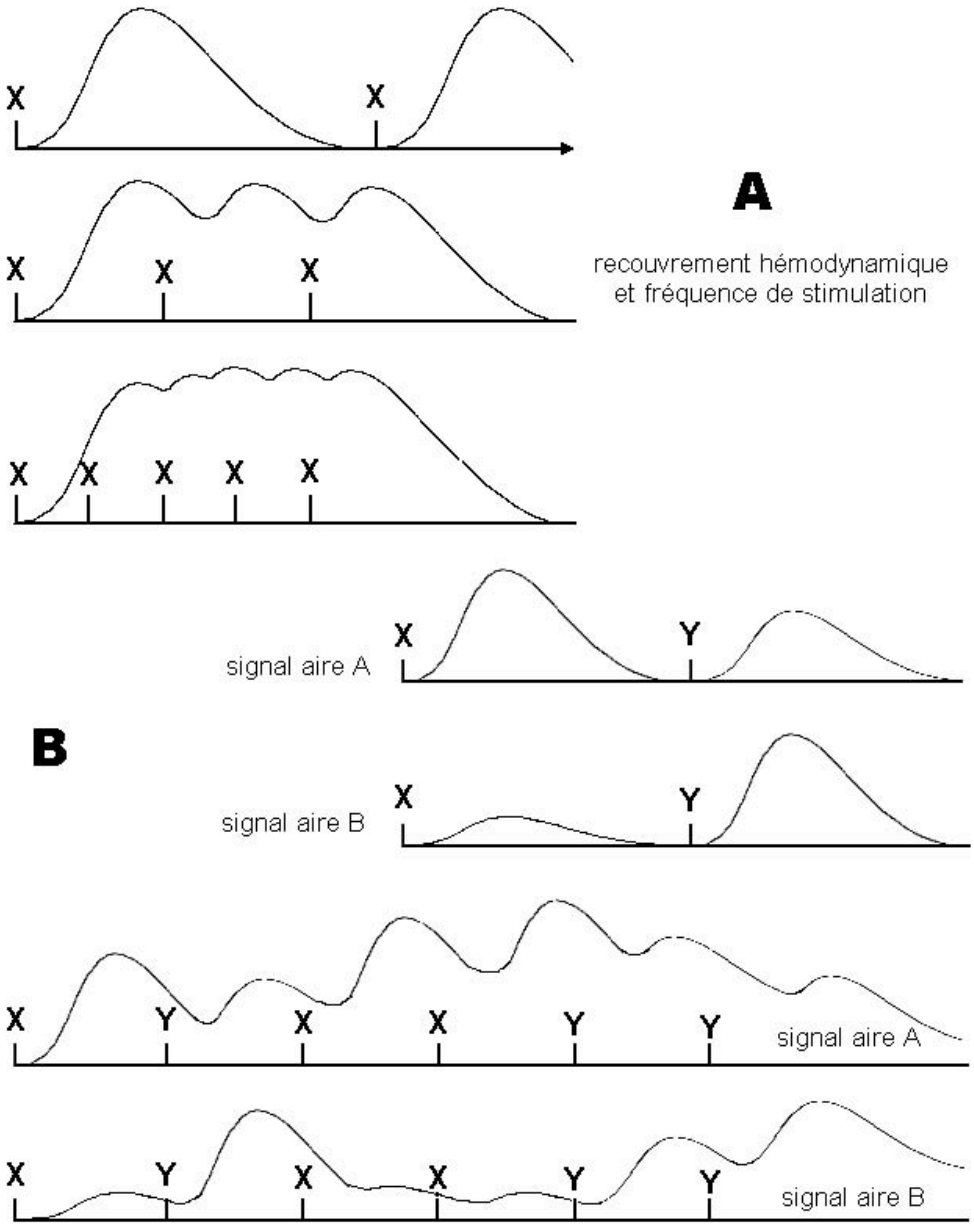


Figure 41 : illustration du recouvrement des réponses hémodynamiques

A. Recouvrement et fréquence de stimulation. Dès que la période de stimulation est inférieure à 15 secondes, la réponse hémodynamique aux stimuli antérieurs est recouverte par les réponses aux stimuli suivants ; la variation d'amplitude du signal est beaucoup plus faible et difficile à détecter.

B. Recouvrement et effet d'ordre. Si deux aires ont des réponses hémodynamiques différentes évoquées par des stimuli X et Y, la variation du signal est directement fonction de l'ordre des stimuli (successions XY, YX, XX et YY). Pour un recouvrement moyen dû à un intervalle interstimulus de 4s par exemple, il importe de maîtriser l'ordre d'apparition des stimuli sur 4 essais successifs.

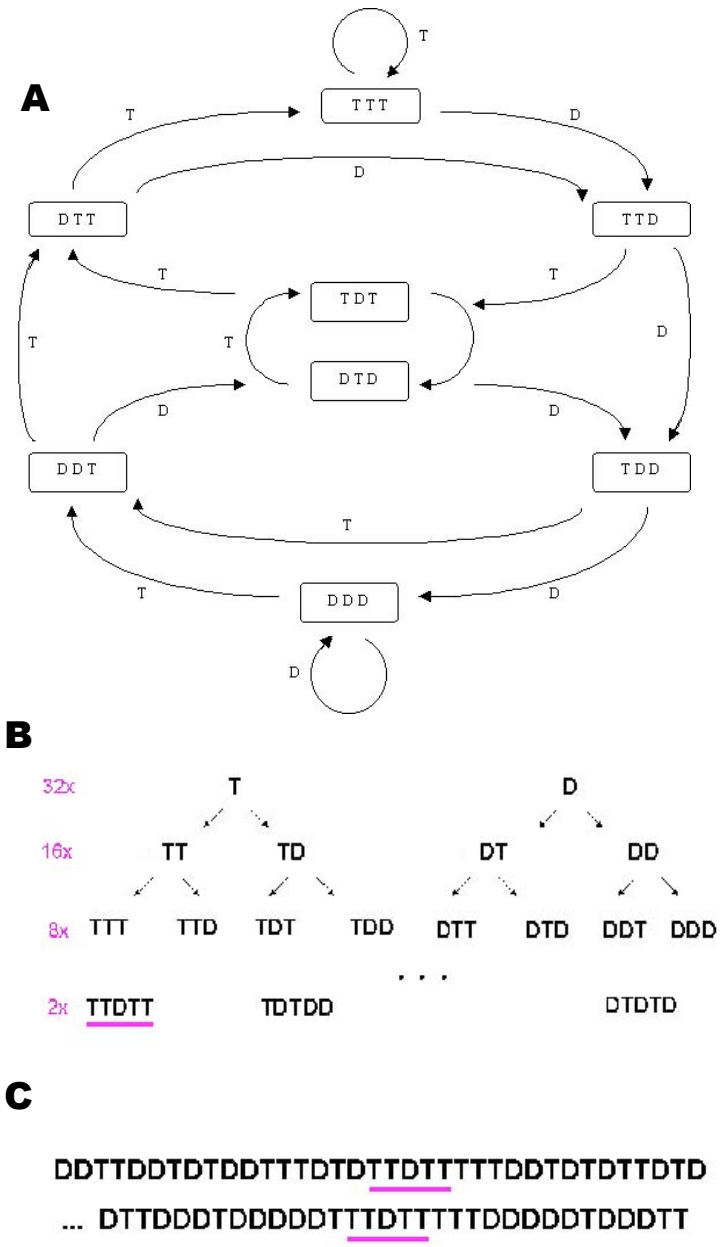


Figure 42 : construction de la séquence de stimulation

A. Graphe binaire de mots de longueurs 3. Sur cet exemple, on considère une succession de trois lettres prises parmi (T,D). On passe de chaque mot possible (8 états) par graphe binaire, chaque flèche représentant la concaténation de la lettre au mot précédent pour former le mot suivant. Nous avons construit un même graphe binaire de mots de longueur 5 pour simuler la succession de 5 stimuli de 2 types, T et D pour cible et distracteur. Un tel graphe à 32 états possède 512 parcours différents, soit 512 mots de longueur 36 (32+4).

B et C. Séquence de stimulation. Dans cette séquence **C**, il y a 32 fois T ou D, 16 fois la succession TT, ou TD, ou DT, ou DD, 8 fois TTT etc, jusqu'à deux fois tous les mots possibles de longueur 5. Cet exemple est la concaténation de deux parcours différents du graphe, donc un mot de longueur 32+32+4.

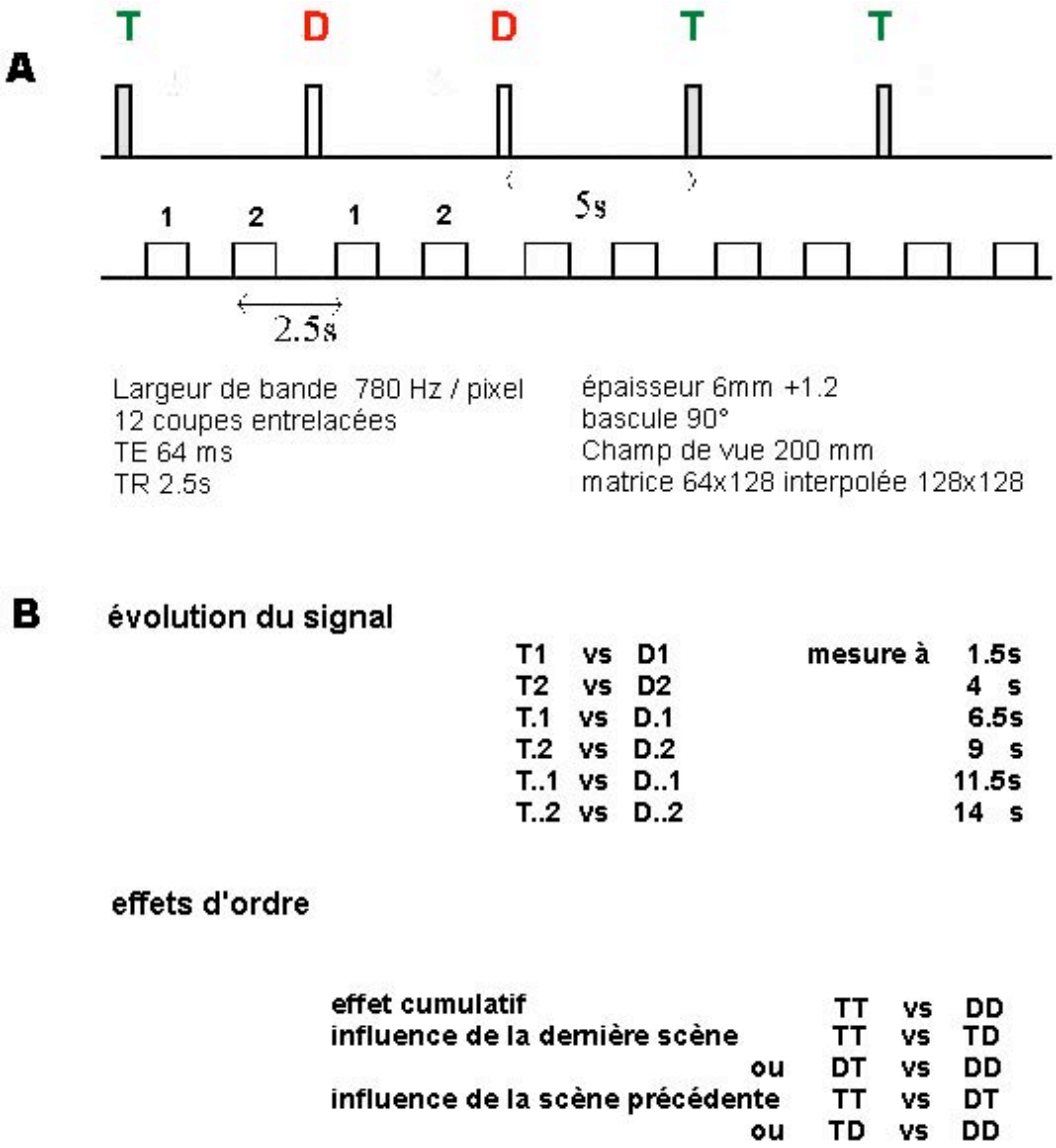


Figure 43 : Protocole IRM et éventail de comparaisons possibles

A. Protocole. La séquence de stimulation TDDTT etc. représente l'ordre de présentation des scènes cibles et distracteurs ; les images sont affichées pendant 33 millisecondes sur des lunettes amagnétiques, chaque 5 secondes. Deux acquisitions d'images IRM suivent chaque présentation de stimulus, à 1.5 et 4s, référencées par les chiffres 1 et 2. Une acquisition peut donc être typée T1, T2, D1 ou D2 en fonction du stimulus qui l'a immédiatement précédée. Elle est en fait typée TTDTD1 par exemple, en fonction du type des 5 stimuli qui l'ont précédée.

B. Possibilités d'analyses. Un tel protocole peut être analysé comme une série temporelle, par corrélation. Il permet aussi une analyse analogue à celle des potentiels évoqués, par comparaison de moyennes. Plusieurs comparaisons sont alors possibles : la mesure d'effets différentiels entre cibles et distracteurs peut être effectuée à des dates différentes : sur la figure un point représente une activation de type indifférent (T ou D, représenté en nombre égal grâce à la séquence de stimulation). Une possibilité est aussi de mesurer le changement de signal en réponse à des successions de stimuli : comparer cible vs distracteur lorsqu'ils ont été suivis d'une cible, ou d'un distracteur, pour mesurer un effet d'ordre.

NeuroImage 11, 634–643 (2000)
doi:10.1006/nimg.2000.0585, available online at http://www.idealibrary.com on IDEAL®

Brain Areas Involved in Rapid Categorization of Natural Images: An Event-Related fMRI Study

Denis Fize,* Kader Boulanouar,† Yvan Chatel,† Jean-Philippe Ranjeva,† Michèle Fabre-Thorpe,* and Simon Thorpe*

*Centre de Recherche Cerveau and Cognition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 Toulouse, France; and †INSERM Unité 455, Hôpital de Purpan, Toulouse, France

Received August 25, 1999

Event-related fMRI was used to investigate brain activation during a visual go/no-go categorization task based on colored photographs of natural scenes, similar to a previous ERP study by Thorpe *et al.* (1996, *Nature* 381: 520–522). Subjects had to press a key when an animal was present in the display. Stimuli were flashed for 33 ms using an intertrial interval of 5 s and a design that carefully balanced targets and distractors in a pseudo-random sequence. Activation produced by targets and distractors was compared with two different techniques, one based on correlations with the stimulation pattern, the other using simple *t* score statistics to compare selected scans. The contralateral primary motor cortex and the ipsilateral cerebellum were both more active following target trials than following distractors, thus confirming the sensitivity of the method. Differential activity was also seen in the posterior cingulate cortex, the fusiform, and the parahippocampic gyri. Activity in such structures could underlie the differential evoked-potentials reported previously in the same task. Surprisingly, in these visual structures, the signal was stronger following distractor trials than target ones. This result could be due to more prolonged processing on distractor trials. Alternatively, it could be that target detection induces strong activation of a small proportion of neurons, which, because of competitive inhibitory mechanisms, could result in a decrease in activity for the population as a whole. We suggest that this kind of mechanism could also account for the decreases in signal observed in perceptual priming experiments. © 2000 Academic Press

Key Words: **event-related fMRI; natural scenes; extra-striate visual cortex; categorization; competitive inhibition; attentional processes.**

INTRODUCTION

In a previous study, we described a go/no-go categorization task in which subjects were asked to decide whether or not briefly displayed color photographs (natural scenes as well as man-made environments) contain an animal (Thorpe *et al.*, 1996). Despite the high demands on the visual system imposed by such a task, subjects performed accurately and rapidly. In addition, simultaneously recorded event-related potentials (ERPs) revealed an early ERP difference between target and distractor trials that started roughly 150 ms after stimulus onset. The effect was particularly visible on frontal electrodes, but the sources of this differential response were unclear.

To help locate the brain areas involved in generating this differential electrical response, we tested human subjects with the same sort of task in a study using functional magnetic resonance imaging (fMRI), which measures local changes in blood oxygenation and provides good spatial resolution. Our approach involved the use of event-related fMRI, a methodology that has become increasingly popular in recent years (Buckner *et al.*, 1996; Schacter *et al.*, 1997; Zarahn *et al.*, 1997; Burock *et al.*, 1998; Friston *et al.*, 1998; Rosen *et al.*, 1998). Such methods allow the use of paradigms that depart from the standard "blocked" testing procedures normally used in fMRI and PET studies by making it possible to isolate brain responses to individual trials. Its flexibility allows different types of trials to be intermixed in a random sequence, a feature that is vital for the go/no-go categorization task used here. A number of empirical studies have made these developments possible by showing that the hemodynamic response sums roughly linearly over trials (Boynton *et al.*, 1996; Dale and Buckner, 1997), and by measuring signal changes in response to brief visual stimuli as short as 34 ms (Savoy *et al.*, 1995). Here we used natural images briefly flashed for 33 ms and an intertrial interval of 5 s; both the interstimulus and presentation durations were in the range of values known to elicit a detectable hemodynamic response related to each stimulus. The relatively short intertrial interval minimized interference from irrelevant cognitive processes, a problem that can occur when long intervals are used to

1053-8119/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

TABLE 1

Behavioral Results

Subject	Age	Sex	Hand	RT ms	Global (% error)	Targets (% error)	Distractors (% error)
cc	19	f	l	546	10.7	12.7	8.8
cm	45	f	l	449	3.4	3.9	2.9
gr	45	f	r	520	14.7	20.6	8.8
bj	30	m	l	580	14.7	17.6	11.8
df	27	m	r	470	12.5	16.7	8
ec	28	m	r	466	12.2	19.6	4.5
Mean	32.3			505	11.4	15.2	7.5

Note. For each of the six subjects indicated by their initials, the table provides their sex (m/f), their age, the used (preferred) hand (r/l), and the mean reaction time for the correct go responses (RT expressed in ms). The table also provides the global percentage of errors (Global % error), and separately the percentage of errors on targets (Targets % error) and the false positives on distractors (Distractors % error).

allow the complete decay of the hemodynamic response.

In the present experiment, we first designed a carefully balanced sequence of target and distractor trials to handle the overlaps between hemodynamic responses and cancel out the effects of surrounding stimuli. To assess the reliability and the sensitivity of the methods used in this study, cerebral activation induced by the motor response was tested separately for left- and right-handers at both individual and group levels. We then focused on the brain structures that were differentially activated during this particularly demanding go/no-go visual categorization task and could be the source of the differential ERP response reported previously.

MATERIALS AND METHODS

Six subjects (three females and three males, three right- and three left-handed subjects, see Table 1) aged between 19 and 45 (mean age: 32.3 years) volunteered in this experiment. Subjects were asked to press the button of a computer mouse when photographs containing an animal were displayed. The stimuli were selected from a very large set of available scenes (4300 color photographs from the Corel photolibrary) that had already been used in previous visual categorization studies in humans and monkeys (Thorpe *et al.*, 1996; Fabre-Thorpe *et al.*, 1998). The photographs containing animals were very varied and included a wide range of species (birds, mammals, reptiles, and fish) in their natural environments. The number of animals, their size and the viewing angle varied widely across the stimulus set. The distractors included pictures of natural scenes as well as man-made environments, and were also very varied in terms of color saliency, spatial frequency and luminance (see examples in Fig. 1). The earlier ERP study showed that VEP components up to 150 ms did not differ significantly between targets and distractors, making it unlikely that there was a systematic difference between the two types of stimulus in terms of low-level factors such as contrast or luminance. Each stimulus was presented once and for each subject the data analysis was based on a minimum of 204 images in 3 blocks of 68 photographs (5 mn 40 s per run). Six different blocks of images were used, none of which was seen by all the subjects. The photographs were flashed in color for 33 ms through a pair of fMRI video goggles (Resonance Technology Inc., California) that provides a visual field of 30°.

Stimuli were flashed every 5 s for the task to be as close as possible to the reference go/no-go task used in the ERP study mentioned above. Reaction times (RT) were recorded with a nonmagnetic mouse linked to a Macintosh computer. Imaging was performed on a 1.5 T MRI system (Siemens Magnetom Vision); each echoplanar scan (EPI) was composed of 12 noncontiguous slices (slice thickness: 6 mm, interslice gap: 1.2 mm, echo time: 64 ms, repetition time: 2.5 s, image size: 128 × 128, field of view: 200 mm). Each set of slices took about 1.3 s to acquire, and during a run, data were continuously collected every 2.5 s, i.e., at delays of roughly 1.5 and 4 s after the onset of the stimulus (Fig. 2A). Because of constraints due to the stimulation sequence described below, image acquisition began after the fifth stimulus.

Stimulation Sequence

The 68 stimuli in each run were presented in a sequence in which targets and distractors were equiprobable. Although the sequence was unpredictable for the subject, it was carefully designed to minimize any possible bias introduced by the overlap of hemodynamic responses due to surrounding stimuli. It was designed so that all combinations of 2 to 5 consecutive stimuli (targets and distractors) were equally represented (Figs. 2B and 2C). Thus any stimulus was not only labeled as a target or distractor but was also characterized by the preceding and subsequent 1–4

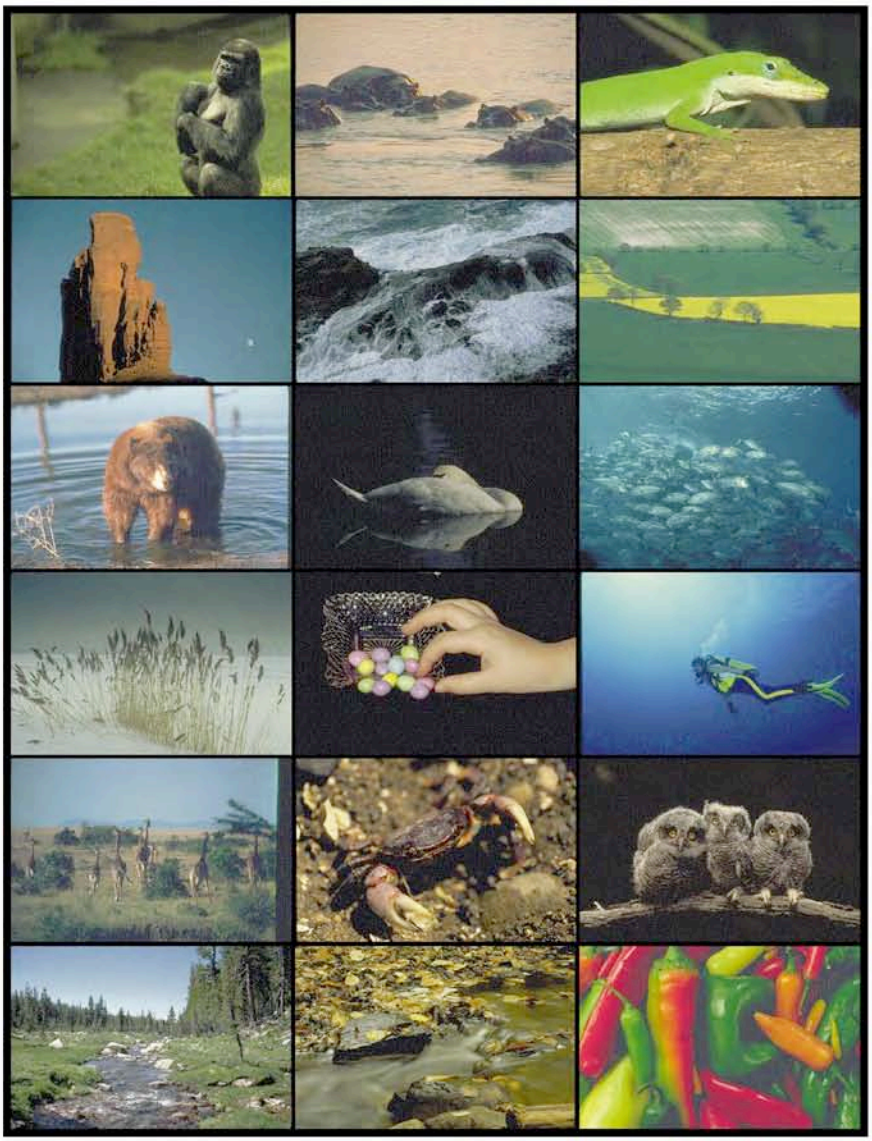


FIG. 1. Examples of stimuli. The 204 color photographs were displayed for 33 ms every 5 s using nonmagnetic video goggles and were presented once. The target scenes contained at least one animal (human beings were considered as distractors). Colors, luminances, and spatial frequencies were very varied for both targets and distractors.

stimuli. Scanning started after the presentation of the fifth stimulus to allow appropriate labeling of the trials. Such a design allowed the mean signals elicited by target and distractor trials to be compared, since any combination surrounding a given trial was equally represented in both target and distractor conditions.

fMRI Data Analysis

The first two scans were discarded to allow signal stabilization, and the remaining 126 scans of each run were realigned to correct for head movement, then normalized and smoothed by a gaussian kernel (10 mm FWHM). After this preprocessing, two different methods of analysis were used. The first involved the detection of differentially activated pixels with the general linear model as implemented in SPM software (Friston *et al.,* 1995). The linear model used low frequency (>120 s) variable adjustment and two covariates for target and distractor activations that were generated by convolving the stimulus pattern with a poissonian hemodynamic response peaking at 6 s delay. The second method was based, as in ERP analysis, on the comparison of scans selected on the basis of correct (go and no-go) trials. Scans were first labeled target or distractor depending on the nature of the preceding trial; the two sets of selected images were then compared with t score statistics using the AFNI (Analysis of Functional NeuroImages) software package (Cox,

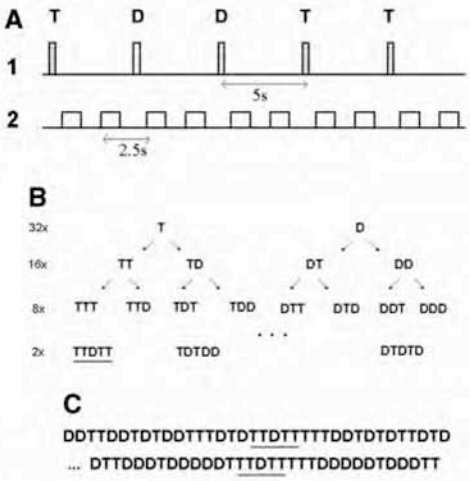


FIG. 2. Stimulation and image acquisition design. (A) In the sequence of stimulation (1), target (T), and distractor (D) stimuli appeared quasi randomly. MRI scans (2) were synchronously acquired 1.5 and 4 s after each stimulus display. (B) The stimulation sequence was designed to equalize the number of target and distractor stimuli and any subsequences composed of 2 to 5 stimuli. It included 32 targets and distractors, 16 subsequences composed by TT, DD, TD, and DT, and 2 subsequences for every possible combination of 5 targets and/or distractors. One of them, TTDTT is underlined for clarity in the sequence (C). (C) The complete stimulation sequence was composed of 68 trials, and MRI acquisitions were performed after the presentation of the fifth stimulus.

1996). The use of t test statistics on selected sets of images was made theoretically possible by the design of the sequence of stimulation since, on average, the effects of surrounding stimuli were cancelled.

The first method that measures the correlation between the recorded data and an expected pattern of response signal is used in most neuroimaging studies. Using this method, we first analyzed separately the data recorded in right- and left-handed subjects in order to assess the sensitivity of the event related fMRI protocol to motor activations. Using the same method, we then performed a group study that included all the subjects (regardless of their hand preference) in order to localize other sites with task-related activity. Finally, the robustness of the results was assessed by comparing individual data obtained using both correlation analysis and t score statistics on selected scans.

RESULTS

Behavioral Results

Despite the physical constraints of the fMRI situation, the subjects performed this challenging task very well. The mean RT was 505 ms on targets, and 568 ms for false detections (Table 1) for a percentage of correct responses that reached 88.6% by run. The subjects were less accurate and less fast than in our previous study (94% correct for a mean RT of 471 ms), but the MRI environment, the subject's position, the use of goggles for stimulus presentation as well as image definition and contrast could all contribute to these differences.

fMRI Results

The results of the analyses performed separately for the right- and left-handed subjects showed an increase in signal strength following target trials in the contralateral motor cortex and ipsilateral cerebellum for the two groups (Figs. 3A and 3B). Similar motor activations were obtained when considering the individual cerebral activations shown by each subject using the two different methods of analysis (Table 2).

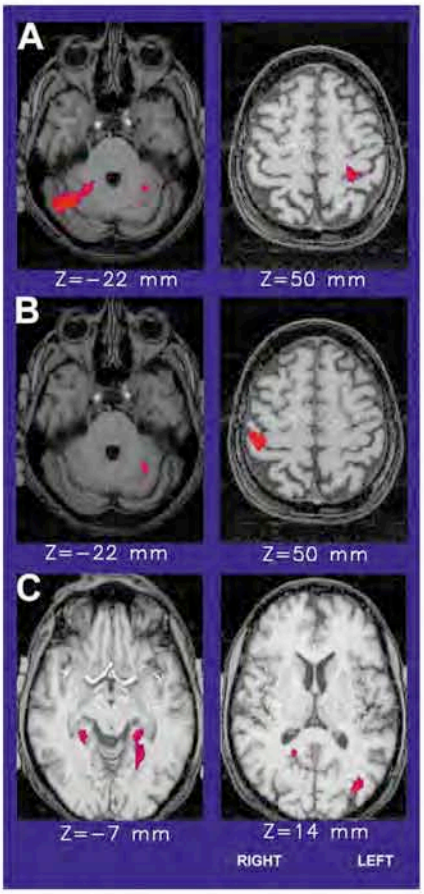


FIG. 3. Statistical parametric maps of correlation analyses performed on groups of subjects using SPM96. Statistical parametric maps are thresholded at $P < 0.001$ for peak height and $P < 0.05$ (corrected) for spatial extent. (A and B) Results of the analysis of 3 right-handed subjects (A) and 3 left-handed subjects (B) performed to assess the sensitivity of the protocol. Activations after target trials (T − D contrasts) were only observed in contralateral Brodmann area 4 and in the cerebellum ipsilateral to the dominant hand. (C) Results from analysis including all 6 subjects. A higher signal was observed after distractor trials when compared to target trials (D − T contrast) in right cingular gyrus near occipitoparietal sulcus (BA 31), left fusiform gyrus (BA 19/37), bilateral parahippocampic gyri, and medial occipital gyrus (BA 19).

TABLE 2

Differentially Activated Areas between Target and Distractor Conditions from Individual and Group Studies

	T − D > 0									D − T > 0								
					Tc									Tc				
Subjects (hand)	Size	Z	t	Lat.	x	y	z	Areas	R/L	Size	Z	t	Lat.	x	y	z	Areas	R/L
CM	102	6,79	5,6	4 s	36	−22	50	BA 4	R	27	4,97	−3,8	9 s	44	−30	7	BA 41/42	R
(L)	57	6,33	5,5	4 s	−24	−52	−22	Cerebellum	L	250	4,96	−3,4	9 s	6	−45	7	(BA 29)	R
											4,15	−3,1	9 s	10	−35	0	Parahipp.	R
											4,53	−3,3	9 s	4	−64	29	BA 31/18	R
											3,34	ns		−6	−76	29	BA 18	L
											3,86	ns		2	−76	22	BA 18	R
										25	4,69	−2,9	9 s	44	−14	7	BA 22	R
GR	152	4,44	4,3	4 s	−36	−42	50	BA 1,2	L	72	4,26	−4,9	6.5 s	−10	−60	7	BA 31	L
(R)	53	3,94	3,2	4 s	−42	−38	43	BA 40	L	18	4,06	−4,6	6.5 s	−36	−74	7	BA 19	L
	34	3,73	2,8	4 s	−36	28	29	Prefrontal	L									
	23	4,37	3,4	4 s	16	−64	−7	Cerebellum	R									
BJ	105	4,52	4,1	4 s	40	−24	50	BA 4	R	25	4,34	−3,2	6,5 s	−12	−82	−7	BA 18	L
(L)										28	3,82	−3,0	6,5 s	10	−86	14	BA 19	R
										23	3,68	ns		30	42	36	BA 9	R
CC	69	5,90	3,4	4 s	−26	−66	−22	Cerebellum	L	109	5,90	−4,1	4 s	−24	−56	14	BA 37	L
(L)											5,47	−4,1	4 s	−22	−52	−7	BA 19	L
										215	4,74	−3,8	6.5 s	−28	−78	36	BA 19	L
										59	5,31	−3,2	4 s	−28	−8	43	BA 6	L
EC	94	5,45	3,7	4 s	−32	−60	−29	Cerebellum	L	118	4,71	−4,1	9 s	12	−58	14	BA 31	R
(R)	21	4,65	2,9	4 s	−32	−26	50	BA 4	R	28	4,21	−4,1	9 s	−4	−62	7	BA 31	L
	62	4,10	2,5	4 s	36	−58	−29	Cerebellum	L	24	4,26	−2,9	6.5 s	−54	−12	−7	BA 22	L
	55	3,84	3,0	4 s	40	8	7	BA 6	R	34	3,94	ns		40	−10	7	BA 22	R
DF	67	4,74	3,6	6.5 s	−50	−34	50	BA 1,2	L	71	3,58	−3,1	9 s	0	−85	22	BA 18	
(R)	51	5,19	4,5	6,5 s	16	−50	−22	Cerebellum	R	14	3,37	ns		46	52	−7	BA 37	R
										23	3,35	ns		−22	34	43	BA 8 (FEF)	L
Left-handed subjects (CM, BJ, CC)										Six subjects								
	112	6,54			36	−16	50	BA 4	R	59	5,21			14	−58	14	BA 31	R
	27	4,26			−24	−52	−22	Cerebellum	L		5,09			12	−54	7	BA 31	R
											4,43			18	−35	−7	Parahipp.	R
Right-handed subjects (GR, EC, DF)										68	5,17			−18	−35	−7	Parahipp.	L
											4,94			−22	−60	−7	BA 19	L
	27	4,34			−26	−28	50	BA 4	L	52	5,05			−30	−82	14	BA 18/19	L
	125	5,30			28	−56	−22	Cerebellum	R		3,74			−26	−80	29	BA 19	L
											3,72			−28	−72	36	BA 19	L

Note. Individual results are presented for each subject indicated on the left by their initials and preferred hand. The results from the group studies (six subjects) are shown in the bottom part of the table (left- and right-handed subjects are considered separately for the activation of the motor structures). The left column lists areas for which the signal difference was observed to be higher after target stimuli (T − D > 0). The right column lists areas for which the signal difference was observed to be higher after distractor stimuli (D − T > 0). In both columns, the results obtained with the two different methods of analysis are presented. The size (number of pixels) of the areas and their maximum Z values were both obtained by SPM analyses run on the overall data of each subject. Within the same areas, the maximum t values and their latency in seconds (t lat.) were obtained with selective comparisons (see Materials and Methods), using only the scans for which the preceding stimulus was correctly categorized. The Talairach coordinates (Tc) of the maxima, the corresponding areas, and the activated hemisphere (L/R) are also indicated. Note that motor structures were more activated after target trials (left column), whereas visual structures were more activated after distractor trials (right column) and generally with a longer latency than motor areas.

Although motor structures were found to be more activated following target trials than following distractors, some visual areas were found to be more activated by distractor trials. These differentially activated structures included the extrastriate visual cortex and the hippocampal gyri (Fig. 3C summarizes the overall results obtained with the six subjects). Four main areas with differential activations can be seen from this group study; they include Brodmann Area 31 (BA 31) in the right occipitoparietal sulcus, the posterior parahippocampic gyri on both sides, a large part of BA 19/37 in the left fusiform gyrus and BA 19 in the median occipital gyrus. This pattern of differential activation was also seen at the individual level of analysis (Table 2) and with the two different methods of analysis. However, some individual differences could be noted; for example, in subjects CM and EC additional differential activity could be seen in BA 22.

In order to determine the temporal course of the responses to distractors and targets in visual areas, we performed *t* test comparisons between images acquired at 1.5 and 4 s poststimulus for targets and distractors separately. As shown in Fig. 4 for subject CM, both targets and distractors elicited robust increases in occipital areas including primary and extrastriate visual cortex. However, further analysis aimed at localizing structures with significantly different levels of activation in response to targets and distractors showed that the main effect, namely a larger signal on distractor trials, was not seen until later (see Fig. 4C).

Figure 5A illustrates the group average signal around target and distractor trials for a differentially activated region in the median occipital gyrus (BA 19). It can be seen that both target and distractor trials induced an initial increase in signal after stimulus presentation with, as a consequence, a modulation of the signal at the presentation rate. But the interesting feature is the divergence between the average response to targets and distractors that lasts for a period of up to 10 s, with distractors producing more activity than targets. Note that both curves are initially roughly equal since the first point on each curve corresponds (thanks to the balanced experimental design) to an equal mix of target and distractor trials. The same sort of averaging was also performed on the differentially activated motor area BA 4. Figure 5B shows that, in this case, signal values in this area increased only around target trials (i.e., when there was a go-response), whereas the signal following distractor trials decreased back to a value below the baseline. This makes sense, given that the baselines were effectively the average of target and distractor signal values. Note that since there was no systematic motor activation on both target and distractor trials, there is no modulation of the signal at the presentation rate.

Figures 5C and 5D show that the differential visual and motor effects have different time-courses: in visual areas the difference in signal lasts longer and reaches its maximum later than in motor cortex.

DISCUSSION

The most surprising result of the present study was that in a go/no-go visual categorization task, extrastriate visual areas, including the right occipitoparietal sulcus, the posterior parahippocampic gyri on both sides, the left fusiform gyrus and the median occipital gyrus are more activated following distractor trials than target ones. The second result concerns the reliability of an event-related fMRI protocol that includes short stimulus presentations and relatively short intertrial intervals, in a counterbalanced sequence of stimuli to handle biases resulting from overlapping hemodynamic responses. This study shows that such protocols allow reliable results to be obtained not only at a group level but also in the case of individual studies.

Protocol

The reliability of the event-related protocol used in this study is clearly demonstrated by the group analysis performed on right and left handed subjects. The results showed clear motor activations (contralateral BA 4 and ipsilateral cerebellum) related to the finger movement required by the task in response to target stimuli. Similar results were obtained with the two different methods, using a linear model and t-tests on selected sets of images. It should be emphasized that this type of design is able to provide a sufficient signal-to-noise ratio (not estimated here), since the pattern of activity resulting from a single subject is close to the pattern exhibited by the group study: motor activations were observed in individual subjects, and differential effects seen in visual areas were also found in anatomically similar sites in different individuals.

The modulations in MR signal elicited by the trials were found to be generally in the range 0.07–0.1% around the mean values and the differential effects between the target trials and distractors were in the range 0.1–0.3%. These small amplitudes may be partly due to the relatively short and fixed intertrial period used here since the overlap between the hemodynamic responses due to preceding stimuli will increase baseline values and lower signal variations (Burock *et al.*, 1998). There are other studies with short ISIs that have reported larger amplitude fluctuations than those reported here, but they randomized the intertrial interval with a third neutral stimulus, a technique that increases the size of signal variations (Buckner *et al.*, 1998; Clark *et al.*, 1998). Another factor may well be the use of very short stimulation durations that prevented any visual exploration of the stimuli. Comparisons of the signal in occipital cortex at 1.5 and 4.5 s poststimulus clearly showed that briefly displayed photographs of both targets and distractors produce statistically significant visual activation. But as Savoy *et al.* (1995) reported, brief 33 ms visual stimulation only produces a relatively small response (less than 1% from the baseline in their case).

Another aspect of the stimulation design is the use of a counterbalanced sequence of stimuli involving up to five successive target and distractor trials. The results suggest that such sequences successfully equated the baselines for target and distractor trials, and the overlap of hemodynamic responses to surrounding stimuli. This is supported by the fact that the averaged signals computed before and 10 s after stimulus presentation are virtually identical for both target and distractor conditions. Such designs can therefore overcome the biases encountered in other rapid mixed designs that use random mixing. For example, Buckner *et al.* (1998)

FIG. 4. Results of t tests performed on subject CM at $Z = -7$ mm. (A and B) t score maps between 1.5 and 4 s poststimulus after correct target trials (A) and after correct distractor trials (B). A large involvement of occipital cortex can be seen for both stimuli types during this period ($t > 2.87$; uncorrected $P < 0.005$; $df = 246$). (C) Differential activations in visual areas were seen in this subject at 9 s poststimulus ($|t| > 2.83$; uncorrected $P < 0.05$; $df = 244$). These visual areas present a higher signal for distractor than target trials (D − T > 0).

have argued that counterbalancing should involve several successive trials. While 30 trials are generally thought to be sufficient to balance a series containing just two types of trial, the number required to equate the four different combinations of two successive trials would be very much longer if it depended on random mixing. In our case, this problem was avoided by defining sequences that balanced target and distractor strings up to fifth order, but remained unpredictable for the subjects.

Taken together, these balanced stimulation designs and the available range of processing methods mean that fMRI studies can be conducted with single-trial protocols very similar to those used in psychophysics and electrophysiological studies.

Relation to the Differential Evoked Potentials

The design of the protocol used in this fMRI experiment is very similar to that used in the ERP study by Thorpe *et al.* (1996) that demonstrated strong differential activity to target and distractor trials starting roughly 150 ms after stimulus presentation. Subsequent source analysis using BESA has shown that this pattern of differential activity is consistent with two dipole sources located in occipitotemporal regions (Fize *et al.*, 1998, and in preparation). These regions are very close to the visual areas BA 19/31 in the fusiform and cingulate cortex that were differentially activated in the present study. The fusiform gyrus has been implicated in the processing of various categories of visual stimuli by a number of recent imaging studies (Martin *et al.*, 1996; Price *et al.*, 1996). For example, the fusiform gyrus contains regions that are strongly activated by faces (Kanwisher *et al.*, 1997; McCarthy *et al.*, 1997; Clark *et al.*, 1998), and recent studies suggest that similar regions are also activated by photographs of animals (Chao *et al.*, 1999a, b). The regions activated by animal photographs were shown to be located in the lateral fusiform gyrus. In contrast, the differentially activated regions reported here appear more medial and anterior, extending as far forward as the parahippocampal region. However, direct comparisons between these different studies is made difficult since we used a categorization task that is clearly different from the passive viewing, matching and naming tasks used in these two studies.

The ERP source analysis also showed clear differential activity in the contralateral motor cortex that started considerably later (250–350 ms). ERPs sources would thus fit with the localization of differential visual and motor fMRI responses seen in the present study.

Underlying Mechanisms

Many visual areas are activated by both target and distractor trials. However, the results show that some of these areas, mainly BA 19 and BA 31, are differentially activated by target and distractor trials and that the signal is smaller in response to targets. When comparing two experimental conditions, many studies have also reported decreases in signal that have been interpreted as reflecting either inhibitory mechanisms of irrelevant sensory modalities (Haxby *et al.*, 1994), suppression of habitual responses during difficult cognitive tasks (Shulman *et al.*, 1997; Vandenberghe *et al.*, 1997), or attentional involvement during simple fixation (Orban *et al.*, 1997). Unlike these studies, our result seems to reflect ongoing visual processing mechanisms during the task since the increases in signal were shorter in duration following the presentations of target stimuli than after distractor ones, and seem to reflect a different kind of dynamic to the more conventional differential activity observed in motor cortex.

What kind of mechanisms could explain the greater signal after distractor trials than following targets? One possibility is that processing lasts longer on distractor trials. This hypothesis is consistent with results from visual-search paradigms that show that subjects have shorter reaction times when a target is present than when the display only contains distractors (Treisman and Gelade, 1980). This has often been explained by supposing that search is a serial process that ends as soon as a target is found, whereas with nontarget displays, analysis has to run to completion. If this were true in the present task, then this might explain why the signal is stronger on distractor trials. But the speed with which both humans and monkeys can perform this task suggests that target detection involves mainly parallel processing (Thorpe *et al.*, 1996; Fabre-Thorpe *et al.*, 1997, 1998). Furthermore, the onset of the differential ERP response did not vary with reaction time. However, it may be that processing related to target search continues for longer when the

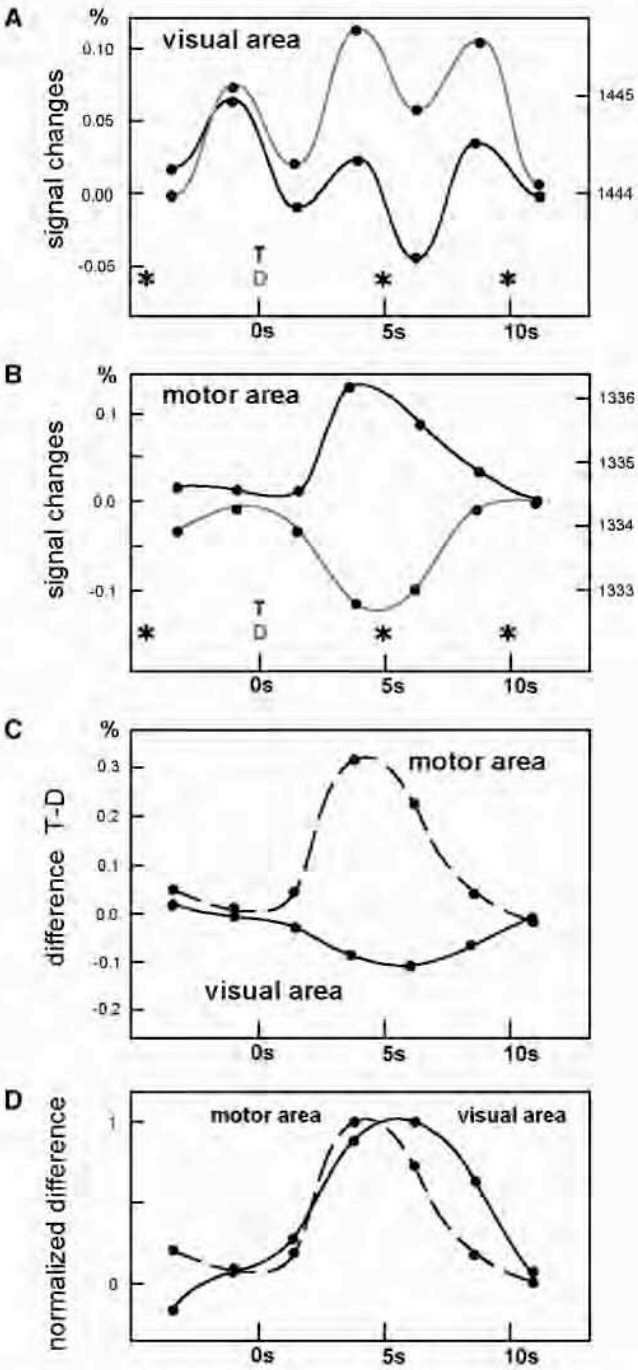


FIG. 5. Timecourses of signal changes from group studies. The marks on the x-axis correspond to stimulus presentations. At time 0 s either a distractor (D) or a target (T) is displayed, whereas for preceding and subsequent stimulations (*), targets, and distractors are equiprobable. In between stimulations, each acquisition of a set of slices is indicated on the result curves by a dot (●). (A) Average of voxels (for which z score >2.5) in the visual area BA 19. The signal increases after target (black line) and distractor trials (grey line) displayed at $t = 0$ s, but this area shows higher activation for distractors than for targets. Note that baselines in (A) and (B) average the responses of the two types of trials. (B) Average of voxels (for which z score >2.5) in the motor area BA 4 in left-handed subjects. This area shows higher signal after target trials (black line), whereas the signal after distractors decreases below the baseline. (C) Average differences between voxel values (targets − distractors) for the data in (A) and (B). Note the larger amplitude difference in motor structures (dashed line) and the change in polarity between motor and visual areas (solid line). (D) Normalized signal differences from (A) and (B). Note that the differential effect lasts longer in the visual area than in the motor one.

stimulus is a distractor, and this could explain the current result.

An alternative explanation might involve the existence of long-lasting competitive inhibition following the detection of a target. This mechanism would also explain why the signal on distractor trials was stronger than with targets. Suppose that within a cortical visual area, the percentage of neurons activated by a particular visual stimulus is relatively small, and that the activation of these neurons provokes a widespread inhibition of the other neurons in the same area, as suggested by a number of experimental studies (Nelson, 1991; Motter, 1994; Maunsell, 1995; Duncan *et al.*, 1997; Borg-Graham *et al.*, 1998; Kastner *et al.*, 1998). Suppose that in addition, when the stimulus is a target, top-down facilitatory processes mean that the relevant neurons have been "primed" before the start of the trial. As a consequence, the responses of the neurons related to the target might be enhanced, which in turn could result in even stronger competitive inhibition (Luck *et al.*, 1997; Caputo and Guerra, 1998; Desimone, 1998). Note that, if the proportion of neurons related to the target is small (5–10%), the *net* effect on the whole population could be an inhibition of overall firing, a result consistent with our finding that the fMRI signal is smaller following targets. Interestingly, if the competitive inhibitory mechanisms activated following the detection of a target were long-lasting, this could explain why the peak difference was only seen at 6.5–9 s after the presentation of the stimulus, much later than the motor related activation, which presumably involves mainly excitatory effects. This sort of prolonged inhibition within the cortex following target detection might well be related to the attentional blink phenomenon, described by a number of recent studies (Chun, 1997; Vogel *et al.*, 1998). It is clear that the temporal resolution of the current experiments is limited and that further studies will be required to determine whether the time-course of posttarget inhibitory effects is consistent with behavioral phenomena such as the attentional blink and single-unit studies of inhibitory mechanisms.

The same sort of mechanism could also account for the somewhat paradoxical finding that perceptual priming, which typically results in increased accuracy and shorter behavioral reaction times, is associated with decreases in signal both in the case of fMRI and electrophysiological studies (Desimone, 1996; Schacter

and Buckner, 1998; Wiggs and Martin, 1998). Here again, an enhanced response in a subset of neurons following priming could potentially result in a decrease in activity for the population as a whole.

Conclusion

Event-related fMRI was used in a go/no-go categorization task adapted from an ERP protocol: complex natural scenes were briefly flashed for 33 ms and subjects were asked to respond when an animal was present in the displayed photograph. Stimuli were presented every 5 s in a carefully counterbalanced intermixed sequence of targets and distractors designed to avoid biased overlaps of hemodynamic responses. We have shown that both correlation analysis using the general linear model and *t* test comparisons close to ERP methods produce similar results: robust activations in motor cortex in response to targets in this active task and large occipital recruitment for both target and distractor stimuli.

Although visual processing of both targets and distractors involved extensive parts of the occipital cortex, only some extrastriate visual areas and parahippocampic gyrus were differentially activated. We propose that these visual areas, which include the fusiform gyrus and cingulate cortex, could be the sources of the early differential effect previously reported using ERPs. Surprisingly, the fMRI signals in these areas were found to be stronger on distractor than on target trials. One interpretation of this result would consider that target detection involves strong activation of a subpopulation of neurons which, because of competitive inhibitory mechanisms, could lead to a reduction in *net* activity across the whole population.

ACKNOWLEDGMENTS

The authors thank the C.N.R.S., the "Action Concertée Sciences du Vivant," and the "G.I.S. Sciences de la Cognition" for financial support. The experimental procedures used were authorized by the ethical committee (CCPPRB No. 9714001).

REFERENCES

Borg-Graham, L. J., Monier, C., and Fregnac, Y. 1998. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393:** 369–373.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16:** 4207–4221.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93:** 14878–14883.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., and Dale, A. M. 1998. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron* **20:** 285–296.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., and Dale, A. M. 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* **9:** 3735–3739.

Caputo, G., and Guerra, S. 1998. Attentional selection by distractor suppression. *Vision Res.* **38:** 669–689.

Chao, L. L., Haxby, J. V., and Martin, A. 1999. Attribute-based substrates in temporal cortex for perceiving and knowing about objects. *Nature Neurosci.* **2:** 913–919.

Chao, L. L., Martin, A., and Haxby, J. V. 1999. Are face-responsive regions selective only for faces? *NeuroReport* **10:** 2945–2950.

Chun, M. M. 1997. Temporal binding errors are redistributed by the attentional blink. *Percept. Psychophys.* **59:** 1191–1199.

Clark, V. P., Maisog, J. M., and Haxby, J. V. 1998. fMRI study of face perception and memory using random stimulus sequences. *J. Neurophysiol.* **79:** 3257–3265.

Cox, R. W. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29:** 162–173.

Dale, A. M., and Buckner, R. L. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain. Mapp.* **5:** 329–340.

Desimone, R. 1996. Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA* **93:** 13494–13499.

Desimone, R. 1998. Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353:** 1245–1255.

Duncan, J., Humphreys, G., and Ward, R. 1997. Competitive brain activity in visual attention. *Curr. Opin. Neurobiol.* **7:** 255–261.

Fabre-Thorpe, M., Richard, G., Fize, D., and Thorpe, S. 1997. Rapid categorization of extrafoveal natural images: Implications for biological models. In *Computational Neuroscience: Trends in Research* (Bower, J. M., Ed.), pp. 7–12. Plenum Press, New York.

Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. 1998. Rapid categorization of natural images by rhesus monkeys. *NeuroReport* **9:** 303–308.

Fize, D., Boulanouar, K., Ranjeva, J. P., Fabre-Thorpe, M., and Thorpe, S. J. 1998. Brain activity during rapid scene categorization: A study using event-related fMRI. *J. Cogn. Neurosci. (Suppl.)* **72.**

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. 1998. Event-related fMRI: Characterizing differential responses. *NeuroImage* **7:** 30–40.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiack, R. S. J. 1995. Statistical Parametric Maps in Functional Imaging: A general linear approach. *Hum. Brain Mapp.* **2:** 189–210.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., and Grady, C. L. 1994. The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* **14:** 6336–6353.

Kanwisher, N., McDermott, J., and Chun, M. M. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17:** 4302–4311.

Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. G. 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* **282:** 108–111.

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77:** 24–42.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. 1996. Neural correlates of category-specific knowledge. *Nature* **379:** 649–652.

Maunsell, J. H. 1995. The brain's visual world: Representation of visual targets in cerebral cortex. *Science* **270:** 764–769.

McCarthy, G., Puce, A., Gore, J. C., and Allison, T. 1997. Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* **9:** 605–610.

Motter, B. C. 1994. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J. Neurosci.* **14:** 2190–2199.

Nelson, S. B. 1991. Temporal interactions in the cat visual system. 1. orientation-selective suppression in the visual cortex. *J. Neurosci.* **11:** 344–356.

Orban, G. A., Dupont, P., Vogels, R., Bormans, G., and Mortelmans, L. 1997. Human brain activity related to orientation discrimination tasks. *Eur. J. Neurosci.* **9:** 246–259.

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. J., and Friston, K. J. 1996. The neural regions sustaining object recognition and naming. *Proc. R. Soc. Lond.* **263:** 1501–1507.

Rosen, B. R., Buckner, R. L., and Dale, A. M. 1998. Event-related functional MRI: Past, present, and future. *Proc. Natl. Acad. Sci. USA* **95:** 773–780.

Savoy, R. L., Bandettini, P. A., O'Craven, K. M., Kwong, K. K., Davis, T. L., Baker, J. R., Weisskoff, R. M., and Rosen, B. R. 1995. Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated changes in noise. *Proc. Soc. Magn. Reson. Med. Third Sci. Meeting Exhib.* **2:** 450.

Schacter, D. L., and Buckner, R. L. 1998. Priming and the brain. *Neuron* **20:** 185–195.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., and Rosen, B. R. 1997. Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* **6:** 259–269.

Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M., and Petersen, S. E. 1997. Top-down modulation of early sensory cortex. *Cerebral Cortex* **7:** 193–206.

Thorpe, S., Fize, D., and Marlot, C. 1996. Speed of processing in the human visual system. *Nature* **381:** 520–522.

Treisman, A. M., and Gelade, G. 1980. A feature-integration theory of attention. *Cogn. Psychol.* **12:** 97–136.

Vandenberghe, R., Duncan, J., Dupont, P., Ward, R., Poline, J. B., Bormans, G., Michiels, J., Mortelmans, L., and Orban, G. A. 1997. Attention to one or two features in left or right visual field: A positron emission tomography study. *J. Neurosci.* **17:** 3739–3750.

Vogel, E. K., Luck, S. J., and Shapiro, K. L. 1998. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* **24:** 1656–1674.

Wiggs, C. L., and Martin, A. 1998. Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.* **8:** 227–233.

Zarahn, E., Aguirre, G., and D'Esposito, M. 1997. A trial-based experimental design for fMRI. *NeuroImage* **6:** 122–138.

1.4 conséquences

a) un protocole très proche des études électrophysiologiques

La validité des mesures effectuées est clairement démontrée par les études de groupes réalisées sur les sujets droitiers et gauchers, qui montrent des activations dans l'aire cérébrale motrice Brodmann 4 controlatérale à la main utilisée, et dans le cervelet ipsilatéral. En plus de leur validité, les mesures semblent être suffisamment sensibles à l'activité cérébrale pour révéler des activations significatives chez chacun des sujets proches des activations trouvées lors de l'étude de groupe. Le rapport signal / bruit dans ce protocole semble être suffisamment grand pour que ces résultats soient observables lors d'une seule série expérimentale. Nous pouvons noter que ces résultats individuels ont été obtenus indépendamment par les deux méthodes d'analyses. Ces résultats obtenus malgré le recouvrement des hémodynamiques de stimuli successifs nous permettent aussi de dire que ces effets de recouvrement ont été correctement compensés par la séquence de stimulation. Cette compensation apparaît clairement dans les figures illustrant le décours du signal au cours du temps autour de stimuli cibles et distracteurs : avant différenciation des types et 10s après, les deux courbes sont identiques dans les deux conditions.

Nous pouvons observer aussi que la brève stimulation visuelle utilisée dans le protocole évoque une forte activité dans l'ensemble du cortex visuel, autant pour les cibles que pour les distracteurs, notamment dans la voie ventrale. La réponse à cette brève stimulation confirme donc les observations de Savoy et al. 1995. Cette activité est mesurée entre 1.5 s et 4 s post stimulus, intervalle pendant lequel toutes les études montrent que le signal lié à l'effet BOLD s'accroît.

b) aires cérébrales activées et sources dipolaires

Le protocole étant construit pour que la tâche effectuée par les sujets soit identique à celle en potentiels évoqués, nous pouvons rechercher les aires cérébrales qui pourraient être la source de l'activité différentielle observée en électrophysiologie. Les résultats d'imagerie suggèrent que cet effet pourrait être lié aux activations observées différentiellement dans les aires visuelles (les aires de Brodmann 19 et 31) du gyrus fusiforme et du cortex cingulaire. Cette proposition est en accord avec les résultats des modélisations par dipôles qui suggèrent une implication des cortex visuels extra-striés. Le gyrus fusiforme et le cortex cingulaire ont été observés dans différentes tâches impliquant des jugements catégoriels dans un certain nombre d'études d'imagerie récentes : le traitement des visages (Kanwisher et al. 1997 ; McCarthy et al. 1997), nommer des animaux (dans le gyrus occipital médian, Martin et al. 1996). D'après Kosslyn, l'aire 19 (d'une surface très importante) est impliquée dans les processus de mémoire associative (p 255 Image and Brain), implication observée dans sa partie la plus dorsale (Kosslyn et al. 1993). D'autre part, Sergent et al. 1992 ont relaté l'activation de l'aire 19 lors de tâches de

catégorisation d'objets ou de discrimination de genre (masculin féminin). Ces mêmes auteurs ont observé l'activation de cette aire lors de tâches de décision 'vivant ou non-vivant' de représentations imagées.

c) la suppression d'activité : compétition entre traits visuels ?

un signal plus fort en réponse aux distracteurs qu'aux cibles

Malgré le plus large recrutement des cortex occipito-temporaux durant le traitement des deux types de stimuli, et l'accroissement de signal qui y correspond, les aires différentiellement activées montrent un signal plus faible en réponse aux cibles qu'en réponse aux distracteurs. Cette diminution est donc relative, et ne peut être rapprochée des diminutions de signal observées lors de contrastes de tâches active vs passive, ce qui peut refléter des mécanismes d'inhibition de modalités sensorielles non-impliquées dans la tâche, des suppressions de réponse automatique lors d'une tâche cognitive difficile, ou un recrutement attentionnel durant une simple fixation (Haxby et al. 1994 ; Dehaene et al. 1996 ; Orban et al. 1997 ; Shulman et al. 1997 ; Vandenberghe et al. 1997 ; Petit et al. 1999). Cette diminution semble refléter au contraire un mécanisme propre au traitement visuel lié à la tâche.

Le décours temporel des signaux montrés dans la Figure 5 de l'article indiquent (comme les comparaisons entre 1.5 et 4s) un accroissement de signal après la présentation des stimuli, mais cet accroissement est rapidement stoppé pour les cibles et suivi d'une baisse rapide, qui se prolonge au-delà de 8s p.s. La durée de cet effet différentiel est notablement plus long que l'effet moteur observé dans l'aire 4.

recherche de cible ou suppression de réponse ?

A quel mécanisme cet effet peut-il être relié ? La première possibilité est que les stimuli distracteurs amènent à un traitement visuel prolongé si le mécanisme sous-jacent implique une recherche sérielle de traits associés à la cible de la tâche. Cette hypothèse émane des résultats des paradigmes de recherche visuelle (*visual search*), qui montrent que les sujets ont des temps de réaction plus courts quand une cible est présente dans le stimulus par rapport à un stimulus qui ne contient que des distracteurs (tâches à double choix Treisman et Gelade 1980). Cet effet est habituellement expliqué par un processus sériel qui se termine dès qu'une cible est trouvée, et qui est mené à terme dans le cas contraire. Si c'était le cas dans notre tâche, ce mécanisme pourrait expliquer la plus grande implication des aires activées différentiellement dans le traitement des distracteurs, et l'activité prolongée. Cependant, la réponse électrophysiologique à 150 ms ne montre pas de variations avec le temps de réponse comportemental, laissant supposer que le mécanisme sous-jacent n'est pas un processus sériel (Eriksen et al. 1986). Tous nos résultats de potentiels évoqués portent à croire qu'il s'agit d'un processus massivement parallèle.

Une autre possibilité est portée par les mécanismes de suppression d'activité neuronale liés à la détection de traits cibles. Plusieurs études électrophysiologiques d'enregistrements unitaires montrent en effet que les réponses de neurones corticaux à des stimulations visuelles peuvent dépendre fortement de l'importance comportementale des stimuli (Nelson 1991 ; Motter 1994 ; Maunsell 1995 ; Duncan et al. 1997 ; Borg-Graham et al. 1998 ; Kastner et al. 1998 ; Reynolds et al. 1999). Maunsell 1995 et Motter 1994 ont notamment décrit des réponses neuronales supprimées ou au contraire renforcées en fonction du trait du stimulus sur lequel porte l'attention d'un singe ; ces neurones répondent modérément à ces mêmes traits dans leur champ récepteur lorsque l'attention est dirigée sur une autre cible. D'autres suppressions de réponse ont pu être observées en électrophysiologie dans des tâches de recherche visuelle, et cet effet semble être d'autant plus fort que les stimuli sont complexes. Ces études suggèrent que cet effet reflète un mécanisme d'attention sélective qui supprime les réponses aux traits ou objets cibles non sélectionnés. Ces descriptions de réponses supprimées pourraient correspondre à la baisse relative du signal lié aux stimuli cibles par rapport aux distracteurs. Mais de quel mécanisme sous-jacent ces effets sont-ils le reflet ?

proposition de mécanisme

Nous proposons que cet effet reflète des mécanismes d'inhibition dès qu'un trait cible est détecté. En effet, certains auteurs ont suggéré que l'attention à la préparation agisse en pré-activant les représentations des traits cibles de la tâche ; lors du traitement du stimulus, la détection de ces traits cibles amènerait conjointement à l'inhibition des neurones non pré-activés par la préparation attentionnelle, mécanisme qui pourrait expliquer les effets de suppression décrits au paragraphe précédent (Desimone et Duncan 1995 ; Caputo et Guerra 1998 ; Fuentes et al. 1999, voir Figure 44).

Appliqué à notre tâche, ce mécanisme supposerait que les aires différentielles observées comportent une sous-population de neurones pré-activés par la préparation à la tâche, dont l'activation post-stimulus conduirait à inhiber l'activité des autres neurones de cette aire. Le signal observé sera moins ample pour les cibles que pour les distracteurs si ces neurones pré-activés sont beaucoup moins nombreux que les autres dans cette aire (et si l'hémodynamique est un indice de l'activité neuronale). De tels mécanismes d'inhibition sont documentés dans d'autres études sur la vision, et ont été décrits comme ayant une durée plus longue que des effets excitateurs (Luck et al. 1997 ; Caputo et Guerra 1998 ; Desimone 1998). C'est ce que nous avons observé dans notre étude, dans laquelle l'effet différentiel visuel est prolongé par rapport à l'effet différentiel moteur. Il est probable que l'activité motrice soit liée principalement à un effet excitateur des neurones dans ces aires.

Nous pouvons noter qu'un tel mécanisme peut rendre compte des résultats d'études sur l'amorçage (*priming*) qui ont montré paradoxalement qu'une meilleure reconnaissance et des temps de réaction plus courts pour des stimuli amorcés sont associés à des baisses de

signal IRM dans des aires perceptives ainsi qu'à des potentiels évoqués moins amples (Desimone 1996; Schacter et Buckner 1998 ; Wiggs et Martin 1998).

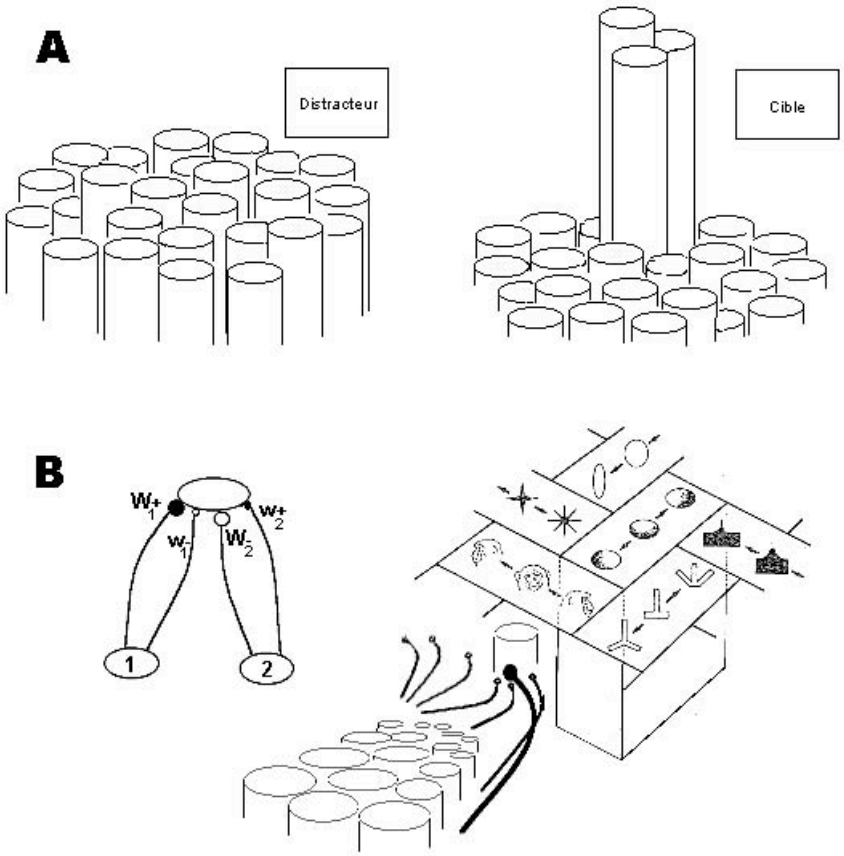


Figure 44 : suppression d'activité et inhibition en réponse aux stimuli attendus

A. Suppression d'activité dans une aire. L'activité d'une aire (mesurée à l'IRM) peut être la moyenne des activités individuelles des neurones. En réponse aux stimuli distracteurs, de nombreux neurones sont activés de façon moyenne ; en réponse à un stimuli cible, la petite assemblée répondant préférentiellement aux traits attendus est fortement activée, amenant une inhibition de la réponse de tout les autres.

B. Mécanisme de compétition. Ce modèle simple (à gauche) est proposé par Reynols, Chelazzi et Desimone 1999, pour rendre compte des effets de la présence de plusieurs traits visuels dans le champ récepteur de neurones de V2 et V4. Lorsque l'attention est portée au trait 1, les neurones répondant préférentiellement à ce trait ont le poids de leurs synapses efférentes qui augmente, qu'elles soient excitatrices ou inhibitrices. Dans notre cas, si ces connections inhibitrices forment un réseau dense (dessin de droite), l'activité des neurones efférents voisins peut être massivement supprimée.

BASES CEREBRALES DE LA CATEGORISATION

1 bilan des résultats

Lors des séries expérimentales, nous avons dégagé trois types de résultats : ceux qui démontrent la rapidité du traitement visuel de scènes naturelles, ceux qui ont trait aux informations utilisées dans ces traitements, et enfin ceux qui concernent les probables mécanismes cérébraux sous-jacents.

rapidité du traitement visuel

Le constat de rapidité avec laquelle est effectuée la reconnaissance visuelle s'impose. Dans les données comportementales, tout d'abord, avec l'extrême brièveté des temps de réaction. Pour chaque expérience, de façon indistincte quant aux groupes de sujets, les temps de réaction médians ont été pour Animal vs Non-Animal 445 ms, 414 ms, 441 ms, 483 ms ; Détection de la présence de couleur dans les scènes : 481 ms ; Détection de Noir et Blanc : 524 ms ; Carrés vs Ronds : 444 ms. Les temps de réaction les plus courts (divergence des distributions des essais réussis et des faux positifs) sont pour Animal vs Non-Animal : 270 ms, 280 ms, 290 ms, 330 ms ; Détection de la couleur dans les scènes **:** 310 ms ; Carrés vs Ronds : 300 ms. Ces mesures globales donnent déjà des indications précieuses : 300 ms sont suffisantes pour analyser une scène visuelle complexe sans aide contextuelle et pour générer une réponse comportementale appropriée. D'autre part, il n'y a pas de différence majeure entre l'analyse visuelle nécessaire pour déceler la présence d'un animal dans une scène naturelle, dire si cette scène est colorée, et faire la différence entre des carrés et des ronds. Si différence il y a, elle atteint au maximum 30 ms dans les temps de réaction comportementaux.

En électrophysiologie ensuite : la divergence à partir de 150 ms entre potentiels évoqués par les cibles et ceux évoqués par les distracteurs de la tâche a été observée tout d'abord dans la tâche Animal vs Non-Animal. Cette divergence a été aussi mise en évidence dans la tâche de détection de Couleur vs Noir et blanc, dans la tâche Noir et Blanc vs Couleur, et dans la tâche Carrés vs Ronds. Ces observations permettent d'attribuer cet effet à un phénomène de type décisionnel plutôt qu'à un encodage de catégorie visuelle. Cette latence de 150 ms à laquelle un tel phénomène débute est compatible avec les 300 ms des temps de réaction les plus précoces – une centaine de millisecondes étant généralement considérée comme suffisante pour déclencher et effectuer une action motrice. Ce résultat implique directement que 150 ms sont suffisantes pour que le système visuel reconnaisse la présence d'un animal dans une scène naturelle.

Pour aussi rapide qu'elle soit, cette tâche ne mobilise pas d'importantes ressources attentionnelles : le fait de partager l'attention visuelle sur trois lieux d'apparition possibles des images sur l'écran, sur une zone du champ visuel dont la surface a été doublée, ne modifie en aucune façon les résultats comportementaux. De même, aucune augmentation de latence n'a été observée en électrophysiologie. De nombreuses études ont pourtant montré les coûts comportementaux qu'induisait l'allocation d'attention supplémentaire dans diverses tâches. Ici, une plus grande attention spatiale a été mise en évidence, mais n'a induit aucun coût comportemental comme si la tâche de reconnaissance était pré-attentive. La reconnaissance semble d'autant plus résulter d'un mécanisme automatique qu'une trace électrophysiologique liée aux animaux dans les scènes est encore présente lorsqu'une autre tâche occupe des sujets non avertis.

Enfin, la rapidité du traitement visuel pour des scènes présentées latéralement (3.5°) semble être le fait d'un parallélisme massif : le coût comportemental est faible, en précision (3%) et en temps de réaction (12 ms), et la trace électrophysiologique de décision est à peine retardée (au maximum 10 ms). Une autre étude menée systématiquement à des excentricités plus grandes (Thorpe et al 99) n'attribue les mêmes coûts qu'au plus faible échantillonnage par la rétine. Le mécanisme de recentrage de l'entrée visuelle proposé par Kosslyn et par le modèle d'Olshausen ne semble donc pas intervenir dans cette tâche.

informations utilisées

Les informations sur lesquelles reposent ces traitements sont peu nombreuses, même pour une tâche d'analyse de scènes naturelles, a priori complexes et riches de détails. Les expériences en vision latérale ont montré que le plus faible échantillonnage de la rétine affecte peu le comportement et la décision visuelle ; une faible acuité est donc suffisante pour la reconnaissance. De même, l'absence de couleur dans les scènes n'affecte que de façon ténue les temps de réaction comportementaux et la latence de décision visuelle. Cette tâche pourrait donc essentiellement reposer sur les informations véhiculées par les voies magnocellulaires, rapides et achromatiques. Il est ainsi probable que la voie ventrale du système visuel soit beaucoup plus influencée par les informations magnocellulaires alors qu'on l'avait pensée principalement parvocellulaire (Desimone et al. 1985).

bases cérébrales

Les mécanismes cérébraux impliqués dans une telle tâche de reconnaissance recrutent l'ensemble des aires de la voie ventrale. Nous n'avons observé aucune supériorité hémisphérique claire dans la tâche 'Animal vs Non-Animal'. Deux types d'activations différenciées ont pu être mises en évidence : des activations précoces entre 100 et 140 ms, observées dans le cas de la couleur et des animaux. Pour la couleur, nous avons pu observer que ces activations ne sont pas modifiées par la tâche à accomplir (cibles ou distracteurs), et sont donc le reflet d'un encodage visuel. Tout laisse à penser que l'activation précoce observée dans le cas des animaux provienne d'un mécanisme

d'encodage similaire. Ensuite, les activations mises en évidence à partir de 150-170 ms différencient cibles et distracteurs quelle que soit la catégorie cible. Nous avons proposé que ce phénomène soit lié à une activité de type décisionnel, étant indépendant de ce qui est cible ou distracteur de la tâche. Cette différence à partir de 150 ms est très proche de la négativité de sélection (SN), observée dans des expériences en potentiels évoqués sur l'attention sélective à des traits visuels simples et à la couleur. Comme nous l'avons mentionné à la fin du deuxième chapitre, il est probable que nous observions cette négativité de sélection dans le cas de l'attention portée à la catégorie Animal. Il est remarquable cependant d'observer que la latence de cette sélection est identique à celles observées dans quelques études pour des traits visuels simples[1], ou comme nous l'avons vu, à des formes comme des carrés ou des ronds.

L'étude en IRMf révèle les gyri occipital médian, cingulaire et fusiforme comme étant le siège d'activités différenciées lors de la tâche. Cette étude en IRMf nous a permis de proposer que ces aires visuelles voient leur activité réduite lors de la détection de cibles, via un mécanisme d'inhibition supprimant l'activité des neurones non sélectionnés. Nous détaillerons cette proposition dans la troisième partie de ce chapitre.

2 contraintes de modélisation

8 étapes corticales

Si 150 ms sont suffisantes pour que le système visuel reconnaisse la présence d'un animal dans une scène naturelle, sachant que l'information visuelle atteint le cortex humain à partir de 70 ms, cela laisse 80 ms pour 8 étapes corticales, à condition que les premières décharges neuronales soit toujours sélectives.

traitements essentiellement ascendants

Si 4 aires corticales suffisent pour atteindre des neurones aux sélectivités élaborées, alors une seule grande boucle est possible : l'essentiel du traitement est effectué de façon ascendante. Ce constat permet d'affirmer que le traitement visuel de reconnaissance n'est pas de type itératif ou de relaxation - dans les tâches mises en œuvre dans cette étude.

parallélisme de représentations et de traitements

La rapidité avec laquelle se réalisent les traitements des entrées latéralisées n'autorise pas le recrutement permanent d'un mécanisme de recentrage. La reconnaissance ne peut

[1] Même pour des attributs visuels simples, les latences sont rarement aussi précoces. Si Smid et al. 1997 observe les premières déflexions à partir de 120 ms, elles ne sont visibles qu'à partir de 200 ms dans toutes les autres études.

s'effectuer dans ce cas que grâce à une architecture massivement parallèle, pouvant converger progressivement vers des représentations invariantes du lieu de présentation.

traitements basés sur la luminance

Nous avons montré que la rapidité du traitement visuel n'est pas affectée par l'absence de couleur. Elle n'est donc pas indispensable à une reconnaissance efficace, et ne participe que marginalement à l'analyse rapide des scènes visuelles.

attention non requise

Nous avons observé que la reconnaissance repose sur des mécanismes automatiques. D'autre part, le temps pris par le traitement ne semble pas compatible avec le recrutement d'un système de contrôle attentionnel guidant post-stimulus les étapes de reconnaissance (les relais frontaux, postulés par Kosslyn).

aide contextuelle non requise

La reconnaissance d'objet n'a pas besoin d'analyse globale de la scène ou d'analyse sémantique pour être efficace. Les modèles mettant en œuvre une 'sélection initiale' de représentations mémorisées et postulant l'émission descendante d'hypothèses après l'apparition du stimulus, ou qui utilisent l'imagerie pour aider à la reconnaissance, ne semblent pas être valides pour expliquer les traitements visuels rapides que nous avons observés.

codage neuronal efficace

80 ms de traitement cortical correspondent à l'émission d'une dizaine (maximum une vingtaine) de potentiels d'actions, à répartir sur 8 étapes corticales. Un code porté par la fréquence de décharge n'est pas envisageable (Gautrais et Thorpe 1998) ; la détection de coïncidence des entrées synaptiques semble beaucoup plus plausible. D'autre part, un mécanisme d'efficacité synaptique sensible à l'ordre d'arrivée des potentiels d'action permettrait de faire parvenir l'information la plus pertinente en premier, comme proposé par Thorpe et Gautrais.

3 la reconnaissance visuelle : propositions

la sélection de cible par inhibition

Dans les aires corticales où la sélectivité des neurones permet l'encodage de représentations élaborées, la reconnaissance d'objets peut être réalisée par un mécanisme de compétition. Nous proposons que cette compétition ne soit pas seulement restreinte aux traits visuels inclus dans le même champ récepteur d'un neurone, comme l'a proposé Desimone, mais que cette compétitivité soit étendue au sein d'une aire cérébrale composée de neurones aux sélectivités proches ; cette compétition va de pair avec l'inhibition des neurones non sélectionnés. Cette proposition est sous-tendue par nos résultats d'IRMf, où les activations ont été trouvées plus amples en réponse aux distracteurs qu'aux cibles, et où le décours de ces activités semble davantage provenir de mécanismes inhibiteurs qu'excitateurs.

Une illustration de cet effet est proposé Figure 44 : nous mesurons l'activité globale d'un grand nombre de neurones aux sélectivités proches. La première hypothèse est que la préparation attentionnelle augmente l'efficacité des synapses des neurones sélectifs à des traits cibles. La seconde hypothèse est que nous mesurons l'activité globale d'une aire activée par la catégorie cible : chaque neurone est sélectif à un exemplaire de cette catégorie, ou à une composante de cet exemplaire.

En réponse aux stimuli distracteurs de la tâche, de nombreux neurones de cette aire sont activés de façon moyenne, aucun neurone de cette aire ne répondant optimalement à un objet de la scène. En réponse à un stimulus cible de la tâche, la petite assemblée répondant préférentiellement aux traits du stimulus actuel est fortement activée. Cette forte activité est causée par l'activité des neurones afférents rendue plus efficace par la préparation attentionnelle. De la même façon, les projections inhibitrices de ce réseau sont aussi plus efficacement activées, et produisent une vaste inhibition de tous les autres neurones voisins. Mesurée globalement, l'aire a donc moins d'activité en réponse à un stimulus cible par rapport à un stimulus distracteur. Lors d'une présentation d'un autre stimulus cible, une autre assemblée sera sélectionnée, mais l'activité globale de l'aire sera réduite comme précédemment.

Ainsi nous proposons que l'effet observé à l'IRMf lors de la tâche soit le reflet d'un tel mécanisme. Ce mécanisme est rendu possible si l'attention de préparation a pour effet d'augmenter l'efficacité des synapses excitatrices et inhibitrices des neurones, comme proposé par Reynolds, Chelazzi et Desimone 1999.

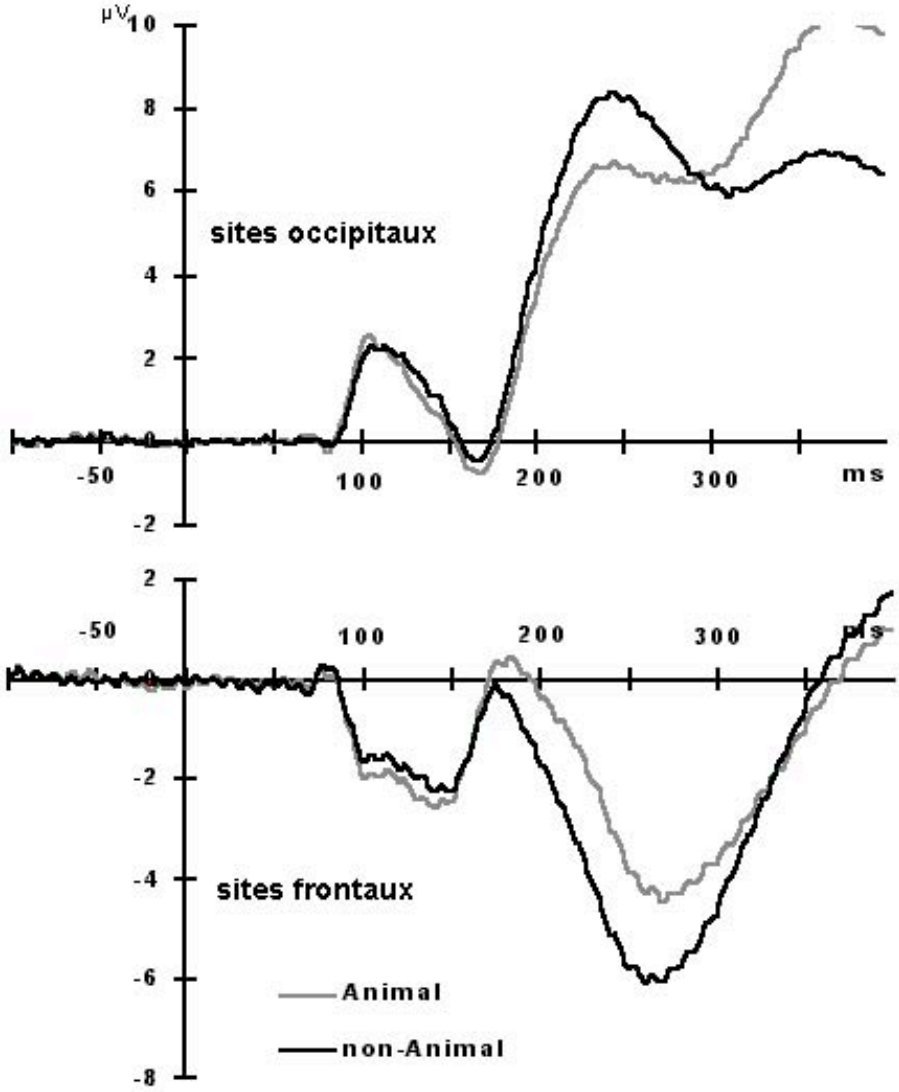


Figure 45 : potentiels 'Animal' et 'Non-Animal' lors de la tâche de catégorisation

Malgré les différences de polarité, les potentiels évoqués par les distracteurs sont toujours plus amples que ceux des cibles, pour l'activité différentielle principale ; cet effet est observable sur tous les sites où cette différence est significative. Le signe de cette différence est analogue à l'effet observé en IRMf. Pour l'activité précoce visible entre 110 et 140 ms, cet effet ne peut être mis en évidence de la même façon. Les données sont celles de l'expérience 'la couleur dans les scènes'.

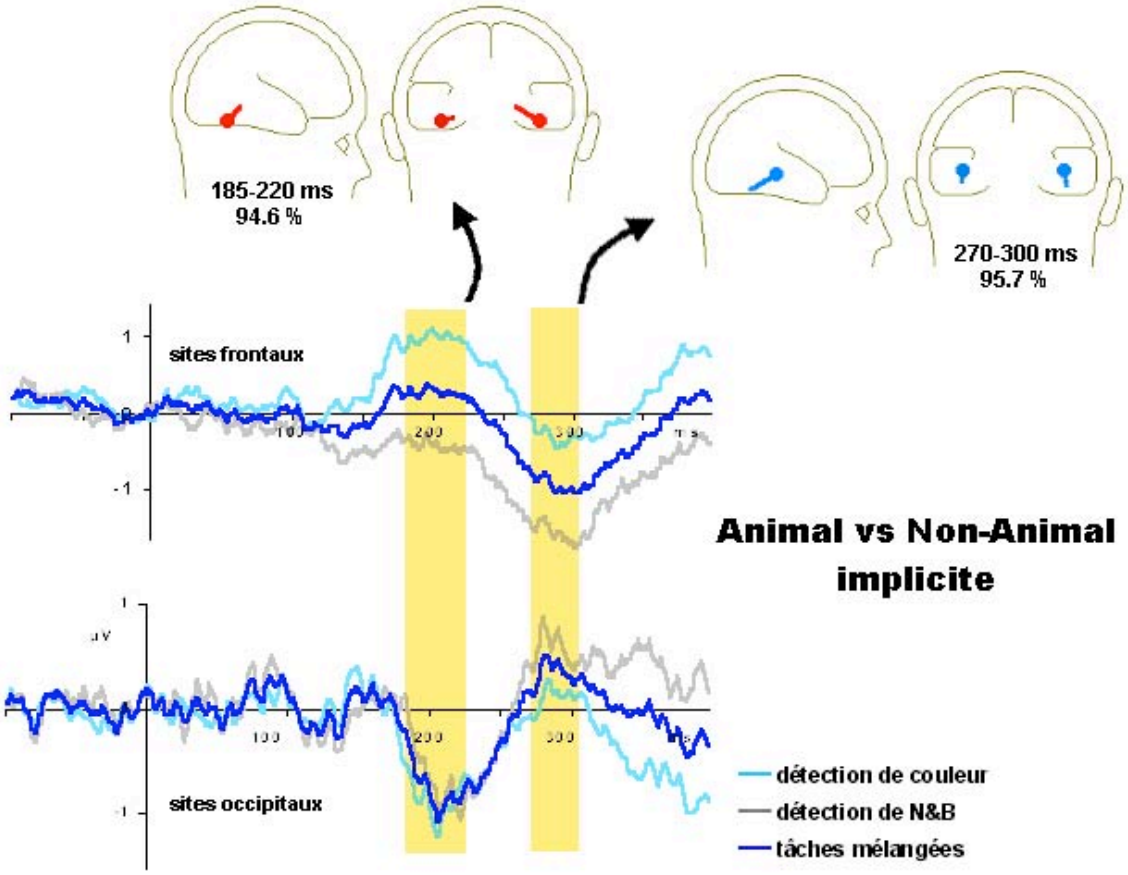


Figure 46 : Modèle des activités 'Animal vs Non-Animal implicite'

Les données sont celles de l'expérience 'détection de couleur dans les scènes'. Les potentiels sont ceux des différences 'Animal vs Non-Animal' dans le cas des deux tâches duales (les sujets sont différents). Sur les sites frontaux, il est remarquable d'observer que malgré des tâches et des sujets différents, les potentiels possèdent les mêmes variations sur 180-400 ms. Cette propriété ne peut provenir que d'une activité propre liée à 'Animal vs Non-Animal'.

Deux conclusions peuvent être tirées de ces modèles : d'abord, les localisations sont identiques aux sources 'Animal vs Non-Animal explicite' de toutes les expériences, laissant penser qu'il s'agit des mêmes aires corticales recrutées dans le cas implicite et lors de la tâche Animal (la première est identique à l'activité 'précoce', la seconde à 'l'activité différentielle à 150 ms'). Ensuite, la polarité de ces sources (le signe de la différence) est l'opposé du cas explicite. Cette propriété est très consistante avec notre hypothèse de suppression d'activité lors de la tâche : une aire plus activée par les scènes d'animaux que par les autres scènes lors d'une reconnaissance implicite verra son activité très réduite lorsque les animaux sont cibles de la tâche, par inhibition de tous les neurones non sélectionnés de la même aire. Cet effet de sélection rend l'activité évoquée par les distracteurs comparativement plus élevée.

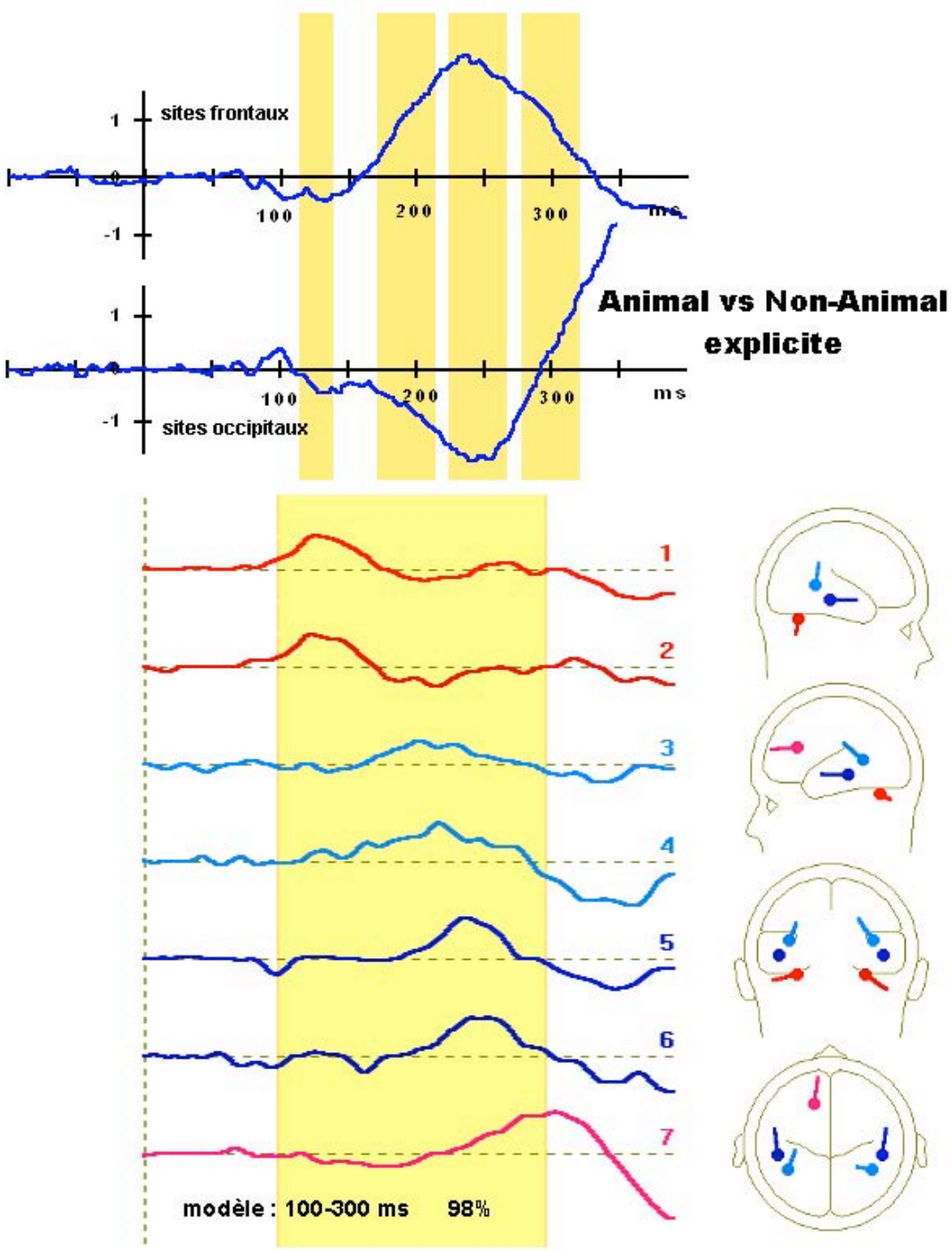


Figure 47 : modèle des activités 'Animal vs non-Animal explicite'

Les données sont celles de l'expérience 'la couleur dans les scènes' : les sujets effectuent la tâche de catégorisation des stimuli animaux vs autres scènes. Nous avons effectué une modélisation des différences sur l'intervalle le plus long possible, contrairement aux autres modèles présentés dans les comptes-rendus expérimentaux (chapitres 2 et 3). Ce modèle est globalement consistant avec tous ceux présentés auparavant. Il indique surtout que l'activité différentielle principale doit être divisée en deux phénomènes, le premier à dynamique lente sur 150-250 ms proche de la jonction occipito-pariétale, le second plus restreint sur 200-250 ms plus antérieur et temporal. L'activité préfrontale gauche a son pic à 310 ms.

suppression visible en potentiels évoqués ?

Par prolongement, nous proposons à titre d'hypothèse d'interpréter les résultats de potentiels évoqués comme un effet du même mécanisme d'inhibition. Un retour sur les potentiels évoqués nous montre en effet un phénomène analogue à celui observé à l'IRMf : sur toutes les électrodes où la différence cible vs distracteur est visible à partir de 150 ms, les potentiels distracteurs sont toujours plus amples que les potentiels cibles[1], comme l'illustre la Figure 45. Cet effet est identique à la 'négativité de sélection', en référence au signe de la différence.

Mais un examen plus détaillé des modèles des activités 'Animal vs Non-Animal' dans les cas explicite et implicite nous montre un phénomène très compatible avec notre proposition de suppression d'activité, lors de la tâche. Comme montré sur les Figures 46 et 47, le modèle 'Animal vs non Animal implicite' possède des sources localisées aux mêmes endroits que le modèle explicite : les mêmes aires sont différentiellement activées par comparaison des mêmes stimuli (stimuli identiques, tâches et sujets différents). Ce premier résultat permet de conclure au partage des bases cérébrales des reconnaissances implicite et explicite de la catégorie Animal dans des scènes. Seulement, le fait crucial est que ces différences 'Animal vs Non-Animal' localisées dans les mêmes aires sont de signe opposé entre les conditions implicite et explicite, ce qui est exprimé par la rotation proche de 180° des dipôles, et par le signe des potentiels. Voici comment nous en envisageons l'interprétation.

Dans la condition implicite, posons l'hypothèse que ces aires peuvent être davantage activées par les scènes d'animaux que par les autres scènes : cette aire rassemble des neurones de sélectivités voisines, sélectivités à des formes plus proches d'être vivants et d'animaux que d'autres formes souvent rencontrées dans des scènes naturelles. L'activation évoquée par les autres scènes est cependant non nulle. Dans le cas d'une reconnaissance implicite, l'activité 'Animal' est plus grande en valeur absolue que celle évoquée par 'Non-Animal', et leur différence est ici positive.

Dans le cas explicite, lors de la tâche de catégorisation Animal, l'attention de préparation portée à cette catégorie a pour effet d'amplifier dans cette aire la sensibilité des neurones à leurs entrées dendritiques. Lorsqu'une scène cible est présentée, l'activation des quelques neurones très sélectifs aux formes de l'animal actuel produit également une vaste inhibition des autres neurones de la même aire. Cette inhibition a pour double effet de lever la sensibilité accrue par l'attention, et d'agir comme une sélection des neurones représentant l'animal 'reconnu'. Si une scène distracteur est présentée, l'activité des neurones afférents est insuffisante, ou incohérente, et ne produit pas de sélection

[1] Cette proposition n'est valide que si la somme des activités sur tous les sites est plus grande pour distracteur que pour cible. Ici la topographie de surface de l'activité différentielle 'à partir de 150 ms' permet d'affirmer sans risque que le champ évoqué par les distracteurs est plus intense. Mais une approche plus rigoureuse serait de projeter l'activité de modèles cible et distracteur sur une sphère fermée et d'intégrer les variations des flux au cours du temps.

d'assemblée neuronale dans cette aire. L'activité globale est cependant non nulle, égale au cas implicite, ou amplifiée par la sensibilité des neurones à leurs entrées. En conséquence, cet effet de sélection de cible rend l'activité évoquée par les distracteurs comparativement plus élevée : dans cette aire, l'activité 'Animal' est plus petite en valeur absolue que celle évoquée par les conditions 'Non-Animal', et leur différence est négative.

Ainsi, nous proposons que ce mécanisme d'inhibition rende compte du changement de signe de la différence 'Animal vs Non-Animal implicite et explicite' observée dans les potentiels évoqués. Il reste à vérifier si le même phénomène est visible à l'IRMf en mettant en œuvre une expérience adaptée : lors d'une catégorisation implicite, lorsque les sujets sont engagés dans une autre tâche, un certain nombre d'aires visuelles doivent être différentiellement activées par comparaison de deux catégories (ou une comparaison d'une catégorie et de scènes, ou de non-objets). Lors de la tâche de catégorisation explicite utilisant ces mêmes types de stimuli, on devrait observer des activations localisées dans les mêmes aires, au moyen des mêmes comparaisons, mais cette fois le signe de leur différence devrait être opposé. De même, des enregistrements électrophysiologiques ou l'imagerie infrarouge devraient permettre de vérifier si des groupes de neurones aux sélectivités proches voient leur activité supprimée dans les mêmes conditions expérimentales.

4 Conclusion

Nous voudrions pour finir souligner le type d'approche et les choix qui ont permis de mener à bien cette étude.

tâche go / no-go

Nous avons dès le début cherché à simplifier au maximum la tâche effectuée par les sujets, afin de réduire le nombre d'hypothèses sur les processus mentaux et les aires recrutées. De ce point de vue, la tâche go / no-go est idéale puisqu'elle est la réalisation la plus simple d'une tâche de choix. Comme nous l'avons constaté, elle rend possible des temps de réaction très courts, ce qui permet d'extraire des contraintes temporelles précises. D'autre part l'analyse des faux positifs est simple, par comparaison d'une tâche de choix où les sujets ont deux boutons réponse : l'erreur ne provient pas d'une confusion sur les gestes à accomplir, elle est indubitablement causée par l'analyse visuelle[1]. La comparaison des essais réussis et des faux positifs permet alors de donner de précieuses indications sur les traitements effectués, soit par analyse des stimuli ayant évoqué ces erreurs, soit par comparaison des temps de réaction, soit par comparaison de mesures conjointes : potentiels évoqués ou images fonctionnelles.

[1] ou par l'anticipation du sujet. Mais les temps de réaction sont alors inférieurs aux temps de réaction simples.

scènes naturelles

Puis l'utilisation d'images naturelles et leur présentation extrêmement brève a permis d'avoir une autre idée de la puissance du système visuel. Le recourt à un grand nombre de photographies numérisées n'a été possible que depuis quelques années ; mais la plupart des auteurs qui n'utilisaient jusqu'à présent que des stimuli simples, n'envisageaient les traitements visuels de stimuli complexes ou naturels que comme une sorte de construction hautement élaborée nécessitant de longues procédures récurrentes, où le contexte des scènes jouait une rôle explicite et déterminant. Nous avons vu que pour expliquer les performances étonnantes du système visuel sur des scènes naturelles, les modèles de la reconnaissance devaient prendre en compte une architecture plus directe.

les différences entre conditions, et leurs latences

A chaque mesure, nous avons cherché à extraire les premiers intervalles de temps où des différences entre conditions expérimentales apparaissent : ces latences expriment une modification des processus associés aux images et à la tâche, et permettent d'obtenir des données temporelles précises. Ce point de vue a été appliqué à la fois aux histogrammes de temps de réaction comportementaux et aux potentiels évoqués, en contraste des analyses classiques qui ne relèvent le plus souvent que les mesures moyennes et les pics d'activité. Ces latences peuvent être utilement comparées lorsque les conditions expérimentales changent, ou lorsqu'elles sont contrebalancées. Nous avons en outre tenté d'établir des modèles dipolaires des différences de potentiels : malgré la justesse des réticences méthodologiques, nous avons observé une grande robustesse des localisations au travers des différentes conditions, des sujets et des tâches, sans prétendre à une explication complète.

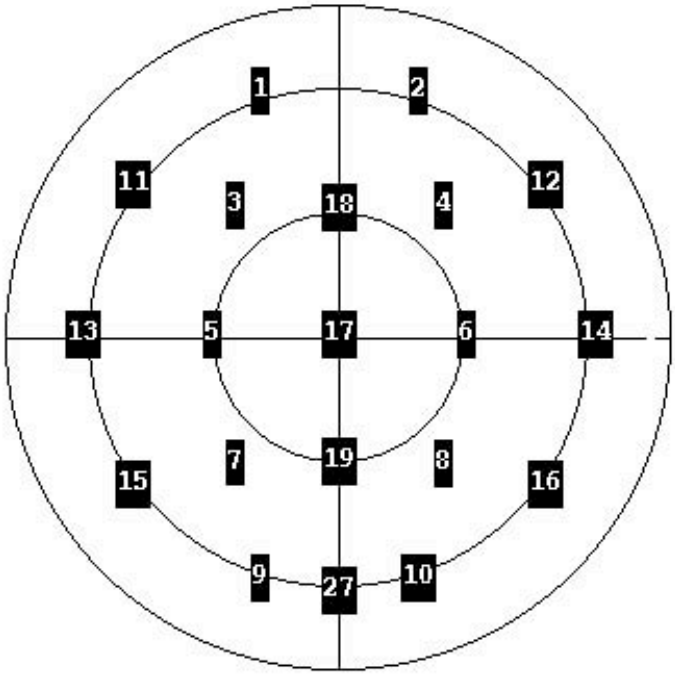
IRMf : un protocole utilisable

Enfin, nous avons cherché à utiliser l'outil IRM de manière à placer les sujets dans des conditions expérimentales moins artificielles que celles des méthodes classiques. Les méthodologies par blocs ne permettent pas des analyses précises des tâches de choix, à cause du mélange des types de stimulation dans un même bloc ; quant aux protocoles événementiels, ils nécessitent de très longues procédures étant donné la quinzaine de secondes que dure l'interstimulus. Nous avons élaboré dans cette étude des séquences de stimulation permettant de contrôler les successions de stimuli, afin de pouvoir adapter directement les protocoles d'électrophysiologie en IRMf. L'intérêt important de cette méthode est de permettre un lien direct entre les diverses mesures fonctionnelles : des expériences peuvent être réalisées au moyen de plusieurs techniques sans modification drastique de l'état des sujets, des stimuli, des tâches et des analyses.

Au terme de cette étude, nous espérons que ces résultats et ces propositions auront permis d'accroître la connaissance du système visuel et de son fonctionnement. Nous espérons avoir contribué par là à la compréhension des phénomènes plus généraux de catégorisation, capacité à notre avis à l'origine des processus mentaux de plus haut niveau, et dont l'étude dans le champ des neurosciences cognitives nous semble prometteuse.

ANNEXE

		théta	phi
1	FP1	-92	-72
2	FP2	92	72
3	F3	-60	-51
4	F4	60	51
5	C3	-46	0
6	C4	46	0
7	P3	-60	51
8	P4	60	-51
9	O1	-92	72
10	O2	92	-72
11	F7	-92	-36
12	F8	92	36
13	T7	-92	0
14	T8	92	0
15	P7	-92	36
16	P8	92	-36
17	Cz	0	0
18	Fz	46	90
19	Pz	46	-90
20	Oz	92	-90



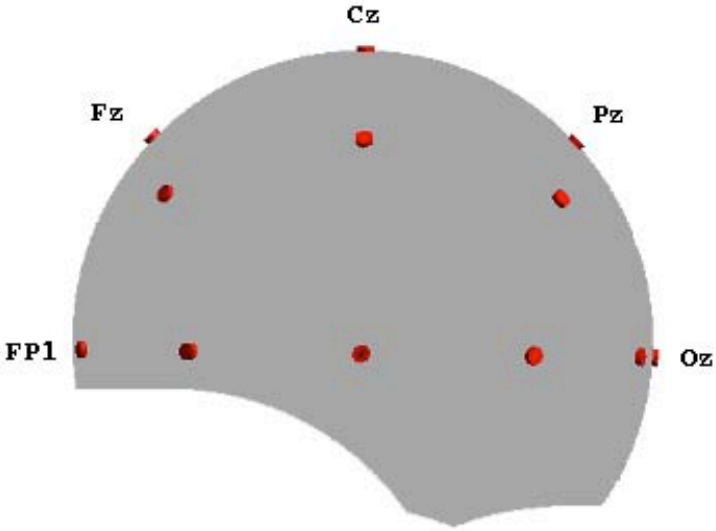
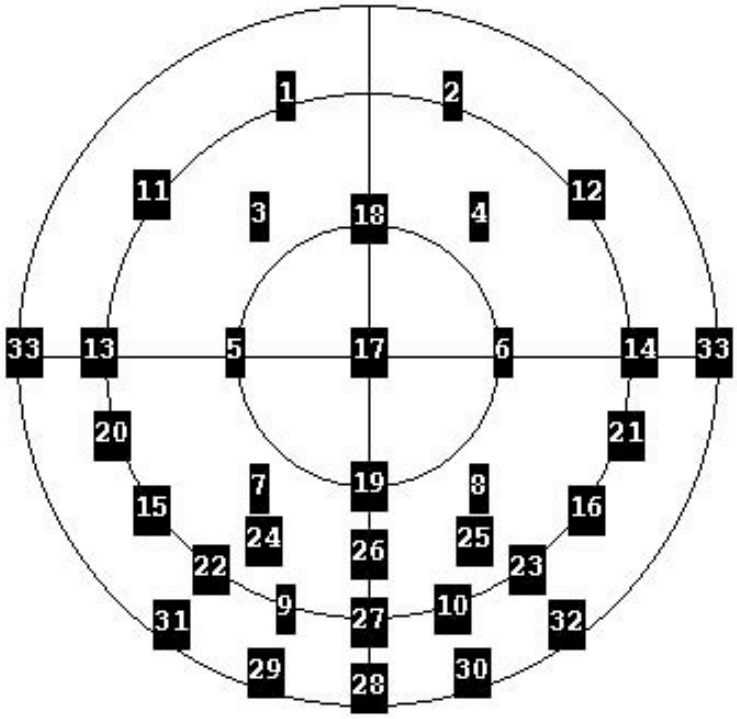


Figure 48 : bonnet à 20 électrodes

Le bonnet à 20 électrodes d'enregistrement a été utilisé lors de l'expérience standard. En colonne est indiqué le numéro des canaux, le nom des électrodes (système 10-20), les coordonnées sphériques. La figure du haut représente une vue de dessus aplatie, la figure du bas une vue du côté gauche. Une électrode FPZ supplémentaire était reliée à masse (entre FP1, FP2 et FZ) ; la référence était donnée par le potentiel oreilles jointes.

		théta	phi
1	FP1	-92	-72
2	FP2	92	72
3	F3	-60	-51
4	F4	60	51
5	C3	-46	0
6	C4	46	0
7	P3	-60	51
8	P4	60	-51
9	O1	-92	72
10	O2	92	-72
11	F7	-92	-36
12	F8	92	36
13	T7	-92	0
14	T8	92	0
15	P7	-92	36
16	P8	92	-36
17	Cz	0	0
18	Fz	46	90
19	Pz	46	-90
20	TP7	-92	18
21	TP8	92	-18
22	PO7	-92	54
23	PO8	92	-54
24	PO3	-74	61
25	PO4	74	-61
26	POz	69	-90
27	Oz	92	-90
28	Iz	115	-90
29		-115	72
30		115	-72
31	PO9	-115	54
32	PO10	115	-54



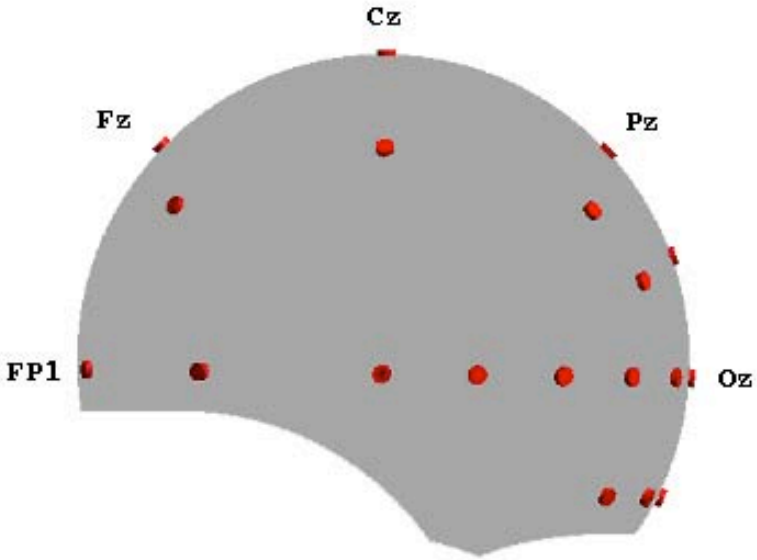


Figure 49 : bonnet à 32 électrodes

Le bonnet à 32 électrodes d'enregistrement a été utilisé lors de toutes les expériences sauf la première. En colonne est indiqué le numéro des canaux, le nom des électrodes (système 10-20 étendu), les coordonnées sphériques. Le montage était identique au précédent : électrode de masse FPZ, référence oreilles jointes.

COMMUNICATIONS

Publications et Chapitres

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

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

Fize D., Boulanouar K., Chatel Y., Ranjeva J.P., Fabre-Thorpe M. and Thorpe S. (2000). Brain areas involved in rapid categorization of natural images : an event-related fMRI study*, NeuroImage*, *sous-presse.*

Fize, D., Fabre-Thorpe, M., Richard, G., & Thorpe, S. Rapid categorisation of natural images without foveal vision : a behavioural and ERP study. *European Journal of Neuroscience, (en préparation)*.

Conférences publiées

18th European Conference on Visual Perception, (Tübingen, Allemagne)
Thorpe S. J., Fize D. & Marlot C. M. (1995). Categorization times for natural images. Perception, 24 (suppl), 8.

5th Cognitive Neuroscience Society Meeting (San Francisco, U.S.A.)
Fize D., Boulanouar K., Ranjeva J. P., Fabre-Thorpe M. & Thorpe S. J. (1998). Brain activity during rapid scene categorisation: A study using event-related fMRI. *J. Cognitive Neuroscience (suppl), 72.*

21st European Conference on Visual Perception (Oxford, G-B)

22nd European Conference on Visual Perception (Trieste, Italie).
Fabre-Thorpe, M., Fize, D., Aubertin, A., & Thorpe, S. (1999). Speed of processing in a go/no-go visual categorisation : low vs. high stimulus complexity. *Perception, 28 (suppl), 100.*

Autres Conférences

Fize D., Thorpe S. & Boulanouar K. (1997). Vers des protocoles communs IRMf-PEV: méthodologie. *Approche des fonctions cognitives par les méthodes de neuro-imagerie, (Cabourg, France).*

Thorpe S. & Fize D. (1997). Visual Categorization of Natural Images in Humans. *Approche des fonctions cognitives par les méthodes de neuro-imagerie, (Cabourg, France).*

BIBLIOGRAPHIE

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

Aine, C. J. and Harter, M. R. 1986. Visual event-related potentials to colored patterns and color names: attention to features and dimension. *Electroencephalogr Clin Neurophysiol* **64**: 228-245.

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M. and Spencer, D. D. 1994. Face recognition in human extrastriate cortex. *J Neurophysiol* **71**: 821-825.

Allison, T., McCarthy, G., Nobre, A., Puce, A. and Belger, A. 1994. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb Cortex* **4**: 544-554.

Anllo Vento, L. 1995. Shifting attention in visual space: the effects of peripheral cueing on brain cortical potentials. *Int J Neurosci* **80**: 353-370.

Anllo Vento, L. and Hillyard, S. A. 1996. Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Percept Psychophys* **58**: 191-206.

Anllo-Vento, L., Luck, S. J. and Hillyard, S. A. 1998. Spatio-temporal dynamics of attention to color: evidence from human electrophysiology. *Hum Brain Mapp* **6**: 216-238.

Antes, J. R. 1974. The time course of picture viewing. *J Exp Psychol* **103**: 62-70.

Bachinski, H. S. and Bachrach, V. R. 1980. Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception and Psychophysics* **28**: 241-248.

Badgaiyan, R. D., Schacter, D. L. and Alpert, N. M. 1999. Auditory priming within and across modalities: evidence from positron emission tomography. *J Cogn Neurosci* **11**: 337-348.

Baker, C. L., Boulton, J. C. and Mullen, K. T. 1998. A nonlinear chromatic motion mechanism. *Vision Res* **38**: 291-302.

Banquet, J. P., Renault, B. and Lesevre, N. 1981. Effect of task and stimulus probability on evoked potentials. Biol Psychol 13: 203-214.

Barrett, S. E. and Rugg, M. D. 1990. Event-related potentials and the phonological matching of picture names. *Brain Lang* **38**: 424-437.

Barrett, S. E. and Rugg, M. D. 1990. Event-related potentials and the semantic matching of pictures. *Brain Cogn* **14**: 201-212.

Barrett, S. E., Rugg, M. D. and Perrett, D. I. 1988. Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia* **26**: 105-117.

Benevento, L. A. and Port, J. D. 1995. Single neurons with both form/color differential responses and saccade- related responses in the nonretinotopic pulvinar of the behaving macaque monkey. *Vis Neurosci* **12**: 523-544.

Bentin, S., Allison, T., Puce, A., Perez, E. and McCarthy, G. 1996. Electrophysiological studies of face perception in humans. *J Cognitive Neurosci* **8**: 551-565.

Bentin, S. and Peled, B. S. 1990. The contribution of task-related factors to ERP repetition effects at short and long lags. *Mem Cognit* **18**: 359-366.

Biederman, I. 1972. Perceiving real-world scenes. *Science* **177**: 77-80.

Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* **94**: 115-147.

Biederman, I., Blickle, T. W., Teitelbaum, R. C., Klatsky, G. J. and Mezanotte, R. J. 1987. Object identification in multi-object, non-scene displays. . *J. Exp. Psychol. Learn. Mem. Cogn.* **in press**: -.

Biederman, I. and Cooper, E. E. 1991. Object recognition and laterality: null effects. *Neuropsychologia* **29**: 685-694.

Biederman, I. and Cooper, E. E. 1991. Priming contour-deleted images: evidence for intermediate representations in visual object recognition. *Cognit Psychol* **23**: 393-419.

Biederman, I. and Ju, G. 1988. Surface versus edge-based determinants of visual recognition. *Cognit Psychol* **20**: 38-64.

Biederman, I., Mezzanotte, R. J. and Rabinowitz, J. C. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cognit Psychol* **14**: 143-177.

Bobes, M. A., Valdes Sosa, M. and Olivares, E. 1994. An ERP study of expectancy violation in face perception. *Brain Cogn* **26**: 1-22.

Boddy, J. and Weinberg, H. 1981. Brain potentials, perceptual mechanisms and semantic categorisation. *Biol Psychol* **12**: 43-61.

Booth, M. C. and Rolls, E. T. 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* **8**: 510-523.

Borg-Graham, L. J., Monier, C. and Fregnac, Y. 1998. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393**: 369-373.

Botzel, K. and Grusser, O. J. 1989. Electric brain potentials evoked by pictures of faces and non-faces: a search for "face-specific" EEG-potentials. *Exp Brain Res* **77**: 349-360.

Boucart, M. 1996. *La reconnaissance des objets,* Presses Universitaires de Grenoble

Boucart, M. and Bruyer, R. 1991. Influence of physical and semantic information in a categorisation task of fragmented forms. *Perception* **20**: 403-414.

Boucart, M. and Humphreys, G. W. 1997. Integration of physical and semantic information in object processing. *Perception* **26**: 1197-1209.

Boynton, G. M., Engel, S. A., Glover, G. H. and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* **16**: 4207-4221.

Braun, J. and Julesz, B. 1998. Withdrawing attention at little or no cost: detection and discrimination tasks. *Percept Psychophys* **60**: 1-23.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. and Movshon, J. A. 1996. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* **13**: 87-100.

Broadbent, D. and Broadbent, M. H. 1990. Human attention: the exclusion of distracting information as a function of real and apparent separation of relevant and irrelevant events. *Proc R Soc Lond B Biol Sci* **242**: 11-16.

Bruce, V., Burton, A. M. and Craw, I. 1992. Modelling face recognition. *Philos Trans R Soc Lond Biol* **335**: 121-127.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E. and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging [see comments]. *Proc Natl Acad Sci U S A* **93**: 14878-14883.

Buckner, R. L. and Koutstaal, W. 1998. Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proc Natl Acad Sci U S A* **95**: 891-898.

Bulthoff, H. H., Edelman, S. Y. and Tarr, M. J. 1995. How are three-dimensional objects represented in the brain? *Cereb Cortex* **5**: 247-260.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R. and Dale, A. M. 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport* **9**: 3735-3739.

Calkins, D. J. and Sterling, P. 1999. Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron* **24**: 313-321.

Caputo, G. and Guerra, S. 1998. Attentional selection by distractor suppression. *Vision Res* **38**: 669-689.

Cattell, J. M. 1886. The time it takes to see and name objects. *Mind* **11**: 63-65.

Carr, T. H., McCauley, R. D., Sperber, R. D. and Parmelee, C. M. 1982. Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance* **8**: 757-777.

Cavanagh, P. and Anstis, S. 1991. The contribution of color to motion in normal and color-deficient observers. *Vision Res* **31**: 2109-2148.

Cavanagh, P., Henaff, M. A., Michel, F., Landis, T., Troscianko, T. and Intriligator, J. 1998. Complete sparing of high-contrast color input to motion perception in cortical color blindness. *Nature Neuroscience* **1**: 242-247.

Celebrini, S., Thorpe, S., Trotter, Y. and Imbert, M. 1993. Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci* **10**: 811-825.

Chelazzi, L., Miller, E. K., Duncan, J. and Desimone, R. 1993. A neural basis for visual search in inferior temporal cortex. *Nature* **363**: 345-347.

Cohen, D. J. 1997. Visual detection and perceptual independence: assessing color and form. *Percept Psychophys* **59**: 623-635.

Colby, C. L., Duhamel, J. R. and Goldberg, M. E. 1993. Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J Neurophysiol* **69**: 902-914.

Cox, R. W. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**: 162-173.

Czigler, I., Balazs, L. and Lenart, A. 1998. Attention to features of separate objects: an ERP study of target- shooters and control participants. *Int J Psychophysiol* **31**: 77-87.

Dale, A. M. and Buckner, R. L. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* **5**: 329-340.

De Graef, P., Christiaens, D. and d'Ydewalle, G. 1990. Perceptual effects of scene context on object identification. *Psychol Res* **52**: 317-329.

de Mattos Pimenta, M. A. and Tiedemann, K. B. 1990. Hemispheric asymmetries in the recognition of animal photographs in different views. *Braz J Med Biol Res* **23**: 1201-1204.

Dehaene, S. 1993. Temporal oscillations in human perception. *Psychological Science* **4**: 264-270.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J. and Mazoyer, B. 1996. Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia* **34**: 1097-1106.

Delorme, A., Fabre-Thorpe, M., Richard, G., Fize, D. and Thorpe, S. 1998. Ultra-Rapid Visual Categorisation of natural scenes is color-blind. *Perception* **(in press)**.

Delorme, A., Richard, G. and Fabre-Thorpe, M. 1999. Rapid processing of complex natural scenes : A role for the magnocellular visual pathways? *Neurocomputing* **In press**.

DeSchepper, B. and Treisman, A. 1996. Visual memory for novel shapes: implicit coding without attention. *J Exp Psychol Learn Mem Cogn* **22**: 27-47.

Desimone, R. 1991. Face-Selective Cells in the Temporal Cortex of Monkeys. *J Cognitive Neurosci* **3**: 1-8.

Desimone, R. 1996. Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A* **93**: 13494-13499.

Desimone, R. 1998. Visual attention mediated by biased competition in extrastriate visual cortex. *Philos Trans R Soc Lond B Biol Sci* **353**: 1245-1255.

Desimone, R. and Duncan, J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* **18**: 193-222.

Desimone, R., Schein, S. J., Moran, J. and Ungerleider, L. G. 1985. Contour, color and shape analysis beyond the striate cortex. *Vision Res* **25**: 441-452.

D'Esposito, M., Zarahn, E. and Aguirre, G. K. 1999. Event-related functional MRI: implications for cognitive psychology. *Psychol Bull* **125**: 155-164.

Dobkins, K. R. and Albright, T. D. 1994. What happens if it changes color when it moves?: the nature of chromatic input to macaque visual area MT. *J Neurosci* **14**: 4854-4870.

Downing, C. J. 1988. Expectancy and visual-spatial attention: effects on perceptual quality. *J Exp Psychol Hum Percept Perform* **14**: 188-202.

Duncan, J., Humphreys, G. and Ward, R. 1997. Competitive brain activity in visual attention. *Curr Opin Neurobiol* **7**: 255-261.

Durnford, M. and Kimura, D. 1971. Right hemisphere specialization for depth perception reflected in visual field differences. *Nature* **231**: 394-395.

Eriksen, B. A., Eriksen, C. W. and Hoffman, J. E. 1986. Recognition memory and attentional selection: serial scanning is not enough. *J Exp Psychol Hum Percept Perform* **12**: 476-483.

Eriksen, C. W. and Hoffman, J. E. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Perception and Psychophysics* **12**: 201-204.

Eriksen, C. W. and Murphy, T. D. 1987. Movement of attentional focus across the visual field: a critical look at the evidence. *Percept Psychophys* **42**: 299-305.

Eriksen, C. W. and Yeh, Y. Y. 1985. Allocation of attention in the visual field. *J Exp Psychol Hum Percept Perform* **11**: 583-597.

Fabre-Thorpe, M., Richard, G., Fize, D. and Thorpe, S. 1997. Rapid categorization of extrafoveal natural images : Implications for biological models. *6th Computational Neuroscience Meeting*, BigSky, Montana.

Farah, M. J. 1984. The neurological basis of mental imagery: a componential analysis. *Cognition* **18**: 245-272.

Felleman, D. J. and Van Essen, D. C. 1987. Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *J Neurophysiol* **57**: 889-920.

Fiorentini, A. 1989. Differences between fovea and parafovea in visual search processes. *Vision Res* **29**: 1153-1164.

Fraisse, P. 1984. Perception and estimation of time. Annu Rev Psychol 35: 1-36.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D. and Turner, R. 1998. Event-related fMRI: characterizing differential responses. *Neuroimage* **7**: 30-40.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D. and Frackowiack, R. S. J. 1995. Statistical Parametric Maps in Functional Imaging: A general linear approach. *Hum. Brain. Map.* **2**: 89-210.

Fuentes, L. J., Boucart, M., Alvarez, R., Vivas, A. B. and Zimmerman, M. A. 1999. Inhibitory processing in visuospatial attention in healthy adults and schizophrenic patients. *Schizophr Res* **40**: 75-80.

Fujita, I., Tanaka, K., Ito, M. and Cheng, K. 1992. Columns for visual features of objects in monkey inferotemporal cortex [see comments]. *Nature* **360**: 343-346.

Fukushima, K. 1975. Cognitron : A self-organizing multilayered neural network. . *Biol. Cybern.* **20**: 121-136.

Gautrais, J. 1997. Théorie et simulations d'un nouveau type de codage impulsonnel pour le traitement visuel rapide : le codage par l'ordre d'activation, E.H.E.S.S.

Gautrais, J. and Thorpe, S. 1998. Rate coding versus temporal order coding: a theoretical approach. *Biosystems* **48**: 57-65.

Gegenfurtner, K. R., Kiper, D. C., Beusmans, J. M., Carandini, M., Zaidi, Q. and Movshon, J. A. 1994. Chromatic properties of neurons in macaque MT. *Vis Neurosci* **11**: 455-466.

Gemba, H. and Sasaki, K. 1989. Potential related to no-go reaction of go/no-go hand movement task with color discrimination in human. *Neurosci Lett* **101**: 263-268.

George, N., Evans, J., Fiori, N., Davidoff, J. and Renault, B. 1996. Brain events related to normal and moderately scrambled faces. *Brain Res Cogn Brain Res* **4**: 65-76.

George, N., Jemel, B., Fiori, N. and Renault, B. 1997. Face and shape repetition effects in humans: A spatio- temporal ERP study. *Neuroreport* **8**: 1417-1423.

Gordon, R. D. and Irwin, D. E. 1996. What's in an object file? Evidence from priming studies. *Percept Psychophys* **58**: 1260-1277.

Gray, C. M., Konig, P., Engel, A. K. and Singer, W. 1989. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**: 334-337.

Grossman, M., Galetta, S. and Desposito, M. 1997. Object recognition difficulty in visual apperceptive agnosia. *Brain Cognition* **33**: 306-342.

Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K. and Clarke, M. 1994. Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes. *J Physiol Paris* **88**: 1-50.

Harter, M. R., Aine, C. and Schroeder, C. 1984. Hemispheric differences in event-related potential measures of selective attention. *Ann N Y Acad Sci* **425**: 210-211.

Harter, M. R. and Guido, W. 1980. Attention to pattern orientation: negative cortical potentials, reaction time, and the selection process. *Electroencephalogr Clin Neurophysiol* **49**: 461-475.

Harter, M. R. and Previc, F. H. 1978. Size-specific information channels and selective attention: visual evoked potential and behavioral measures. *Electroencephalogr Clin Neurophysiol* **45**: 628-640.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P. and Grady, C. L. 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* **14**: 6336-6353.

Heinze, H. J. and Mangun, G. R. 1995. Electrophysiological signs of sustained and transient attention to spatial locations. *Neuropsychologia* **33**: 889-908.

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

Hellige, J. B. and Michimata, C. 1989. Categorization versus distance: hemispheric differences for processing spatial information. *Mem Cognit* **17**: 770-776.

Henderson, J. M. 1992. Identifying objects across saccades: effects of extrafoveal preview and flanker object context. *J Exp Psychol Learn Mem Cogn* **18**: 521-530.

Henderson, J. M. and Hollingworth, A. 1999. High-level scene perception. *Annu Rev Psychol* **50**: 243-271.

Henderson, J. M., McClure, K. K., Pierce, S. and Schrock, G. 1997. Object identification without foveal vision: evidence from an artificial scotoma paradigm. *Percept Psychophys* **59**: 323-346.

Henderson, J. M., Pollatsek, A. and Rayner, K. 1987. Effects of foveal priming and extrafoveal preview on object identification. *J Exp Psychol Hum Percept Perform* **13**: 449-463.

Henderson, J. M., Pollatsek, A. and Rayner, K. 1989. Covert visual attention and extrafoveal information use during object identification. *Percept Psychophys* **45**: 196-208.

Hillyard, S. A. and AnlloVento, L. 1998. Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci USA* **95**: 781-787.

Hillyard, S. A. and Mangun, G. R. 1986. The neural basis of visual selective attention: a commentary on Harter and Aine. *Biol Psychol* **23**: 265-279.

Hillyard, S. A. and Mangun, G. R. 1987. Sensory gating as a physiological mechanism for visual selective attention. *Electroencephalogr Clin Neurophysiol Suppl* **40**: 61-67.

Hillyard, S. A. and Munte, T. F. 1984. Selective attention to color and location: an analysis with event- related brain potentials. *Percept Psychophys* **36**: 185-198.

Hillyard, S. A., Teder-Salejarvi, W. A. and Munte, T. F. 1998. Temporal dynamics of early perceptual processing. *Curr Opin Neurobiol* **8**: 202-210.

Hillyard, S. A., Vogel, E. K. and Luck, S. J. 1998. Sensory Gain-Control (Amplification) as a Mechanism of Selective Attention - Electrophysiological and Neuroimaging Evidence. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES* **353**: 1257-1270.

Holcomb, P. J. 1993. Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* **30**: 47-61.

Holcomb, P. J. and McPherson, W. B. 1994. Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn* **24**: 259-276.

Hubel, D. H. and Livingstone, M. S. 1987. Segregation of form, color, and stereopsis in primate area 18. *J Neurosci* **7**: 3378-3415.

Hummel, J. E. and Biederman, I. 1992. Dynamic binding in a neural network for shape recognition. *Psychol Rev* **99**: 480-517.

Humphreys, G. W. and Boucart, M. 1997. Selection by color and form in vision. *J Exp Psychol Hum Percept Perform* **23**: 136-153.

Intraub, H. 1979. The role of implicit naming in pictorial encoding. *J. Exp. Psychol. Hum. Learn. Mem.* **5**: 78-87.

Intraub, H. 1980. Presentation rate and the representation of briefly glimpsed pictures in memory. *J Exp Psychol [Hum Learn]* **6**: 1-12.

Intraub, H. 1981. Rapid conceptual identification of sequentially presented pictures. *J Exp Psychol Hum Percept Perform* **7**: 604-610.

Intraub, H. and Hoffman, J. E. 1992. Reading and visual memory: Remembering scenes that were never seen. *American Journal of Psychology* **105**: 101-114.

Jeffreys, D. A. 1989. A Face-Responsive Potential Recorded from the Human Scalp. *Experimental Brain Research* **78**: 193-202.

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

Jeffreys, D. A., Tukmachi, E. S. and Rockley, G. 1992. Evoked potential evidence for human brain mechanisms that respond to single, fixated faces. *Exp Brain Res* **91**: 351-362.

Jemel, B., George, N., Olivares, E., Fiori, N. and Renault, B. 1999. Event-related potentials to structural familiar face incongruity processing. *Psychophysiology* **36**: 437-452.

Jodo, E. and Kayama, Y. 1992. Relation of a negative ERP component to response inhibition in a Go/No-go task [published erratum appears in Electroencephalogr Clin Neurophysiol 1992 Oct;83(4):270]. *Electroencephalogr Clin Neurophysiol* **82**: 477-482.

Jolicoeur, P., Gluck, M. A. and Kosslyn, S. M. 1984. Pictures and names: making the connection. *Cognit Psychol* **16**: 243-275.

Kahneman, D., Treisman, A. and Burkell, J. 1983. The cost of visual filtering. *J Exp Psychol Hum Percept Perform* **9**: 510-522.

Kant, J. D. 1996. Modélisation et mise en oeuvre de processus cognitifs de catégorisation à l'aide d'un réseau connexionniste. Informatique, Université de Rennes.

Kanwisher, N., McDermott, J. and Chun, M. M. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* **17**: 4302-4311.

Karayanidis, F., Andrews, S., Ward, P. B. and McConaghy, N. 1991. Effects of inter-item lag on word repetition: an event-related potential study. *Psychophysiology* **28**: 307-318.

Karayanidis, F. and Michie, P. T. 1996. Frontal processing negativity in a visual selective attention task. *Electroencephalogr Clin Neuro* **99**: 38-56.

Kastner, S., De Weerd, P., Desimone, R. and Ungerleider, L. G. 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* **282**: 108-111.

Katz, D. 1953. *Studien zur experimentellen Psychologie,* Basel: Schabe

Kenemans, J. L., Kok, A. and Smulders, F. T. 1993. Event-related potentials to conjunctions of spatial frequency and orientation as a function of stimulus parameters and response requirements. *Electroencephalogr Clin Neurophysiol* **88**: 51-63.

Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M. and Spitzer, M. 1998. The time course of brain activations during response inhibition: evidence from event-related potentials in a go/no go task. *Neuroreport* **9**: 765-770.

Kleiber, G. 1990. *La sémantique du prototype,* P.U.F., Paris.

Klein, R. 1988. Inhibitory tagging system facilitates visual search. *Nature* **334**: 430-431.

Knight, R. 1996. Contribution of human hippocampal region to novelty detection. *Nature* **383**: 256-259.

Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics* **43**: 59-62.

Kosslyn, S. M. 1996. *Image and Brain - The Resolution of the Imagery Debate,* MIT Press, Cambridge, Massachusetts.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L. and Anderson, A. K. 1994. Identifying objects seen from different viewpoints. A PET investigation. *Brain* **117**: 1055-1071.

Kosslyn, S. M., Daly, P. F., McPeek, R. M., Alpert, N. M., Kennedy, D. N. and Caviness, V. S., Jr. 1993. Using locations to store shape: an indirect effect of a lesion. *Cereb Cortex* **3**: 567-582.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J. and Gabrieli, J. D. 1989. Evidence for two types of spatial representations: hemispheric specialization for categorical and coordinate relations. *J Exp Psychol Hum Percept Perform* **15**: 723-735.

Kounios, J. and Holcomb, P. J. 1994. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J Exp Psychol Learn Mem Cogn* **20**: 804-823.

Krose, B. J. and Julesz, B. 1989. The control and speed of shifts of attention. *Vision Res* **29**: 1607-1619.

Kwak, H. W., Dagenbach, D. and Egeth, H. 1991. Further evidence for a time-independent shift of the focus of attention. *Percept Psychophys* **49**: 473-480.

Kwak, H. W. and Egeth, H. 1992. Consequences of allocating attention to locations and to other attributes. *Percept Psychophys* **51**: 455-464.

LaBerge, D. 1995. *Attentional Processing, The Brain's Art of Mindfulness,* Harvard University Press, Cambridge, London.

Lavie, N. 1997. Visual feature integration and focused attention: response competition from multiple distractor features. *Percept Psychophys* **59**: 543-556.

Lavie, N. and Tsal, Y. 1994. Perceptual load as a major determinant of the locus of selection in visual attention. *Percept Psychophys* **56**: 183-197.

Leiphart, J., Rosenfeld, J. P. and Gabrieli, J. D. 1993. Event-related potential correlates of implicit priming and explicit memory tasks. *Int J Psychophysiol* **15**: 197-206.

Levine, S. C. and Banich, M. T. 1982. Lateral asymmetries in the naming of words and corresponding line drawings. *Brain Lang* **17**: 34-45.

Livingstone, M. S. and Hubel, D. H. 1987. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J Neurosci* **7**: 3416-3468.

Loftus, G. R. and Mackworth, N. H. 1978. Cognitive determinants of fixation location during picture viewing. *J Exp Psychol [Hum Percept]* **4**: 565-572.

Logothetis, N. K. and Sheinberg, D. L. 1996. Visual object recognition. *Annu Rev Neurosci* **19**: 577-621.

Luck, S. J., Chelazzi, L., Hillyard, S. A. and Desimone, R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* **77**: 24-42.

Luck, S. J., Girelli, M., McDermott, M. T. and Ford, M. A. 1997. Bridging the gap between monkey neurophysiology and human perception: an ambiguity resolution theory of visual selective attention. *Cognit Psychol* **33**: 64-87.

Luck, S. J. and Hillyard, S. A. 1995. The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. *Int J Neurosci* **80**: 281-297.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P. and Hawkins, H. L. 1994. Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *J Exp Psychol Hum Percept Perform* **20**: 887-904.

Luzzatti, C. and Davidoff, J. 1994. Impaired retrieval of object-colour knowledge with preserved colour naming. *Neuropsychologia* **32**: 933-950.

Mangun, G. R. 1995. Neural mechanisms of visual selective attention. *Psychophysiology* **32**: 4-18.

Manolas, M. G., Stamoulos, T. D. and Anninos, P. A. 1999. Differences in human visual evoked potentials during the perception of colour as revealed by a bootstrap method to compare cortical activity. A prospective study. *Neurosci Lett* **270**: 21-24.

Marr, D. and Nishihara, H. K. 1978. Representation and recognition of three-dimensional shapes. . *Proc. Roy. Soc. Lond. B* **200**: 269-294.

Marsolek, C. J. 1995. Abstract visual-form representations in the left cerebral hemisphere. *J Exp Psychol Hum Percept Perform* **21**: 375-386.

Martin, A., Wiggs, C. L., Ungerleider, L. G. and Haxby, J. V. 1996. Neural correlates of category-specific knowledge. *Nature* **379**: 649-652.

Maunsell, J. H. 1995. The brain's visual world: representation of visual targets in cerebral cortex. *Science* **270**: 764-769.

Maunsell, J. H. and Gibson, J. R. 1992. Visual response latencies in striate cortex of the macaque monkey. *J Neurophysiol* **68**: 1332-1344.

McCarthy, G., Puce, A., Gore, J. C. and Allison, T. 1997. Face-specific processing in the human fusiform gyrus. *J Cognitive Neurosci* **9**: 605-610.

McClurkin, J. W. and 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**: 481-495.

McKeever, W. F. and Jackson, T. L., Jr. 1979. Cerebral dominance assessed by object- and color-naming latencies: sex and familial sinistrality effects. *Brain Lang* **7**: 175-190.

Mel, B. W. 1997. SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput* **9**: 777-804.

Menon, V., Ford, J. M., Lim, K. O., Glover, G. H. and Pfefferbaum, A. 1997. Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport* **8**: 3029-3037.

Michie, P. T., Karayanidis, F., Smith, G. L., Barrett, N. A., Large, M. M., O'Sullivan, B. T. and Kavanagh, D. J. 1999. An exploration of varieties of visual attention: ERP findings. *Brain Res Cogn Brain Res* **7**: 419-450.

Mikami, A., Nakamura, K. and Kubota, K. 1994. Neuronal responses to photographs in the superior temporal sulcus of the rhesus monkey. *Behav. Brain Res* **60**: 1-13.

Mikami, A., Newsome, W. T. and Wurtz, R. H. 1986. Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol* **55**: 1308-1327.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E. and et al. 1991. Perception and action in 'visual form agnosia'. *Brain* **114**: 405-428.

Missal, M., Vogels, R. and Orban, G. A. 1997. Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb Cortex* **7**: 758-767.

Mountcastle, V. B., Motter, B. C., Steinmetz, M. A. and Sestokas, A. K. 1987. Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *J Neurosci* **7**: 2239-2255.

Motter, B. C. 1994. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J Neurosci* **14**: 2190-2199.

Musen, G. and Treisman, A. 1990. Implicit and explicit memory for visual patterns. *J Exp Psychol Learn Mem Cogn* **16**: 127-137.

Naatanen, R., Teder, W., Alho, K. and Lavikainen, J. 1992. Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport* **3**: 493-496.

Nakamura, K. and Kubota, K. 1995. Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J Neurophysiol* **74**: 162-178.

Nakamura, K., Matsumoto, K., Mikami, A. and Kubota, K. 1994. Visual response properties of single neurons in the temporal pole of behaving monkeys. *J Neurophysiol* **71**: 1206-1221.

Nakayama, K. and Mackeben, M. 1989. Sustained and transient components of focal visual attention. *Vision Res* **29**: 1631-1647.

Nasman, V. T. and Rosenfeld, J. P. 1990. Parietal P3 response as an indicator of stimulus categorization: increased P3 amplitude to categorically deviant target and nontarget stimuli. *Psychophysiology* **27**: 338-350.

Neisser, U. 1978. Anticipations, images, and introspection. *Cognition* **6**: 169-174.

Nelson, S. B. 1991. Temporal Interactions in the Cat Visual System .1. Orientation-Selective Suppression in the Visual Cortex. *Journal of Neuroscience* **11**: 344-356.

Nelson, W. W. and Loftus, G. R. 1980. The functional visual field during picture viewing. *J Exp Psychol [Hum Learn]* **6**: 391-399.

Nettleton, N. C. and Bradshaw, J. L. 1983. Name and face matching in one or two visual fields: a test of models of hemispheric specialization. *Brain Cogn* **2**: 103-118.

Newsome, W. T., Mikami, A. and Wurtz, R. H. 1986. Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *J Neurophysiol* **55**: 1340-1351.

Nobre, A. C., Allison, T. and McCarthy, G. 1994. Word recognition in the human inferior temporal lobe. *Nature* **372**: 260-263.

Nobre, A. C., Allison, T. and McCarthy, G. 1998. Modulation of human extrastriate visual processing by selective attention to colours and words. *Brain* **121**: 1357-1368.

Nowak, L. G. and Bullier, J. 1997. The timing of information transfer in the visual system. In *Extrastriate cortex in primates* (J. Kaas, K. Rockland and A. Peterss, Eds.), pp. 205-241 . Plenum, New York.

Nowak, L. G., Munk, M. H., Girard, P. and Bullier, J. 1995. Visual latencies in areas V1 and V2 of the macaque monkey. *Vis Neurosci* **12**: 371-384.

Oldfield, R. C. and Wingfield, A. 1965. Response latencies in naming objects. . *Q. J. Exp. Psychol.* **17**: 273-281.

Oliva, A. and Schyns, P. G. 1997. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognit Psychol* **34**: 72-107.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci* **13**: 4700-4719.

Olshausen, B. A., Anderson, C. H. and Vanessen, D. C. 1995. A multiscale dynamic routing circuit for forming size- and position-invariant object representations. *J Comput Neurosci* **2**: 45-62.

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

Orban, G. A., Dupont, P., Vogels, R., Bormans, G. and Mortelmans, L. 1997. Human brain activity related to orientation discrimination tasks. *Eur J Neurosci* **9**: 246-259.

Ostergaard, A. L. and Davidoff, J. B. 1985. Some effects of color on naming and recognition of objects. *J Exp Psychol Learn Mem Cogn* **11**: 579-587.

Paller, K. A. and Gross, M. 1998. Brain potentials associated with perceptual priming vs explicit remembering during the repetition of visual word-form. *Neuropsychologia* **36**: 559-571.

Palmer, S. E. and Bucher, N. M. 1981. Configural effects in perceived pointing of ambiguous triangles. *J Exp Psychol Hum Percept Perform* **7**: 88-114.

Papathomas, T. V., Gorea, A. and Julesz, B. 1991. Two carriers for motion perception: color and luminance. *Vision Res* **31**: 1883-1892.

Perrett, D. I., Mistlin, A. J. and Chitty, A. J. 1987. Visual neurons responsive to faces. *Trends Neurosci.* **10**: 358-364.

Perrett, D. I., Oram, M. W. and Ashbridge, E. 1998. Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* **67**: 111-145.

Perrett, D. I., Rolls, E. T. and Caan, W. 1982. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* **47**: 329-342.

Petit, L., Dubois, S., Tzourio, N., Dejardin, S., Crivello, F., Michel, C., Etard, O., Denise, P., Roucoux, A. and Mazoyer, B. 1999. PET study of the human foveal fixation system. *Hum Brain Mapp* **8**: 28-43.

Pfefferbaum, A., Ford, J. M., Weller, B. J. and Kopell, B. S. 1985. ERPs to response production and inhibition. *Electroencephalogr Clin Neurophysiol* **60**: 423-434.

Plendl, H., Paulus, W., Roberts, I. G., Botzel, K., Towell, A., Pitman, J. R., Scherg, M. and Halliday, A. M. 1993. The time course and location of cerebral evoked activity associated with the processing of colour stimuli in man. *Neurosci Lett* **150**: 9-12.

Poffenberger, A. T. 1912. Reaction time to retinal stimulation with special reference to the time lost in conduction throught nerve centers. *Archives of Psychology* **23**: 1-73.

Posner, M. I. 1978. *Chronometric Explorations of Mind,* Englewood Cliffs, NJ: Erlbaum

Posner, M. I. and Dehaene, S. 1994. Attentional networks. *Trends Neurosci* **17**: 75-79.

Posner, M. I., Snyder, C. R. and Davidson, B. J. 1980. Attention and the detection of signals. *J Exp Psychol* **109**: 160-174.

Potter, M. C. and Faulconer, B. A. 1975. Time to understand pictures and words. . *Nature* **253**: 437-438.

Potter, M. C. and Intraub, H. 1999. *Fleeting memories,* MIT Press, Cambridge.

Potter, M. C., Kroll, J. and Harris, C. 1979. Comprehension and memory in rapid sequential memory. In *Attention and performance. VIII.* (R. S. Nickersons, Eds.), pp. . Erlbaum, Hillsdale.

Posner, M. I., Snyder, C. R. and Davidson, B. J. 1980. Attention and the detection of signals. *J Exp Psychol* **109**: 160-174.

Pratarelli, M. E. 1994. Semantic processing of pictures and spoken words: evidence from event- related brain potentials. *Brain Cogn* **24**: 137-157.

Previc, F. H. and Harter, M. R. 1982. Electrophysiological and behavioral indicants of selective attention to multifeature gratings. *Percept Psychophys* **32**: 465-472.

Ragot, R. and Lesevre, N. 1986. Electrophysiological study of intrahemispheric S-R compatibility effects elicited by visual directional cues. *Psychophysiology* **23**: 19-27.

Ragot, R. and Renault, B. 1981. P300, as a function of S-R compatibility and motor programming. *Biol Psychol* **13**: 289-294.

Ramachandran, V. S. and Gregory, R. L. 1978. Does colour provide an input to human motion perception? *Nature* **275**: 55-56.

Rayner, K. and Bertera, J. H. 1979. Reading without a fovea. *Science* **206**: 468-469.

Renault, B., Fiori, N. and Giami, S. 1988. Latencies of event related potentials as a tool for studying motor processing organization. *Biol Psychol* **26**: 217-230.

Renault, B., Ragot, R., Lesevre, N. and Remond, A. 1982. Onset and offset of brain events as indices of mental chronometry. *Science* **215**: 1413-1415.

Reynolds, J. H., Chelazzi, L. and Desimone, R. 1999. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* **19**: 1736-1753.

Richmond, B. J. and Optican, L. M. 1990. Temporal Encoding of 2-Dimensional Patterns by Single Units in Primate Primary Visual Cortex .2. Information Transmission. *Journal of Neurophysiology* **64**: 370-380.

Ritter, W., Ford, J. M., Gaillard, A. W., Harter, M. R., Kutas, M., Naatanen, R., Polich, J., Renault, B. and Rohrbaugh, J. 1984. Cognition and event-related potentials. I. The relation of negative potentials and cognitive processes. *Ann N Y Acad Sci* **425**: 24-38.

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B. and Varela, F. J. 1999. Perception's shadow: long-distance synchronization of human brain activity . *Nature* **397**: 430-433.

Rolls, E. T. 1991. Neural organization of higher visual functions. *Curr Opin Neurobiol* **1**: 274-278.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E. and Nalwa, V. 1989. The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Exp Brain Res* **76**: 153-164.

Rolls, E. T., Burton, M. J. and Mora, F. 1976. Hypothalamic neuronal responses associated with the sight of food. *Brain Res* **111**: 53-66.

Rolls, E. T., Perrett, D., Thorpe, S. J., Puerto, A., Roper-Hall, A. and Maddison, S. 1979. Responses of neurons in area 7 of the parietal cortex to objects of different significance. *Brain Res* **169**: 194-198.

Rolls, E. T., Perrett, D. I., Caan, A. W. and Wilson, F. A. 1982. Neuronal responses related to visual recognition. *Brain* **105**: 611-646.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L. and Azzopardi, P. 1994. The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp Brain Res* **101**: 473-484.

Rosch, E. 1975. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance* **1**: 303-322.

Rosch, E. 1978. Principles of categorization. In *Cognition and categorization* (N. J. Hillsdales, Eds.), pp. . Erlbaum

Rosen, B. R., Buckner, R. L. and Dale, A. M. 1998. Event-related functional MRI: past, present, and future. *Proc Natl Acad Sci U S A* **95**: 773-780.

Rudell, A. P. 1992. Rapid stream stimulation and the recognition potential. *Electroencephalogr Clin Neurophysiol* **83**: 77-82.

Rugg, M. D. 1983. Further study of the electrophysiological correlates of lexical decision. *Brain Lang* **19**: 142-152.

Rugg, M. D. and Coles, M. G. H. 1995. *Electrophysiology of mind : event-related brain potentials and cognition,* Oxford University Press, Oxford ; New York.

Rugg, M. D., Milner, A. D., Lines, C. R. and Phalp, R. 1987. Modulation of visual event-related potentials by spatial and non- spatial visual selective attention. *Neuropsychologia* **25**: 85-96.

Saarinen, J. and Julesz, B. 1991. The speed of attentional shifts in the visual field. *Proc Natl Acad Sci U S A* **88**: 1812-1814.

Sagi, D. and Julesz, B. 1985. Fast noninertial shifts of attention. *Spat Vis* **1**: 141-149.

Sagi, D. and Julesz, B. 1985. "Where" and "what" in vision. *Science* **228**: 1217-1219.

Sasaki, K., Gemba, H., Nambu, A. and Matsuzaki, R. 1993. No-go activity in the frontal association cortex of human subjects. *Neurosci Res* **18**: 249-252.

Savoy, R. L., Bandettini, P. A., O'Craven, K. M., Kwong, K. K., Davis, T. L., Baker, J. R., Weisskoff, R. M. and Rosen, B. R. 1995. Pushing the temporal resolution of fMRI: studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated changes in noise. *Proc. Soc. Magn. Reson. Med. Fourth Sci. Meeting Exhib.*

Schacter, D. L. and Buckner, R. L. 1998. Priming and the brain. *Neuron* **20**: 185-195.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M. and Rosen, B. R. 1997. Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *Neuroimage* **6**: 259-269.

Schein, S. J. and Desimone, R. 1990. Spectral properties of V4 neurons in the macaque. *J Neurosci* **10**: 3369-3389.

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

Schendan, H. E., Kanwisher, N. G. and Kutas, M. 1997. Early brain potentials link repetition blindness, priming and novelty detection. *Neuroreport* **8**: 1943-1948.

Scherg, M. and Berg, P. 1991. Use of prior knowledge in brain electromagnetic source analysis. *Brain Topogr* **4**: 143-150.

Schiller, P. H. 1995. Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature* **376**: 342-344.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D. and Leventhal, A. G. 1998. Signal timing across the macaque visual system. *J Neurophysiol* **79**: 3272-3278.

Schyns, P. G. and Oliva, A. 1999. Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* **69**: 243-265.

Seeck, M. and Grüsser, O. J. 1992. Category-related components in visual evoked potentials: Photographs of faces, persons, flowers and tools as stimuli. *Exp. Brain Res.* **92**: 338-349.

Sereno, A. B. and Maunsell, J. H. 1998. Shape selectivity in primate lateral intraparietal cortex [see comments]. *Nature* **395**: 500-503.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R. and Tootell, R. B. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging [see comments]. *Science* **268**: 889-893.

Sergent, J. 1987. Information processing and laterality effects for objects and face perception. In *Visual object processing : a cognitive neuropsychology approach* (H. e. Riddochs, Eds.), pp. . Erlbaum

Sergent, J., Ohta, S. and MacDonald, B. 1992. Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* **115 Pt 1**: 15-36.

Shimoyama, I., Morita, Y., Uemura, K., Kojima, Y., Nakamura, T., Nishizawa, S., Yokoyama, T., Ryu, H., Murayama, M. and Uchizono, K. 1992. Verbal versus non-verbal visual evoked potentials: Kanji versus line drawings. *Brain Topogr* **5**: 35-39.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E. and Petersen, S. E. 1997. Common blood flow changes across visual tasks .1. Increases in subcortical structures and cerebellum but not in nonvisual cortex - .2. Decreases in cerebral cortex. *J Cognitive Neurosci* **9**: 648-663.

Simson, R., Vaughan, H. G., Jr. and Ritter, W. 1977. The scalp topography of potentials in auditory and visual Go/NoGo tasks. *Electroencephalogr Clin Neurophysiol* **43**: 864-875.

Smid, H. G., Bocker, K. B., van Touw, D. A., Mulder, G. and Brunia, C. H. 1996. A psychophysiological investigation of the selection and the use of partial stimulus information in response choice. *J Exp Psychol Hum Percept Perform* **22**: 3-24.

Smid, H. G., Jakob, A. and Heinze, H. J. 1997. The organization of multidimensional selection on the basis of color and shape: an event-related brain potential study. *Percept Psychophys* **59**: 693-713.

Smid, H. G., Jakob, A. and Heinze, H. J. 1999. An event-related brain potential study of visual selective attention to conjunctions of color and shape. *Psychophysiology* **36**: 264-279.

Sperling, G. and Melchner, M. J. 1978. The attention operating characteristic: examples from visual search. *Science* **202**: 315-318.

Srinivas, K. 1995. Representation of rotated objects in explicit and implicit memory. *J Exp Psychol Learn Mem Cogn* **21**: 1019-1036.

Sugita, Y. 1995. Electrophysiological correlates of visual search asymmetry in humans. *Neuroreport* **6**: 1693-1696.

Sutton, S., Braren, M., Zubin, J. and John, E. R. 1965. Evoked-potential correlates of stimulus uncertainty. *Science* **150**: 1187-1188.

Syrkin, G. and Gur, M. 1997. Colour and luminance interact to improve pattern recognition. *Perception* **26**: 127-140.

Tallon-Baudry, C., Bertrand, O., Delpuech, C. and Permier, J. 1997. Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *J Neurosci* **17**: 722-734.

Tallon-Baudry, C., Bertrand, O., Peronnet, F. and Pernier, J. 1998. Induced gamma-band activity during the delay of a visual short-term memory task in humans. *J Neurosci* **18**: 4244-4254.

Tanaka, K. 1996. Inferotemporal cortex and object vision. *Annu Rev Neurosci* **19**: 109-139.

Tanaka, K., Saito, H., Fukada, Y. and Moriya, M. 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* **66**: 170-189.

Tassinari, G., Aglioti, S., Chelazzi, L., Peru, A. and Berlucchi, G. 1994. Do peripheral non-informative cues induce early facilitation of target detection? *Vision Res* **34**: 179-189.

Thorpe, S., Fize, D. and Marlot, C. 1996. Speed of processing in the human visual system. *Nature* **381**: 520-522.

Thorpe, S., Gegenfurtner, K., Fabre-Thorpe, M. and Bülthoff, H. H. 1999. Categorisation of complex natural images in extreme peripheral vision. *Perception* **28 (supplement)**: 61.

Thorpe, S. J. 1988. Identification of rapidly presented images by the human visual system. *Perception* **17**: A77-.

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

Tobimatsu, S., Tomoda, H. and Kato, M. 1996. Human VEPs to isoluminant chromatic and achromatic sinusoidal gratings: separation of parvocellular components. *Brain Topogr* **8**: 241-243.

Tootell, R. B., Hamilton, S. L. and Switkes, E. 1988. Functional anatomy of macaque striate cortex. IV. Contrast and magno- parvo streams. *J Neurosci* **8**: 1594-1609.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I. and Dale, A. M. 1997. Functional analysis of V3A and related areas in human visual cortex. *J Neurosci* **17**: 7060-7078.

Treisman, A. 1992. Perceiving and re-perceiving objects. *Am Psychol* **47**: 862-875.

Treisman, A., Kahneman, D. and Burkell, J. 1983. Perceptual objects and the cost of filtering. *Percept Psychophys* **33**: 527-532.

Treisman, A. and Paterson, R. 1984. Emergent features, attention, and object perception. *J Exp Psychol Hum Percept Perform* **10**: 12-31.

Treisman, A. M. 1969. Strategies and models of selective attention. *Psychol Rev* **76**: 282-299.

Treisman, A. M. and Gelade, G. 1980. A feature-integration theory of attention. *Cognit Psychol* **12**: 97-136.

Treisman, A. M. and Kanwisher, N. G. 1998. Perceiving visually presented objects: recognition, awareness, and modularity. *Curr Opin Neurobiol* **8**: 218-226.

Treue, S. and Maunsell, J. H. R. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**: 539-541.

Troscianko, T., Montagnon, R., Le, C. J., Malbert, E. and Chanteau, P. L. 1991. The role of colour as a monocular depth cue. *Vision Res* **31**: 1923-1929.

Trotter, Y. and Celebrini, S. 1999. Gaze direction controls response gain in primary visual-cortex neurons. *Nature* **398**: 239-242.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. and Imbert, M. 1996. Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *J Neurophysiol* **76**: 2872-2885.

Tzourio, N., Crivello, F., Mellet, E., Nkanga-Ngila, B. and Mazoyer, B. 1998. Functional anatomy of dominance for speech comprehension in left handers vs right handers. *Neuroimage* **8**: 1-16.

Ullman, S. 1989. Aligning pictorial descriptions: an approach to object recognition. *Cognition* **32**: 193-254.

Ullman, S. 1992. Low-level aspects of segmentation and recognition. *Philos Trans R Soc Lond B Biol Sci* **337**: 371-378.

Ullman, S. 1995. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb Cortex* **5**: 1-11.

van der Heijden, A. H. 1993. The role of position in object selection in vision. *Psychol Res* **56**: 44-58.

Van Essen, D. C. 1979. Visual areas of the mammalian cerebral cortex. *Annu Rev Neurosci* **2**: 227-263.

Van Rullen, R., Gautrais, J., Delorme, A. and Thorpe, S. 1998. Face processing using one spike per neurone. *Biosystems* **48**: 229-239.

Van Rullen, R. and Thorpe, S. 1999. Spatial attention in asynchronous neural networks. *Neurocomputing* **26-27**: 911-918.

Vandenberghe, R., Duncan, J., Dupont, P., Ward, R., Poline, J. B., Bormans, G., Michiels, J., Mortelmans, L. and Orban, G. A. 1997. Attention to one or two features in left or right visual field: a positron emission tomography study. *J Neurosci* **17**: 3739-3750.

Vaughan, H. G., Arezzo, J. R. and Arezzo, J. C. 1988. The neural basis of event-related potentials. In *Handbook of electroencephalography and clinical neurophysiology* (T. W. Pictons, Eds.), pp. 45-98 . Elsevier, Amsterdam - New York - Oxford.

Vetter, T., Hurlbert, A. and Poggio, T. 1995. View-based models of 3D object recognition: invariance to imaging transformations. *Cereb Cortex* **5**: 261-269.

Vitkovitch, M. and Underwood, G. 1992. Visual field differences in an object decision task. *Brain Cogn* **19**: 195-207.

Vogels, R. 1999. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur J Neurosci* **11**: 1239-1255.

Vogels, R. and Orban, G. A. 1993. Responses of monkey infero-temporal units in an orientation discrimination task. *Prog Brain Res* **95**: 427-444.

von der Heydt, R. and Peterhans, E. 1989. Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neurosci* **9**: 1731-1748.

Walley, R. E. and Weiden, T. D. 1973. Lateral inhibition and cognitive masking: a neuropsychological theory of attention. *Psychol Rev* **80**: 284-302.

Wallis, G. and Rolls, E. T. 1997. Invariant face and object recognition in the visual system. *Prog Neurobiol* **51**: 167-194.

Wang, J., Jin, Y., Xiao, F., Fan, S. and Chen, L. 1999. Attention-sensitive visual event-related potentials elicited by kinetic forms. *Clin Neurophysiol* **110**: 329-341.

Warrington, E. K. and James, M. 1986. Visual object recognition in patients with right-hemisphere lesions: axes or features? *Perception* **15**: 355-366.

Warrington, E. K. and Taylor, A. M. 1978. Two categorical stages of object recognition. *Perception* **7**: 695-705.

Wiesel, T. N. and Hubel, D. H. 1966. Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J Neurophysiol* **29**: 1115-1156.

Wiggs, C. L. and Martin, A. 1998. Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol* **8**: 227-233.

Wijers, A. A., Lange, J. J., Mulder, G. and Mulder, L. J. M. 1997. An ERP study of visual spatial attention and letter target detection for isoluminant and nonisoluminant stimuli. *Psychophysiology* **34**: 553-565.

Wijers, A. A., Mulder, G., Okita, T., Mulder, L. J. and Scheffers, M. K. 1989. Attention to color: an analysis of selection, controlled search, and motor activation, using event-related potentials. *Psychophysiology* **26**: 89-109.

Wolfe, J. M., O'Neill, P. and Bennett, S. C. 1998. Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Percept Psychophys* **60**: 140-156.

Wurm, L. H., Legge, G. E., Isenberg, L. M. and Luebker, A. 1993. Color improves object recognition in normal and low vision. *J Exp Psychol Hum Percept Perform* **19**: 899-911.

Young, M. P. and Yamane, S. 1993. An analysis at the population level of the processing of faces in the inferotemporal cortex. In *Brain mechanisms of perception and memory: From neuron to behaviour.* (L. Squire, T. Ono, M. Fukuda and D. Perretts, Eds.), pp. . Oxford University Press, New York.

Yvert, B., Bertrand, O., Thevenet, M., Echallier, J. F. and Pernier, J. 1997. A systematic evaluation of the spherical model accuracy in EEG dipole localization. *Electroencephalogr Clin Neurophysiol* **102**: 452-459.

Zarahn, E., Aguirre, G. and D'Esposito, M. 1997. A trial-based experimental design for fMRI. *Neuroimage* **6**: 122-138.

Zani, A. and Proverbio, A. M. 1995. ERP signs of early selective attention effects to check size. *Electroencephalogr Clin Neurophysiol* **95**: 277-292.