



HAL
open science

Contraintes temporelles des traitements visuels dans une tâche de catégorisation de scènes naturelles

Nadège M. Bacon-Macé

► **To cite this version:**

Nadège M. Bacon-Macé. Contraintes temporelles des traitements visuels dans une tâche de catégorisation de scènes naturelles. Neurosciences [q-bio.NC]. Université Paul Sabatier - Toulouse III, 2006. Français. NNT: . tel-00102319

HAL Id: tel-00102319

<https://theses.hal.science/tel-00102319>

Submitted on 29 Sep 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.

UNIVERSITÉ TOULOUSE III - PAUL SABATIER

UFR SVT - Sciences de la Vie et de la Terre -

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ TOULOUSE III

Discipline : Neurosciences

présentée et soutenue

par

Nadège M. Bacon-Macé

le 2 juin 2006

Titre :

**Contraintes temporelles des traitements visuels
dans une tâche de catégorisation de scènes naturelles**

JURY

Dr. Muriel Boucart - Hôpital Salengro - CNRS, Lille

Dr. Thierry Hasbroucq - LNC - CNRS, U. Provence, Marseille

Pr. Jean-Michel Lassalle - CRCA - U. Paul Sabatier, Toulouse

Dr. Aude Oliva - CVCL - MIT, Boston

Dr. Manuel Samuelides - ONERA, Toulouse

Dr. Simon Thorpe - CERCO - CNRS, UPS, Toulouse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Directeur de thèse

Centre de recherche CERveau et COgnition (CerCo)

UMR 5549 CNRS - Université Paul Sabatier

Faculté de médecine de Rangueil

31062 Toulouse Cedex 9

à Marc,
à mes enfants.

- REMERCIEMENTS -

Je tiens tout d'abord à remercier Simon Thorpe de m'avoir accueilli dans son équipe qui m'a tant apporté (sur le plan scientifique et personnel) ! J'avoue maintenant que ça n'était pas facile de te suivre tout le temps. Je n'ai pas réussi à soutenir le rythme des cinq nouvelles manips par semaine... Ton enthousiasme scientifique est un vrai souffle dans l'équipe, qu'on se le dise !

Un grand merci à Michèle pour son soutien et sa gentillesse. Ces quatre années passées au Cerco ont été riches sur tellement de plans que je ne te remercierai jamais assez pour ton écoute et ton ouverture. Et puis entre nous... heureusement que tu étais là recadrer mon encadrement !

Je tiens aussi à remercier Manuel Samuelides, mon co-directeur de thèse (hum hum...) ainsi que tous les membres du jury : les rapporteurs Muriel Boucart et Thierry Hasbroucq, le président Jean-Michel Lassalle et l'examinatrice Aude Oliva. C'est sur ce dernier entretien que je quitte le monde scientifique !

Guillaume Rousselet... notre phénomène, notre gibbon, notre ami. J'admire tellement ta persévérance, ta frénésie dans la science et dans la vie. Merci de m'avoir donné ton avis avisé sur le dernier jet du mémoire ...

Merci à Rufin VanRullen et Holle Kirchner pour leurs conseils et leur perspicacité scientifique.

Merci à Rudy, Olivier, Magali, Marianne, Jong, Nicolas, Tim et tous les djeun's du labo pour leur bonne humeur. Accrochez-vous ! Il viendra un jour où tous les étudiants français auront un poste de chercheur (mais voyons, ne parlons pas d'échéance...).

Et puis, un grand merci chaleureux à Catherine, Ghislaine, les Pascaux, Youn, Lionel et tous les membres du labo fidèles au rendez-vous de midi au coin café. Boutades et calembours, interrogations passionnées et réponses sur Google : tellement de moments mémorables !

Enfin Marc... merci.

- PUBLICATIONS -

1 – Article publié :

Bacon-Macé N, Macé MJM, Fabre-Thorpe M & Thorpe SJ (2005) The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res* 45, 1459-69.

2 – Articles en révision ou en préparation :

Bacon-Macé N, Kirchner H, Fabre-Thorpe M & Thorpe SJ (en révision) Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *J Exp Psychol [Hum Learn]*.

Bacon-Macé N & Thorpe SJ (en préparation) Investigating early masking effects using the RSVP paradigm. *Psychol Sci*.

Macé MJM, **Bacon-Macé N**, Nespoulous JL & Fabre-Thorpe M (en préparation) What's seen first: The animal or the bird? *J Cogn Neurosci*.

3 – Résumés de conférence publiés :

Thorpe SJ, **Bacon N**, Rousselet GA, Macé MJM & Fabre-Thorpe M (2002) Rapid categorization of natural scenes: feedforward vs. feedback contribution evaluated by backward masking. (ECP) *Perception*, 31(suppl.): 132.

Bacon-Macé N, Macé MJM, Kirchner H, Fabre-Thorpe M & Thorpe SJ (2003) Dynamics of rapid scene categorization: Backward masking and RSVP studies. (EBBS) *Acta Neurobiologica Experimentalis*: 24.

Kirchner H, **Bacon-Macé N** & Thorpe SJ (2003) In which of two scenes is the animal ? Ultra-rapid object detection and saccadic eye movements. (ECP) *Perception*, 32(suppl.): 170.

Thorpe SJ, **Bacon-Macé, N** (2005) Rapid scene categorisation and RSVP paradigms. (ECP) *Perception*, 34(suppl): 10.

Kirchner H, Barba I, Guyonneau R, **Bacon-Macé N**, Thorpe SJ (2005) Stimulus dependent variations in processing time revealed by the choice saccade task. (ECP) *Perception*, 34(suppl): 120.

Fabre-Thorpe M, **Bacon-Macé N**, Macé MJM, Thorpe SJ (2005) Coarse to fine processing in natural scene categorisation (EBBS) *Acta Neurobiologica Experimentalis*.

- TABLE DES MATIÈRES -

Chapitre 1 : Catégorisation visuelle et paradigme du masquage

1.1 Mettre en évidence la rapidité des traitements sensoriels :	
Tâche de catégorisation visuelle ultra-rapide _____	19
1.2 Les latences dans le système visuel :	
Données électrophysiologiques et comportementales _____	22
1.2.1 Électrophysiologie chez l'animal _____	22
1.2.1.1 De la rétine à V4 : extraction des propriétés bas et moyen-niveau _	22
1.2.1.2 Le cortex inféro-temporal _____	23
1.2.1.3 De l'encodage visuel à la réponse motrice _____	26
1.2.2 Électrophysiologie chez l'homme _____	28
1.2.2.1 Enregistrements intracrâniens _____	29
1.2.2.2 Enregistrements de l'activité à la surface du scalp _____	29
1.2.2.3 Quels traitements entre 50-130 ms ? _____	31
1.2.2.4 Activité visuelle spécifique après 150 ms _____	33
1.2.3 Approche comportementale chez l'homme _____	34
1.3 Étudier le déroulement temporel des traitements visuels:	
Paradigme du masquage _____	41
1.3.1 Terminologie du masquage _____	41
1.3.2 Psychophysique du masquage _____	44
1.3.2.1 Masquage par pattern _____	44
1.3.2.2 Masquage paracontraste / métacontraste _____	46
1.3.2.3 Modèles de masquage faisant intervenir des mécanismes plus tardifs _____	50
1.3.3 Neurophysiologie du masquage _____	50
1.4 Étudier la dynamique temporelle des traitements de catégorisation :	
L'intérêt d'utiliser des stimuli masqués _____	55
1.4.1 Présentation du contexte expérimental _____	55
1.4.2 Mise au point du protocole de masquage : quelques règles _____	56

Chapitre 2 : Masquage rétrograde dans une tâche de catégorisation visuelle ultra-rapide

2.1 Objectifs	59
2.2 Résultats	60
2.3 Analyses complémentaires	63
2.3.1 Propriétés du masque dynamique : expérience contrôle	63
2.3.2 Effet du masquage sur les potentiels	65
2.4 Discussion	70
2.4.1 Décours temporel des traitements visuels	70
2.4.2 Vers un modèle d'accumulation des informations	70
2.4.3 Quels mécanismes d'interférence entre la cible et le masque ?	74
2.4.4 Transmission continue ou discrète des informations ?	78
Conclusion	81

Chapitre 3 : Catégorisation visuelle : Traitements sensoriels, décisionnels et moteurs

3.1 Tâches manuelle et saccadique	83
3.2 Objectifs	90
3.3 Protocole expérimental	91
3.4 Résultats	93
3.4.1 Pourcentage correct	93
3.4.2 d'	93
3.4.3 Temps de réaction	94
3.4.4 Compromis précision-vitesse	94
3.5 Analyses complémentaires	97
3.5.1 Effet de séquence	97
3.5.2 Le cas particulier des saccades express	98
3.6 Discussion	102
3.6.1 Quelle que soit la tâche, les traitements visuels commencent par une phase commune d'analyse sensorielle	102
3.6.2 En fonction de la tâche et du mode de réponse, les composantes décisionnelles et motrices se dissocient dans le temps	102

3.6.3 Validation du protocole oculaire pour l'étude de la reconnaissance d'objet _____	110
3.6.4 Vision subliminale et conscience _____	111
3.7 Conclusion _____	117

Chapitre 4 : Catégorisation visuelle et protocole RSVP

4.1 Que se passe-t-il quand le masque a un sens ? _____	119
4.2 Catégorisation d'images naturelles dans un flux RSVP :	
Expériences préliminaires _____	121
4.2.1 Catégorisation animal / non-animal _____	121
4.2.2 Effet de l'entraînement dans la catégorisation animal / non-animal _____	125
4.2.3 Changement de catégories : catégorisation animal–humain–visages _____	126
4.3 Catégorisation d'images naturelles dans un flux RSVP :	
Comparaison des effets perturbateurs de trois masques _____	129
4.3.1 Objectifs _____	129
4.3.2 Protocole expérimental _____	130
4.3.3 Résultats _____	132
4.3.4 Résultats complémentaires _____	139
4.3.4.1 Moyennage des sets d'images _____	139
4.3.4.2 Analyse des images : taille relative des animaux dans l'image _____	140
4.3.5 Discussion _____	143
4.3.5.1 Interférences de bas-niveau _____	143
4.3.5.2 Supériorité du traitement des visages ? _____	144
4.3.5.3 Des mécanismes de compétition haut-niveau ? _____	146
4.3.5.4 Nature des informations rapides _____	149
4.4 Détection et identification d'images naturelles dans un flux RSVP :	
Expérience de dénomination _____	150
4.4.1 Protocole expérimental _____	150
4.4.2 Résultats _____	152
4.4.3 Discussion _____	159
4.4.3.1 De la détection à l'identification _____	159
4.4.3.2 « J'ai vu une femme aux cheveux longs ! » :	
Propriétés de la catégorie Visage _____	162
4.4.3.3 Nature des indices diagnostiques _____	163

Émergence :

« Apparition dont ne peut rendre compte un système de causalité. Apparition d'un état ou d'un être qualitativement différent et irréductible à l'état dont il procède. »

Définition ATILF



Introduction

La vision constitue un sujet de recherche fascinant de par son importance biologique : en un coup d'œil, nous sommes capables de comprendre le sens d'une scène visuelle, même dans les conditions de perception les plus bruitées, en construisant une représentation complexe à partir d'indices aussi variés que la luminance, la texture, la couleur, les contours... C'est maintenant un des systèmes d'analyse sensorielle le plus connu chez l'homme. Malgré tout, la grande efficacité des traitements pour intégrer l'information de manière à la fois *fiable* et *rapide* soulève encore de grandes questions sur la façon dont les messages rétiniens sont transmis et interprétés tout au long de la hiérarchie corticale. A l'échelle neuronale, cette problématique se réfère à la nature même des processus mis en jeu, et en particulier à trois aspects fonctionnels de l'encodage : comment sont représentés, transformés et transmis les signaux perceptifs depuis leur format rétinien brut jusqu'à leur format le plus élaboré permettant de faire émerger une interprétation consciente de la scène visuelle (deCharms & Zador, 2000; Miller, 1988; Mouret & Hasbroucq, 2000) ? Dans ce cadre conceptuel, chaque étape de traitement reçoit l'information dans une représentation initiale, applique des transformations qui lui sont propres et transmet ces informations sous leur nouvelle forme aux étapes ultérieures. La mise en jeu de ces représentations successives (voire simultanées) doit permettre d'extraire les caractéristiques pertinentes du monde visuel qui se traduiront à terme par le déclenchement de processus cognitifs et moteurs adaptés.

Mes travaux de DEA puis de thèse se sont inscrits dans cette problématique, en essayant d'appréhender, par une approche principalement comportementale, comment l'ensemble des processus impliqués dans la boucle sensori-motrice permettent d'expliquer notre excellente performance visuelle. Plus particulièrement, nous nous sommes intéressés à la dynamique des traitements grâce à un protocole de catégorisation ultra-rapide de scènes naturelles, qui a déjà montré sa pertinence pour mettre en évidence l'étonnante rapidité du système visuel (Fabre-Thorpe, Richard & Thorpe, 1998; Thorpe, Delorme & Van Rullen, 2001; Thorpe, Fize & Marlot, 1996). Plusieurs méthodes peuvent être envisagées pour étudier le décours temporel des traitements lorsque l'on présente un stimulus dans une condition donnée. L'analyse des temps de réaction fournit des informations à l'issue des processus engagés dans la boucle sensori-motrice. Un autre paradigme –dit de masquage– consiste à essayer d'interrompre les traitements à différents niveaux du parcours cortical et à évaluer l'impact de ces perturbations sur les performances. C'est l'outil qui a été choisi pour analyser le décours temporel des mécanismes de catégorisation visuelle dans l'ensemble des travaux que je présente dans cette thèse.

Dans le chapitre 1, après avoir passé en revue de manière non exhaustive les données de la littérature sur les latences enregistrées dans les systèmes visuel et moteur selon différentes approches, je présenterai brièvement les protocoles expérimentaux qui ont permis de valider le paradigme du masquage comme un puissant outil d'analyse des traitements perceptifs.

Dans le chapitre 2, je rentrerai alors dans la description du protocole de masquage que nous avons adapté à la tâche de catégorisation visuelle d'images naturelles et dont les résultats ont été publiés

dans *Vision Research*. Cette expérience a non seulement permis de mettre en évidence une nouvelle fois à quel point les traitements visuels sont robustes et rapides, mais a aussi posé des contraintes temporelles supplémentaires sur la manière dont l'information est intégrée et transmise tout au long de la hiérarchie corticale.

Le chapitre 3 s'intéresse plus globalement à l'enchaînement des étapes perceptives, décisionnelles et motrices à travers la description d'une série de quatre expériences de masquage dont les résultats varient selon la manipulation de ces différents paramètres. En effet, une expérience récente de discrimination de scènes naturelles a montré que l'utilisation de l'effecteur oculaire permet de générer des réponses beaucoup plus rapides qu'avec l'effecteur manuel. D'autre part, il est important de souligner que la performance s'explique à la fois par la capacité du système à extraire de manière efficace l'information disponible mais aussi par sa capacité à prendre des décisions sur la base de cette information limitée, d'où l'importance de considérer la boucle sensori-motrice dans son ensemble lorsque l'on analyse la performance dans des tâches visuelles. Enfin, je présenterai dans le chapitre 4 les résultats de deux séries d'expériences utilisant un protocole de masquage particulier : le paradigme RSVP (Rapid Serial Visual Presentation). Dans ces expériences, la cible est non seulement suivie mais aussi précédée de masques. Nous avons étudié la nature des informations perturbées par le masquage, à travers la manipulation du contenu des masques et de la position de la cible dans la séquence. L'ensemble de ces résultats peut être replacé dans un modèle où les informations s'accablent rapidement et de manière progressive grâce aux premières vagues de traitements qui parcourent le système visuel. Une perturbation de plus en plus tardive de ces traitements permet d'élaborer une représentation de la scène visuelle de plus en plus précise, comme le confirment les résultats d'une expérience requérant des processus de détection ou d'identification du contenu des images cibles.

Catégorisation visuelle et paradigme du masquage

1

1.1 Mettre en évidence la rapidité des traitements sensoriels : tâche de catégorisation visuelle ultra-rapide

Quand je suis arrivée au sein de leur équipe, Simon et Michèle Thorpe travaillaient depuis plusieurs années sur un protocole de catégorisation visuelle ultra-rapide de type go/no-go, qui avaient déjà prouvé toute son utilité et sa pertinence dans l'étude de la reconnaissance visuelle d'objet. Dans ce protocole établi chez l'homme et le singe, les sujets doivent relâcher un bouton réponse lorsqu'une photographie de scène naturelle contient un animal. L'image est flashée sur l'écran pendant seulement 20 ms, une durée qui empêche toute exploration active de l'image. Avec une précision supérieure à 94 %, les temps de réaction moyens sont particulièrement courts, autour de 420 ms. Les réponses les plus précoces impliquant malgré tout un traitement poussé de l'image apparaissent dès 250-260 ms (Thorpe, Fize & Marlot, 1996). Chez le singe macaque, les réponses sont encore plus rapides puisqu'elles commencent à 170-180 ms pour une précision de 87-90% (Fabre-Thorpe, Richard & Thorpe, 1998). Or pour l'homme comme pour le singe, les images utilisées sont très variées et les sujets n'ont aucune information sur l'animal à chercher (mammifère, reptile, oiseau, poisson...), ni sur la taille, la position, l'orientation ou le nombre de cibles dans l'image. Les images pour lesquelles les sujets ont répondu avec des temps de réaction très courts sont suffisamment complexes et variées et ne semblent pas plus faciles à catégoriser que les autres (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001).

Les sujets ont toujours pour consigne de répondre le plus rapidement et le mieux possible, et les données comportementales recueillies témoignent de la grande efficacité des traitements visuels mis en jeu. L'analyse des potentiels évoqués (PEV) met en évidence une activité différentielle entre les réponses cérébrales pour les cibles et celles pour les distracteurs, démarrant significativement à 150 ms et prenant naissance dans les aires visuelles extrastriées (Fize et al., 2000). Un délai de 150 ms semble donc être suffisant pour commencer à voir apparaître les

traitements reflétant la dissociation entre les images cible et distracteur. Cette latence est extrêmement courte, ce qui fait naître un certain nombre de questions. Comment transporter et traiter l'information entre la rétine et le cortex extrastré en un temps aussi court ? C'est cette question qui a motivé l'ensemble de mes travaux, d'abord en DEA puis en thèse, et qui a sollicité le développement d'un protocole annexe : la catégorisation d'images naturelles masquées. Mais avant d'entrer dans le détail de ce protocole, faisons le point sur les données de la littérature concernant les latences dans le système visuel, que ce soit par des méthodes électrophysiologiques ou comportementales, chez l'animal et chez l'homme.

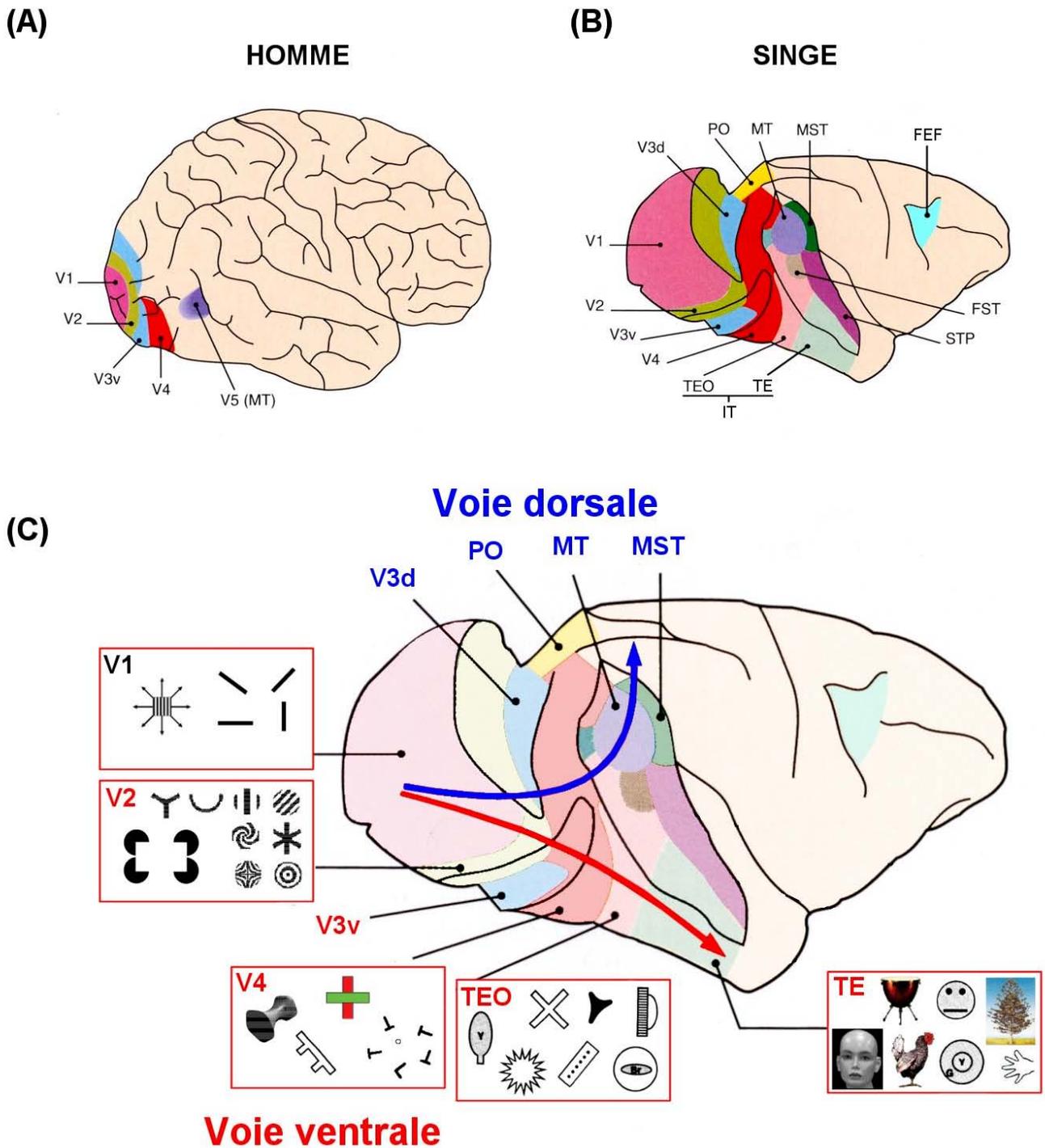


Figure 1.1 : Représentation des différentes aires du système visuel, chez l'homme (A) et le singe (B et C). Les informations concernant la scène visuelle sont projetées de la rétine vers les aires extrastrées en se divisant en deux sous-système : une *voie ventrale* qui s'étend vers le lobe temporal, impliquée dans la reconnaissance d'objet, et une *voie dorsale* s'étendant vers le lobe pariétal et assurant le contrôle visuo-moteur sur les objets. Les aires corticales contribuant au système ventral se définissent par des représentations de plus en plus complexes et spécialisées de la scène naturelle. Il est cependant remarquable de constater que l'intégration de ces propriétés complexes, aboutissant à des représentations aussi spécialisées que celles des visages ou des arbres dans IT, se fait très rapidement puisque les premières réponses neuronales apparaissent autour de 40-50 ms dans V1 et 70-80 ms dans TE.

FEF : Frontal Eye Field, FST : Fundus du sulcus temporal supérieur, IT : cortex inféro-temporal, MT : aire temporal médiane, MST : aire temporale médiane supérieure, PO : aire pariéto-occipitale, STP : Aire temporale supérieure polysensorielle, TE : cortex inféro-temporal antérieur, TEO : cortex inféro-temporal postérieur. Adapté de (Bullier, 1998).

1.2 Les latences dans le système visuel : données électrophysiologiques et comportementales

Dans un article pour la revue *La recherche* en 1992, Jean Bullier comparait le déroulement temporel de l'activité corticale à « un feu de forêt qui, par fort mistral, débute au même instant dans des endroits fort éloignés, et embrase de façon quasi-systématique une énorme étendue de forêt » (Bullier, Salin & Girard, 1992). La complexité des connexions reliant les aires visuelles et leurs interactions multiples ne permettent pas d'établir une distribution unidirectionnelle des latences dans le système visuel. Cependant, si on se limite aux régions de la voie ventrale impliquées dans la reconnaissance et l'identification de stimuli statiques, on peut effectivement constater un transfert de l'information des aires primaires vers les aires extrastriées permettant l'intégration de propriétés de plus en plus complexes (Felleman & Van Essen, 1991). Que nous apportent les données électrophysiologiques sur le déroulement temporel de ces informations dans la voie ventrale ?

1.2.1 Électrophysiologie chez l'animal

Dans le modèle hiérarchique de la transmission d'information, la fixation d'une scène visuelle initie une vague d'entrées rétiniennes, qui parcourt progressivement les différentes étapes de traitement pour aboutir à une représentation de plus en plus détaillée du stimulus.

1.2.1.1 De la rétine à V4 : extraction des propriétés bas et moyen-niveau

L'enregistrement de cellules ganglionnaires de la rétine chez le chat a montré que les latences absolues des réponses dépendent fortement du contraste entre le stimulus et le fond, mais même pour de forts contrastes, les réponses n'apparaissent pas avant **15-20 ms** après le début de présentation du stimulus (Bolz, Rosner & Wässle, 1982). Au niveau du CGL, les latences apparaissent à partir de 30 ms, avec un décalage de l'ordre de 10 à 20 ms entre les couches magnocellulaires et parvocellulaires (Irvin, Norton, Sesma & Casagrande, 1986). Les premières réponses corticales enregistrées dans le cortex visuel primaire (V1) émergent autour de **40-50 ms** (Celebrini, Thorpe, Trotter & Imbert, 1993; Foxe & Simpson, 2002). À partir de V1, l'activation se répand aux autres aires corticales et chaque aire fonctionnelle semble présenter une gamme de latences d'activation qui lui est propre (Nowak & Bullier, 1997). Si certaines aires occipito-temporales de la voie dorsale peuvent être activées très brièvement après V1 (Schmolesky et al., 1998; Schroeder, Mehta & Givre, 1998), en revanche dans la voie ventrale, l'activation successive des aires V1, V2, V4 se fait avec **un décalage de 10 à 15 ms** entre chaque étape, chez le primate (Raiguel, Lagae, Gulyas & Orban, 1989) comme chez le chat (Dinse & Kruger, 1994). Cette valeur comprend à la fois le temps de conduction axonale qui dans certains cas peut être assez court (par exemple de l'ordre de la milliseconde entre V1 et V2), mais aussi les délais synaptiques et le temps d'intégration (Nowak & Bullier, 1997). En effet, une fois que le potentiel d'action a atteint la terminaison axonale, il faut considérer le temps de dépolarisation nécessaire pour atteindre le potentiel de membrane des cellules visées, en sachant que ce délai peut varier de

quatre à plusieurs dizaines de millisecondes. Ce temps d'intégration doit donc être pris en compte lorsque l'on s'intéresse au déroulement temporel de la transmission d'information le long de la voie ventrale, et correspond au délai nécessaire à l'extraction et à l'encodage des propriétés physiques afin de construire une représentation des contours et bordures séparant les surfaces de contraste, couleur ou texture différentes.

1.2.1.2 Le cortex inféro-temporal

Le cortex inféro-temporal (IT) correspond à une région du cortex relativement large, qui se sous-divise en 2 parties chez le singe, TE dans sa partie antérieure et TEO en postérieure (Iwai & Mishkin, 1969; Von Bonin, 1947). La topographie et l'organisation laminaire des projections ont permis d'établir une autre sous-division couramment utilisée, définissant 3 régions selon qu'elles sont postérieure, centrale ou antérieure: PIT, CIT et AIT (Felleman & Van Essen, 1991).

Gross et collaborateurs ont été les premiers à enregistrer des réponses évoquées dans IT chez le singe vigile et anesthésié (voir (Gross, 1994) et (Logothetis & Sheinberg, 1996) pour des revues très complètes). Leurs résultats ont été depuis largement vérifiés, ce qui a permis d'établir que IT était la dernière aire majoritairement visuelle de la voie ventrale, avec 85% des neurones répondant à des stimuli plus ou moins complexes (Desimone, Albright, Gross & Bruce, 1984). Les cellules de IT sont en effet sélectives à de nombreuses propriétés, comme la couleur, l'orientation, la texture, la direction du mouvement et surtout la forme. Il faut cependant noter que la plupart des neurones de IT sont sélectifs à des stimuli plus complexes que des barres, disques ou simples textures (Tanaka, Saito, Fukada & Moriya, 1991). Mais quels que soient les stimuli utilisés, on retrouve dans la littérature des gammes de latences de réponses relativement homogènes, avec un départ d'activité **après 70 ms** (Table 1.1).

De nombreux enregistrements ont également été recueillis au niveau du Sulcus Temporal Supérieur (STS), aire polymodale qui reçoit à la fois des afférences visuelles de IT, mais aussi auditives du cortex temporal supérieur et somesthésiques du cortex pariétal postérieur (Gross, Schiller, Wells & Gerstein, 1967; Mountcastle et al., 1975). Avec l'amygdale, le gyrus fusiforme et le cortex frontal orbital et médial, le STS semble faire partie du réseau formant le « cerveau social » (Adolphs, 2003). Les enregistrements ont montré que ses cellules répondaient au mouvement biologique, comme une vidéo de personne qui marche, ou à des photographies de visage, ce qui fait du STS une aire en grande partie visuelle. Les latences enregistrées sont comparables à celles obtenues dans IT (Table 1.1).

Si ces latences d'activation sont particulièrement rapides, est-ce que la partie précoce de la réponse est déjà sélective aux propriétés du stimulus ou est-ce que la sélectivité apparaît au fur et à mesure que la réponse se développe (Thorpe & Imbert, 1989) ? En fait, à partir du moment où l'information arrive dans le cortex inféro-temporal, il semble que très peu de temps soit nécessaire pour encoder spécifiquement les propriétés du stimulus. En analysant l'information contenue dans la réponse neuronale du cortex temporal chez le singe vigile, Tovée et Rolls ont montré que presque 65 % de l'information estimée sur une période de 400 ms était disponible dans les 20 premières millisecondes, et presque 87% dans les 50 premières millisecondes après le

début de la réponse (Rolls & Tovee, 1994; Rolls, Tovee & Panzeri, 1999; Tové & Rolls, 1995). Ces résultats sont d'une importance fondamentale car si l'on considère qu'un délai de 10 à 15 millisecondes se rajoute à chaque étape du parcours visuel des informations, le fait de pouvoir encoder la majorité des propriétés pertinentes sur le stimulus dans les 20 premières millisecondes constitue un aspect important de l'efficacité d'une transmission rapide des informations dans le système visuel. Il faut par ailleurs noter que la première partie de la réponse neuronale semble avoir des propriétés spécifiques qui diffèrent de la partie plus tardive de la réponse (Amarasingham et al., 2006; Muller, Metha, Krauskopf & Lennie, 2001), ce qui a d'ailleurs conduit certains auteurs à parler d'un encodage en deux temps, dépendant du niveau de précision des informations sur le stimulus (Sugase, Yamane, Ueno & Kawano, 1999).

Stimuli		Latences	Analyse	Où ?	Référence
Simples	Flashes lumineux	A partir de 49 ms	Courant de source	IT	Schroeder, Mehta & Givre, 1998
	Gratings	70 à 130 ms (110)	Enregistrements unitaires	TE	Vogels & Orban, 1994
	Stimuli de Walsh	75 à 120 ms	Enregistrements unitaires	TE	Richmond, Optican, Podell & Spitzer, 1987
	Gratings	A partir de 70 ms	Enregistrements unitaires	TE	Richmond, Wurtz & Sato, 1983
Complexes	Aliments et objets	100 à 230 ms	Enregistrements unitaires	TE	Rolls, Judge & Sanghera, 1977
	Objets et visages	80 à 180 ms	Enregistrements unitaires	STS	Perrett, Rolls & Caan, 1982
	Visages	70 à 150 ms	Enregistrements unitaires	STS	Baylis, Rolls & Leonard, 1985
	Visages	A partir de 70 ms	Enregistrements unitaires	STS	Tovée, Rolls, Treves & Bellis, 1993
	Visages	A partir de 75-85 ms	Enregistrements unitaires	STS	Rolls, Tovee & Panzeri, 1999
	Objets naturels et artificiels	(100)	Enregistrements unitaire	TE	Kovacs et al., 2003

Table 1.1 : Latences des réponses neuronales enregistrées au niveau de IT et du STS, pour différentes catégories de stimuli. Les chiffres entre parenthèses correspondent aux valeurs médianes, si indiquées. Toutes ces données ont été recueillies chez le singe vigile.

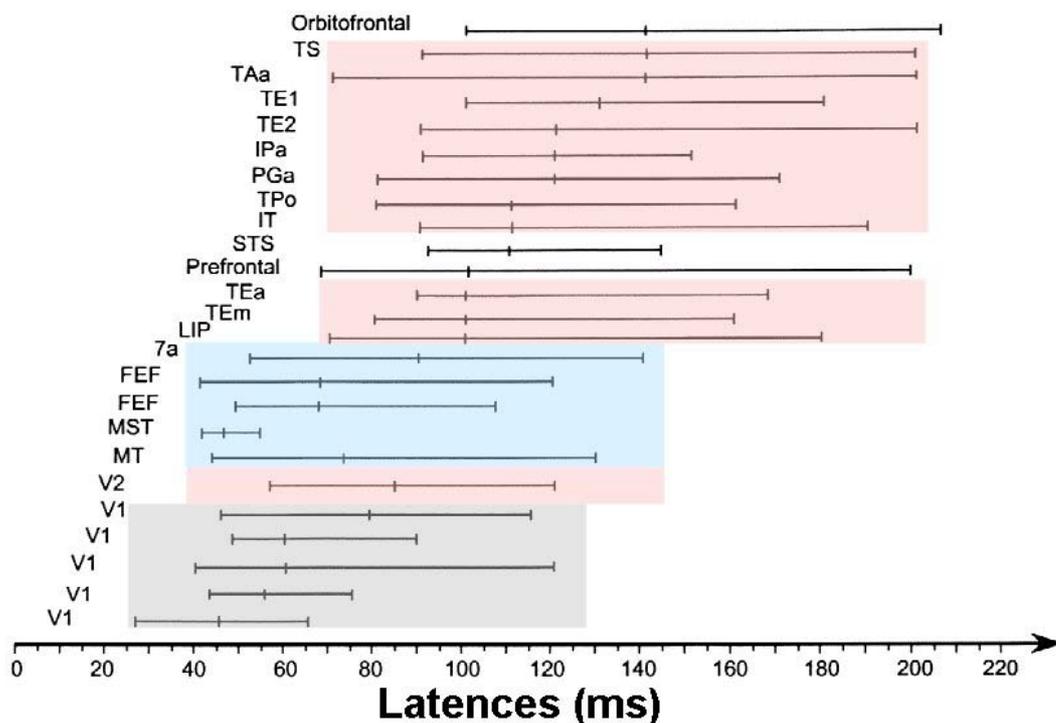


Figure 1.2 : Distribution des latences dans différentes aires du système visuel chez le singe macaque. Les aires de la voie ventrale ont été soulignées en rouge, de la voie dorsale en bleue, et V1 en gris. Adapté de (Nowak & Bullier, 1997).

1.2.1.3 De l'encodage visuel à la réponse motrice

La réalisation de mouvements dirigés vers un objectif dépend des informations relatives à la connaissance de la situation du corps dans l'espace, des objectifs à atteindre et de la sélection d'une stratégie pour les atteindre. Un ensemble de régions et structures organisées de manière hiérarchique interviennent dans la réalisation du mouvement volontaire. Comment sont connectées les aires motrices permettant de générer une réponse manuelle ou oculaire, et dans quelle gamme de latences sont-elles activées ?

Réponse manuelle

De nombreuses études de connectivité mettent clairement en évidence des connexions directes entre la voie visuelle dorsale et le cortex prémoteur situé en avant du cortex moteur, dans le lobe frontal (Boussaoud, di Pellegrino & Wise, 1995; Cavada & Goldman-Rakic, 1989; Chavis & Pandya, 1976). La partie médiane du cortex prémoteur semble être spécialisée dans la localisation des cibles, et la partie latérale d'avantage impliquée dans le traitement des formes et des objets dans le but de les manipuler. Cependant, même si la voie dorsale reçoit des afférences directes du cortex visuel primaire, il reste difficile de savoir si les informations intégrées par le cortex inféro-temporal dans des tâches complexes de reconnaissance visuelle peuvent être transmises aux régions motrices par l'intermédiaire de la voie dorsale. En revanche, les informations intégrées par la voie visuelle ventrale sont probablement transmises aux aires motrices par l'intermédiaire du cortex préfrontal (Baizer, Ungerleider & Desimone, 1991;

Ungerleider, Gaffan & Pelak, 1989; Webster, Bachevalier & Ungerleider, 1994), lui même connecté aux aires prémotrices à la fois médiane et latérale (Arikuni, Watanabe & Kubota, 1988; Barbas & Pandya, 1987; Lu, Preston & Strick, 1994). Une autre possibilité de trajet, bien que peu rentable en terme de temps de transmission, pourrait faire intervenir le cortex périrhinal, aire polymodale située dans la partie ventromédiane du lobe temporal recevant des afférences du cortex inféro-temporal, et connectée à de nombreuses aires frontales (Buckley, Gaffan & Murray, 1997; Lavenex, Suzuki & Amaral, 2002; Murray & Richmond, 2001).

Le cortex préfrontal s'occuperait de définir les stratégies motrices, c'est à dire les objectifs du mouvement et les comportements à mettre en œuvre pour atteindre ces objectifs. Les aires motrices secondaires travaillent alors avec le cervelet pour spécifier une séquence précise de contractions des différents muscles requis pour la réalisation de l'acte moteur. Enfin le cortex moteur primaire, le tronc cérébral et la moelle épinière entrent en jeu pour produire la contraction de tous les muscles nécessaires.

Les premières latences de réponse issues d'un signal visuel apparaissent **autour de 120-140 ms** dans le cortex pré moteur, avec des valeurs médianes autour de 160 ms. Elles précèdent les réponses enregistrées dans le cortex moteur de 20 à 40 ms, dont les premières valeurs autour de **160 ms** et les latences médianes vers 190 ms (Cisek & Kalaska, 2005; Crutcher & Alexander, 1990; Kurata & Tanji, 1986).

Réponse oculaire

Les réponses saccadiques sont en général bien plus courtes que les réponses manuelles, elles peuvent être déclenchées dès 80-100 ms dans des tâches visuelles simples (Busettoni, Masson & Miles, 1997; Masson, Rybarczyk, Castet & Mestre, 2000), et nous avons récemment observé dans l'équipe que des réponses oculaires correctes peuvent être générées en seulement 120 ms dans une tâche plus complexe de catégorisation de scènes naturelles (Kirchner & Thorpe, 2005). Nous y reviendrons dans le Chapitre 3.

L'aire oculomotrice frontale (en anglais, Frontal Eye Field, FEF) est une région du cortex préfrontal participant à la transformation du signal visuel en commande motrice saccadique. Des enregistrements chez le singe ont permis d'observer que près de la moitié des neurones de cette région ont une réponse visuelle et interviennent dans la sélection de la cible visuelle vers laquelle s'oriente la saccade (Schall, 1995; Schall, 2002). Une autre population de neurones du FEF décharge spécifiquement avant et pendant la saccade, et innerve le colliculus supérieur ainsi que le circuit nerveux préoculomoteur du tronc cérébral responsable de la génération des saccades (Bruce & Goldberg, 1985; Sommer & Wurtz, 2001). Le colliculus supérieur (CS) est aussi une région impliquée de façon prépondérante dans le contrôle des mouvements oculaires. Cette région contient un groupe de neurones dit de fixation, qui empêche la génération de saccades lorsqu'ils sont activés (Munoz & Wurtz, 1993a; Munoz & Wurtz, 1993b). La génération des mouvements oculaires est sous l'influence d'un équilibre entre l'activité des neurones du mouvement dans le FEF et celui des neurones de fixation du CS. Un mouvement oculaire est déclenché lorsque l'activité des neurones de mouvement augmente et celle des neurones de fixation diminue de façon significative, suggérant l'existence d'un seuil d'activation à franchir pour déclencher la

réponse saccadique (Segraves, 1992). Ce seuil semble être dépendant de la quantité d'informations accumulées par le système sensoriel, et la variabilité observée dans les mesures de temps de réaction saccadiques résulterait de la vitesse à laquelle ces informations s'accumulent, se reflétant par une augmentation plus ou moins rapide du signal dans le FEF (Dorris, Pare & Munoz, 1997; Hanes & Schall, 1996; Schall, 2001; Schall, 2003).

Le FEF est connecté à de nombreuses aires visuelles, allant de V2, V3, V4 au cortex inféro-temporal pour la voie ventrale, et reçoit des afférences massives de la voie dorsale, de MT, MST et LIP (Schall, Morel, King & Bullier, 1995). Les latences de réponses dans le FEF sont particulièrement courtes, avec les premières réponses visuelles à un flash lumineux apparaissant **dès 40 ms** chez le singe, et à 70 ms en moyenne (Nowak & Bullier, 1997; Schmolesky et al., 1998; Thompson, Hanes, Bichot & Schall, 1996). Si l'on considère un délai minimum de 20-25 ms pour le déclenchement moteur de la saccade au niveau du circuit du tronc cérébral, ces latences neuronales expliquent les temps de réaction saccadiques particulièrement courts mesurés dans des tâches comportementales. Une question centrale porte sur le niveau de traitement sensoriel nécessaire pour orienter les réponses saccadiques majoritairement vers des réponses correctes. Nous reviendrons plus largement sur ce point dans le Chapitre 3 (Kirchner & Thorpe, 2005).

1.2.2 Électrophysiologie chez l'homme

Les données électrophysiologiques chez l'animal permettent de définir les latences neuronales à différents niveaux de la hiérarchie corticale tout en apportant des données fondamentales sur la nature des propriétés pour lesquelles ces neurones sont sélectifs. L'analogie entre la dynamique temporelle des traitements chez l'homme et l'animal est difficile à établir car peu d'études à ma connaissance se sont penchées sur une comparaison directe des enregistrements obtenus chez les deux espèces pour un même type de tâche. Les réponses motrices fournies par les sujets humains sont tout de fois bien plus longues que celles mesurées chez le singe, avec un rapport de un tiers environ (Fabre-Thorpe, Richard & Thorpe, 1998; Goffart, Quinet, Chavane & Masson, 2006). Les explications avancées sont diverses et peuvent faire intervenir des différences de stratégie comme des compromis rapidité-vitesse (les hommes sont généralement plus précis dans leur réponse, ce qui se compense par une augmentation des temps de réaction), ou des différences plus physiologiques, comme une augmentation de la distance cortico-corticale (le cerveau humain est plus volumineux, les distances à parcourir entre aires corticales sont donc plus longues).

Chez l'homme, la majorité des données électrophysiologiques sont obtenues lorsque les sujets sont engagés dans une tâche dont on cherche à isoler les corrélats neuronaux. Les enregistrements fournissent donc des gammes de latences beaucoup plus vastes, et il est important de bien faire la distinction entre ces différents travaux selon le type de stimuli utilisé et la tâche concernée (Table 1.2). Depuis une dizaine d'années, certaines études ont suggéré que les indices de reconnaissance apparaissent relativement tôt, parfois même dès 30-50 ms après la présentation

du stimulus (George, Jemel, Fiori & Renault, 1997; Mouchetant-Rostaing et al., 2000a; Seeck et al., 1997)...

1.2.2.1 Enregistrements intracrâniens

L'analogie avec les données recueillies chez le singe est rendue difficile car les enregistrements intracrâniens sont rares. Ces données sont obtenues chez des patients présentant une forme sévère d'épilepsie pharmaco-résistante que l'on opère pour pouvoir supprimer le foyer épileptogène. Afin de le localiser, le patient est préalablement hospitalisé et des électrodes lui sont implantées dans différents sites cérébraux. Les chercheurs profitent de cette occasion unique pour recueillir des enregistrements de l'activité au niveau de ces électrodes lors de tâches cognitives. En 1994, Halgren a enregistré l'activité chez 33 patients sur presque 1200 sites allant des aires occipitales aux aires temporales, pariétales et limbiques (Halgren et al., 1994). Alors que certains sujets sont engagés dans une tâche de reconnaissance de visages inconnus répétés, d'autres effectuent la tâche avec des mots de 4 à 7 lettres. La première composante évoquée par les visages apparaît **entre 75 et 105 ms**, et semble prendre naissance au niveau de V1/V2. En revanche, rien n'apparaît avant 190 ms pour les mots, mais les auteurs justifient cette absence par une présentation centrale des mots, contrairement aux visages, alors que les sites d'enregistrement sont situés loin de la projection fovéale. En 1997, Seeck et collaborateurs enregistrent les potentiels évoqués en EEG chez des sujets sains, ainsi que l'activité intracrânienne dans une tâche identique chez des patients épileptiques dans différentes régions comprenant l'amygdale, l'hippocampe, les lobes frontaux et temporaux (Seeck et al., 1997). Les sujets devaient reporter manuellement si deux images (visages ou objets courants) présentées successivement avaient déjà été vues ou non. Les deux images étaient séparées entre elles de 150 ms, et les paires d'images de 2 secondes. Une activité différentielle entre les visages nouveaux et répétés est observée **dès 50 ms** -avec une tendance non significative démarrant à 30 ms- et semble être majoritairement présente au niveau du cortex temporal. Les auteurs concluent que cette activité différentielle est le reflet d'un traitement avancé de reconnaissance des visages. Nous reviendrons sur cette conclusion un peu plus loin.

1.2.2.2 Enregistrements de l'activité à la surface du scalp

Le calcul des potentiels évoqués (PE) correspond au moyennage de l'activité électroencéphalographique (EEG) ou magnétoencéphalographique (MEG) lors d'un événement répété de nombreuses fois, comme la présentation d'un stimulus visuel. De par la grande résolution temporelle et leur facilité de mise en oeuvre, les études en PE sont particulièrement nombreuses et certains auteurs se sont intéressés à l'activité enregistrée dans les 200 premières millisecondes.

La même année que Seeck, George et collaborateurs enregistrent les potentiels évoqués visuels lors de la présentation de visages Mooney (George, Jemel, Fiori & Renault, 1997). Les Mooneys sont des représentations de stimuli contrastées à 100%, donc formés uniquement par des plages de noir et de blanc (figure 1.3). En comparant l'activité générée par leur première et leur deuxième présentation, les auteurs mettent en évidence un effet de répétition au niveau des

régions inféro-temporales **entre 50 et 110 ms**, que ce soit pour les stimuli présentés à l'endroit et perçus comme des visages ou pour des stimuli non reconnus, présentés à l'envers. Cette différence est interprétée comme un effet de répétition ou de familiarité et ne semble pas refléter des processus spécifiques à la reconnaissance des visages. Afin de tester cette hypothèse sur les expériences de Seeck présentées ci-dessus, Debruille a comparé l'activité générée par des visages célèbres reconnus en tant que tels



Figure 1.3 : Exemple de visages Mooney, présentée à l'endroit et à l'envers.

par des sujets lors d'une tâche comportementale et des visages inconnus pour ces mêmes sujets (Debruille, Guillem & Renault, 1998). Une différence apparaît **autour de 110 ms**, au niveau des aires fronto-centrales. En revanche, rien n'est observé avant cette latence, ce qui pour les auteurs permet d'exclure l'hypothèse de traitements visuels avancés spécifiques aux visages avant 70 ms, comme le suggérait Seeck. Par ailleurs comme le remarque Rufin VanRullen, les résultats obtenus par Seeck ne peuvent pas être complètement validés à cause d'une erreur conceptuelle d'ordre 1 dans la réalisation du protocole expérimental (VanRullen, 2000). En résumé, dans presque 10% des essais, les potentiels évoqués moyennés pour une cible (c'est à dire pour un stimulus répété) sont contaminés par le signal d'un distracteur qui le précède de 150 ms seulement. Or de nombreux travaux, à commencer par ceux de notre équipe (Thorpe, Fize & Marlot, 1996), ont montré que le signal entre cible et distracteur diffère entre 150 et 200 ms, soit de 0 à 50 ms après la présentation du stimulus cible concerné dans leur tâche...

Une autre étude soutient l'idée que des traitements avancés de catégorisation peuvent avoir lieu avant 70 ms (Mouchetant-Rostaing et al., 2000a). Dans une première partie, les auteurs présentent des photographies de visages d'hommes ou de femmes à des sujets dont la tâche est de détecter les visages avec des lunettes. Certaines séries contiennent uniquement des visages de femmes ou uniquement des visages d'hommes (Condition 1), et certaines à la fois des visages d'hommes et de femmes (Condition 2). Dans cette dernière condition, les sujets peuvent donc réaliser une catégorisation du genre qui est dite *implicite*, puisqu'elle n'est pas l'objectif de la tâche. En comparant l'activité enregistrée en potentiels évoqués entre la condition où hommes et femmes sont mélangés et les conditions où les deux genres sont séparés, les auteurs observent une différence dans les signaux **entre 45 et 90 ms**, interprétée comme le reflet de mécanismes automatiques de catégorisation du genre. En revanche, lorsque les sujets ont pour tâche de catégoriser *explicitement* le genre des stimuli (dans le cas bien sûr où les deux types de stimuli sont mélangés - Condition 3), le signal ne diffère plus de celui où les genres sont séparés. Cette catégorisation précoce est aussi retrouvée avec des stimuli de mains, et des stimuli non-biologiques comme des formes hachurées (Mouchetant-Rostaing et al., 2000b). D'une part, il peut paraître étrange que seuls les mécanismes de catégorisation implicite soient visibles sur cette période, alors que les mécanismes liés à la catégorisation explicite ne le sont pas. D'autre part, une cinquantaine de millisecondes est une période vraiment très courte pour voir apparaître des

indices de catégorisation avancée comme celle du genre d'un visage. On peut se demander si l'activité différentielle enregistrée ne reflète pas d'autres phénomènes visuels, et en particulier un effet de séquence lorsque des stimuli appartenant à une même catégorie – et physiquement similaires- sont répétés de nombreuses fois. Il faut remarquer que dans la Condition 3 où les sujets doivent reporter explicitement le genre des visages, les stimuli sont répartis dans une proportion 20% de cibles - 80% de distracteurs dans une même série. Cette condition se rapproche de la Condition 1 où 100% des stimuli sont du même genre, alors que la Condition 2, avec 50% de chaque genre, s'en éloigne. Pour définitivement rejeter l'hypothèse d'un effet de séquence, il faudrait pouvoir montrer que l'activité différentielle à 50 ms n'est pas présente lorsque les sujets réalisent la tâche explicite dans des séries où 50% de chaque genre est représenté.

Si les études montrant une activité spécifique à 50 ms peuvent être contestées sur le plan expérimental, les travaux révélant une activité précoce différentielle **autour de 100 ms** sont déjà plus nombreux, que ce soit entre des stimuli linguistiques et non linguistiques (Schendan, Ganis & Kutas, 1998), entre des visages à l'endroit et des visages à l'envers (Linkenkaer-Hansen et al., 1998) ou entre des visages et d'autres catégories comme des têtes d'animaux ou des objets courants (Halgren et al., 2000). Concernant cette dernière référence, Halgren remarque que l'activité observée en MEG à partir de 110 ms est particulièrement sensible aux propriétés physiques des stimuli (émotions positives vs. neutres – photographies vs. schémas de visages...). Comment interpréter l'activité spécifique entre 50 et 130 ms ?

1.2.2.3 Quels traitements entre 50-130 ms ?

Est-il raisonnable de penser que des traitements avancés de catégorisation sont visibles dès 50 ms ? Nous avons vu que les études présentant ces résultats pouvaient être critiquées sur le plan expérimental, ce qui est plutôt rassurant au vu des données électrophysiologiques chez l'animal montrant que les informations visuelles ne quittent pas la rétine avant 15-20 ms. Cependant il est quand même possible d'envisager que des aires moins avancées dans la hiérarchie corticale soient génératrices d'un signal précoce visible en électrophysiologie chez l'homme, comme le suggère Halgren (Halgren et al., 2000).

Tout stimulus visuel -qu'il soit damier, images d'objets ou de visage- génère un signal caractéristique, avec différentes composantes exogènes. La question posée par les différentes études décrites ci-dessus se réfèrent à l'analyse différentielle des signaux entre plusieurs catégories de stimuli, et la sensibilité de certaines parties du signal à une propriété physique voire à une catégorie donnée suggère que les traitements sont spécifiques à cette propriété ou catégorie. Maintenant, il est important de s'interroger sur les différentes sources de ce signal en fonction du temps. Les outils de localisation de source et le couplage des enregistrements aux images anatomiques obtenues en IRM permettent de répondre en partie au problème de résolution spatiale dont souffrent les méthodes d'enregistrements à la surface du scalp. Le débat n'est pas clos entre les différentes équipes qui cherchent à déterminer le profil spatio-temporel des activations corticales. Cependant, il semble que la première composante générée autour de 50 ms (appelée aussi C1) soit principalement le reflet de l'activité dans V1 (Di Russo et al., 2001; Vanni

et al., 2004b) alors que la composante suivante, entre 75 et 130 ms (correspondant à P1), est générée par de multiples aires visuelles, sachant que la majorité des aires rétinitopiques ont déchargé à cette période (Clark, Fan & Hillyard, 1995; Foxe & Simpson, 2002; Vanni et al., 2004a).

Alors comment considérer l'activité enregistrée autour de 100 ms ? Faut-il penser qu'elle reflète uniquement des différences de propriétés physiques entre les groupes d'images ou est-elle le reflet de traitements déjà plus avancés ?

Certaines données de notre équipe appuient la première idée. Dans une expérience de double tâche, les sujets devaient catégoriser en alternance soit des animaux soit des moyens de transport, et les cibles d'une tâche étaient les distracteurs de l'autre (VanRullen & Thorpe, 2001). Ainsi, un même groupe d'images pouvaient avoir deux statuts différents, soit cible soit distracteur. En comparant les signaux moyennés sur les essais cible et distracteur (que ce soit pour les animaux ou pour les moyens de transport), une activité différentielle classique apparaît à partir de 150 ms, reflet des traitements visuels de catégorisation. Or cette activité à 150 ms est aussi précédée d'une activité différentielle de moindre amplitude, autour de 100ms. Mais lorsque l'on compare les signaux d'un même groupe d'images dans des statuts différents (par exemple des animaux vus soit en cible soit en distracteur), cette composante précoce disparaît contrairement au signal différentiel à 150 ms (même si son amplitude diminue), ce qui suggère fortement que l'activité à 100 ms est liée aux différences physiques entre les images et indépendante de la tâche. De la même manière, les modifications de luminance ou de contraste affectent principalement cette composante précoce, laissant l'activité différentielle à 150 ms quasiment intacte (Macé, Rousselet & Fabre Thorpe, en préparation).

Cependant d'autres études reportent des données plus ambiguës. En particulier, la double tâche décrite ci-dessus apporte des résultats différents lorsque les sujets catégorisent non pas des moyens de transport et des animaux, mais des visages et des animaux (Rousselet, Macé, Thorpe & Fabre-Thorpe, soumis). En soustrayant les signaux obtenus sur les visages en statut cible et les animaux en statut distracteur, une activité différentielle est obtenue dès 120 ms, suggérant que cette composante puisse déjà représenter la signature d'une activation précoce codant pour les propriétés diagnostiques des images. Mais l'argument le plus pertinent pour affirmer que des traitements complexes peuvent déjà survenir à ces latences est issu du protocole introduit par Holle Kirchner (Kirchner & Thorpe, 2005). Dans ce protocole de choix forcé, deux images sont flashées simultanément de part et d'autre du point de fixation, et le sujet doit orienter son regard vers l'image contenant une catégorie cible, par exemple un animal. Il est particulièrement surprenant d'observer que les sujets sont capables de fournir des réponses oculaires en moins de 130 ms (premières réponses significatives à 120 ms), délai pendant lequel des traitements sensoriels, décisionnels et moteurs ont du être effectués. Nous reviendrons plus largement sur cette expérience dans le chapitre 3.

Ces résultats ne signifient pas que les traitements visuels de catégorisation sont complètement achevés en 100 ms, mais ils suggèrent fortement que les populations neuronales ont déjà commencé à extraire et intégrer des informations pertinentes pour la tâche à réaliser. Ainsi, l'intégration rapide et rudimentaire de la scène visuelle à partir des indices saillants

permettrait d'aider le système à orienter ses traitements plus détaillés (Bar, 2004; Bar et al., 2006). Une question sous-jacente à cette remarque porte sur la nature de ces informations pertinentes...

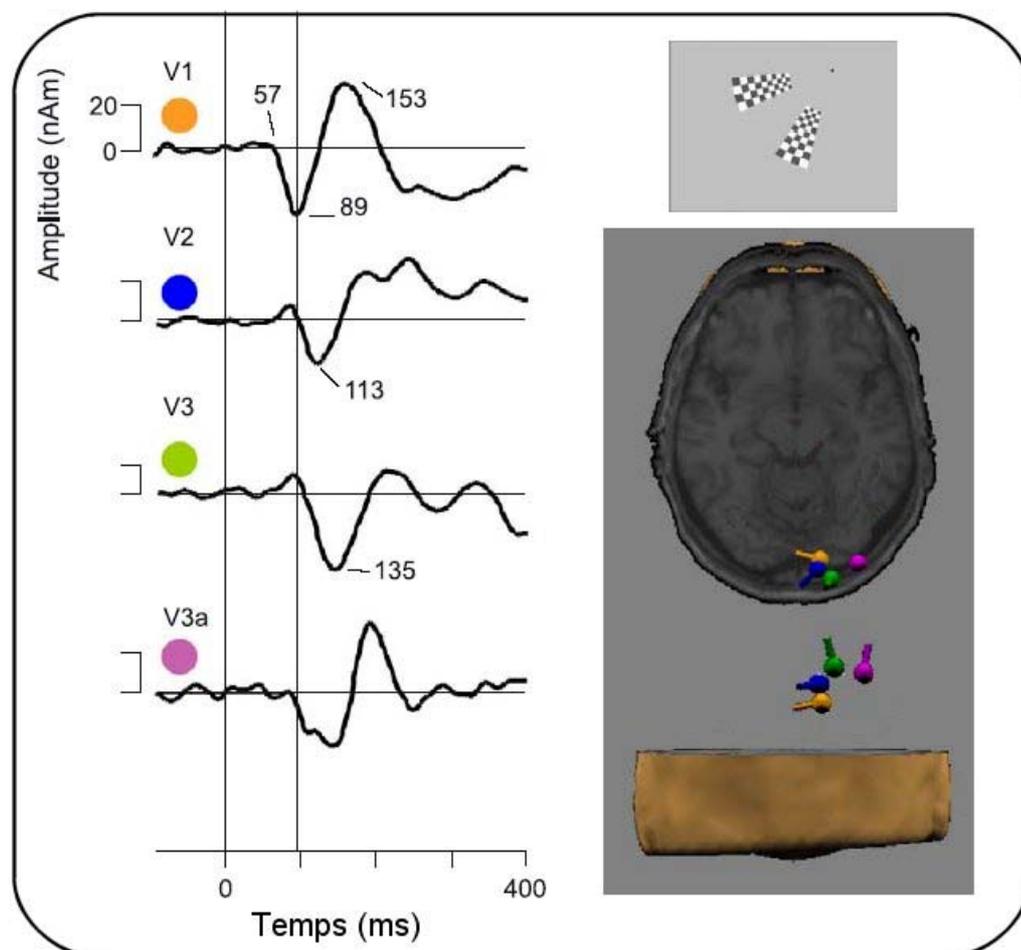


Figure 1.4 : Potentiels évoqués et localisation des sources associées lorsque l'on présente à un sujet humain un stimulus pattern. Dans cette condition particulière, le stimulus est composé de deux damiers et présenté dans le quadrant inférieur gauche, pendant 500 ms. Les dipôles déterminés par l'analyse de sources des données EEG sont placés sur un plan anatomique IRM, et mis en correspondance avec les régions activées lors de cette même tâche en IRM fonctionnelle. Adapté de Vanni et al., 2004b.

1.2.2.4 Activité visuelle spécifique après 150 ms

La littérature concernant le signal EEG ou MEG après 150 ms est particulièrement vaste, et nous ne chercherons pas à faire de revue exhaustive de ces différents travaux. Mais il paraît clair que l'activité enregistrée après 150 ms reflète des traitements relativement avancés, comme le montrent toutes les études travaillant avec des stimuli de visages. En particulier, la composante N170 ou M170 apparaît entre 150 et 200 ms au niveau des électrodes occipito-temporales. Elle est souvent interprétée comme le reflet d'un encodage spécifique des visages, correspondant à la génération de leur représentation globale et configurale (Bentin et al., 1996; Carmel & Bentin, 2002; Eimer, 1998; Sagiv & Bentin, 2001). De manière plus générale, d'autres auteurs soutiennent que la composante N1 apparaissant dans cette gamme de latence n'est pas spécifique

aux visages, mais est également sensible à d'autres catégories de stimuli comme des fleurs ou des voitures, pour lesquels son amplitude augmente (Curran, Tanaka & Weiskopf, 2002; Itier, Latinus & Taylor, 2005; Rousselet, Mace & Fabre-Thorpe, 2004). Elle reflèterait d'avantage le niveau d'expertise des sujets pour les catégories de stimuli présentés, ce qui justifierait entre autres sa sensibilité aux visages (Rossion, Curran & Gauthier, 2002; Rossion et al., 2002; Tanaka & Curran, 2001).

Nous reviendrons plus largement sur la particularité des stimuli visages dans le Chapitre 4, dans lequel je présenterai une série d'expériences réalisées à la fois sur des images d'humains, de visages et d'animaux.

Quelle que soit l'interprétation finale, le signal après 150 ms semble sensible aux effets d'inversion des stimuli (Bentin et al., 1996; Eimer, 2000b; Rossion et al., 2000; Rousselet, Mace & Fabre Thorpe, 2004), à la charge attentionnelle (Eimer, 1997; Eimer, 2000a) (Mangun & Hillyard, 1991) ou à d'autres facteurs cognitifs liés à la tâche (Bentin & Golland, 2002; Mangun & Hillyard, 1991).

1.2.3 Approche comportementale chez l'homme

Dans un premier temps, nous avons passé en revue les travaux électrophysiologiques chez l'animal fournissant des données précises sur les latences des réponses neuronales à des endroits précis de la hiérarchie corticale visuelle. Puis nous avons décrit un certain nombre d'études électrophysiologiques chez l'homme qui se sont intéressées à définir un profil spatio-temporel général d'activation neuronale lorsque les sujets sont engagés dans une tâche cognitive.

Nous terminons cette revue bibliographique par une description (non-exhaustive) des travaux comportementaux chez l'homme. Les études comportementales ont largement contribué à développer nos connaissances sur la dynamique temporelle des traitements sensoriels, et les premières données sur ce sujet sont même largement antérieures aux travaux décrits précédemment.

Pourquoi orienter cette revue dans ce sens, sans respecter l'ordre chronologique et en partant des résultats électrophysiologiques les plus précis aux données comportementales plus générales et plus variables ? D'une part, la plupart de mes travaux de thèse s'appuient sur des résultats comportementaux. D'autre part l'approche comportementale, de part son manque de visibilité précise sur les activations spatio-temporelles cérébrales générées lors des tâches, a poussé les auteurs à développer des protocoles expérimentaux adaptés et à décrire des modèles généraux sur les mécanismes de transfert et d'intégration des informations sensorielles pouvant concilier la majorité des résultats.

A la base de ces travaux sur l'étude de la dynamique des traitements sensoriels, l'hypothèse principale considère que les processus mentaux peuvent être observés par l'intermédiaire de mesures comportementales, comme le temps de réaction, la précision des réponses ou des compromis précision-vitesse (« speed-accuracy tradeoffs »). Différents paradigmes

expérimentaux de « chronométrie mentale » ont ainsi été développés, dans lesquels la performance des sujets est mesurée lors de nombreux essais (Figure 1.5 et voir (Meyer, Osman, Irwin & Yantis, 1988; Posner, 1986) pour de très bonnes revues de ces paradigmes). Un essai est en général constitué d'une période de fixation d'un signal d'alerte (comme une croix de fixation) suivi de la présentation du stimulus et de la période de réponse qui doit être rapide et précise.

Dès le 19^{ème} siècle, Helmholtz établit que la vitesse de conduction neuronale des nerfs périphériques est d'approximativement 50 m/s en utilisant un protocole simple de temps de réaction (Helmholtz, 1853). Un peu plus tard, Donders développe sa « méthode de soustraction » : en soustrayant le temps mis pour répondre à un seul ou à deux stimuli, Donders cherche à isoler le temps mis pour discriminer deux stimuli (Donders, 1969). Plus tard, il soustrait le temps mis pour répondre à deux stimuli lorsque une ou deux réponses sont possibles, dissociant cette fois le délai de sélection de la réponse. A la suite de ces travaux, différents modèles ont été décrits pour essayer d'isoler les différentes étapes intervenant dans la chaîne totale des traitements, de la stimulation à la réponse motrice. Les enjeux principaux de ces recherches ont été de déterminer si les différentes étapes de traitements avaient lieu de manière sérielle ou parallèle, et si les informations étaient transmises d'une étape à l'autre de manière continue ou discrète (figure 1.5). Ces questions n'ont toujours pas trouvé de réponse unanime, et nous reviendrons plus largement sur cette problématique dans le chapitre 3.

Dans les années 50, l'étude des temps de réaction est combinée à une approche plus mathématique, celle de la théorie de l'information, afin de déterminer les vitesses de transmission de l'information dans le système nerveux. A partir des années 60, l'arrivée des technologies informatiques, des enregistrements unitaires chez l'animal et des enregistrements des potentiels évoqués chez l'homme ont permis d'observer les traitements neurophysiologiques et de confirmer ou rejeter les modèles développés à partir des observations comportementales.

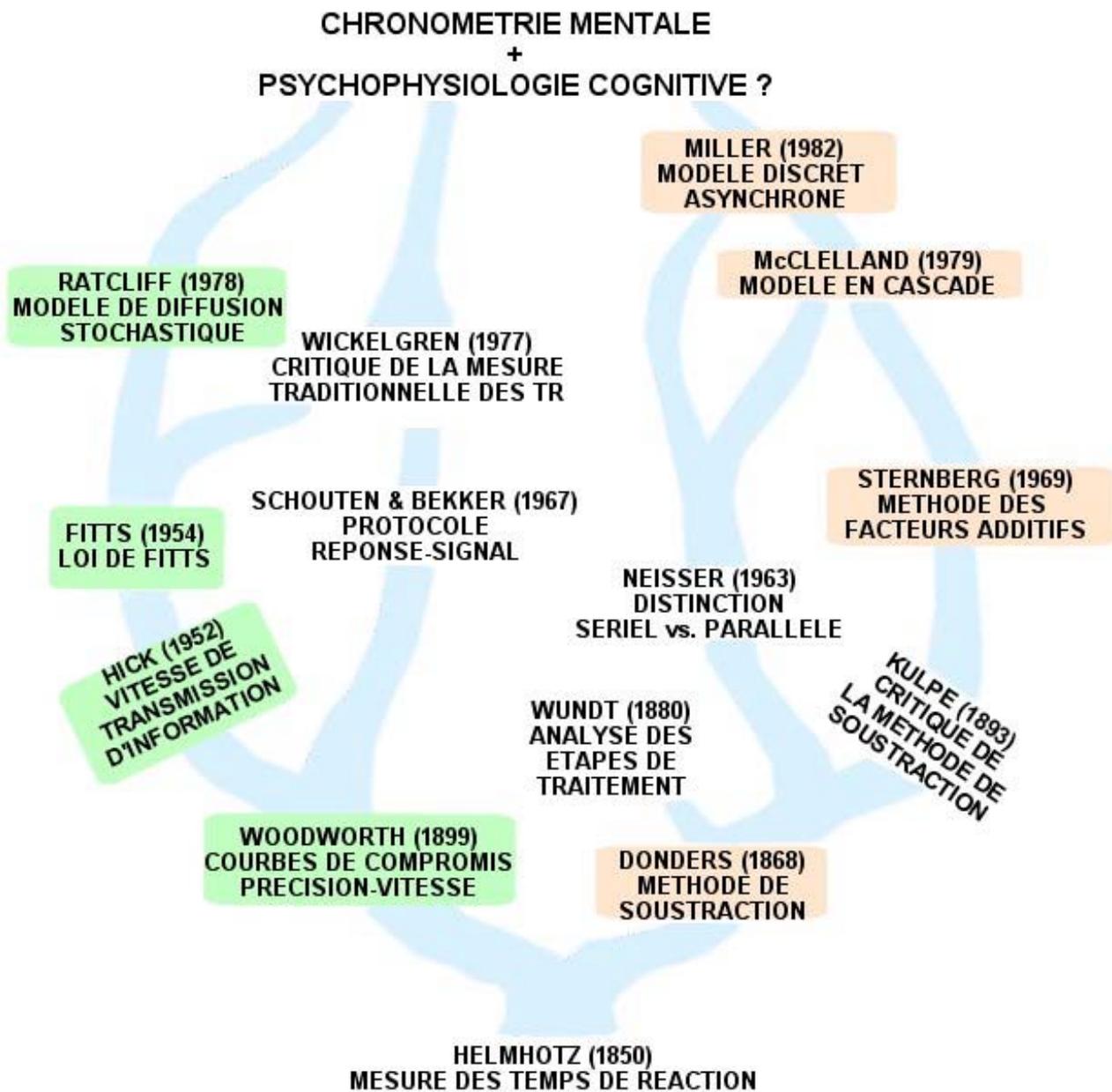


Figure 1.5 : *Arbre chronologique des principaux courants de la chronométrie mentale, définie comme l'étude du déroulement temporel des informations dans le système nerveux. A la base de ces travaux, Helmholtz développe un protocole simple de mesure du temps de réaction. La branche de droite oriente les travaux sur les modèles décrivant les différentes étapes de traitement de l'information. La branche de gauche réunit les travaux et modèles s'appuyant sur les composantes statistiques de prise de décision. Adapté de Meyer, Osman, Irwin & Yantis, 1988.*

Voici quelques descriptions des courants ou modèles surlignés sur l'arbre :

Méthode de soustraction de Donders (Donders, 1969). Donders analyse différentiellement les données en variant les tâches produites par les sujets. Par exemple, il mesure les temps de réaction lorsque les sujets répondent à un stimulus unique (tâche A) ou à des stimuli multiples avec des réponses uniques ou multiples (tâche B). Donders propose alors une estimation du temps nécessaire aux processus de discrimination et de sélection des réponses en soustrayant les temps de réaction obtenus dans la tâche A et B. Cette méthodologie suppose que les différents processus estimés sont strictement successifs et qu'ils contribuent à des délais indépendants dans la mesure du temps de réaction global.

Méthodes des facteurs additifs (Sternberg, 1969). En manipulant différents facteurs de la tâche, comme les propriétés du stimulus visuel, la longueur de listes mémorisées, la probabilité de réponses alternatives, Sternberg recherche les éléments qui entrent en interaction ou qui s'additionnent dans le temps de réaction global., le but étant de déterminer le nombre et le rôle des processus cognitifs différents engagés dans une tâche. Sternberg en isole au moins 3 : l'encodage sensoriel, le stockage de l'information sous une représentation mnésique, et la sélection de la réponse.

Modèle en cascade (McClelland, 1979). Contrairement au modèle de Sternberg qui supposait que les différentes étapes de traitement s'engagent de manière sérielle et rajoutent une quantité discrète d'information à la chaîne globale des traitements, McClelland propose que ces processus se déroulent plutôt de manière parallèle, grâce au flux continu d'activation qui transmet en permanence l'information d'un processus au suivant.

Modèle discret asynchrone (Miller, 1982). Comme le modèle en cascade proposé par McClelland, le modèle discret asynchrone suppose que les traitements sont massivement parallèles, suggérant que les processus de sélection et de préparation motrice sont déjà engagés alors que les traitements sensoriels sont encore en cours. La différence entre les deux modèles porte sur le format de l'information transmise d'un processus à l'autre. Le modèle discret asynchrone s'appuie sur des quantités discrètes d'information, séparées dans le temps, relatives à quelques propriétés du stimulus, et non pas sur un flux continu d'activation comme le propose le modèle en cascade.

Vitesse de transmission de l'information (Hick, 1952). Les travaux de Hick appuient l'idée que les temps de réaction sont linéairement corrélés à la quantité d'information nécessaire pour une tâche cognitive. Cette relation n'est compatible qu'avec un modèle sériel de transmission de l'information. Cependant il reste difficile de concilier l'idée d'une capacité fixe de transmission de l'information avec la grande variabilité observée sur les temps de réaction selon les différentes manipulations de la tâche.

Courbes de compromis rapidité-vitesse (Woodworth, 1899). Woodworth a été le premier à mettre en évidence la relation complexe qui existe entre vitesse et précision : plus le mouvement demande à être précis, plus le délai nécessaire pour répondre est important.

Loi de Fitts (Fitts, 1954). Le temps mis pour pointer une cible est proportionnel à sa distance et inversement proportionnel à sa taille. Ainsi les cibles les plus petites, qui demandent une certaine précision, nécessitent plus de temps pour être atteintes.

Modèle de diffusion stochastique (Ratcliff, 1978). Ratcliff propose un modèle dans lequel l'activation s'accumule progressivement et indépendamment pour chacune des entrées du système jusqu'à atteindre les seuils spécifiques aux représentations mnésiques encodées.

En dehors de l'étude des temps de réaction, une autre approche psychophysique pour étudier la dynamique des traitements consiste à essayer d'isoler le temps d'intégration des informations de la scène visuelle. Dès la deuxième moitié du 19^{ème} siècle, grâce au développement du tachistoscope qui permet de varier avec une très grande précision la durée de présentation de stimuli visuels, de nombreux travaux ont cherché à démontrer les limites du système pour percevoir et mémoriser des stimuli présentés brièvement (Cattell, 1885-1886; Erdman & Dodge, 1898).

A la base de ces recherches, la compréhension de nos capacités de lecture a constitué une grande motivation (Coltheart, 1999; Rayner & Pollatsek, 1987). Pendant des périodes de temps très courtes, le système visuel est capable de se fixer sur une partie du texte, d'en extraire l'information, de décider et de planifier le programme moteur pour déclencher une saccade oculaire vers le point suivant avec une localisation précise. Ce temps de fixation dure approximativement 250-350 ms. La question de ce qui a été perçu et mémorisé durant cette période a amené à développer des protocoles expérimentaux divers.

Par exemple à la fin des années 50, Sperling présente un set de caractères alphanumériques à des sujets pendant un temps très court et leur demande de reporter soit l'ensemble de ce qu'ils ont vu, soit seulement une partie, comme le haut ou le bas du set (Sperling, 1960). Il remarque que lors du rapport partiel, les sujets voient 3 à 4 fois plus de lettres que lors du rapport complet, suggérant qu'une quantité importante d'informations visuelles sont intégrées après l'affichage du stimulus, mais que ces informations sont rapidement éliminées ou remplacées par d'autres informations entrantes. Cette limitation dans l'encodage à long-terme des informations visuelles explique pourquoi les sujets se disent souvent gênés de ne pas pouvoir rapporter oralement ou par écrit une quantité d'informations qu'ils ont pourtant l'impression d'avoir perçu. En contrepartie, il note que le temps pendant lequel les informations concernant un stimulus sont disponibles ne correspond pas au délai d'affichage du stimulus, mais peut être largement supérieur grâce aux propriétés de persistance du système visuel.

Cette impression que nous intégrons beaucoup de détails de la scène visuelle mais que nous ne pouvons pas tout reporter *a posteriori*, a aussi été testé avec des stimuli complexes, comme des photographies d'images naturelles. Dans les années 70, Potter (Potter & Levy, 1969) et Forster (Forster, 1970) définissent le paradigme RSVP (Rapid Serial Visual Presentation) qui consiste à présenter une séquence de stimuli, images ou mots, à différentes fréquences et à tester ce qui a été vu ou mémorisé.

Dans la première étude de Potter et Levy (Potter & Levy, 1969), les auteurs présentaient successivement 16 photographies parmi 128, avec des fréquences variant de 2 à 10 images par seconde. Les photographies étaient aussi variées que des images d'objets, de visages, de scènes d'intérieur ou d'extérieur, et n'étaient vues qu'une fois. Les sujets devaient réaliser un test de reconnaissance, en indiquant si l'image test présentée après le flux faisait partie de la séquence ou non. Les auteurs remarquent que la performance est étonnamment basse dans ce contexte de présentation puisque lorsque les images du flux sont affichées moins de 125 ms (soit 8 images par seconde), la performance dans la tâche de reconnaissance est proche du niveau chance. En revanche, lorsque ces mêmes images sont présentées isolément mais suivies d'un masque

(constitué de bouts de photographies mélangés aléatoirement), la performance reste correcte même avec des durées de présentation autour de 90 à 120 ms. Dans une autre série d'études (Potter, 1975; Potter, 1976), Molly Potter compare les performances dans des tâches de détection ou de reconnaissance d'images. Deux tâches de détection sont testées : les sujets doivent repérer dans la séquence une image cible soit présentée avant la séquence (Condition « Détection Image»), soit décrite préalablement par un titre court (par exemple « un bateau » ou « une femme assise sur un lit », Condition « Détection Nom »). La tâche de reconnaissance était identique à celle présentée au dessus (Condition « Reconnaissance Image »). Comme le montre la figure 1.6, la performance dans les tâches de détection est nettement supérieure à la performance de reconnaissance, suggérant que les sujets peuvent extraire le sens d'images n'ayant aucun rapport entre elles même lorsqu'elles ne sont séparées que de 100 à 200 ms. En revanche, des délais supplémentaires sont nécessaires pour consolider ces informations en mémoire et réaliser la tâche de reconnaissance. Dans les deux cas, des phénomènes de masquage interviennent, c'est à dire que la présentation d'images successives génère des interférences dans les traitements liés à chaque image. Potter fait la distinction entre *masquage perceptuel*, responsable de la perturbation des traitements précoces et le *masquage conceptuel*, qui intervient sur les traitements plus avancés, au moment de l'encodage mnésique des informations visuelles (Potter, 1976). Si le masquage conceptuel explique la performance particulièrement basse des sujets dans la tâche de reconnaissance en limitant l'encodage mnésique des informations, le masquage perceptuel lui, expliquerait la baisse la performance pour la durée d'affichage la plus courte à 113 ms, même pour les tâches de détection (Figure 1.6). Dans ce dernier cas, l'extraction des propriétés du stimulus est limitée dans le temps et contrainte par l'apparition du stimulus suivant.

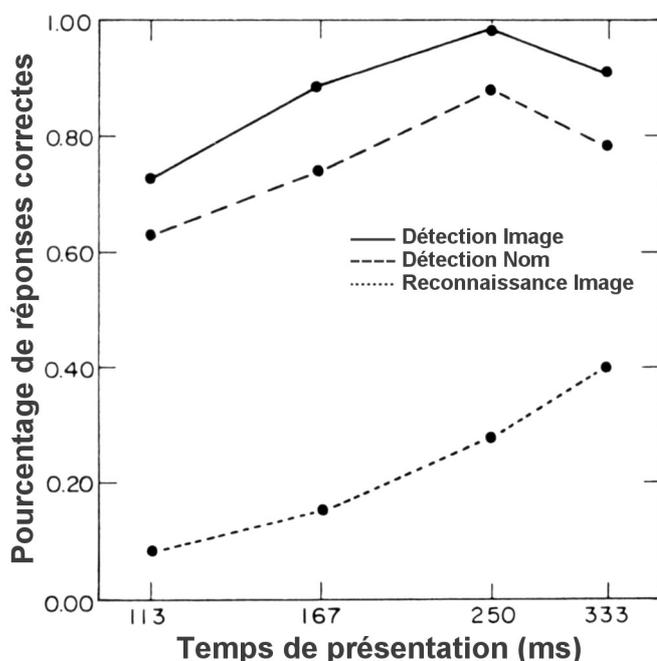


Figure 1.6 : Pourcentage de réponses correctes dans les tâches de détection et de reconnaissance en fonction de la fréquence de présentation du flux RSVP. Le pourcentage correct ne prend en compte que les réponses sur les cibles. L'image cible était présentée avant la séquence dans la tâche de détection d'image, et après la séquence dans la tâche de reconnaissance. Dans la tâche Détection Nom, l'image cible est décrite avant la séquence par un titre court. Adapté de Potter, 1976.

L'observation de l'impact du masquage perceptuel sur les performances est un moyen très efficace pour observer la quantité et la nature de l'information intégrée pendant une période de temps donnée. Le paradigme RSVP rejoint en cela les nombreuses études qui se sont développées en parallèle en utilisant le phénomène de masquage comme un outil d'observation de la dynamique des traitements visuels. Nous nous sommes particulièrement intéressés à ces paradigmes expérimentaux, masquage et RSVP, afin d'aller plus loin dans notre analyse de la vitesse des traitements lors de la tâche de catégorisation visuelle de scènes naturelles. En effet, les expériences réalisées depuis 1996 ont déjà apporté un certain nombre d'indications montrant que les temps de traitement semblent être particulièrement courts :

- la présentation des stimuli pendant 20 ms seulement confirme que l'extraction des informations peut être très rapide et reposer sur les propriétés de persistance du système visuel,
- les temps de réaction associés à la précision comportementale apportent des indices sur l'efficacité des traitements dans ce type de tâche (Fabre-Thorpe, Richard & Thorpe, 1998; Thorpe, Fize & Marlot, 1996),
- l'analyse de l'activité différentielle mesurée entre les potentiels cibles et distracteurs à partir de 150 ms a contraint les modèles de transmission de l'information dans le système visuel (Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2002).

Cependant, des questions restent posées sur les délais d'intégration minimum nécessaires à la tâche et de manière plus générale, sur la dynamique des traitements sous-jacents. En quoi l'utilisation du protocole de masquage dans la tâche de catégorisation visuelle peut répondre à cette question (Chapitre 2) ? Est-ce que cet outil peut nous renseigner sur l'organisation temporelle des composantes sensorielles, décisionnelles et motrices engagées dans la tâche (Chapitre 3) ou nous renseigner sur la nature des informations visuelles utilisées lors de la catégorisation de scènes naturelles (Chapitre 4) ?

Dans la section suivante, nous allons passer en revue différentes études aussi bien comportementales qu'électrophysiologiques ayant utilisé le masquage comme un outil d'analyse du déroulement temporel des informations dans le système visuel, ainsi que les modèles associés à ces observations.

1.3 Étudier le décours temporel des traitements visuels: paradigme du masquage

Le masquage est un outil expérimental aujourd'hui couramment utilisé, dans des travaux de psychophysique comme en imagerie cérébrale. Enns et Di Lollo font cependant une observation intéressante : en survolant un volume récent du journal *Perception and Psychophysics* (vol.61, 1999), ils remarquent que 15% des articles utilisent le masquage rétrograde dans un but pratique, afin de limiter l'accès visuel à une cible sur une période de temps donnée. En revanche seulement 5% sont intéressés par les dimensions spatiales et temporelles du masquage lui-même. Ce sont ces dernières études qui vont maintenant nous intéresser. Les expériences de masquage utilisent des protocoles parfois très différents, et après avoir résumé la terminologie spécifique des travaux de masquage, nous établirons une revue non exhaustive des travaux ayant conduit à développer des théories fondées sur les interactions précoces entre cible et masque. En effet, dans le cadre de la catégorisation visuelle ultra-rapide, nous avons relevé un certain nombre d'indices suggérant que les traitements permettant de détecter un animal dans une scène naturelle pouvaient être produits dans les 150 premières millisecondes après la présentation du stimulus. Nous nous attacherons donc à essayer de synthétiser les différents modèles proposés pour expliquer les interférences précoces du masquage.

1.3.1 Terminologie du masquage

Le masquage visuel consiste à réduire la visibilité d'un premier stimulus, la cible, en manipulant la proximité spatio-temporelle d'un deuxième stimulus, le masque. Ce protocole a joué un rôle fondamental dans l'étude des propriétés spatiales et temporelles de la perception visuelle. Dans son livre « *Visual masking, an integrative approach* », Breitmeyer dresse une revue très complète de l'histoire des travaux de masquage et pose la définition des différents phénomènes de blocage de la perception ainsi que leurs mécanismes sous-jacents (Breitmeyer, 1984).

Plusieurs dissociations doivent être faites, en fonction des relations spatiales et temporelles qui existent entre les contours de la cible et du masque (Figure 1.7) Lorsque le masque et la cible se superposent spatialement, on distingue trois types de masquage selon la nature du masque : *masquage par flash lumineux* (le masque est un écran homogène de lumière généralement plus large que la cible pour éviter les interactions entre contours), *masquage par pattern* (le masque est constitué d'un ensemble de patterns non homogènes comme des lignes, des lettres ou diverses formes) et *masquage par bruit* (le masque est composé de bruit blanc). Sur l'échelle temporelle, on distingue le masquage *antérograde* (masque présenté avant la cible), ou *rétrograde* (masque présenté après la cible). Dans le cas où masque et cible sont adjacents mais ne se recouvrent pas, on parle de *masquage métacontraste* si le masque est présenté avant la cible et *masquage*

paracontraste si le masque intervient après. Le masquage est dit *métacontraste simultané* si les deux stimuli sont présentés en même temps.

Dans les parties qui suivent, nous présenterons principalement les données de la littérature ayant utilisé le masquage par pattern (rétrograde et antérograde) ainsi que le masquage métacontraste/paracontraste pour essayer de définir la nature des interférences qui conduisent à des baisses importantes de performance.

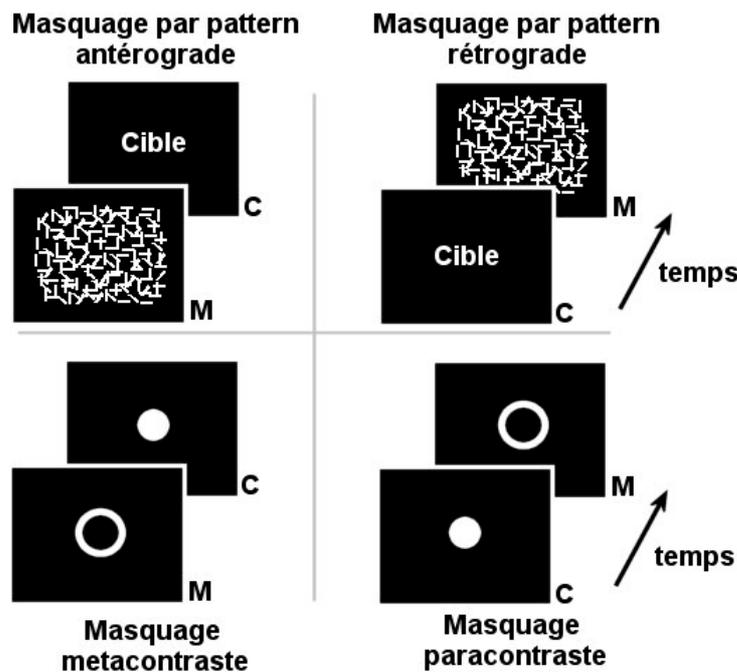


Figure 1.7 : Différents protocoles de masquage selon les relations spatiales et temporelles entre masque et cible. La lettre C à côté des écrans représente un stimulus Cible, la lettre M un stimulus Masque.

L'impact du masquage sur les traitements visuels est en général mesuré par la diminution du pourcentage de réponses correctes lors de tâche de détection ou d'identification. Les variables manipulées incluent le délai entre cible et masque, ou d'autres paramètres comme l'intensité, le contraste, la fréquence spatiale, la similarité cible-masque, ou la complexité du masque. Il existe plusieurs façons de manipuler le délai entre cible et masque, selon que l'on prenne en compte ou non la durée des stimuli (Figure 1.8). On utilise alors le délai entre fin de la cible et début du masque (Inter Stimulus Interval **ISI**), le délai entre début de la cible et début du masque (Stimulus Onset Asynchrony **SOA**) et le délai entre fin de la cible et fin du masque (Stimulus Termination Asynchrony **STA**). Le **SOA** est la mesure la plus couramment utilisée dans la littérature bien que de récents travaux aient cherché à dissocier plus précisément l'impact de ces trois mesures lors du masquage paracontraste et métacontraste (Macknik & Livingstone, 1998).

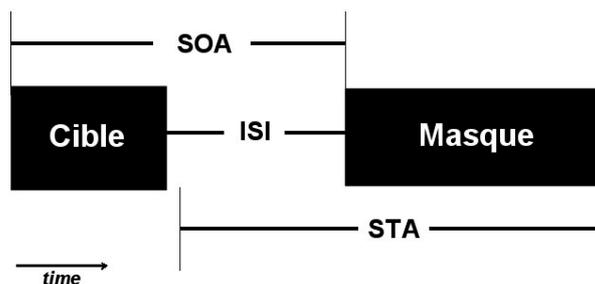


Figure 1.8 : Relations temporelles entre la cible et le masque. Les SOA (Stimulus Onset Asynchrony), ISI (Inter Stimulus Interval) et STA (Stimulus Termination Asynchrony) peuvent être manipulés afin de varier le degré de perturbation des traitements visuels.

En fonction de la nature du masque, on obtient différents profils de dégradation de la perception lorsque l'on varie le SOA (Figure 1.9). On parle de **fonction de Type-I** (ou Type-A ou fonction monotone ou encore forme en « J ») lorsque les effets de masquage les plus forts interviennent pour des SOA réduits, puis diminuent au fur et à mesure que l'on augmente le SOA. Il s'agit d'une **fonction de Type-II** (ou Type-B ou forme en « U ») lorsque les effets de masquage sont faibles avec un SOA = 0, optimaux avec un SOA intermédiaire, et diminués pour des SOA importants.

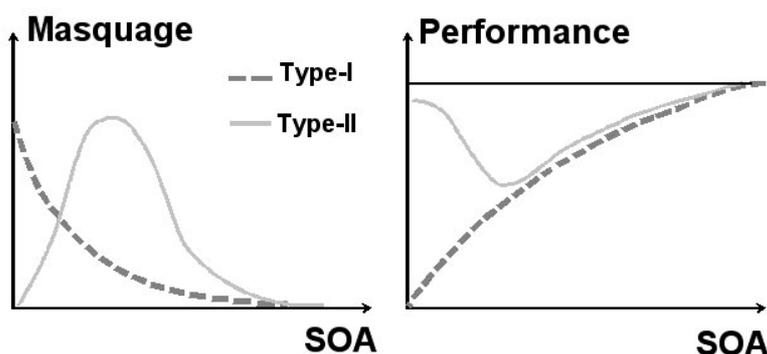


Figure 1.9 : Effets du masquage et impact sur la performance en fonction du SOA. Deux profils de masquage, de Type-I ou de Type-II, sont définis selon la latence de leur impact maximum sur les performances.

Le champ expérimental des travaux de masquage est particulièrement vaste et a nécessité de poser la définition de nombreux termes, que ce soit pour distinguer les protocoles entre eux (masquage par flash lumineux, pattern, bruit, paracontraste / métacontraste), la relation temporelle entre cible et masque (antérograde ou rétrograde) ou leur impact sur les performances (fonction de Type-I ou II). La description des mécanismes sous-jacents reflète aussi cette complexité, et de nombreux modèles ont été proposés avec chacun leur termes spécifiques. Dans la partie suivante, nous allons essayer d'établir une revue non exhaustive de ces différents travaux, afin de comprendre les principaux modèles établis sur les mécanismes d'interférence entre cible et masque.

1.3.2 Psychophysique du masquage

1.3.2.1 Masquage par pattern

Dans les années 70, Turvey réalise toute une série d'expériences pionnières pour la compréhension des mécanismes d'intégration de l'information visuelle, en utilisant un protocole de masquage par pattern (Turvey, 1973). Au cours de ces tâches, les sujets doivent identifier les 3 lettres d'un trigramme de consonnes (par exemple « TMX »), suivi par un masque composé de barres à différentes orientations, avec un SOA variable (Figure 1.10A). Turvey rapporte que la performance est encore affectée lorsque le masque est présenté environ 160 ms après le trigramme. Mais il centralise surtout ses travaux autour de la manipulation de l'énergie de la cible et du masque, à travers des changements de contraste. Au cours des 60 premières millisecondes, le rapport d'énergie entre cible et masque semble jouer un rôle critique (Figure 1.10B). Si la cible a un contraste plus important que le masque, la courbe de masquage respecte une fonction de Type-II : la présentation simultanée des deux stimuli n'empêche pas l'identification de la cible alors que l'effet est maximal pour un SOA d'environ 50 ms. Au contraire, si le masque a un contraste plus important, on retrouve une fonction de masquage de Type-I, plus classique dans les expériences de masquage par pattern. D'autres études ont reporté ce phénomène (Hellige, Walsh, Lawrence & Prasse, 1979; Kolers, 1962; Schiller, 1965; Spencer & Shuntich, 1970; Weisstein, 1972). Turvey explique ces différences en décrivant deux mécanismes de masquage : le masquage par intégration et le masquage par interruption :

- *Masquage par intégration* : ce phénomène est à relier aux propriétés de « persistance » des neurones : la réponse d'un neurone à un stimulus est prolongée d'un délai plus ou moins fixe après la disparition de ce stimulus. Le phénomène de persistance neuronale a été observée à différents niveaux du système visuel, des cellules ganglionnaires de la rétine (Levick & Zacks, 1970) aux plus hautes aires corticales (Keysers, Xiao, Foldiak & Perrett, 2001; Rolls & Tovee, 1994; Rolls, Tovee & Panzeri, 1999) et dure en général de 60 à 100 ms après la disparition de la cible. Or, Levicks et Sacks ont montré que la présentation de deux flashes de lumière pendant 2 ms produisait une seule réponse fusionnée au niveau des cellules rétinienne ON chez le chat anesthésié pour des ISI de 30 à 50 ms (Levick & Zacks, 1970). La réponse est identique à celle produite par un seul stimulus dont la durée correspond à celle de l'ISI, suggérant que les deux réponses ont été intégrées et ne sont pas distinguables de la présentation d'un seul stimulus.
- *Masquage par interruption* : la présentation d'un nouveau stimulus interrompt les traitements liés au stimulus précédent. Ce type de masquage pourrait reposer sur des mécanismes d'inhibition entre les groupes de neurones concernés par la cible et le masque. Les processus liés au premier stimulus étant incomplets, la perception est affectée.

Pour Turvey, les mécanismes de perturbation dépendent largement du rapport d'énergie entre cible et masque. Ainsi, lorsque le rapport d'énergie est en faveur du masque, la fonction de masquage obtenue est de Type-I, et s'explique en grande partie par les mécanismes d'intégration (Figure 1.10B). Lorsqu'au contraire l'énergie de la cible est supérieure à celle du masque, les mécanismes d'intégration sont moins efficaces pour des SOA nuls ou faibles et la baisse de

performance pour des SOA compris entre 30 et 120 ms s'explique d'avantage par les processus d'interruption. La manipulation du rapport d'énergie entre cible et masque devrait par conséquent permettre de dissocier les mécanismes d'intégration et d'interruption, mais on comprend vite que le problème reste complexe car en augmentant le contraste du masque par exemple, on influe à la fois sur la capacité d'intégrer les informations de la cible et de les interrompre. D'autre part, la répartition temporelle de ces mécanismes de masquage fait l'objet d'un large débat, et si les processus d'intégration n'interviennent probablement plus après 60 ms de ISI, il reste difficile d'interpréter la réduction de performance au-dessus de cette valeur, semblant inclure à la fois intégration et interruption.

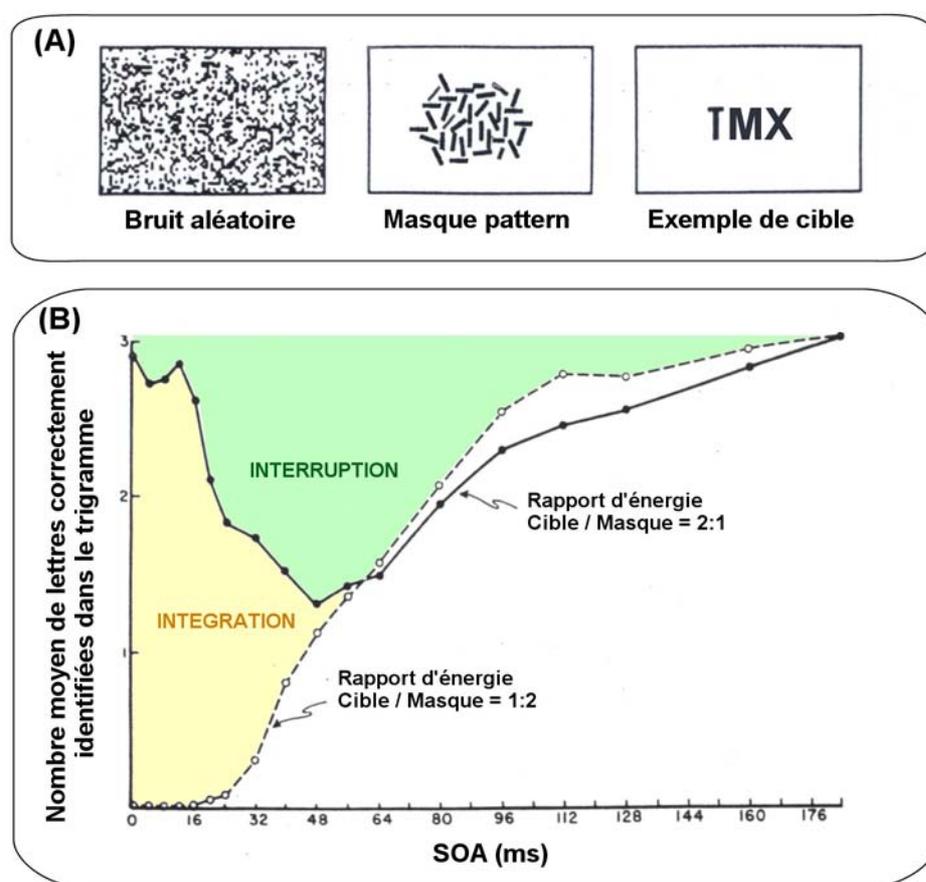


Figure 1.10 : Mécanismes d'interférence du masquage selon Turvey, en fonction du SOA et du rapport d'énergie entre la cible et le masque. (A) Exemple de stimuli utilisés lors de cette expérience de masquage par pattern. (B) Précision moyenne des sujets en fonction du SOA et du rapport d'énergie Cible-Masque. Si le rapport d'énergie est en faveur du masque, la courbe de masquage est de Type I, faisant principalement intervenir des processus d'intégration pour les courts SOA. Dans le cas contraire, elle est de Type II, et la baisse de performance tardive dans l'échelle du SOA s'expliquerait d'avantage par des mécanismes d'interruption. Adapté de Turvey, 1973.

Il est intéressant d'observer que la similarité entre cible et masque est aussi un facteur déterminant pour l'impact du masquage rétrograde (Hellige, Walsh, Lawrence & Prasse, 1979), et son effet varie avec le SOA. Hellige utilise des lettres curvilignes comme cibles, et des lettres soit curvilignes soit rectilignes comme masques. Pour des SOA courts (au dessous de 32 ms), un masque curviligne, qui partage d'avantage de propriétés bas-niveau avec la cible, a moins d'effet qu'un masque rectiligne, quel que soit le rapport d'énergie entre cible et masque. Les mécanismes d'intégration sont visiblement plus perturbateurs si le masque diffère de la cible de par ses propriétés spatiales. A l'inverse pour des SOA longs (au dessus de 32 ms), un masque curviligne a un effet perturbateur plus grand qu'un masque composé de lettres rectilignes.

D'autres protocoles de masquage par pattern ont cherché à manipuler les propriétés bas-niveau du masque pour comprendre les mécanismes sous-jacents du masquage. Dans une tâche de détection, Mitov et collaborateurs ont manipulé le contenu en fréquences spatiales de réseaux utilisés à la fois comme cible et masque (Mitov, Vassilev & Manahilov, 1981). Lorsque la cible et le masque ont une fréquence spatiale inférieure à 6 c/deg, la performance est fortement affectée par le masquage et marquée par une baisse ample et transitoire. Avec un masque à 6 c/deg et une fréquence de cible plus haute à 18 c/deg, ces effets sont beaucoup moins marqués. Dans le cas contraire, faible fréquence spatiale de cible et haute fréquence spatiale de masque, plus aucun effet n'est observé sur la performance. Ces résultats suggèrent que le système magnocellulaire, plus sensible aux basses fréquences spatiales est à l'origine de ces baisses de performances transitoires et qu'il agit sur son propre système, et non pas sur le système parvocellulaire sensible à des plus hautes fréquences. Nous reviendrons sur cette hypothèse un peu plus loin.

1.3.2.2 Masquage paracontraste / métacontraste

Le masquage métacontraste a la particularité de respecter une fonction de masquage de Type-II. Les effets sont peu présents pour des SOA courts et optimaux pour des SOA intermédiaires. Contrairement au masquage par pattern présenté précédemment, les contours de la cible et du masque ne se superposent pas, ce qui suppose que les mécanismes d'intégration correspondant à une fusion des informations spatialement cohérentes de la cible et du masque sont en partie limités.

Breitmeyer propose un modèle basé sur trois types d'interactions intervenant à un stade précoce entre les voies magnocellulaire et parvocellulaire. Le système magnocellulaire a la propriété de transmettre les informations plus rapidement que le système parvocellulaire, et le décalage entre les premières réponses enregistrées pour chaque système atteint 20 ms dans V1 (Nowak, Munk, Girard & Bullier, 1995). Pour Breitmeyer, les interactions précoces entre ces deux canaux permettent d'expliquer la majorité des données dans la littérature du masquage métacontraste et peuvent aussi s'appliquer au masquage par pattern et flash lumineux (Breitmeyer, 1984; Breitmeyer & Ganz, 1976). Dans des conditions de masquage simultané ou pour des SOA courts, deux types d'interaction sont mises en jeu entre les activités respectives magnocellulaires et parvocellulaires engendrées par la cible et le masque : *l'intégration intra-*

canal et *l'inhibition intra-canal*. En revanche, lorsqu'un certain délai sépare la cible du masque, les activités magnocellulaires et parvocellulaires entrent en compétition à travers une troisième forme d'interférence, *l'inhibition inter-canal*. Par le terme *intégration*, Breitmeyer se réfère aux mécanismes d'intégration des informations concernant la cible et le masque décrits dans le paragraphe précédent, qui ne peuvent avoir lieu qu'entre informations de même nature. L'*inhibition* se réfère d'avantage à des mécanismes d'interruption ou de compétition (Figure 1.11).

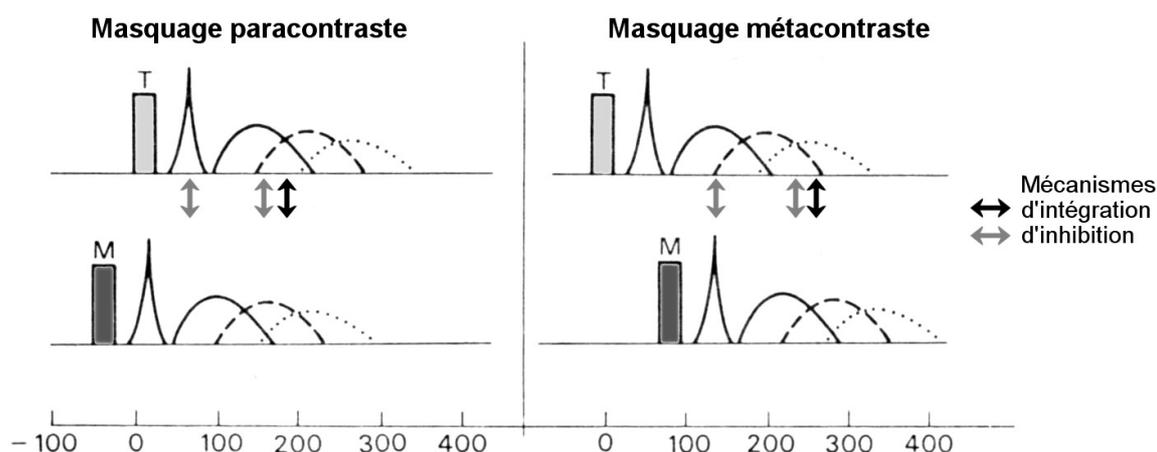


Figure 1.11 : Déroulement temporel des canaux magnocellulaires et parvocellulaires activés par la cible (T) et par le masque (M), selon le modèle de Breitmeyer. Le canal magnocellulaire est représenté par la fonction conique. Le canal parvocellulaire activé pour des gammes croissantes de fréquences spatiales est représenté par les trois fonctions arrondies. Les mécanismes d'intégration ne peuvent avoir lieu qu'entre informations d'un même canal, contrairement aux mécanismes d'inhibition. Adapté de Breitmeyer, 1984.

Suivant le modèle de Breitmeyer, Ogmen fait une dissociation intéressante entre l'impact du masquage sur les traitements déterminant la visibilité de la cible et les traitements nécessaires à sa localisation (Ogmen, Breitmeyer & Melvin, 2003). Dans une tâche évaluant la visibilité de la cible, les sujets devaient donner une estimation de la luminosité de la cible sur une échelle de 1 à 5. Dans une tâche de localisation, les sujets devaient indiquer le plus rapidement possible si la cible apparaissait à gauche ou à droite du méridien vertical, en sachant que le masque était présenté simultanément des deux côtés avec un SOA négatif (paracontraste) ou positif (métacontraste). Étonnamment, dans le cas du masquage métacontraste, c'est la visibilité qui est affectée et non la localisation de la cible. Au cours du masquage paracontraste, à la fois visibilité et localisation sont perturbées. Les auteurs font l'hypothèse que les informations permettant de localiser la cible sont principalement véhiculées par le système magnocellulaire, alors que les informations parvocellulaires jouent un rôle majeur pour sa visibilité. En suivant cette dissociation, les mécanismes d'interférence inter-canal proposés par le modèle de Breitmeyer expliquent l'ensemble des résultats (Figure 1.12), en préservant les informations magnocellulaires

lorsque le masquage est rétrograde, mais en perturbant les deux types d'information lorsque le masque est présenté avant la cible.

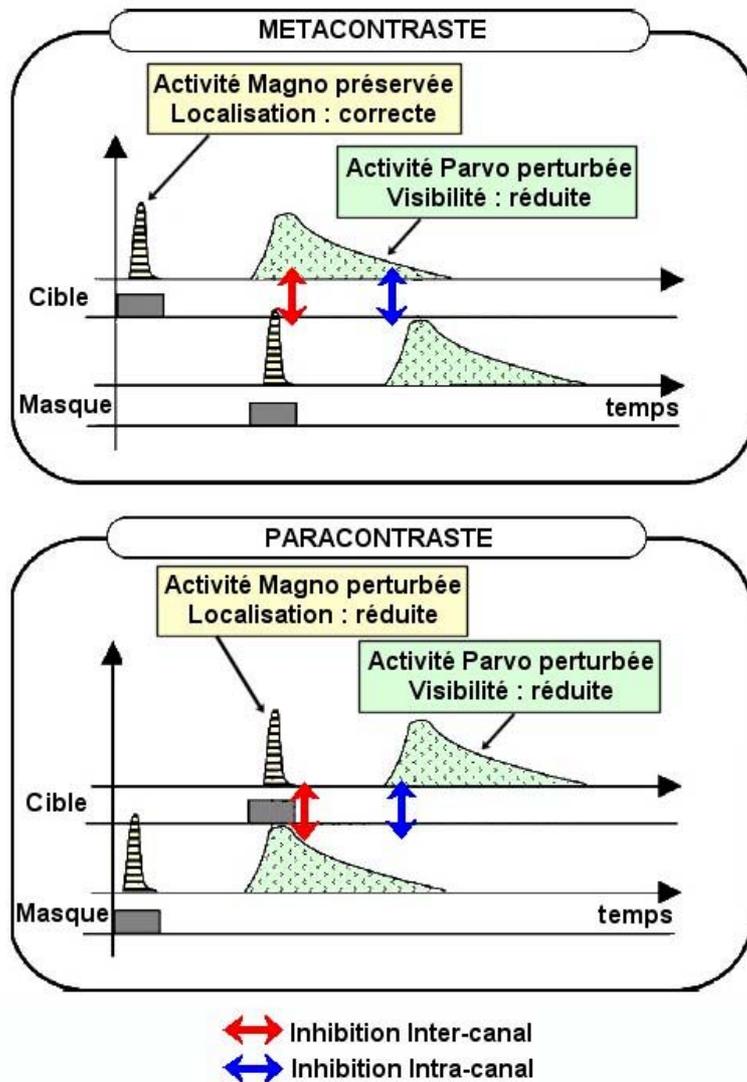


Figure 1.12 : Mécanismes d'interférence du masquage para/métacontraste dans des tâches comparant la visibilité et la localisation d'une cible. Ce modèle suppose que les traitements permettant la localisation d'une cible sont principalement basés sur les informations magnocellulaires, alors que la visibilité repose d'avantage sur les informations parvocellulaires. Adapté de Ogmen, Breitmeyer & Melvin, 2003.

Dans une autre étude de masquage métacontraste, Lachter et Durgin montrent que la capacité à discriminer la localisation d'une séquence Cible-Masque de celle d'une séquence Masque-Masque présentée simultanément, dépend des consignes données aux sujets (Lachter & Durgin, 1999). Lorsque l'on demande aux sujets de répondre le plus rapidement possible, la fonction de masquage obtenue est de Type-I. En revanche, si les sujets n'ont pas reçu de consigne de rapidité, la fonction de masquage est de Type-II. A partir de ces résultats, Ogmen et Breitmeyer suggèrent un accès modulable aux informations magnocellulaires ou parvocellulaires, dépendant des consignes données par les expérimentateurs aux sujets (Ogmen, Breitmeyer & Melvin, 2003). Or la consigne de rapidité nécessiterait que le système visuel donne plus d'importance aux informations magnocellulaires plus rapides, faisant d'avantage intervenir les mécanismes

d'interférence intra-canal définis par le modèle de Breitmeyer. Cette hypothèse d'un accès modulable aux deux canaux selon les contraintes de la tâche est très avancée. Mais il est intéressant de faire le parallèle de ces résultats avec le protocole de catégorisation ultra-rapide utilisé par notre équipe, qui suppose que les informations magnocellulaires de la voie ventrale ont un rôle majeur dans les traitements mis en jeu pour détecter une cible dans une scène naturelle (Delorme, Richard & Fabre Thorpe, 1999; Macé, Thorpe & Fabre-Thorpe, 2005). En suivant ce raisonnement, si l'on présente un masque après la cible (équivalent d'un masquage rétrograde), seuls les mécanismes d'interférence intra-canal résultant de la superposition des informations magnocellulaires de la cible et magnocellulaires du masque sont supposés avoir un impact sur la performance. En revanche, la perturbation des informations parvocellulaires devrait avoir un impact limité. Cette remarque conditionne la gamme de SOA sur laquelle travailler, puisque les interférences intra-canal supposées ne sont compatibles qu'avec des délais cible-masque particulièrement courts.

Enfin, il est important d'essayer de comprendre en quoi le protocole de masquage métacontraste diffère des autres formes de masquage, et notamment du masquage par pattern. En effet, la majorité des études rapporte des fonctions de masquage de Type-I lorsque le masque superpose spatialement la cible (masquage pattern) avec une amplitude maximale décroissante dans les 60 premières millisecondes. Lorsque les contours ne se superposent pas (masquage métacontraste), la fonction est plutôt de Type-II, avec un effet maximal entre 60 et 100 ms. Nous avons vu que cette dissociation n'est pas automatique et dépend de nombreux facteurs, ce qui a conduit certains auteurs à proposer des modèles ne s'appuyant pas sur les différences entre les protocoles mêmes, mais s'appuyant sur ces différents facteurs. Certains se basent sur le rapport d'énergie entre cible et masque (Anbar & Anbar, 1982; Weisstein, 1972), ou sur la distance du masque à la cible (Bridgeman, 1978). Or comme l'a récemment relevé Francis (Francis & Herzog, 2004), il ne semble pas y avoir eu d'études comparant directement l'amplitude et la fonction de masquage en fonction du type de masque, pattern ou métacontraste, pour des rapports d'énergie équivalents. Dans son modèle, Breitmeyer propose que les mécanismes d'intégration intra-canal sont principalement responsables de la baisse de performance pour des SOA courts dans le cas du masquage par pattern, alors que les processus plus tardifs d'inhibition intra et inter-canal sont majoritaires pour le masquage métacontraste. Dans un article récent (Bugmann & Taylor, 2005), Bugmann propose que le recouvrement spatial des informations cible-masque dès la rétine explique les effets précoces du masquage par pattern. En revanche pour le masquage métacontraste, il faut attendre que les informations appartenant à des zones spatiales distinctes soient combinées pour que les mécanismes d'interférence prennent place, ce qui justifierait que l'amplitude du masquage soit maximale pour un SOA important.

Globalement, les différents modèles proposés pour expliquer les mécanismes d'interférence précoces entre les informations de la cible et du masque, ne se contredisent pas directement. Les auteurs sont généralement d'accord pour supposer l'existence de phénomènes d'intégration des informations au plus bas niveau du système visuel, et de mécanismes d'interruption – inhibition – compétition (inter ou intra-canal) intervenant un peu plus tard et un peu plus haut dans la

hiérarchie des traitements. Tout le débat porte sur ces notions de localisation dans l'architecture visuelle. Les données électrophysiologiques présentées ci-après apportent des réponses complémentaires.

1.3.2.3 Modèles de masquage faisant intervenir des mécanismes plus tardifs

Dans le modèle de Breitmeyer, comme dans tous les mécanismes de masquage présentés jusqu'à présent, on suppose que les interactions entre les propriétés physiques de la cible et du masque se font lors de la transmission feed-forward des informations le long des voies visuelles. D'autres modèles ont été développés pour prendre en compte le caractère dynamique du système visuel, en faisant intervenir la connectivité feed-forward et feedback de la rétine au cortex (le Retino-Cortical Dynamics Model : RECOD, (Breitmeyer & Ogmen, 2000; Ogmen, 1993), ou la convergence des informations rétino-geniculo-striée avec des informations non-spécifiques d'origine thalamique (Perceptual Retouch theory : PR, (Bachmann, 1984; Bachmann, 1997).

Il est important de noter que ces modèles font intervenir des mécanismes d'interférence que l'on peut qualifier de tardifs. A la fin des années 70, Michaels et Turvey propose en effet que les interférences entre cible et masque peuvent prendre place à différents niveaux (Michaels & Turvey, 1979). Le niveau « périphérique », c'est à dire rétinien et précoce, inclut les mécanismes d'intégration, alors que le niveau « central » servant de base à la construction d'une représentation iconique du stimulus, inclut à la fois les processus d'intégration, d'inhibition et de compétition attentionnelle. L'étude des mécanismes d'interférence attentionnelle a d'ailleurs motivé Enns et Di Lollo à développer un nouveau protocole expérimental de masquage, dans le but de s'affranchir des processus d'interférence plus précoces (Enns, 2004; Enns & Di Lollo, 2000). Dans ce « modèle de substitution d'objet », le stimulus cible est présentée parmi des stimuli distracteurs similaires, et est suivi d'un masque ne le recouvrant pas. Ce protocole diffère du masquage métaconstraste de par l'incertitude spatiale de la cible qui requiert que l'attention du sujet soit diffuse plutôt que centralisée sur une cible bien définie spatialement. La fonction de masquage obtenue en manipulant la distance entre la cible et le masque a permis aux auteurs d'écarter l'implication de processus précoces basés sur les interactions de contours (Di Lollo, Enns & Rensink, 2000).

1.3.3 Neurophysiologie du masquage

En enregistrant l'activité des cellules à différents niveaux du parcours rétino-cortical des informations visuelles, certains auteurs ont cherché à établir les corrélats neurophysiologiques de la baisse de performance due aux effets de masquage.

La première étape du système visuel à laquelle des données ont été reportées se situe au niveau des cellules ganglionnaires de la rétine. En dessous de 48 ms de ISI, les réponses codant pour la présentation de deux flashes lumineux présentés très brièvement (2ms) commencent à se recouvrir, reflétant les mécanismes d'intégration décrits ci-dessus (Levick & Zacks, 1970). Or si la réponse du deuxième stimulus est légèrement diminuée même pour des ISI plus grands, le

deuxième flash n'interrompt jamais la réponse du premier flash, ce qui laisse supposer qu'au niveau ganglionnaire, seuls les mécanismes d'intégration et non d'interruption peuvent être observés. Au niveau du CGL, des enregistrements chez le chat ont montré que les cellules centre-OFF montrent aussi un profil d'intégration des réponses à deux stimuli lumineux au dessous de 150 ms de ISI (Schiller, 1968).

A un plus haut niveau dans le système visuel, Rolls et collaborateurs ont recherché l'impact du masquage par pattern sur les enregistrements unitaires de cellules dans IT et STS chez le singe macaque (Rolls & Tovee, 1994; Rolls, Tovee & Panzeri, 1999; Tovee & Rolls, 1995). Les cibles étaient composées d'un visage en niveau de gris et présentées pendant 16 ms. Les masques contenaient soit des visages soit une structure composées de lettres N et O, mais dans les deux cas, les sujets humains avaient une performance d'identification au niveau chance lors d'une présentation simultanée de la cible et du masque, et au dessus du niveau chance pour un SOA de 16 ms, ce qui suppose une fonction de masquage de Type-I. Au niveau neuronal, les enregistrements cellulaires reflètent clairement l'impact du masquage par une diminution progressive de la durée de réponse parallèlement à la diminution du SOA, suggérant cette fois-ci des mécanismes d'interruption, que les auteurs interprètent comme issus d'inhibition latérale entre les cellules corticales adjacentes. Cependant, comme le remarque Keysers (Keysers, 2000), la durée de leurs réponses neuronales diminue avec le SOA, mais jamais en dessous de 60 ms.

De la même manière, Kovacs et collaborateurs ont réalisé une série d'expériences similaires, en utilisant des stimuli plus simples composés de lignes, ce qui permettait un recouvrement spatial parfait du masque (Kovacs, Vogels & Orban, 1995). Au niveau comportemental, les performances de détection d'une cible donnée étaient comparables entre les sujets humains et singes. De la même manière, l'enregistrement de cellules dans IT indique que lorsque le SOA diminue, c'est d'abord la durée de réponse qui est affectée, avec une amplitude constante, mais en dessous de 40 ms de SOA, la durée de réponse reste inchangée autour de 60 ms alors que l'amplitude décroît. Même si les auteurs interprètent ces données en faveur de mécanismes d'interruption, l'utilisation d'un masque partageant exactement les mêmes propriétés spatiales que la cible ne permet pas d'exclure totalement l'intervention de mécanismes d'intégration (Keysers, 2000).

Dans une expérience de RSVP dans laquelle des sujets singes et humains devaient détecter un visage cible dans une séquence de visages, Keysers et collaborateurs ont enregistré l'activité de neurones du STS en manipulant à la fois la fréquence, la durée et l'intervalle de présentation des stimuli dans la séquence (Keysers, Xiao, Foldiak & Perrett, 2001). Il apparaît encore une fois que l'activité neuronale spécifique persiste environ 60 ms après la disparition du stimulus, et pour des SOA courts, c'est l'amplitude de décharge qui est affectée par le masquage pendant cet intervalle. Plutôt que de dissocier phénomène d'intégration et d'interruption, Keysers et Perret préfèrent parler de mécanismes de *compétition*, pour lesquels les forces contradictoires générées par deux stimuli déterminent la réponse neuronale dans les aires visuelles avancées (Keysers & Perrett, 2002). En s'appuyant sur ces travaux de masquage et de rivalité binoculaire, ils proposent que la compétition dure environ 60 ms, pendant lesquelles la représentation neuronale du nouveau stimulus devient progressivement plus saillante alors que celle du stimulus précédent décroît

(Figure 1.13). Cette hypothèse a l'avantage de prendre en compte l'influence des rapports d'énergie entre cible et masque sur les performances comme l'avait suggéré d'autres travaux purement comportementaux. Elle est d'ailleurs tout à fait compatible avec d'autres données électrophysiologiques chez le singe montrant que l'augmentation du contraste sur un stimulus comparativement à un deuxième stimulus augmente la réponse des neurones de V4 codant pour le premier parallèlement à une diminution de la réponse neuronale sur le deuxième pendant une centaine de millisecondes (Reynolds & Desimone, 1999; Reynolds & Desimone, 2003). D'autre part à travers cette hypothèse, à la fois la cible et le masque exercent un effet de masquage l'un sur l'autre, permettant de concilier les données sur le masquage rétrograde et antérograde. Cependant, cette hypothèse de compétition, même si elle permet de généraliser un ensemble de données comportementales et neurophysiologiques sur le masquage, ne spécifie pas explicitement le niveau de l'architecture visuelle où les interférences prennent place, ni par quel mécanisme puisque selon les auteurs, la compétition peut prendre les trois formes : latérale, feed-forward ou feedback.

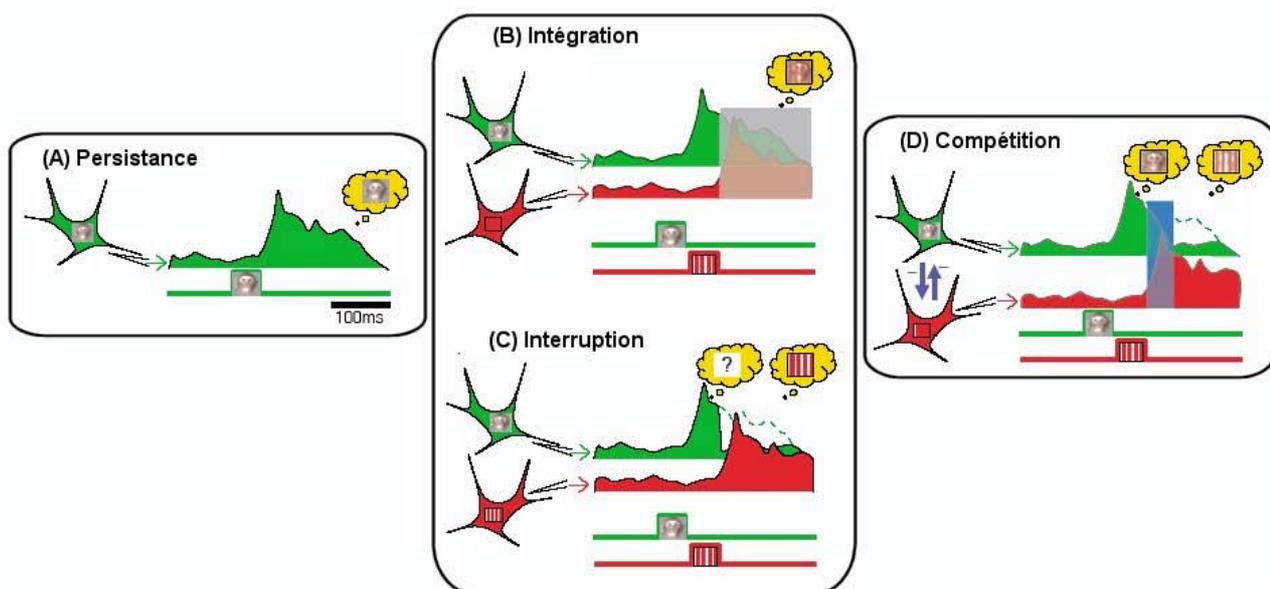


Figure 1.13 : *Interprétation neurophysiologique du masquage*. Deux cellules hypothétiques sont enregistrées simultanément : une cellule sélective aux visages et l'autre à un réseau vertical. (A) La persistance neuronale prédit que la présentation brève d'un stimulus de visage induit une réponse de la cellule (verte) plus longue que la seule présentation du stimulus, d'environ 60 ms. (B) Les mécanismes d'intégration supposent que la présentation d'un visage suivi d'un réseau induit une réponse combinée des deux stimuli, conduisant à la perception d'un seul stimulus contenant les représentations à la fois du visage et du réseau. (C) Les mécanismes d'interruption en revanche impliquent que la réponse neuronale sélective au visage soit brutalement interrompue par la présentation du réseau, affectant plus ou moins la perception du stimulus visage. (D) Enfin, les mécanismes de compétition suggèrent que les réponses neuronales des stimuli entrent en contradiction. La réponse au stimulus visage persiste après sa disparition. Mais quand les réponses des deux stimuli sont actives en même temps, si la cellule du réseau a un profil phasique supérieure à l'activité de la cellule visage, alors elle gagne la compétition. Au bout de 60 ms, seule la cellule réseau reste active. Adapté de Keysers & Perrett, 2002.

Il faut cependant souligner que les travaux de Perret et collaborateurs, comme ceux de Rolls auparavant, sont particulièrement intéressants dans le sens où les stimuli utilisés -photographies de visages ou de têtes de singe- sont complexes et ont une valeur biologique bien plus importante que les stimuli composés de barres ou de formes simples.

Les travaux chez l'homme ont une autre approche de la localisation des effets de masquage. Dans une étude en IRM fonctionnelle (Imagerie par Résonance Magnétique), Grill-Spector s'est intéressée à l'impact du masquage rétrograde lorsque l'on varie la durée de présentation de la cible, en l'occurrence une scène naturelle (Grill-Spector, Kushnir, Hendler & Malach, 2000). Les sujets devaient identifier verbalement l'objet contenu dans la scène suivie par un masque composé de fragments d'images naturelles (« scrambled pictures », on découpe l'image en petits carrés que l'on mélange, comme dans un taquin). Il apparaît que l'amplitude du signal est directement corrélée à la performance des sujets, moins importante pour des présentations d'images courtes, et plus ample pour des durées plus longues ou après des séances d'entraînement. Les auteurs observent que la corrélation entre signal et performance est plus marquée au niveau des aires occipitales, et particulièrement au niveau du complexe occipital latéral (LOC), impliqué dans la reconnaissance d'objets complexes (Grill-Spector et al., 1998a; Grill-Spector et al., 1998b). En contrepartie, l'activité dans V1 est peu affectée par la diminution de présentation de l'image, puisque avec 20 ms de durée, l'activité de V1 est toujours à 70 % de son activation maximale pour une présentation non masquée pendant 500 ms. D'autres auteurs ont cherché à corréler l'amplitude du signal en imagerie avec la performance dans des tâches de masquage, afin notamment d'isoler les aires qui joueraient un rôle important dans la construction d'une représentation consciente des objets (Bar et al., 2001; Dehaene et al., 2001; Vanni, Revonsuo, Saarinen & Hari, 1996).

Dans une étude récente, Green et collaborateurs ont cherché non pas à moduler la force du signal cible en augmentant sa durée de présentation comme dans l'expérience de Grill-Spector, mais à moduler l'efficacité du masque en gardant des durées de présentation de cible et masque constantes et en variant le SOA entre 34 et 102 ms (Green et al., 2005). La cible était constituée d'un cercle ouvert sur un côté (soit en haut, en bas, à gauche ou à droite), et le masque était composé de 12 cercles complets recouvrant totalement la cible. Les sujets devaient localiser l'ouverture du cercle. Les auteurs ont ainsi isolé un ensemble de régions sensibles à la diminution du SOA, comprenant le lobe latéral occipital, les régions inféro-temporales, thalamiques et cingulaires antérieures. En revanche, encore une fois, les aires plus précoces ne semblent pas sensibles à la diminution du SOA. Cependant il faut souligner que ces différentes études travaillent avec une gamme de SOA relativement lente (à partir de 20 ms dans l'expérience de Grill Spector et 34 ms dans l'expérience de Green) et qui n'est pas optimisée puisque dès le premier pas de SOA, les sujets sont au dessus du niveau chance, ce qui suppose qu'un certain nombre de traitements ont déjà permis d'intégrer suffisamment d'informations pour répondre à la tâche. D'autre part, ces études n'excluent pas la possibilité que les interférences du masquage aient lieu au niveau des aires précoces, mais elles ne sont probablement pas visibles car les populations neuronales continuent de décharger, que ce soit pour les informations des cibles ou

pour celles du masque. Les corrélations observées dans les aires plus avancées reflèteraient d'avantage le résultat de ces interférences, responsables d'une reconnaissance dégradée des cibles.

Assez peu d'études en revanche se sont intéressées à corréler les performances dans des tâches de masquage avec les paramètres du signal enregistré en EEG (Rieger, Braun, Bulthoff & Gegenfurtner, 2005). Une des premières études à ce sujet a été menée par Donchin et collaborateurs dans une expérience de masquage par flash lumineux (Donchin, Wicke & Lindsley, 1963). Pour des SOA courts, les sujets étaient incapables de dissocier les deux flashes, et l'activité mesurée au niveau des aires occipitales ne différaient pas du signal obtenu en ne présentant qu'un seul flash. D'autres travaux ont été réalisés ensuite avec différents protocoles de masquage, mais la corrélation entre performance et signal y apparaît moins clairement (Kaitz, Monitz & Neshet, 1985; Schwartz & Pritchard, 1981; Vaughan & Silverstein, 1968). Une étude récente a cependant retrouvé des indices de corrélations particulièrement intéressants sur leur aspect temporel en analysant le signal enregistré en magnétoencéphalographie (MEG) lors d'une tâche de discrimination de scènes naturelles (Rieger, Braun, Bulthoff & Gegenfurtner, 2005). Les sujets devaient effectuer une tâche de discrimination de scènes naturelles suivies d'un masque « scrambled », alors que la durée de présentation de l'image était variable. Les auteurs ont analysé l'activité occipitale additionnelle enregistrée sur les potentiels évoqués dans la condition Cible + Masque en comparaison avec la condition Cible seule, et concluent que la performance est affectée seulement lorsque cette activité additionnelle recouvre les 170 ms d'activité initiale liée aux traitements de la cible, qu'ils supposent relativement bas-niveau. Si leurs conclusions indiquent qu'au moins 60 à 90 ms de SOA sont nécessaires pour obtenir de bonnes performances dans cette tâche, ce qui est cohérent avec les données électrophysiologiques obtenues chez le singe par exemple (Keysers, Xiao, Foldiak & Perrett, 2001; Kovacs, Vogels & Orban, 1995; Rolls, Tovee & Panzeri, 1999), ces travaux pose un problème théorique majeur dans l'analyse de l'activité cérébrale puisqu'à travers la soustraction du signal (Cible + Masque) – (Cible), ils reposent sur l'hypothèse très incertaine d'une additivité linéaire des signaux enregistrés à la surface du scalp. Cependant, ces résultats confirment l'existence d'une corrélation entre activité mesurée en EEG et performance dans une tâche de reconnaissance de scènes naturelles masquées, comme nous le proposons aussi dans l'article présenté dans le chapitre suivant.

En conclusion, certains travaux supposent que des interférences existent au niveau des aires primaires et qu'elles se font majoritairement sous forme d'intégration des informations de la cible et du masque, résultant par une réponse fusionnée des deux réponses initiales. Au niveau des aires plus avancées, les mécanismes d'interférence sont moins évidents (intégration – interruption – compétition) mais il paraît clair que les 60 premières millisecondes de décharge neuronale correspondent à une période critique pendant laquelle les performances comportementales sont significativement affectées si la réponse neuronale est altérée.

1.4 Étudier la dynamique temporelle des traitements de catégorisation : l'intérêt d'utiliser des stimuli masqués

1.4.1 Présentation du contexte expérimental

La tâche de catégorisation animal/non-animal a montré que les sujets humains étaient particulièrement efficaces pour détecter la présence d'une cible dans une scène naturelle. Les temps de réaction minimaux apparaissent dès 270-280 ms (180 ms chez le singe), et l'activité cérébrale mesurée en potentiels évoqués séparément sur les essais cibles et distracteurs, diffère à partir de 150 ms. Or ces valeurs posent des limites « supérieures » aux latences nécessaires pour réaliser la tâche. Les temps de réaction incluent à la fois les délais dus aux traitements sensoriels, décisionnels et moteurs, alors que la mesure de l'activité différentielle semble refléter des processus cognitifs déjà avancés, même s'il est encore difficile de les définir précisément (Rousselet, Thorpe & Fabre-Thorpe, 2004). Entre la stimulation et la réponse manuelle, les informations visuelles doivent traverser une dizaine d'aires de la rétine aux aires motrices, en passant par le cortex inféro-temporal impliqué dans la reconnaissance d'objet. Sachant que les vitesses de conduction des connections cortico-corticales sont relativement lentes, un minimum de 10 ms seraient nécessaires pour que l'information soit disponible d'une étape à l'autre chez le singe (Nowak & Bullier, 1997). Vu que les distances inter-étapes sont un peu plus courtes chez le singe, 10 à 15 ms serait une valeur plus appropriée chez l'homme. La figure 1.14 donne des indications temporelles approximatives chez le singe tout au long des étapes de traitements engagées lors de la catégorisation rapide d'images, en se basant sur les différentes données de la littérature répertoriées dans la partie précédente. Cependant, ce parcours des informations reste très hypothétique, et nous ne savons pas dans quelle mesure l'information doit être intégrée dans les aires extrastriées pour pouvoir répondre correctement à la tâche. D'autre part, l'analyse des temps de réaction dans la tâche de catégorisation ou de l'activité différentielle en potentiels évoqués ne permet pas de rendre compte de la dynamique temporelle de ces traitements sensoriels, décisionnels et moteurs. L'utilisation d'un protocole de masquage paraissait donc complètement adapté pour observer le décours temporel des traitements, le but étant de bloquer l'intégration des informations visuelles à différents niveaux de leur parcours cortical et d'en mesurer la répercussion sur les performances.

Évidemment, les interférences produites par le masquage ne sont pas toujours simples à interpréter, comme l'a montré la revue de la littérature sur les principaux travaux de masquage. Un des enjeux de cette méthode consiste à déterminer le profil spatio-temporel de ces interférences (à quel niveau de la hiérarchie corticale sont-elles maximales, et par conséquent dans quelle gamme de latences ?). Un autre enjeu porte sur les mécanismes d'interférence et son impact sur les réponses neuronales (intégration - interruption – compétition ?). Le chapitre suivant décrit notre première expérience de masquage développée dans la tâche de catégorisation de scènes naturelles, dont les résultats ont été publiés dans *Vision Research* (Bacon-Macé, Mace, Fabre-Thorpe & Thorpe, 2005).

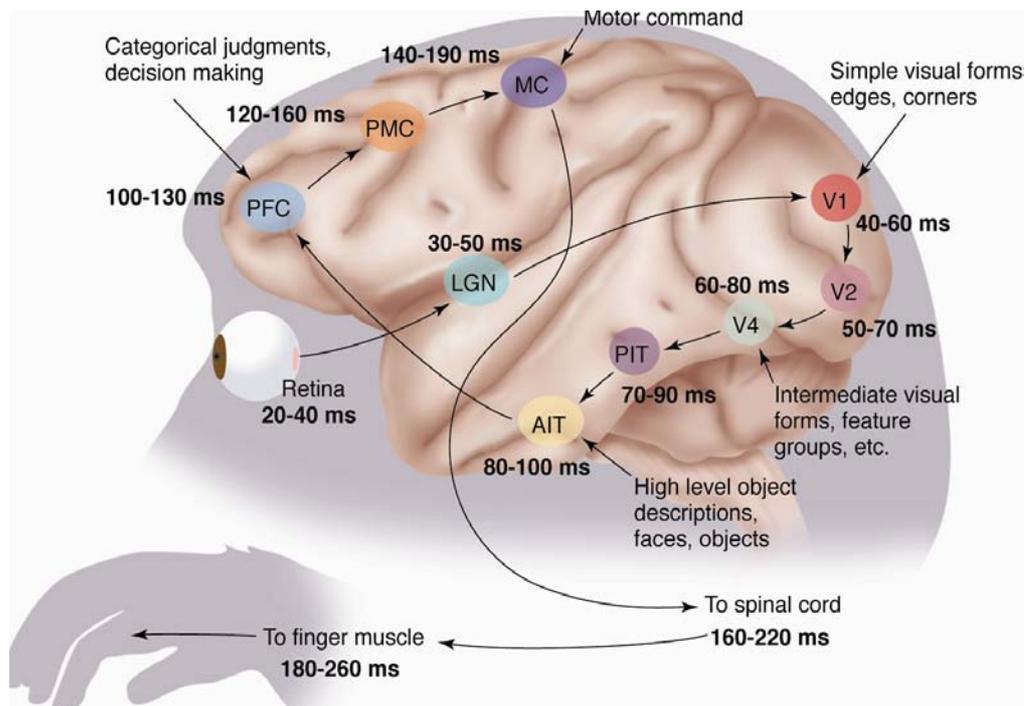


Figure 1.14 : Indications approximatives des latences pour chaque étape de traitement lors d'une tâche de catégorisation chez le singe. L'information part de la rétine et doit traverser le thalamus, puis les aires V1, V2, V4 avant d'arriver au cortex inféro-temporal supérieur (PIT) et inférieur (AIT). D'autres traitements prennent le relais au niveau du cortex préfrontal (PFC), prémoteur (PMC), et moteur (MC) pour finalement activer une réponse de la main. D'après les résultats de la tâche, la totalité de ces traitements peut être effectuée en 180 ms chez le macaque. Il est important de noter que cette représentation anatomique des connexions a été conçue pour rendre compte de la vitesse des traitements qui permettent de générer des réponses rapides et correctes. Il faut bien sûr considérer que le réseau de connexions est bien plus complexe et dynamique : les connexions latérales et feedback jouent probablement un rôle important dans la préparation du système à recevoir les informations véhiculées en feedforward. Adapté de Thorpe & Fabre-Thorpe, 2001.

1.4.2 Mise au point du protocole de masquage : quelques règles

Il est important de noter que la détermination psychophysique d'un seuil temporel au-dessous duquel les traitements visuels sont perturbés et au-dessus duquel les effets de masquage sont réduits ne permet pas de rendre compte précisément des latences dans le système visuel. Il doit être presque toujours possible de trouver un masque ou un protocole encore plus efficace pour perturber les traitements visuels relatifs à une catégorie de cibles. En plus du délai cible-masque, de nombreux paramètres influent sur le degré de masquage, que ce soit le type de protocole, le rapport d'énergie entre cible et masque, la similarité entre cible et masque etc ... Alors pourquoi et comment utiliser la technique du masquage lorsque l'on s'intéresse à la dynamique des traitements dans le système visuel ? Implicitement, nous nous sommes basés sur un ensemble de règles et hypothèses dans l'ensemble des travaux qui vont suivre :

- Nous nous sommes toujours attachés à travailler sur une gamme de SOA optimisée pour observer une amplitude maximale d'effet perturbateur. Dans tous les travaux qui suivent, nous avons rapporté des performances allant du niveau chance à un niveau comparable aux performances sans effet de masquage (articles 1, 2 et 3).

- Les données de la littérature sur les travaux de masquage permettent de cibler certains paramètres du protocole. Le rapport d'énergie stimulus / masque est en faveur du masque, contrasté à 100%, afin d'optimiser les mécanismes d'interférence dans des gammes de latence courtes (Turvey, 1973). D'autre part, le masque est composé de formes aléatoires à différentes fréquences spatiales, afin d'interférer avec un maximum de caractéristiques bas-niveau contenues dans les images naturelles (Hellige, Walsh, Lawrence & Prasse, 1979; Torralba & Oliva, 2003).
- Si le protocole de masquage ne permet pas d'extraire des valeurs absolues sur la dynamique des traitements dans le système visuel, l'observation de la fonction de masquage doit permettre de poser des hypothèses sur les mécanismes d'intégration de l'information. Nous l'avons déjà vu à travers la définition des fonctions de masquage de Type-I et de Type-II (cf 1.2.1), et d'autres problématiques sont soulevées dans l'article 1, relatives aux modèles de transmission de l'information discrets ou continus.
- Deux types de seuils temporels peuvent néanmoins être déterminés : le point à partir duquel la précision commence à être significativement au-dessus du niveau chance, et le point à partir duquel la précision atteint son plateau maximal. Si l'obtention d'une condition au niveau chance pour un masque donné ne permet pas d'exclure l'existence d'un masque plus efficace qui décalerait cette valeur vers une latence plus longue, nous supposons que nous avons déjà une bonne estimation du délai minimal nécessaire pour extraire une quantité d'information suffisante pour les traitements. Comme confirmé dans les articles 2 et surtout 3, la manipulation du masque ou du protocole intervient d'avantage sur la pente de la fonction de masquage, et donc sur le délai nécessaire pour atteindre le plateau de précision maximale.
- Dans le chapitre 2, nous nous sommes effectivement intéressés aux fonctions de masquage. Dans les chapitres 3 et 4, nous avons utilisé le protocole de masquage comme un outil permettant de dégrader la perception de cibles, alors que d'autres paramètres étaient manipulés : le type de décision et la réponse motrice (Chapitre 3), le contenu des cibles et masques (Chapitre 4A), le degré d'analyse des cibles (Chapitre 4B).

Masquage rétrograde dans une tâche de catégorisation visuelle ultra-rapide

2

2.1 Objectifs

Les études réalisées jusqu'à présent dans l'équipe donnent des informations limitées sur la dynamique temporelle des traitements sensoriels engagés dans la tâche de catégorisation de scènes naturelles, ce qui a conduit à développer un protocole de masquage afin d'interrompre les processus à différentes étapes du parcours cortical des informations visuelles.

Cette étude avait donc plusieurs objectifs :

- Étudier la dynamique temporelle des traitements visuels consistant à extraire et intégrer suffisamment d'informations sur la scène naturelle pour pouvoir décider s'il s'agit d'une cible ou d'un distracteur
- Tester l'hypothèse d'une transmission feedforward des informations, développée à la suite de précédents travaux (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001). Le protocole habituel se basait sur la présentation de la scène naturelle pendant une vingtaine de millisecondes. Même si ce délai est court, il ne permet pas d'exclure que le système visuel accède aux propriétés de l'image par un système de « rebouclage » permanent vers les aires bas-niveau, favorisé par les propriétés de persistance rétinienne et neuronale. La présentation d'un masque après la cible doit permettre de remplacer en grande partie les informations de la scène naturelle et limiter l'utilisation des connexions feedback.
- Tester l'hypothèse d'une transmission parallèle et continue des informations, par opposition au modèle sériel qui impliquerait qu'à chaque étape les traitements sont terminés avant d'envoyer les informations aux aires suivantes.
- Évaluer l'impact du masquage sur l'activité cérébrale mesurée en potentiels évoqués, et notamment sur la composante d'activité différentielle calculée à de nombreuses reprises dans les précédents travaux de l'équipe.

2.2 Résultats

En variant le délai entre le début de la présentation de la scène et du masque entre 6 et 106 ms (Stimulus Onset Asynchrony – SOA), nous avons pu obtenir une fonction de masquage rapportant l'augmentation de précision comportementale avec un intervalle de temps de plus en plus grand. Plusieurs remarques peuvent être faites au sujet de ces latences : la précision est au niveau chance à 6 ms de SOA, au dessus du niveau chance dès 12 ms, atteint un plateau à 44 ms et n'augmente plus après 81 ms. La tâche étant de type go/no-go (les sujets doivent relever le doigt du bouton réponse uniquement s'ils voient une cible), il apparaît que les sujets produisent de moins en moins de réponses lorsque les conditions de perception deviennent difficiles. Ainsi les effets de masquage n'influencent pas ou peu le pourcentage d'erreurs sur les distracteurs (False Alarms), mais augmentent le nombre de cibles manquées (Missed), suggérant que le système visuel n'a pas pu extraire suffisamment d'informations sur la scène naturelle pour déclencher une réponse motrice. Les temps de réaction augmentent significativement avec la diminution du SOA, ce qui suppose que le système visuel a besoin de plus de temps pour intégrer les informations dans les conditions difficiles de perception.

L'enregistrement simultané des potentiels évoqués a permis de mettre en évidence une diminution significative de l'amplitude de l'activité différentielle occipitale, calculée comme la soustraction des signaux sur les essais cibles et distracteurs corrects, avec la diminution du SOA. Une très forte corrélation peut être calculée entre cette amplitude et la précision comportementale des sujets, particulièrement au niveau de l'hémisphère droit. En revanche, les latences de cette composante restent inchangées avec la difficulté de la tâche.

ARTICLE 1

The time course of visual processing: Backward masking and natural scene categorisation

Bacon-Macé N, Macé MJ-M, Fabre-Thorpe M & Thorpe SJ
Vision Res 45, 1459-69 (2005).

The time course of visual processing: Backward masking and natural scene categorisation

Nadège Bacon-Macé *, Marc J.-M. Macé, Michèle Fabre-Thorpe, Simon J. Thorpe

*Centre de Recherche Cerveau et Cognition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil,
133, Route de Narbonne, 31062 Toulouse, France*

Received 23 April 2004; received in revised form 23 December 2004

Abstract

Human observers are very good at deciding whether briefly flashed novel images contain an animal and previous work has shown that the underlying visual processing can be performed in under 150 ms. Here we used a masking paradigm to determine how information accumulates over time during such high-level categorisation tasks. As the delay between test image and mask is increased, both behavioural accuracy and differential ERP amplitude rapidly increase to reach asymptotic levels around 40–60 ms. Such results imply that processing at each stage in the visual system is remarkably rapid, with information accumulating almost continuously following the onset of activation.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Natural images; Backward masking; Early processing; Information integration; Event-related potentials (ERP)

1. Introduction

Human subjects are very quick and efficient at analysing briefly viewed natural scenes, an ability that has obvious survival value. We can determine whether a briefly flashed image contains an animal and make a behavioural response in as little as 250 ms, and this ability extends to other categories of visual stimulus such as faces or means of transport (Macé & Fabre-Thorpe, 2003; Rousselet, Macé, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a). Simultaneously recorded event-related potentials (ERP) diverge sharply between correct target and dis-

tractor trials just 150 ms after stimulus onset (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe et al., 1996) which imposed even more severe temporal constraints. Extensive training failed to reduce this 150 ms latency, indicating that even with images never seen before, the system is operating virtually optimally and with a minimal number of processing stages (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001).

This sort of behavioural and electrophysiological evidence imposes an upper limit on the amount of time required for animal detection but provides relatively little direct information about the dynamics of the underlying processing. With only a 150 ms delay between the onset of activation in the retina and a cerebral differentiation between target and distractor pictures, it is a challenge to explain how visual information is processed and transmitted through the visual pathways. A distinction is often made between discrete or continuous models of information transmission (Eriksen & Schultz, 1979; Hasbroucq, Burle, Bonnet, Possamai, & Vidal, 2002;

* Corresponding author. Tel.: +33 5 62 17 37 75; fax: +33 5 62 17 28 09.

E-mail address: nadege.bacon-mace@cerco.ups-tlse.fr (N. Bacon-Macé).

McClelland, 1979), the first implying that there is a fixed minimum processing time at each stage before information can be sent to the next level, while the latter supposes that it can be transmitted continuously as soon as information becomes available. Both are consistent with a pipeline processing scheme in which every step can operate simultaneously and in parallel. Indeed, some form of pipeline processing seems necessary to account for the results of a recent study using RSVP (rapid serial visual presentation) showing that human subjects can detect images in sequences presented at rates of up to 75 images per second (Keysers & Perrett, 2002). Such data imply that less than 15 ms are enough to process a sufficient amount of information concerning each picture of the sequence.

RSVP experiments can be integrated into the broader approach of masking, which involves two or more temporally close stimuli to reduce the associated perception (Breitmeyer, 1984). Masking protocols are very useful to study the timing of information processing in the visual system since they allow processing to be interrupted at different times. Electrophysiological studies on monkeys have shown that the intensity and duration of neuronal responses are more and more affected as the mask gets closer to the stimulus, but that considerable information is available in monkeys from the first 30 ms of the neuronal responses (Rolls, Tovee, & Panzeri, 1999; Tovee & Rolls, 1995). In human subjects, there are many experiments that concern the influence of stimulus/mask interval on behavioural responses (Breitmeyer, 1984; Enns & Di Lollo, 2000), but few of them were used in the context of high-level tasks, such as categorisation. Moreover, few masking experiments have investigated the associated changes in cerebral activity, and most of those have involved fMRI methods. Nevertheless, there are reports of a correlation between the ability to detect or to name objects and the activation in occipital regions (Dehaene & Naccache, 2001; Grill-Spector, Kushnir, Hendler, & Malach, 2000; Vanni, Revonsuo, Saarinen, & Hari, 1996). As image and mask get temporally closer, both performance and cerebral activation decrease. This type of correlation can be particularly useful to understand the signals recorded from the brain during perceptual processing.

We present here the results of a backward masking experiment in a go/no-go categorisation task, in which natural scenes were followed by a very strong dynamic mask after a varying stimulus onset asynchrony (SOA). By interrupting processing after different delays, we could determine how information accumulates over time during the task. One of the novel features of the experiment was the use of a high screen refresh rate (160 Hz) that allowed us to present the test image for a single 6.25 ms frame and to vary the SOA by small 6.25 ms steps, a much higher resolution than is typically used in masking experiments.

2. Experimental procedure

2.1. Task

Sixteen subjects participated in this experiment (8 females, 8 males, with a mean age of 29 ranging from 21 to 50). They all volunteered for the study and gave their written informed consent. The go/no-go categorisation task was based on an experimental procedure introduced by Thorpe et al. (1996). Subjects were seated in a dimly lit room, at 1 m from a screen adjusted to an 800×600 pixel resolution and a 160 Hz refresh rate. Natural scene pictures (600×400 pixels in size) were flashed on the monitor for a single frame, which corresponds to 6.25 ms. Subjects were asked to release a button within 1 s if the picture contained an animal and maintain pressure otherwise.

Each subject was tested on 16 series of 90 trials, each of which contained the same number of target and distractor images. All subjects had previously completed at least 3 training blocks of 90 trials. They were asked to try to release the button on 50% of the trials, whatever the masking condition.

A trial began with the display of a white fixation cross in the middle of the black screen for 600–900 ms at random. Then the picture—target or distractor—was flashed, followed by the mask stimulus. Eight different values were used for the stimulus onset asynchrony (SOA) between the picture and the mask (6.25, 12.50, 18.75, 25.00, 31.25, 43.75, 81.25 and 106.25 ms) and display latencies were verified with a photodiode connected to an oscilloscope. Furthermore, we added a control condition in which only the mask was displayed after the fixation cross, without any picture presentation. These 9 conditions were counter-balanced in each series, with 10 trials per condition presented at random, producing a total of 90 trials per block. Any given subject only saw each picture once.

χ^2 tests were used to evaluate if behavioural accuracy was above chance level for each SOA condition. Masking effects between the conditions were assessed with analysis of variance (ANOVA) and post-hoc analyses were performed by using paired *t*-tests with a Bonferroni correction or Mann–Whitney tests.

2.2. Stimuli

A total of 1280 grey level natural images were used in this experiment. As demonstrated in previous work, ultra-rapid categorisation does not rely on colour cues, as performance is almost unaltered when stimuli are presented in grey level (Delorme, Richard, & Fabre-Thorpe, 2000). Moreover, masking effects were easier to obtain and control without colour information in the natural scenes. Half the images contained animals, and were as varied as possible (fish, insects, mammals,

birds or reptiles). The subjects had no knowledge about the size, position and number of the targets in a single picture. The other half of the images were distractors with a wide range of material including natural landscapes, indoor or outdoor scenes, man-made objects, etc... None of the pictures had been seen previously by the subjects and training pictures were not used in the test series.

2.3. Mask

To construct the mask, a white noise image was filtered at four different spatial scales, and the resulting images were thresholded to generate high contrast binary patterns. For each of the 4 spatial scales, 4 different versions were generated by mirroring and rotating the original image. A pool of 16 images was thus available for masking. The mask used in this experiment was a sequence of 8 images - a so-called “dynamic mask” (Fig. 1). The 8 images were chosen randomly from the pool, with each of the four spatial scales presented once during the first 4 images and again during the last four images. Thus, a pattern at each of the spatial scales appeared twice in the “dynamic” mask (see Fig. 1). All the images in the mask were presented for 2 refresh cycles, so that overall, the masking stimuli were displayed for 16 frames (around 100 ms).

2.4. ERP recordings

EEG data were recorded from a 32-electrode cap. Electrode locations were defined using the standard 10–20 Oxford system with 12 additional electrodes over

occipital sites. Electrical activity was amplified by a NeuroScan Synamps amplifier linked to a PC computer, digitized at 1000 Hz, corresponding to a sample bin of 1 ms, and low-pass filtered at 100 Hz. Each recording epoch began 100 ms before the stimulus display on the screen and continued for 1000 ms after the stimulus onset. A baseline correction was carried out for each epoch using the 100 ms of pre-stimulus activity. Trials with artefacts related to ocular movements were rejected, by using a criterion of $[-80; +80 \mu\text{V}]$ on two frontal electrodes (FP1 and FP2) between -100 and $+400$ ms. Within this time period, another artefact rejection was performed on trials with a strong alpha frequency activity, by using a $[-40; +40 \mu\text{V}]$ criterion on parietal electrodes (Oz and Pz). Signals were then low-pass filtered at 40 Hz before the analysis. We were particularly interested in the occipito-temporal electrodes (standard O1, O2, OZ, IZ and non-standard PO7, PO8, PO9, PO10, O9, O10, P7, P8) and the frontal electrodes (standard FP1, FP2, F3, F4, Fz). Epochs corresponding to correct responses were averaged separately for targets and distractor trials on each masking condition.

Differential activities were determined by subtracting the average signal on correct distractor trials from the signal on correct target trials. Eight different curves were obtained, one for each SOA condition. The differential activity amplitude was calculated by a Matlab program that determined the most negative point between 150 and 250 ms after the onset of stimulus presentation (Rousselet, Thorpe, & Fabre-Thorpe, 2004; Thorpe et al., 1996). It was measured separately for each individual and also using the average signal across all

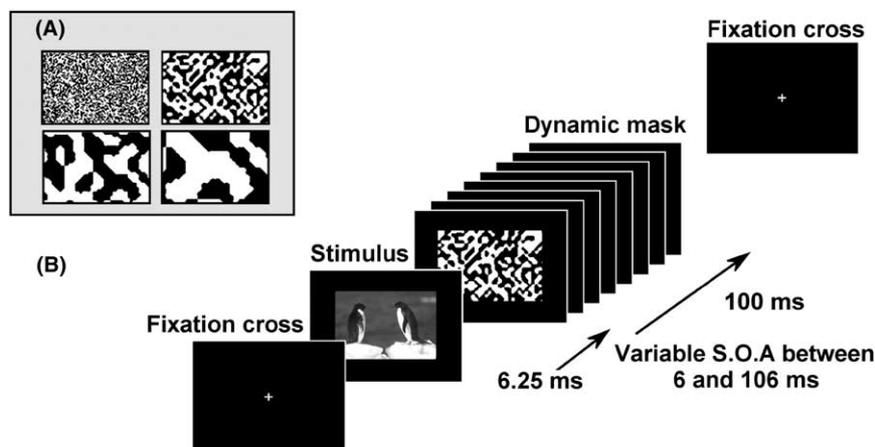


Fig. 1. Behavioural paradigm. (A) Four pictures with different spatial scales that constitute the dynamic mask. Each could be presented at 4 different orientations making a total of 16 different patterns. The images were intermixed to reduce the risk of generating retinal after-effects with the restriction that each spatial scale was used once during the first 4 pictures of the mask and once again during the 4 last ones. (B) In each series, subjects were tested on 90 trials organised as follows: first the fixation point is displayed on the centre of the screen for a random delay to avoid anticipated responses. Then the grayscale picture is flashed for only one frame using a monitor set at 160 Hz. After a variable 6.25–106.25 ms SOA, a dynamic mask is displayed, composed of eight 100% contrasted mask patterns at the four different spatial scales. The subjects then have 1000 ms to release the button if the picture contains an animal. Eight time steps were chosen for the SOA: 6.25, 12.50, 18.75, 25, 31.25, 43.75, 81.25 and 106.25 ms.

subjects. Differences in latencies and amplitude among SOA conditions were statistically evaluated by ANOVAs and post-hoc analyses were performed using *t*-tests with a Bonferroni correction. The correlations between electrophysiological measurements and behavioural data were performed with Pearson tests.

3. Results

3.1. Behavioural performance

3.1.1. Mask efficiency

We evaluated behavioural performance in terms of accuracy and reaction time as a function of the SOA. For each condition, a χ^2 test between correct and incorrect responses determined if accuracy was above chance level, set at 50% because targets and distractors were equally likely. Only the very shortest SOA interval (6 ms) resulted in performance at chance level, with a mean value of 51.9% (Fig. 2A). This result emphasizes the high efficiency of the mask, which effectively prevented visual processing when presented close to the stimulus. However for the next SOA (12 ms) condition, accuracy was already above chance level ($p < 0.01$) and rapidly increased to reach 85.6% with a 44 ms SOA. Accuracy then stabilized at a maximum value of 91.4% for the last condition (106 ms). However, increasing processing time above 44 ms had only a minor effect on performance since accuracy scores in the last three SOA conditions 44–81–106 ms were not significantly different ($p > 0.33$). Note that the maximum accuracy was very close to the accuracy obtained in a previous study (DeLorme et al., 2000), using achromatic natural images flashed for 20 ms in the same go/no-go categorisation task without masking, and where subjects averaged 93% correct. Thus the mask has relatively little effect when it appears 40–60 ms after the image presentation onset, and visual processing remains extremely good de-

spite the fact that the stimulus picture was flashed for only 6 ms.

3.1.2. Response inhibition with increasing difficulty

We noticed a strong reduction in response rate with the most difficult masking conditions. Before the experiment, subjects were asked to try to release the button on about half of the trials in each series. This instruction has been respected since the mean response rate, including correct and incorrect responses, was about 47% when grouped across all conditions. However, the response rate varied considerably with the SOA, as it exceeds 50% from 106 to 25 ms SOA, and drops strongly with SOAs below 25 ms (Fig. 2B).

Interestingly, these variations only affected the proportion of correct go responses. In contrast, the proportion of erroneous go responses to distractors was remarkably stable across all SOA conditions. It would appear that short SOAs did not lead subjects to make more false positives to distractor pictures but did prevent them extracting enough information to make a response on target trials. Another interesting result is given by comparing the response rate obtained with the shortest SOA (22.3%) with the control condition when only the mask was displayed without any picture (19.6%). These two conditions were not significantly different ($p > 0.24$) which suggests that with a 6 ms SOA, subjects behaved as if no image had been presented at all.

3.1.3. Reaction times

Mean reaction time decreased with longer SOAs, particularly for values over 44 ms ($F(7, 127) = 2.591$, $p < 0.02$). In the conditions where the mask was close to the stimulus (SOA 6, 12, 19, 25 and 31 ms), reaction times were significantly longer ($p < 0.01$) than when the mask appeared later (44, 81 and 106 ms). A maximum difference of 54 ms was found between the 12 and 81 ms SOA conditions. This suggests that when the mask interrupts the visual processing earlier, the amount of

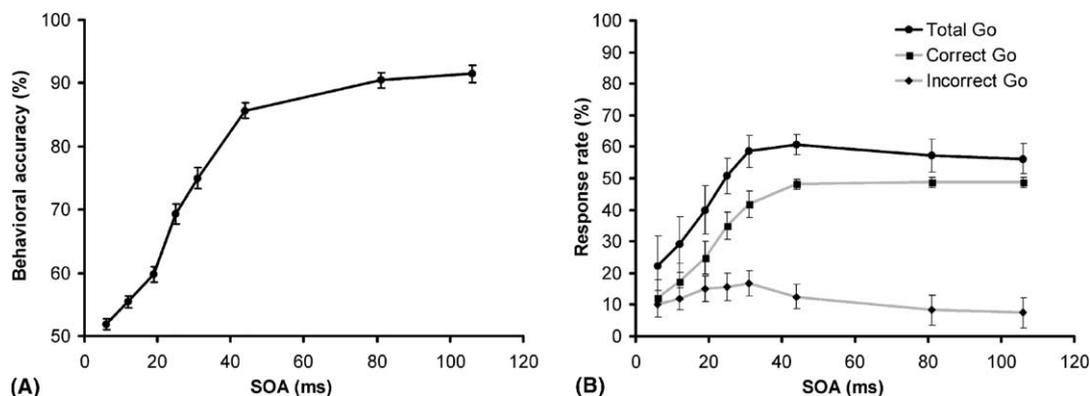


Fig. 2. Behavioural performance as a function of the SOA, averaged above 16 subjects. (A) Behavioural accuracy (\pm s.e.m.). (B) Mean percentage of go responses (\pm SD).

information is reduced and perceptual decisions require more time.

Masking effects can be observed throughout reaction time distributions by comparing the condition of optimal perception (106 ms) with each of the others (Fig. 3A). Above 44 ms, distributions are remarkably similar, but with 31 ms SOA, the median part of the distribution shows a pronounced plateau. Thus, there is a strong effect of the mask on the median reaction time, but the initial part of the distribution, corresponding to early responses, is not affected. With an SOA of 25 ms and less, early responses are also disrupted and the early part of the reaction time distributions no longer superimpose.

3.2. Electrophysiology

Disruptive effects on visual processing can also be observed on the ERP data. For each subject we averaged separately signals on distractor and target trials and subtracted one from the other to calculate a differential

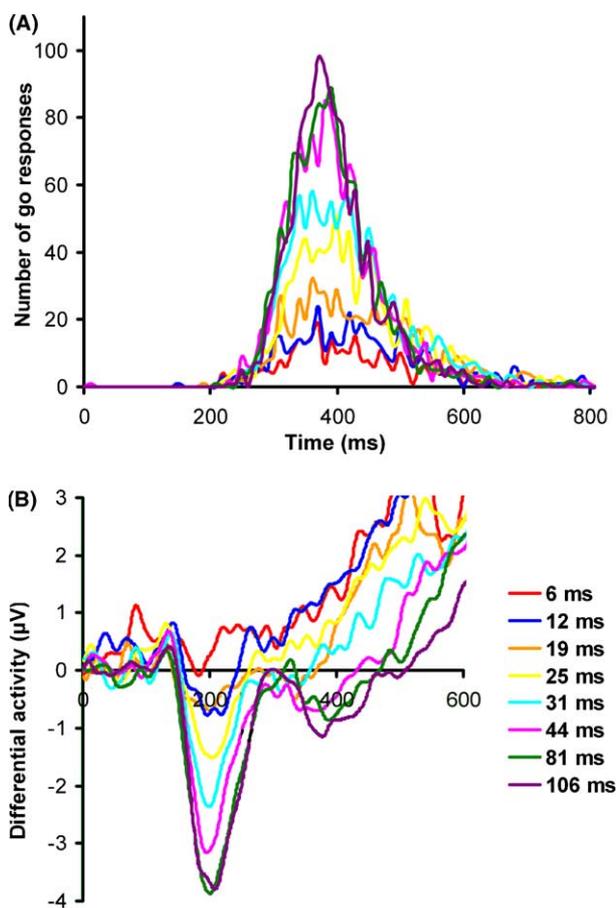


Fig. 3. Masking effects on behavioural reaction time and occipital cerebral activity. (A) Reaction time distribution of correct go responses as a function of the SOA (10 ms bin width), averaged on 16 subjects. (B) Differential activity averaged on 16 subjects for each SOA. The activity is calculated as the difference between signals on correct target and distractor trials, obtained from PO8 electrode.

activity curve. As signals on targets and distractors contain information about the response to both the picture and the mask, subtracting these two signals is a good way to cancel out the activity associated with the physical encoding of the mask. The effects of the mask on image categorisation processing remain clearly observable on the residual signal (Fig. 4). Therefore, we will not present here a detailed analysis of the shape of the underlying ERP signals but rather focus on an analysis of the differential effects.

We analyzed the differential activity with respect to the SOA. Fig. 3B shows averaged signals recorded on a representative occipital electrode (PO8). The onset of the differential activity appears to start at around the same latency (150 ms) but it is clear that the signal amplitude decreases with shorter SOAs ($F(7, 1535) = 77.13$, $p < 0.001$). Moreover, with the exception of the two shortest SOAs (6 and 12 ms) for which the activity was rather weak, peak latencies are remarkably stable between 200 and 215 ms. In other words, when the mask is closer and closer to the picture, the reduction of perceptual differences between target and distractor stimuli strongly affects the amplitude of differential activity.

A lateralization effect can be observed in this task, as the mean amplitude of the differential activity was significantly larger for the electrodes over the right hemisphere compared to the left hemisphere (respectively $2.56 \mu\text{V}$ for the average of electrodes O2, PO10, PO8, O10, P8 versus $2.22 \mu\text{V}$ for the average of O1, PO9, PO7, O9, P7, all SOA conditions grouped; $F(1, 1279) = 15.45$, $p < 0.0001$). This was the case for each of the SOA conditions except for the shortest one at 6 ms.

Although the masking effect is particularly visible on occipital electrodes, similar effects can be seen at most electrode sites (Fig. 5). At frontal sites, shortening the SOA induced a significant diminution of the maximal amplitude of the differential activity ($F(7, 639) = 22.305$, $p < 0.0001$), appearing around 200 ms. But in contrast no lateralization effect could be observed at these sites ($p = 0.145$).

These electrophysiological results can be directly related to behavioural data, which also showed a clear diminution of performance with decreasing SOA. Differential activity amplitude and behavioural accuracy variations are in fact strongly correlated, as showed on Fig. 6. This observation reinforces the idea that the differential activity reflects the result of a perceptual decision and that differential ERP responses provide a powerful investigative tool.

3.3. Control experiment

The choice of the dynamic mask was made after a number of pilot experiments, and appeared to be very

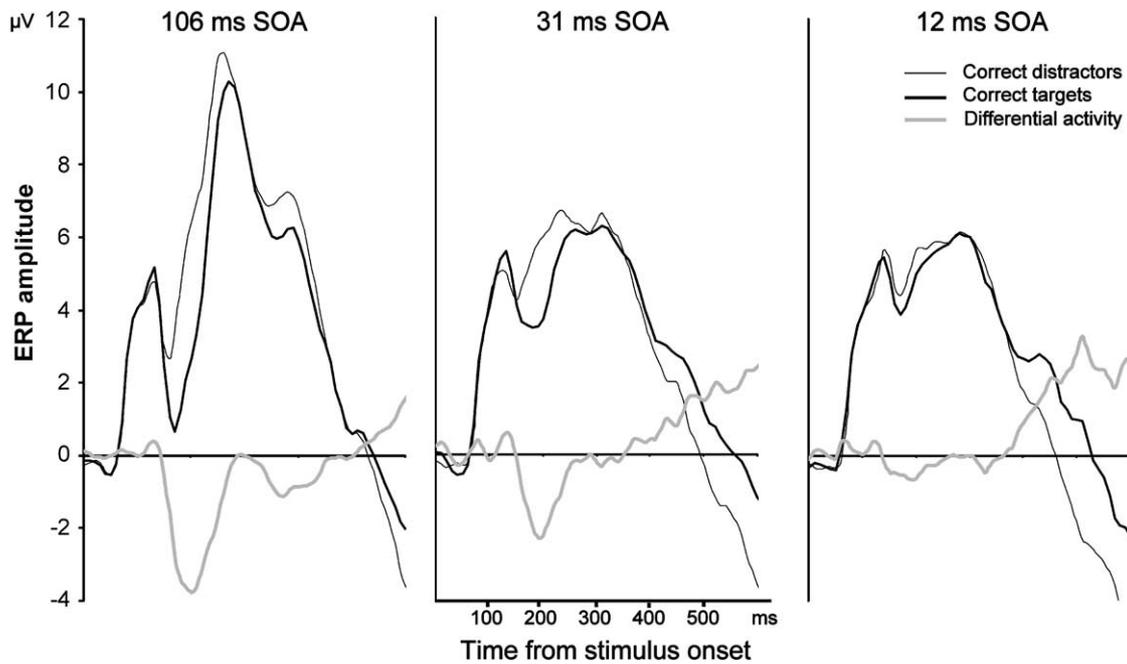


Fig. 4. Grand-average ERP on electrode PO8 for three SOA conditions. Differential activity is obtained by subtracting the signals on correct target and distractor responses.

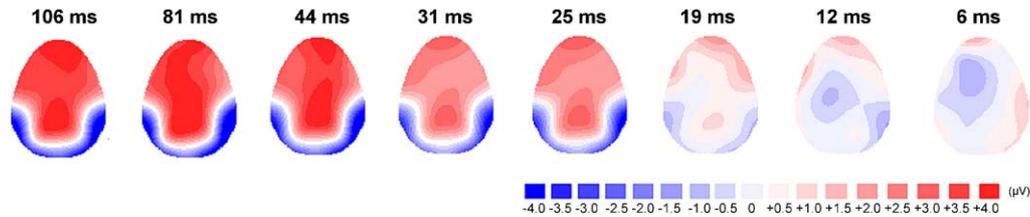


Fig. 5. Differential activity over the scalp at the maximal point of the amplitude, 200 ms after the image onset. Activity was averaged over 16 subjects.

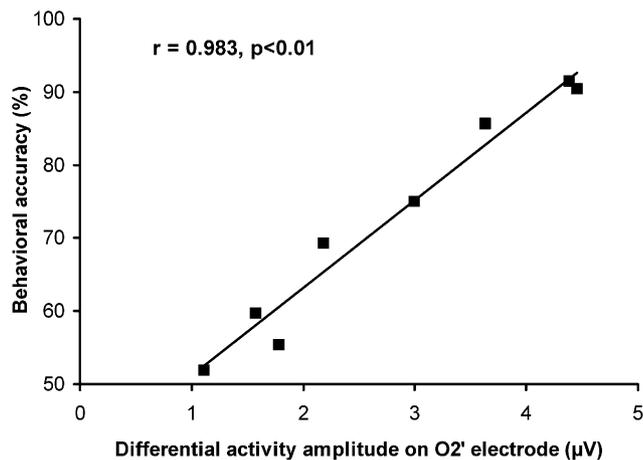


Fig. 6. There is a linear relation between behavioural accuracy and the amplitude of the occipital differential activity, particularly on the right hemisphere. Differential activity from PO8 electrode has been averaged over the 16 subjects, and correlated by a Pearson test ($p < 0.01$) to the mean behavioural accuracy among all SOA conditions.

efficient, as demonstrated by the fact that performance was at chance level with the shortest SOA. However, one possible problem of using a mask with multiple-frames is that one cannot be sure at what exact point the masking becomes totally effective. This could potentially add considerable uncertainty to the SOA considered as the disruptive latency.

We therefore have made a behavioural control experiment where 10 subjects performed the categorisation task at four different SOA values (4 conditions: 6, 12, 44 and 106 ms) and in which we varied the number of pictures in the mask from 1 to 8 (5 conditions: 1, 2, 3, 4 and 8 pictures). Across all the subjects, a total of 640 trials was performed for each of the 20 conditions. Furthermore, we encoded the spatial scale pattern used for each picture of the mask and particularly for the first one. These patterns were randomly chosen for each trial. Fig. 7 shows behavioural accuracy as a function of the first mask pattern. With only one image in the mask

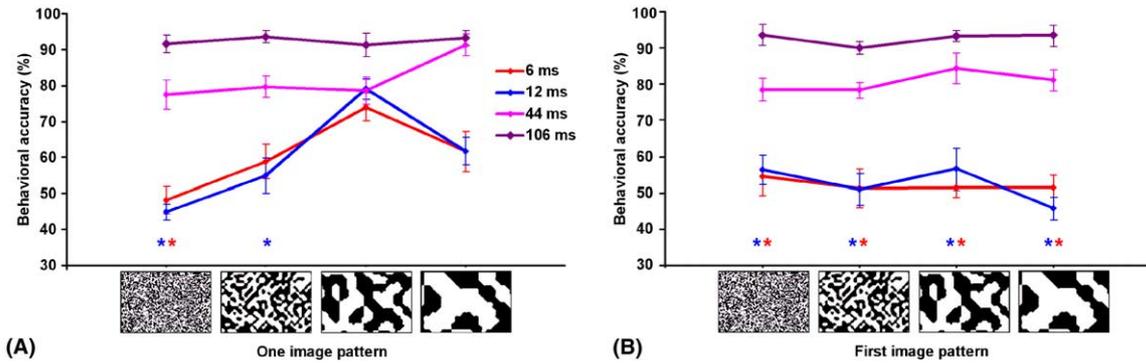


Fig. 7. Behavioural accuracy as a function of the spatial scale of the mask, for different SOAs (\pm s.e.m.) and as a function of the number of images in the mask. (A) Results when only one image is used in the mask. (B) Results when a sequence of two images is used in the mask. The pattern of the first image is illustrated here, all 3 other spatial patterns were used equally as the second image of the mask. The different curves represent behavioural accuracy for different SOAs. Ten subjects performed the experiment. Coloured asterisks indicate when the condition is at chance-level ($p < 0.01$).

(Fig. 7A), disruption effects depended strongly on the spatial scale of the mask at 6 and 12 ms SOA. When the finest spatial scale was used, masking was effectively complete since performance was at chance-level. Even with the second finest scale, performance was still very poor. This means that on virtually half the trials, just the first mask pattern was enough to completely disrupt processing, limiting the number of problematic trials. In contrast, at coarser spatial scales, the masking was less effective and subjects were able to perform more than 70% correct when just the mid-to-coarse scale mask image was used. However, when the mask contained two different images in succession (Fig. 7B), accuracy was no better than chance for both 6 and 12 ms SOA, irrespective of which spatial scale was flashed first. This result may appear to contradict the data from the previous experiment, in which performance was significantly above chance with an SOA of 12 ms, despite having used an even longer 8 image dynamic mask. However, it should be noted that there were much more trials per condition in the original experiment (2560 vs. 640 trials), which increases the statistical sensitivity of the test.

Together, the results of this control experiment demonstrate conclusively that the masking effects were indeed very strong and occurred very rapidly from the onset of the mask. Given that for the main experiment, the 8 image dynamic mask was continued for 100 ms, we can safely conclude that the disruption was complete throughout a critical period for target processing.

4. Discussion

4.1. Time course of information extraction

4.1.1. Visual information is extracted before masking

As might be expected, the behavioural data shows a strong masking effect on the visual processing involved

in this high-level categorisation task, with both a drop in accuracy and an increase in reaction times. With a 6 ms SOA, processing appears to be completely blocked since the subjects were unable to perform significantly above chance. However, from 12 ms onwards, accuracy is already above chance level and performance increases until a ceiling effect is reached between 44 and 81 ms.

The high vertical refresh rate of the monitor (160 Hz), allowed us to observe a large range of masking levels, which leads to make several remarks about visual information extraction. First, it is noteworthy that the maximum accuracy reached by the subjects was close to the accuracy obtained in the same categorisation task used without masking (Delorme et al., 2000), indicating that the mask has no major effect after 81 ms. This also means that the presentation time of the picture, reduced to 6 ms in the present experiment from 20 ms in most of our previous experiments, had no appreciable effect on precision. It therefore appears that this 6 ms stimulation period is sufficient for the retina to extract enough information from the picture for animal detection to occur. Finally, the control experiment confirmed that the masking effects were strong even with just the first mask image, and that by the time the second mask pattern was presented, the disruption was complete. Given that the dynamic mask was maintained for 100 ms, we can be very confident that a long period of target processing is affected by the effective masking. This supposes that the delay available to extract relevant features from the stimulus is limited before masking takes place, and the visual system should base its analysis on a restricted amount of information to perform the task.

4.1.2. Visual information accumulates over time

The electrophysiological data also argue for a progressive accumulation of information. The differential activity, calculated by taking the difference between ERPs on correct target and distractor trials, is strongly affected by the SOA reduction. Its amplitude decreases

when the mask gets closer to the picture. We found a very high correlation between this reduction and behavioural accuracy. This can be related to another analysis of the differential activity in a previous categorisation task (Rousselet et al., 2004), where the status of the trials (Correct / False Alarm / Missed) could also be linked to the amplitude of the differential effects at occipital, frontal and parietal sites. The predictability of the behavioural outcome on the basis of the differential ERP signals constitutes a striking demonstration of a strong link between perception and cerebral activity, as previously shown with both fMRI (Dehaene & Naccache, 2001; Grill-Spector et al., 2000; Vanni et al., 1996) and unitary recordings in monkeys (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Leopold & Logothetis, 1996; Thompson & Schall, 1999).

The analysis of the differential signals between roughly 150 and 250 ms demonstrated that the more the activity averaged on targets differs from the distractor activity, the more subjects are able to detect the animals. This result could be related to the analysis of response rate since the mask has a higher effect on subjects' decisions for targets than for distractors (Fig. 2B). The overall data show that the difference between target and distractor responses is maximized when the mask is presented far from the stimulus, as if there were more and more cues accumulating to dissociate these two groups of images. The results are in accordance with the model of sensory information accumulation proposed by Schall (Schall, 2001) and derived from earlier work by Shadlen, Newsome and colleagues (Gold & Shadlen, 2000; Kim & Shadlen, 1999; Salzman & Newsome, 1994; Shadlen & Newsome, 1996). In their experiments, monkeys were trained to judge the main direction of motion in a collection of moving dots. The monkeys reported their responses by making an eye movement to one of two points, each indicating a given direction. The authors proposed that the decision depends on the accumulated signal corresponding to the increasing discrimination of the main direction of motion. In the same way in our experiment, we can suppose that the greater the separation between stimulus and mask, the greater the amount of processing that can be performed. Relevant information concerning the presence of an animal in the picture is accumulated until reaching a decisional threshold.

This model of cue accumulation over time fits with our data on reaction time. With SOAs between 25 and 44 ms, the early part of the reaction time distribution did not appear to be affected by the mask, but we observed a saturation effect on correct responses with median reaction times (Fig. 3A). This suggests that when pictures contain particularly salient cues, extensive information accumulation is not necessary and the subjects are capable of executing fast responses. However, when the target discrimination requires more analysis,

information would not be available because processing is disrupted by the mask. Below 25 ms, the integration time was probably insufficient to process as much information because the mask affected even the earliest responses.

If the visual system bases the results of its analysis on the accumulated information, what does it imply for information encoding? How can the mechanisms of information extraction at different steps of the visual pathways be decomposed and what determines the impact of masking interference?

4.2. Information encoding in the visual pathways

4.2.1. Interference between stimulus and mask information: the where and how issues

Behavioural performance does not increase much with SOAs longer than 40 ms. We may relate this result with the latencies obtained from macaque neurophysiology showing that the first 30–40 ms includes the most selective part of the neuronal responses (Kovacs, Vogels, & Orban, 1995; Rolls et al., 1999; Tovee & Rolls, 1995). This data suggests that there is an upper limit on the time required at each processing stage to extract the relevant information that needs to be transmitted to the next step. Any processing that would take longer would be obliterated or smothered by the mask information.

Where would these masking effects take place? A first model would propose that the effects are more likely to occur relatively early in the visual pathways, for instance at the level of V1, depending directly on the structure where mask information could be encoded. Recordings in monkey infero-temporal cortex have demonstrated that the majority of neurons are maximally activated by stimuli more complex than bars or simple textures (Tanaka, Saito, Fukada, & Moriya, 1991), and showed a high degree of sensitivity to image scrambling, with activation decreasing together with performance (Vogels, 1999a, 1999b). Other studies, using functional imaging in humans, have compared the activation produced by objects and textures and found that a region of lateral occipital cortex was preferentially activated by objects even when the spatial frequencies and contrast of the object stimuli matched those of the texture stimuli (Grill-Spector, Kushnir, Edelman, Itzhak, & Malach, 1998; Malach et al., 1995). All these studies suggest that the mask, as a kind of texture stimulus, should mainly interact with picture information in earlier areas.

The next question concerns the *mechanism* by which the masking effect occurs. One simple explanation of masking assumes that the mask produces interference when neural responses to the mask and the test image overlap in time, and this effect is all the more important when it concerns spatially overlapping information related to critical features of the stimulus. There is good

evidence that the activation of sensory inputs to areas such as the striate cortex results in strong intracortical inhibition that could well interfere with the processing of subsequent inputs. The disrupting effects will thus depend on the spatio-temporal overlap between the neural responses to the test and mask stimuli.

Neurophysiological studies have shown that the onset latencies of neurons within a given visual structure vary from neuron to neuron, even when the visual stimulus is unchanged. For example, in primate visual cortex, onset latencies can vary from as little as 30 ms to 70 ms or more. The reasons for this variability are diverse, but one of the most important factors is undoubtedly stimulus contrast. It is notable that the shortest latencies ever seen have been obtained with very high contrast and high luminance stimuli. Given that the mask stimuli used in our experiments all have maximal contrast, we can assume that many neurons in V1 will respond to the mask with particularly short latencies (Albrecht, Geisler, Frazor, & Crane, 2002; Albrecht & Hamilton, 1982; Foxe & Simpson, 2002; Nowak & Bullier, 1997; Reich, Mechler, & Victor, 2001; Sestokas & Lehmkuhle, 1986). In contrast, the neural responses to the natural images used as test patterns are likely to be substantially more variable. Indeed, if we suppose that any given photograph of an animal will contain many different features that can be diagnostic for the presence of an animal, it is clear that the contrast associated with each feature will vary a lot. Thus, much of the critical information about the stimulus will be conveyed less rapidly than information about the mask, strengthening the effects of inhibitory mechanisms. Only information that can survive this spatio-temporal overlap would then be transmitted to the next step, and contribute to accumulate cues about the test stimulus. The first interpretation is thus based on the disruption of feed-forward processes, due to mask processing catching up with stimulus processing.

Another interpretation is based on the difference in transmission rates along the fast magnocellular (M) and the slower parvocellular (P) visual pathways. Detailed chromatic representation in the P stream reaches visual cortex roughly 20 ms after the M inputs that mainly transmit motion and coarse luminance-based information (Nowak & Bullier, 1997; Nowak, Munk, Girard, & Bullier, 1995). Taking into account this 20 ms delay between the two streams of information, the mask might have little effect on the magnocellular processing of the test image but would strongly interfere with its parvocellular processing. However, magnocellular information may be sufficient to allow good accuracy in the fast categorisation task used here (Delorme et al., 2000; Macé, Thorpe, & Fabre-Thorpe, *in press*), and the interference of the highly contrasted mask with the feed-forward processing of magnocellular information (Macé et al., *in press*) may still be significant.

Finally, mask processing could interrupt feedback processing of the stimulus, at least at two levels. Iterative loops are thought to be important for segmentation, and involve the convergence of feedback from higher areas to areas like V1 or V2 (Hupe et al., 1998; Lamme, Super, & Spekreijse, 1998). In such pattern masking experiments, the processing of feedback information is probably made difficult with a mask closely following the stimulus. Moreover, subjects often reported that they released the button without explicit understanding of the photograph, which is in accordance with the common idea that feedback processing may be crucial for conscious image perception (Bullier, 2001; Lamme & Roelfsema, 2000; Pascual-Leone & Walsh, 2001).

These interpretations are not mutually exclusive and could even explain the striking differences between the effect of short SOAs on the initial part of the RT distribution and the plateau effect obtained with the 31 ms SOA (Fig. 3A). In fact, two different kinds of perturbations may be reflected here. The plateau effect may result from disruption of feedback processing related to the detection of the target, or disruption of direct parvocellular inputs. In contrast, the shift observed in the initial part of the RT distribution with shorter SOAs could reflect the disruption of the initial wave of processing.

4.2.2. A pipeline architecture

If we assume that the visual system accumulates sensory information until a decision threshold is reached, the very progressive masking effect over time is another point of interest. Presumably, this sort of task requires information processing at several different levels of the visual system including the retina, LGN, V1, V2, V4 and inferotemporal cortex. We have argued in the past that this sort of fast processing may leave only a short time at each processing stage before the next level has to respond, maybe as little as 10 ms or so (Bullier & Nowak, 1995; Fabre-Thorpe et al., 2001; Nowak et al., 1995; Thorpe & Fabre-Thorpe, 2002). These results confirm that visual processing can rely on such short latencies and challenges traditional views that use firing rate codes to convey information (Thorpe, Delorme, & VanRullen, 2001; VanRullen & Thorpe, 2001b; VanRullen & Thorpe, 2002). Further investigations will be necessary to understand how visual processing can be performed in such temporally constrained conditions. In the case of a serial model of information transmission, we might have expected sharply contrasted responses in which performance and ERPs are strongly affected below the decision threshold and less disrupted above this threshold, although the averages of performance across trials and subjects may obscure some types of more discrete transition (Miller, 1988). In contrast, in the case of a continuous model, masking effects may be progressively lessened with increased SOA. Our data showed that as SOA is increased above 12 ms, subjects

gradually increased their ability to detect an animal in the picture, which suggests that information can be conveyed to extrastriate areas in a continuous and asynchronous way. The lack of a clear threshold may also indicate that the information does not need to be fully processed at each stage (that is for all points of the space at the same time), but could be forwarded to the next stage and processed more progressively. Although the results do not argue conclusively in favour of a continuous model, they nevertheless strongly suggest that information transfer occurs progressively at each stage using a form of pipeline architecture.

A final point concerns the nature of the information used to perform the animal/non-animal task. While, in principle, we think that this should be considered as a true high-level visual task, there have been suggestions that even relatively high-level categorisations such as “natural vs man-made scenes” can be made on the basis of relatively low-level information. For example, Torralba and Oliva have reported that a linear combination of the outputs of a series of orientation and spatial frequency tuned channels can allow performance at over 80% correct (Torralba & Oliva, 2003). We certainly cannot exclude the possibility that our subjects are using this sort of information. However, so far at least, none of these purely low-level strategies has succeeded in achieving performance levels of above 90%, nor have they been used to differentiate between classes of objects. We therefore feel that other more complex visual processing strategies are probably at work. The current set of experiments does not allow us to distinguish between these possibilities. Nevertheless, they do demonstrate that, whatever the nature of the information used to perform the task, it is information that the visual system can extract extremely rapidly.

Acknowledgements

This work was supported by the CNRS, the ACI Integrative and Computational Neuroscience. Financial support was provided to both N. Bacon-Macé and M.J.-M. Macé by a Ph.D. grant from the French government. We thank G.A. Rousselet for technical assistance and helpful discussions on the results of the experiment.

Supplementary data

Correlation between behavioural accuracy and the amplitude of the occipital differential activity, on the right hemisphere occipital electrodes. The values shows individual correlation calculated with a Pearson test ($p < 0.01$) between behavioural accuracy and the differential activity amplitude on five occipital electrodes. Differential activity amplitude was determined by the most

negative point between 150 ms and 250 ms on averaged signals by condition, for each subject. The bottom line indicates an even stronger r -value by averaging the parameters for all 16 subjects before correlating them. Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.visres.2005.01.004](https://doi.org/10.1016/j.visres.2005.01.004).

References

- Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *Journal of Neurophysiology*, *88*(2), 888–913.
- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, *48*(1), 217–237.
- Breitmeyer, B. G. (1984). *Visual masking: an integrative approach* (p. 454). Oxford, New York: Oxford University Press.
- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*(1), 87–100.
- Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences*, *5*(9), 369–370.
- Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research*, *40*(16), 2187–2200.
- Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Perception and Psychophysics*, *25*(4), 249–263.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*(2), 171–180.
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. *Experimental Brain Research*, *142*(1), 139–150.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390–394.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*(1), 191–202.
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837–843.
- Hasbroucq, T., Burle, B., Bonnet, M., Possamai, C. A., & Vidal, F. (2002). Dynamique du traitement de l'information sensorimotrice: apport de l'électrophysiologie. *Canadian Journal of Experimental Psychology*, *56*(2), 75–97.
- Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787.

- Keyser, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences*, 6(3), 120–125.
- Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, 2(2), 176–185.
- Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of National Academic Science USA*, 92(12), 5587–5591.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23(11), 571–579.
- Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535.
- Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, 379(6565), 549–553.
- Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (in press). Rapid categorisation of achromatic natural scenes: how robust at very low contrasts? *European Journal of Neuroscience*.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of National Academic Science USA*, 92(18), 8135–8139.
- McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, 86(4), 287–330.
- Miller, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychologica*, 67(3), 191–257.
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. S. Rockland, J. H. Kaas, & A. Peters (Eds.), *Extrastriate visual cortex in primates* (Vol. 12, pp. 205–241). New York: Plenum Press.
- Nowak, L. G., Munk, M. H., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neurosciences*, 12(2), 371–384.
- Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516), 510–512.
- Reich, D. S., Mechler, F., & Victor, J. D. (2001). Temporal coding of contrast in primary visual cortex: when, what, and why. *Journal of Neurophysiology*, 85(3), 1039–1050.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. *Journal of Cognitive Neuroscience*, 11(3), 300–311.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5(7), 629–630.
- Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3(6), 440–455.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Research*, 44(9), 877–894.
- Salzman, C. D., & Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science*, 264(5156), 231–237.
- Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *National Review of Neuroscience*, 2(1), 33–42.
- Sestokas, A. K., & Lehmkuhle, S. (1986). Visual response latency of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. *Vision Research*, 26(7), 1041–1054.
- Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of National Academic Science USA*, 93(2), 628–633.
- Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66(1), 170–189.
- Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience*, 2(3), 283–288.
- Thorpe, S., Delorme, A., & VanRullen, R. (2001). Spike-based strategies for rapid processing. *Neural Network*, 14(6–7), 715–725.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.
- Thorpe, S. J., & Fabre-Thorpe, M. (2002). Fast visual processing and its implications. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed.). Cambridge, MA: MIT Press.
- Torralla, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, 14(3), 391–412.
- Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition*, 2(1), 35–58.
- Vanni, S., Revonsuo, A., Saarinen, J., & Hari, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. *Neuroreport*, 8(1), 183–186.
- VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird. Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30(6), 655–668.
- VanRullen, R., & Thorpe, S. J. (2001b). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6), 1255–1283.
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23), 2593–2615.
- Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part1: behavioural study. *European Journal of Neuroscience*, 11(4), 1223–1238.
- Vogels, R. (1999b). Effect of image scrambling on inferior temporal cortical responses. *Neuroreport*, 10(9), 1811–1816.

2.3 Analyses complémentaires

Dans chacun des chapitres présentant des résumés d'article publié ou soumis, nous rajouterons un certain nombre d'analyses complémentaires qui n'ont pas pu être développé lors de la rédaction des manuscrits.

2.3.1 Propriétés du masque dynamique : expérience contrôle

Une des grandes difficultés de cette expérience consistait à trouver un masque le plus efficace possible pour bloquer totalement la perception des photographies de scènes naturelles. Nous avons choisi d'utiliser des images en noir et blanc pour faciliter la recherche d'un masque, car une précédente expérience de catégorisation sans masquage a montré que les réponses sont aussi rapides et précises pour des images en niveau de gris que des images en couleur (Delorme, Richard & Fabre-Thorpe, 2000). Les scènes naturelles sont composées de multiples fréquences spatiales, extrêmement variables d'une image à l'autre et difficilement contrôlables. Comme nous l'avons vu dans les chapitres précédents, une majorité des études de masquage rétrograde utilise des stimuli visuels simples : lettres de l'alphabet, formes géométriques ou dessins d'objets et l'utilisation de bruit blanc ou d'un masque pattern superposant les contours du stimulus permet d'obtenir des résultats très satisfaisants (Delord, 1998). Dans le cas des scènes naturelles, ces masques ne suffisent pas. Notre masque a donc été obtenu à partir d'un bruit blanc filtré à différentes échelles spatiales et contrasté à 100% pour générer des formes aléatoires noires et blanches plus ou moins étendues. Nous avons utilisé plusieurs images-masques successives, chacune étant composée de fréquences spatiales plus ou moins élevées, ce qui a justifié son appellation de « masque dynamique »

En utilisant un masque dynamique, on se pose évidemment le problème de l'image du masque à partir de laquelle les mécanismes d'interférence sont maximaux. S'il faut attendre 2 voire 3 images de masque (présentées chacune 12.5 ms) pour que le masquage soit optimal, il faut alors rajouter ce délai à tous les SOA que l'on considère comme latence d'interruption. Nous avons donc effectué une expérience contrôle, dont les résultats sont en partie publiés dans l'article qui suit. Dans cette expérience, 10 sujets ont effectué la tâche de catégorisation pour 4 conditions de SOA (6, 12, 44 et 106 ms). Le nombre d'images qui composent le masque variait aléatoirement entre 1 et 8, et la séquence de pattern était contrôlée (4 pattern possibles avec des motifs noirs et blancs plus ou moins gros, pour ne pas parler de « fréquence spatiale »). Les résultats montrent en effet qu'une seule image n'est pas suffisante pour masquer complètement les traitements lorsque le SOA est réduit à 6 ms (Figure 2.1B). Mais si l'on regarde plus particulièrement la précision obtenue avec une seule image de masque en fonction du pattern d'images de masque, il apparaît que deux patterns seulement posent problème à 6 ms : ceux avec les motifs les plus grands¹ (Figure 2.1C). Cette remarque permet déjà de réduire le nombre d'essais problématiques. D'autre

¹ Il faut remarquer que ces masques, qui contiennent de grands motifs, contiennent aussi des échelles spatiales plus élevées, de par leur structure à bord net.

part, lorsque le masque contient deux images successives, la précision reste au niveau chance pour un SOA de 6 ms, quels que soient les patterns composant ces deux images de masque (Figure 2.1D). Sachant que la durée du masque dynamique composé de 8 images dans l'expérience initiale s'étendait sur une centaine de millisecondes en tout, on peut de toutes façons conclure aisément que les effets perturbateurs interviennent à un moment critique de l'analyse sensorielle de l'image.

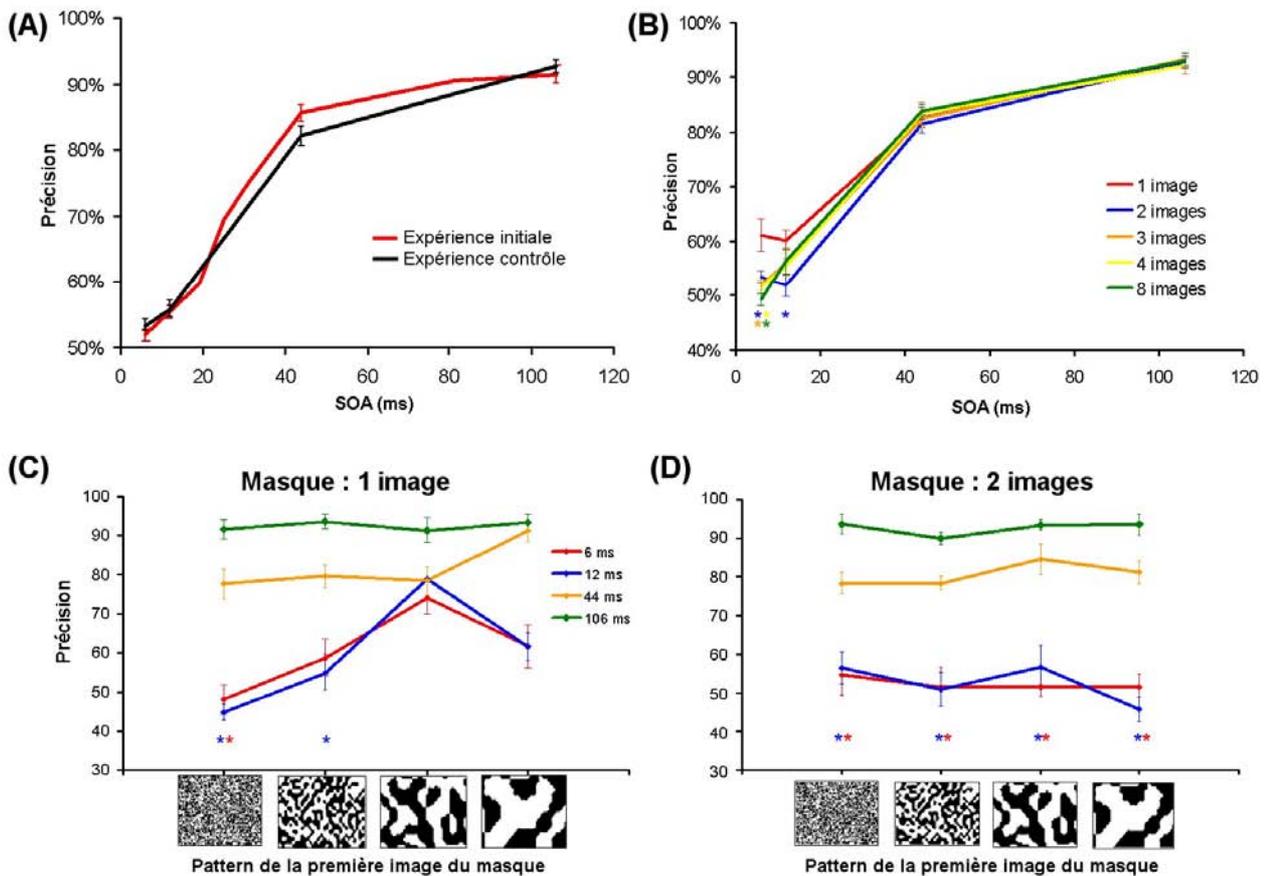


Figure 2.1 : Résultats de l'expérience contrôle. (A) Précision en fonction du SOA dans l'expérience initiale et dans l'expérience contrôle. Les valeurs de performance sont similaires. (B) Si l'on décompose les courbes de précision selon le nombre d'images qui composaient le masque, les résultats sont très similaires entre eux, excepté pour deux conditions. Lorsqu'il n'y a qu'une image de masque, le premier pas de SOA à 6 ms commence au dessus du niveau chance contrairement aux autres conditions de masquage. Au contraire, lorsque deux images composent le masque, la précision reste au niveau chance à 12 ms de SOA. Précision comportementale lorsqu'une seule image (C) ou deux images (D) composent le masque, selon le pattern que constitue la première image du masque. Les étoiles de couleur indiquent les conditions au niveau chance.

2.3.2 Effet du masquage sur les potentiels

L'étude de l'activité différentielle permet de s'affranchir indirectement de la composante du signal induite par la présentation du masque en tant que stimulus. Les signaux moyennés séparément sur les essais cibles et distracteurs corrects sont à la fois induits par la présentation de la scène naturelle et par les 8 images constituant le masque. La soustraction des signaux cibles et distracteurs permet de se focaliser sur l'activité différentielle qui représente à la fois les opérations visuelles et cognitives liées à la tâche, ainsi que leurs modifications par les effets de masquage selon les différentes conditions de SOA (Article 1). La figure 2.1 montre le profil classique des signaux moyennés sur les essais cibles et distracteurs sur les aires frontale, pariétale et occipitale.

L'activité différentielle prend naissance au moment où le signal moyenné sur les essais distracteurs corrects atteint une plus grande amplitude que le signal cible, commençant sur le pic de la N1 pour se finir après le pic de la P2. Certains auteurs argumentent que l'activité différentielle représente des traitements relativement précoces liés aux différences physiques entre les groupes d'images, notamment parce qu'il n'existe pas de corrélation claire entre cette composante et les temps de réaction (Johnson & Olshausen, 2003; Johnson & Olshausen, 2005). Or l'absence de corrélation avec les temps de réaction n'est pas en soi un facteur suffisant pour exclure une interprétation plus cognitive de cette composante. Les auteurs du papier original de notre équipe avait d'ailleurs proposé que l'activité différentielle résulterait d'une inhibition massive des traitements par les aires frontales sur les essais no-go (Thorpe, Fize & Marlot, 1996). Sachant que la latence de cette activité ne varie pas avec les TR, il était proposé que le signal différentiel ne dépende pas des essais « go », mais des essais no-go sur les distracteurs, pour lesquels les sujets n'effectuaient pas de réponse motrice. D'autres expériences avaient déjà mis en évidence des effets semblables d'inhibition de réponse (Gemba & Sasaki, 1989; Sasaki, Gemba, Nambu & Matsuzaki, 1993). Or, il paraît maintenant clair que cette interprétation était fautive et que la variabilité des temps de réponses comportementales ne peut pas être reliée aux traitements purement sensoriels. Il est intéressant de citer ici les travaux de Di Carlo et Maunsell, qui ont montré dans une analyse essai par essai que les neurones de AIT ont des latences de réponses très peu corrélées avec les temps de réaction moteurs, confirmant que cette aire est une région essentiellement sensorielle et que la variabilité des TR a une origine post-perceptuelle (DiCarlo & Maunsell, 2005).

En revanche, si la latence de l'activité différentielle n'est pas corrélée aux temps de réaction, l'expérience de masquage montre que son amplitude est fortement corrélée à la performance comportementale des sujets. Ce résultat est à mettre en relation avec l'absence d'activité différentielle sur les essais cibles incorrects (Missed) alors qu'elle est faible mais existante sur les essais distracteurs incorrects (False Alarms), suggérant que le signal différentiel pourrait être le reflet de processus avancés de sélection de la cible. Il est possible d'envisager que les premières vagues de traitement feedforward permettent d'intégrer suffisamment d'informations pour déclencher une réponse motrice. Après cette activation feedforward, la représentation de la scène pourrait être affinée par un flux d'informations feedback en provenance des aires frontales – ce que refléterait l'activité différentielle - permettant par exemple de faire correspondre la

représentation du stimulus encodé avec les représentations mnésiques, ou de spécifier d'avantage la précision spatiale dans le but de sélectionner la cible...

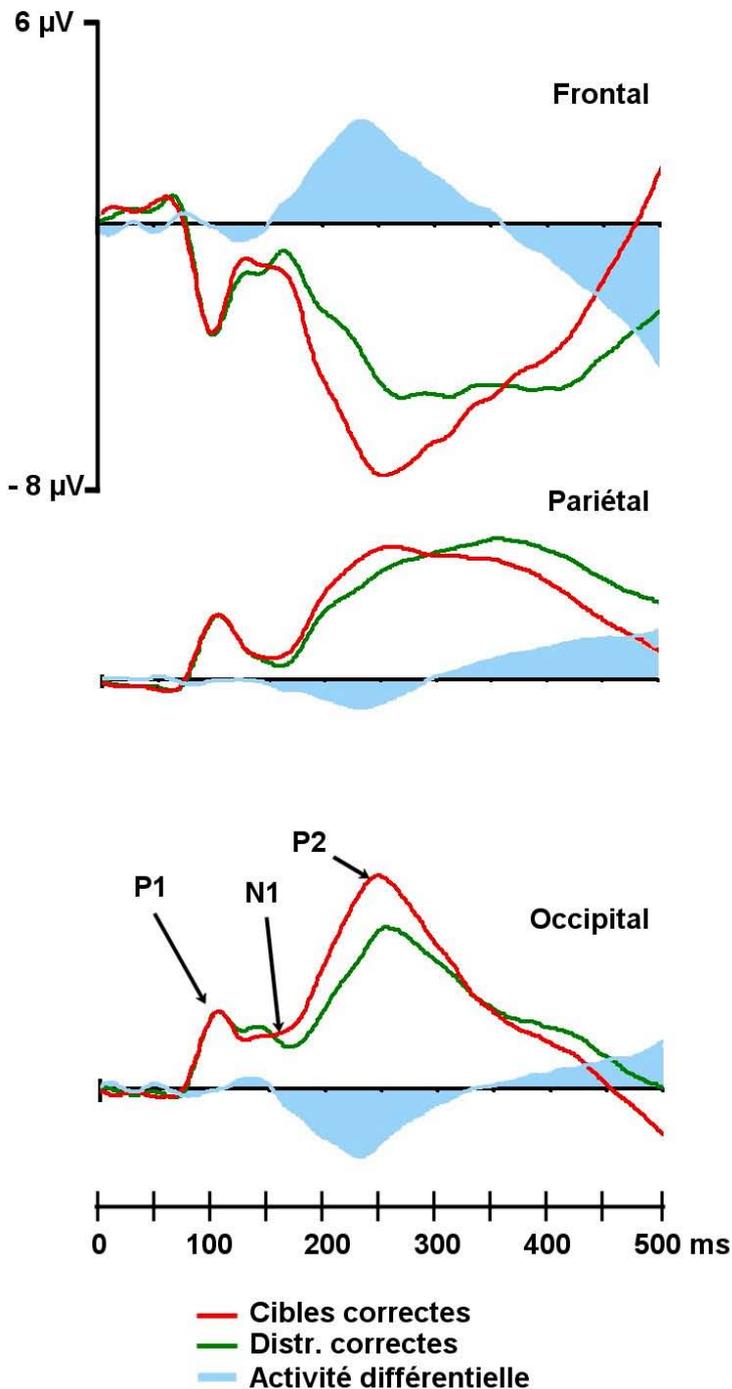


Figure 2.2 : Potentiels évoqués et activité différentielle enregistrés sur différentes régions cérébrales. Les potentiels sont moyennés séparément sur les essais corrects cibles (vert) et distracteurs (rouge) dans une tâche de catégorisation animal/non-animal où les scènes naturelles sont présentées en niveau de gris (d'après l'étude publiée dans Macé, Thorpe & Fabre-Thorpe, 2005). L'activité cérébrale est calculée comme la soustraction des signaux distracteurs aux signaux cibles, et correspond à une plus forte amplitude du signal potentiel des distracteurs à partir du pic de la N1 jusqu'à la phase descendante de l'onde P2. Les signaux frontaux sont moyennés sur 4 électrodes (FP1, FP2, F3 et F4), les signaux pariétaux sur 4 électrodes (P3, P4, P03 et P04) et les signaux occipitaux sur 8 électrodes (O1, O2, O9, O10, P07, P08, P09 et P010).

Nous avons maintenant de bons indices pour penser qu'en 150 ms, les traitements de catégorisation sont déjà bien avancés, puisqu'un protocole de discrimination de scènes naturelles développé récemment dans l'équipe a permis de montrer que les réponses saccadiques (plutôt que manuelles) pouvaient être initiées correctement vers les images cibles dès 120 ms, avec une

moyenne autour de 220 ms² (Kirchner & Thorpe, 2005). Bien que cette tâche diffère un peu de la tâche classique de catégorisation (les sujets doivent effectuer une saccade à gauche ou à droite pour déterminer laquelle de deux images contient un animal), les temps de réaction mesurés comprennent à la fois les délais engagés pour les traitements sensoriels, décisionnels et moteurs ! Nous reviendrons sur cette expérience dans le chapitre suivant. Ces temps de réaction sont inférieurs aux latences de l'activité différentielle à 150 ms, suggérant définitivement que cette composante est le reflet de processus tardifs (pour une plus ample discussion sur l'origine de l'activité différentielle, se référer à Rousselet, Thorpe & Fabre-Thorpe, 2004).

Si l'activité différentielle liée aux traitements de catégorisation voit son amplitude diminuer alors que sa latence reste inchangée avec les effets du masquage, il peut être intéressant de regarder comment évolue le signal des potentiels mêmes. Or une condition particulière avait été introduite dans le protocole expérimental afin de comparer la performance des sujets dans les conditions les plus difficiles de masquage avec le cas où la scène naturelle était complètement absente : dans un essai sur 10, le masque était présenté seul, sans aucune scène auparavant et les sujets n'étaient pas informés de cette condition. Au niveau comportemental en effet, les sujets répondaient aussi peu qu'il y ait une scène suivie du masque avec 6 ms de SOA ou qu'il n'y ait pas de scène du tout (resp. 22,3% et 19,6%, la légère différence observée dans ce taux de réponse indique seulement que le système visuel a commencé à traiter une partie de l'information sur l'image, sans pour autant faire de différence entre les essais cibles et distracteur puisque la précision est au niveau chance pour une SOA de 6 ms). La mise au point de cette condition présente aussi l'avantage de pouvoir isoler le signal EEG du masque seul. En décalant ce signal au cours du temps pour faire correspondre son apparition au SOA utilisé dans chaque condition de masquage, on peut soustraire le signal du masque aux potentiels cibles et distracteurs moyennés pour chaque condition de SOA (Figure 2.2A). Pour toutes les conditions, les potentiels se superposent dans leur partie initiale jusqu'à un point de décrochement qui intervient de plus en plus tard dans le déroulement des potentiels, en fonction de l'arrivée du masque (Figure 2.2C). Au-delà de ce point, les signaux sont très perturbés, comme tronqués. La forme des potentiels laisse penser que les traitements sont interrompus brusquement. Pour la condition 6 ms, il intervient au tout début de l'augmentation du signal, même si quelques millisecondes semblent suffisantes pour voir apparaître une faible partie de l'activité. Or les potentiels apparaissent vers 70-80 ms avec la composante endogène positive P1, et reflètent à ce moment des traitements visuels précoces (Allison, Puce, Spencer & McCarthy, 1999; Di Russo et al., 2001), appuyant l'idée que le blocage des informations intervient très tôt dans les étapes de traitement. Cependant, nous n'insisterons pas plus sur cette analyse qui repose sur deux hypothèses qui ne sont pas totalement fiables :

- Le décalage et la soustraction du signal lié à la présentation isolée du masque supposent que l'on observe uniquement des mécanismes d'interférence par interruption et non pas par

² Des résultats récents, non publiés, ont montré qu'il était possible d'obtenir des temps de réaction moyen encore plus courts, lors de cette tâche, avec des valeurs médianes en-dessous de 150 ms pour certains sujets !

intégration des informations de l'image et du masque. En effet, dans l'hypothèse du masquage par intégration, certains groupes de neurones intègrent les informations venant des deux stimuli, induisant un signal moins important que si les activités sont produites par des groupes de neurones différents qui entrent en compétition par des mécanismes d'interruption (Jeffreys & Musselwhite, 1986). Or en soustrayant les signaux enregistrés de manière isolée, on néglige complètement les mécanismes d'intégration qui sont supposés être majeurs dans ce type d'interférence.

- Cette remarque est à relier au problème théorique plus général concernant l'additivité temporelle des signaux potentiels. Rieger et al. ont basé leur analyse des effets de masquage sur l'activité cérébrale en soustrayant le signal moyenné sur les scènes naturelles présentées seules au signal moyenné sur la scène plus le masque (Rieger, Braun, Bulthoff & Gegenfurtner, 2005). Les auteurs posent l'hypothèse que le départ de l'activité liée au masque pose une limite temporelle supérieure au délai à partir duquel le masque commence à interférer sur les traitements de la scène. Ils supposent donc que le problème est réduit si l'on s'attache uniquement à ce point d'interférence et non pas aux propriétés ultérieures du signal soustrait. Il en reste néanmoins que ce type d'analyse reste très controversé, de par les résultats contradictoires qu'elles ont soulevés et qui ont parfois conduit à conclure qu'aucun impact du masquage ne pouvait être observé sur le signal (Kaitz, Monitz & Neshet, 1985; Schiller & Chorover, 1966; Schwartz & Pritchard, 1981).

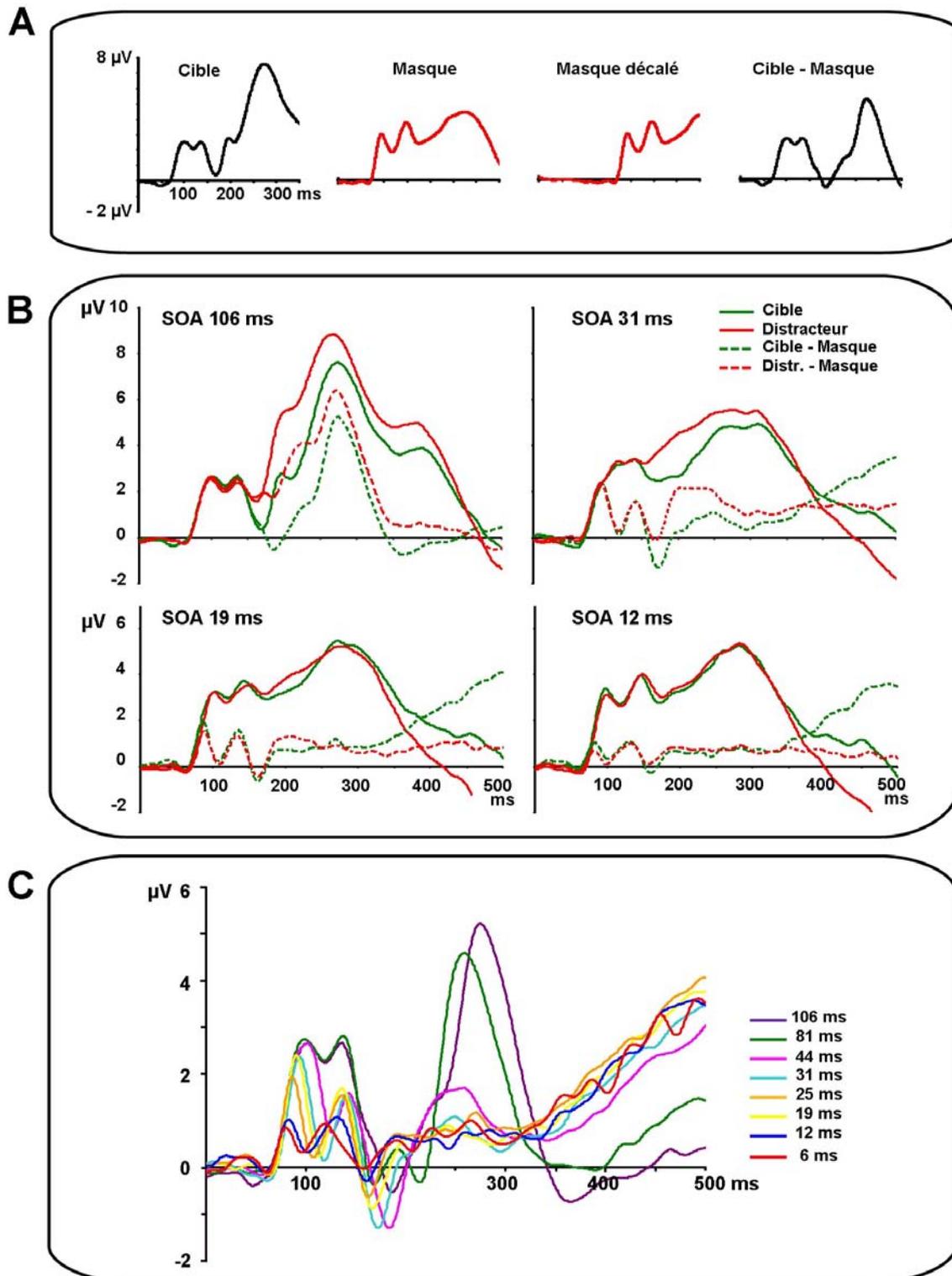


Figure 2.2 : Potentiels évoqués après soustraction du signal lié à la présentation du masque seul. (A) L'activité moyennée sur les essais cibles contient à la fois le signal de la scène et du masque (Cible), mais le signal du masque peut être isolé grâce aux essais dans lesquels le masque était présenté seul (Masque). En décalant ce signal d'un délai correspondant au SOA de chaque condition (Masque décalé), on soustrait l'activité obtenue au signal initial de la scène (Cible - Masque). (B) Signal moyenné sur les 16 sujets pour 4 conditions de SOA sur les essais cibles (vert) et distracteurs (rouge), avant (traits pleins) et après (traits hachurés) soustraction du signal décalé du masque. (C) Signal moyenné sur les essais cibles corrects pour les différents conditions de SOA, après soustraction du signal décalé du masque. Plus le SOA diminue, plus les composantes précoces du signal sont affectées.

2.4 Discussion

2.4.1 Décours temporel des traitements visuels

Les données comportementales montrent un fort effet du masquage sur les traitements mis en jeu dans la tâche de catégorisation visuelle animal/non-animal. Le masquage induit une baisse de la précision et une augmentation des temps de réaction. Les expériences préliminaires ont permis de couvrir un intervalle de SOA suffisant pour observer différents degrés de perturbation de la perception. Il est intéressant de noter que la précision atteint un maximum comparable aux résultats obtenus dans d'autres tâches de catégorisation sans masquage (Delorme, Richard & Fabre-Thorpe, 2000), malgré la courte présentation des scènes naturelles de 6.25 ms (notons d'ailleurs que cette valeur est surestimée, puisque les 6.25 ms de rafraîchissement d'écran correspondent au délai mis par le faisceau lumineux pour parcourir la totalité de l'écran. Or les dimensions verticales de l'image ne dépasse pas les deux tiers de l'écran, ce qui correspondrait plutôt à des durées d'affichage de l'ordre de 4 ms). Ces premières observations permettent de faire la distinction entre temps d'extraction et temps d'intégration des informations. Quelques millisecondes de présentation du stimulus sont largement suffisantes pour extraire les caractéristiques bas-niveau des images, de par les propriétés de persistance neuronale qui permettent au système visuel d'accéder à l'information pendant une plus longue durée³ (Coltheart, 1980; Keyser & Perrett, 2002; Loftus & Irwin, 1998; Sperling, 1960). En revanche, tout l'enjeu du masquage porte sur le délai laissé au système visuel pour **intégrer** ces informations afin d'en inférer une représentation suffisamment développée qui permettra de répondre à la tâche. S'il est difficile de prouver qu'il n'existe pas de masque plus efficace pour bloquer l'intégration de ces informations, on peut cependant remarquer que les 40 ms nécessaires pour atteindre le plateau de performance sont tout à fait cohérentes avec les données électrophysiologiques chez le singe, qui montrent que la majorité de l'information est contenue dans les 30-50 premières millisecondes de réponse neuronale (Kovacs, Vogels & Orban, 1995; Rolls, Tovee & Panzeri, 1999; Tovee & Rolls, 1995).

2.4.2 Vers un modèle d'accumulation des informations

Avant ces 40 ms de SOA, la précision comportementale augmente presque linéairement avec l'augmentation du SOA, et un certain nombre d'indices laissent penser que les informations concernant le stimulus s'accumulent progressivement pour pouvoir déclencher une réponse motrice. D'une part, l'amplitude de l'activité différentielle est directement corrélée à la précision

³ Nous avons d'ailleurs essayé d'augmenter la fréquence de rafraîchissement de l'écran, et nous avons pu constater qu'à 200Hz, soit seulement 3,3 ms d'affichage d'image, les performances restent tout à fait similaires. Pour des raisons techniques (proportion d'erreurs trop importante dans la latence d'affichage sur l'ensemble des essais), nous avons dû rester à 160 Hz. Cependant, il est étonnant d'observer qu'avec 20 ms comme avec moins de 5 ms, le système peut extraire ses informations avec une grande efficacité. Il serait intéressant d'utiliser une autre technique (tachistoscope ?) pour diminuer encore plus le délai d'affichage, probablement en dessous de la milliseconde pour observer une diminution significative des performances. Le facteur limitant n'est probablement pas le temps d'exposition de l'image mais la quantité de photons qui parvient sur la rétine.

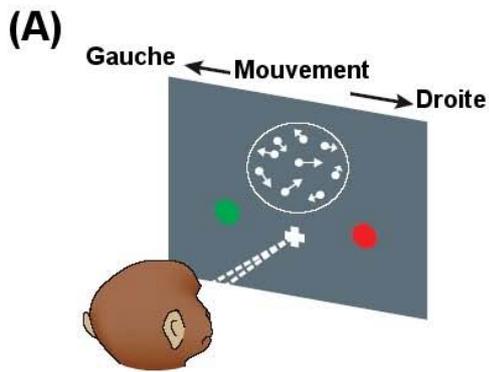
comportementale : plus le sujet est capable de détecter la présence d'un animal, plus l'activité cérébrale liée aux cibles est différente de celle liée aux distracteurs. D'autre part, le masque perturbe plus le déclenchement des réponses sur les cibles que celui des réponses distracteurs. En d'autres termes, lorsque les conditions de perception deviennent difficiles, les sujets ont plus de difficultés à décider qu'ils ont « vu » un animal, alors que la probabilité de répondre correctement sur les distracteurs est quasi-constante. Lorsque les traitements sont bloqués par la présence d'un second stimulus proche temporellement de l'image, tout se passe comme si on limitait la quantité d'informations permettant au sujet de prendre la décision de lever le doigt du bouton réponse. Ces résultats peuvent être reliés à la théorie de l'accumulation de l'information développée ces dernières années par les auteurs qui se sont intéressés aux corrélats neuronaux des processus décisionnels, étape indispensable pour comprendre le lien entre traitements sensoriels et moteurs.

A l'origine de cette série de travaux, Newsome et ses collègues ont cherché à savoir si certains groupes de neurones dans MT peuvent refléter la décision perceptuelle (Newsome, Britten & Movshon, 1989; Salzman, Britten & Newsome, 1990). Dans ces expériences, le singe devait déterminer le sens majoritaire d'un ensemble de points en mouvement dans des directions variables (Figure 2.3A et B). La réponse était donnée par une déviation du regard sur un des deux points indiquant le sens majoritaire des points en mouvement. Les auteurs remarquent que le taux de décharge neuronal corrèle presque parfaitement avec la quantité de points qui bougent de manière cohérente, indice utilisé par le singe pour décider la direction de son mouvement. En 1996, Shadlen propose un modèle basé sur ces observations, dans lequel une centaine de neurones analysant le mouvement à une position précise était moyennée pendant 2 s afin d'estimer la direction majoritaire de ces signaux au cours du temps (Shadlen, Britten, Newsome & Movshon, 1996). Cette direction était ensuite choisie pour déterminer la direction donnée pour la réponse saccadique. Shadlen et collègues ont vérifié les prédictions du modèle en montrant qu'en moyenne, les neurones de la région pariétale LIP produisent un taux de décharge croissant pendant l'affichage des points en mouvement, et la pente de cette activité croissante est corrélée à la cohérence du mouvement (Figure 2.3C) (Shadlen & Newsome, 1996; Shadlen & Newsome, 2001). L'aire LIP semble donc être un site interface entre les systèmes sensoriels et moteurs sachant qu'elle reçoit des projections de l'aire MT et projette vers le FEF. Au cours de cette même période, Schall et collaborateurs se sont intéressés à la transformation du signal sensoriel en signal permettant de générer une saccade au niveau du FEF (Hanes & Schall, 1996; Schall, 2001; Schall, 2002; Schall, 2003; Thompson, Bichot & Schall, 1997). Dans une tâche dite de « oddball », un singe devait fixer un point central alors que 8 stimuli apparaissaient sur un cercle, 7 étant d'une même couleur et le 8^{ème} d'une couleur différente (Figure 2.3D). La tâche du singe était d'effectuer une saccade rapide vers cet élément différent. Afin de comprendre comment les cellules du FEF, qui ont une organisation rétinotopique, effectuent une sélection correcte parmi les différents signaux représentant les huit stimuli, les auteurs ont moyenné les signaux séparément pour les réponses rapides, moyennes et lentes. Or comme le montre la figure 2.3D, le taux de décharge sur ces signaux moyennés augmente beaucoup plus rapidement pour les essais rapides que pour les essais lents, suggérant que la réponse motrice est déclenchée avec un délai fixe après que le taux de décharge ait atteint un seuil. Ces expériences ont une importance

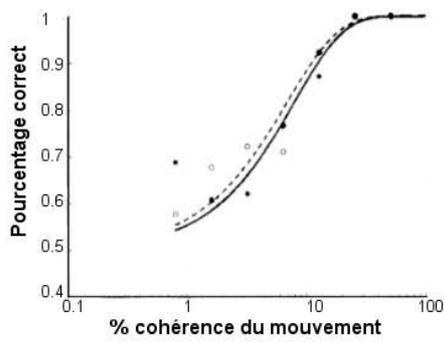
fondamentale pour notre compréhension des mécanismes de transformation du signal sensoriel en signal moteur, et appuie l'idée que les informations s'accumulent progressivement jusqu'à atteindre un seuil décisionnel. La saillance des éléments sensoriels, le bruit dans les différents systèmes, les mécanismes de compétition / inhibition entre les différentes alternatives, les fluctuations du seuil par des contraintes environnementales sont autant de paramètres à prendre en compte dans les modèles développés à partir des données expérimentales. Cependant, il est important de noter que ces données ont été observées sur un certain type de neurones. Par exemple le FEF possède clairement deux types de neurones, les uns sélectifs aux propriétés sensorielles dont les latences de décharge ne sont pas corrélées aux réponses comportementales, alors que le groupe de neurones concernés par les expériences décrites ci-dessus ont d'avantage de valeur « motrice » (DiCarlo & Maunsell, 2005; Schall, 2002). La compréhension des mécanismes de transition entre les traitements sensoriels et moteurs doit maintenant constituer un enjeu important, à la fois sur l'échelle temporelle et spatiale. Nous reviendrons sur ce point dans le chapitre suivant.

Dans notre expérience de catégorisation d'images masquées, les différentes observations citées ci-dessus suggèrent que plus on éloigne le masque de l'image, plus les traitements ont de temps pour s'effectuer et augmenter la quantité d'informations pertinentes pour détecter un animal dans l'image. A partir d'un certain seuil d'activation, l'analyse sensorielle deviendrait suffisante pour atteindre le seuil décisionnel du sujet et lui permettre de lever le doigt du bouton réponse. Cette idée sera de nouveau soulevée dans le chapitre suivant.

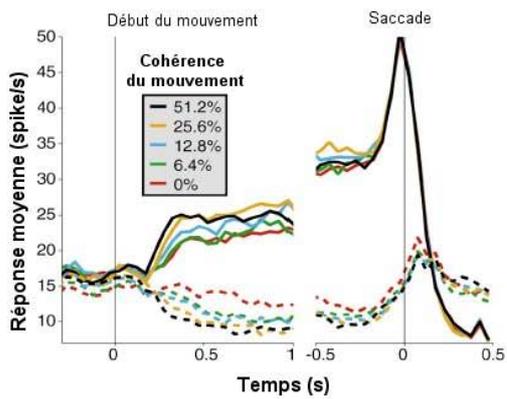
Figure 2.3 (page suivante) : *Corrélat neuronal des processus décisionnels.* (A) Tâche de discrimination utilisée dans les travaux de Newsome et al., dans laquelle un singe doit reporter le sens majoritaire de points en mouvement dans des directions variables. La réponse est donnée par une saccade sur un des deux points indiquant le sens majoritaire des points en mouvement. (B) Relation entre l'activité neuronale dans MT et la perception du mouvement. La courbe pointillée accompagnée par les points vides indique la probabilité qu'un animal déclenche une réponse dans une direction donnée, en fonction de la cohérence du stimulus en mouvement, c'est à dire en fonction de la proportion de points allant dans le même sens. La courbe pleine indique la réponse moyenne d'une cellule de MT lorsque l'on varie la cohérence du stimulus. Les deux courbes ont des profils particulièrement proches. (C) Réponse moyenne d'une population de 104 neurones situés dans la région LIP pendant la tâche de discrimination de points en mouvement. Les deux parties du graphique représentent les périodes de présentation du stimulus et d'attente avant la réponse. Les courbes en trait plein et pointillé représentent l'activité au cours du temps alors que le singe évalue la réponse vers ou en direction inverse du champ récepteur respectivement. Les différentes courbes de couleur représentent cette activité en fonction de la cohérence du stimulus en mouvement. A la fois les latences et les amplitudes sont affectées par la cohérence du stimulus. (D) Profil d'activité durant la tâche oddball utilisée par Schall et al. Les temps de réaction saccadiques présentent une variabilité naturelle, et sont donc divisés par les auteurs en trois parties : rapides, intermédiaires et lents. L'activité des neurones du FEF est mesurée pendant cette tâche et moyennée selon les catégories de temps de réaction engagés. La pente de l'activité démarre après 100 ms varie selon les catégories de temps de réaction. Il faut noter que l'activité a été moyennée sur les essais où la position de la cible engageait la meilleure réponse du neurone. Ces différents schémas ont été adaptés de Glimcher, 2001; Glimcher, 2003; Shadlen & Newsome, 2001.



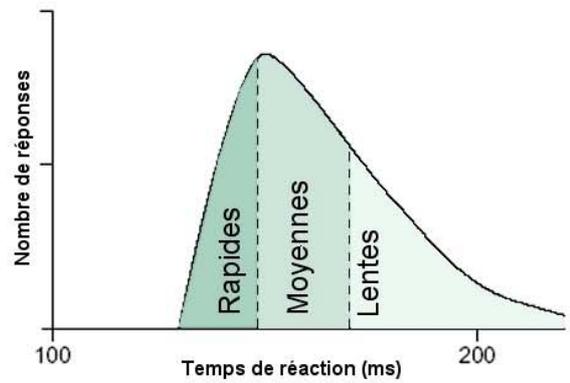
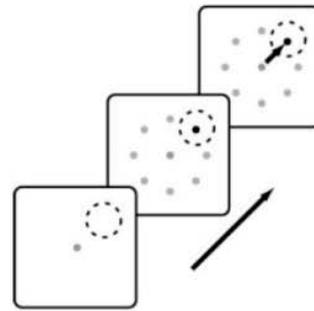
(B) MT



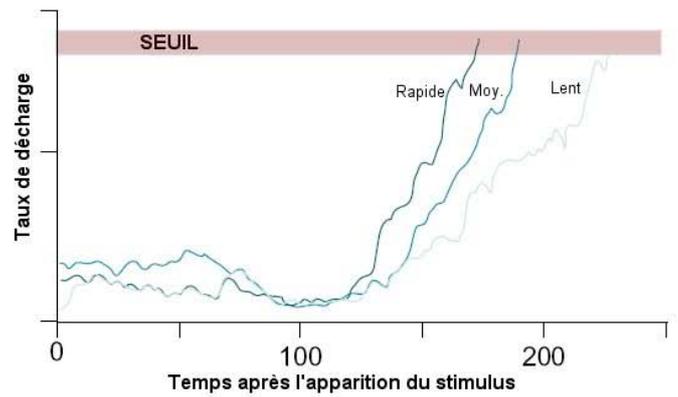
(C) LIP



(D)



FEF



2.4.3 Quels mécanismes d'interférence entre la cible et le masque ?

La compréhension des données rapportées dans la littérature du masquage a conditionné le type de masque et la gamme de SOA choisis pour bloquer les traitements mis en jeu dans la tâche de catégorisation de scènes naturelles. Les scènes naturelles étant des stimuli complexes et très variables, le choix d'un masque efficace a reposé sur de nombreuses expériences pilotes. Le masquage ne pouvait être que de type pattern, car le masquage par bruit blanc ne permettait pas d'interférer suffisamment avec les informations de contraste utilisées pour segmenter la cible dans l'image. En contrepartie, un masque composé de formes aléatoires à basse échelle spatiale n'était pas non plus suffisant comme le montre l'expérience contrôle décrite dans l'article 1. Nous avons donc mis au point un masque pattern composé de 8 images à différentes échelles spatiales et le niveau chance obtenu pour le SOA le plus court confirme son efficacité. La question se pose de savoir quand et où dans la hiérarchie visuelle ont lieu les mécanismes d'interférence entre informations de la scène et du masque.

Comme nous l'avons vu dans la partie introductive, la littérature du masquage par pattern rapporte largement deux types d'interférence : les mécanismes précoces d'intégration et les mécanismes d'interruption (Breitmeyer, 1984; Breitmeyer & Ganz, 1976; Keysers & Perrett, 2002; Turvey, 1973). On observe que la fonction de masquage est de type I, avec un impact plus important pour les plus courts SOA. Il faut aussi noter que le rapport d'énergie entre les scènes naturelles et le masque est certainement en faveur de ce dernier, de par ses motifs à 100% de contraste utilisés pour les différentes échelles spatiales. Le recouvrement spatial entre les informations de la cible et du masque implique des mécanismes d'intégration intervenant certainement très tôt dans le parcours cortical, ce qui justifie que les SOA concernés soient particulièrement courts, avant 40 ms. Lors de la rédaction de l'article décrivant cette expérience, nous avons proposé que ces interférences pouvaient avoir lieu avant V4. En effet, un certain nombre d'études ont montré que la majorité des neurones du cortex inféro-temporal sont sélectifs à des stimuli plus complexes que des barres ou des textures (Tanaka, Saito, Fukada & Moriya, 1991; Vogels, 1999a; Vogels, 1999b), résultat à relier à l'activation préférentielle du cortex occipital pour les objets par rapport aux textures (Grill-Spector et al., 1998; Malach et al., 1995). Il est donc peu probable que les images du masque, comparables à des textures, soient représentées très haut dans la hiérarchie visuelle et les mécanismes d'interférence, qu'ils soient de type intégratif ou interruptif par des connexions feed-forward ou latérales, doivent intervenir très tôt à la fois dans le parcours cortical et dans le temps. Les régions encodant les propriétés bas et moyen niveau comme V1 ou V4 sont probablement le lieu où se déroule une grande partie des mécanismes d'interférence. Sachant que les masques contiennent toutes sortes d'orientations, mais à des contrastes maximaux, il est tout à fait possible d'envisager un recouvrement spatial des informations contenues dans les champs récepteurs de certains groupes de neurones qui devraient être impliqués dans les processus de segmentation de la cible. Ce recouvrement est d'autant plus important qu'il a été montré que les latences les plus courtes sont obtenues avec des stimuli très contrastés dans le LGN ou V1 (Albrecht, Geisler, Frazor & Crane, 2002; Sestokas & Lehmkuhle, 1986). On peut alors imaginer que les informations du masque « rejoignent » celles de la cible avec un délai encore inférieur à la latence déterminée par le SOA. Cependant, nous avons réalisé

qu'il était aussi tout à fait possible de concevoir des mécanismes d'interruption intervenant encore plus tôt dans le système : dès la couche des cellules ganglionnaires de la rétine. Les réponses enregistrées dans ces cellules ont des latences de décharge qui ne démarrent pas avant 15-20 ms après la présentation du stimulus (Bolz, Rosner & Wassle, 1982), ce qui veut dire que dans la gamme de SOA de 6 à 12 ms, les informations du masque et de la cible peuvent être encodées simultanément au niveau de ces cellules, avant même que les premiers potentiels d'action ne soient générés dans le nerf optique. D'autre part, sachant que l'information encodée par 125 millions de photorécepteurs doit être compressée pour être transmise par le nerf optique qui ne contient qu'un million de fibres, cette zone semble être privilégiée pour faire intervenir des mécanismes d'interférence. Rappelons notamment que Levicks et Sacks ont montré que la présentation de deux flashes de lumière présentés pendant 2 ms produisait une seule réponse fusionnée au niveau des cellules ganglionnaires chez le chat anesthésié (Levick & Zacks, 1970), mettant ainsi en évidence un phénomène d'intégration des réponses neuronales.

Il est intéressant de faire le rapprochement entre cette hypothèse et le modèle de codage par rang proposé par Simon Thorpe et collaborateurs, qui met en avant l'importance des premiers potentiels d'action dans la transmission de l'information le long de la hiérarchie corticale. En résumé, ce modèle de reconstruction d'images se base sur un codage asynchrone des informations dans lequel ce n'est pas la fréquence de décharge mais l'ordre de décharge des neurones qui détermine leur importance relative. Le modèle fonctionne donc comme un convertisseur transformant les données qui sont à l'origine sous un format analogique (le niveau d'activation des neurones) en un format temporel. Dans un article de 2001, Rufin VanRullen et Simon Thorpe ont proposé de reconstruire ce qu'un observateur idéal pourrait observer s'il était situé à l'entrée de V1, à partir de l'information transmise dans le nerf optique (VanRullen & Thorpe, 2001a). En utilisant ce codage par rang, les auteurs ont simulé le fonctionnement des cellules ganglionnaires de la rétine en calculant l'intensité des contrastes locaux grâce à des filtres à différentes polarités et échelles spatiales (cellules de type ON-OFF). Ainsi la reconstitution des scènes naturelles à partir d'une vague de potentiels asynchrones montre que l'image reconstruite contient la grande majorité de l'information de l'image initiale lorsque seulement 1% de neurones utilisés pour simuler cette rétine artificielle ont déchargé. Plus récemment, Rudy Guyonneau a repris cette analyse et a montré dans une tâche psychophysique que des sujets pouvaient catégoriser des animaux sur la quasi-totalité des scènes présentées pour des images reconstruites avec moins de 2 % des neurones (Figure 2.4) (données de thèse non publiées). L'analogie de cette courbe du taux de réponse en fonction du pourcentage de décharge avec notre courbe de précision en fonction du SOA dans l'expérience de masquage reste bien sûr anecdotique, mais il est intéressant de replacer ces expériences dans un contexte où l'information disponible s'accumule au cours du temps. Ces deux courbes, l'une basée sur des données expérimentales, l'autre basée sur un modèle de codage par rang, montrent qu'il est possible d'obtenir de bonnes performances après des délais de traitements très courts.

Figure 2.4 (page précédente): *Modèle de codage par rang appliqué à la rétine.* (A) Chacune des scènes naturelles est filtrée dans un modèle de rétine calculant l'intensité des contrastes locaux à 9 échelles spatiales et 2 polarités. Ces filtres sont équivalents à des différences de gaussiennes qui effectuent une sommation spatiale linéaire entre le centre et le pourtour. Les échelles spatiales utilisées vont de 1 à 64 (1, 2, 4, 8, 16, 32, 64, 128, 256). Pour traiter les différentes échelles spatiales, l'image est redimensionnée au facteur d'échelle correspondant (par exemple, l'image est divisée par 4 pour l'échelle 4). Un filtre OFF-center est une simple inversion du filtre ON-center. L'application des filtres à chaque position et échelle spatiale donne un ensemble de valeurs correspondant aux niveaux d'activation des cellules ganglionnaires du modèle. Au total, 1 400 000 neurones sont simulés pour représenter la sortie de ses filtres rétinien. (B) Dans le modèle de codage par rang, les neurones agissent comme des convertisseurs d'un format analogique à un format temporel. Chaque neurone se voit attribuer un rang : le plus activé décharge en premier et ainsi de suite. Dans cet exemple, l'ordre de décharge serait : B-A-H-C-G-D-F-E. Ainsi avec seulement 8 neurones, plus de 40 000 combinaisons sont possibles. (C) Exemple d'une scène cible et distracteur à différents niveaux de reconstruction (le pourcentage de neurones qui ont déchargé un potentiel d'action unique est indiqué dans chaque image). A chaque fois qu'un potentiel est reçu d'une cellule ganglionnaire, le filtre à la polarité et l'échelle correspondante est ajouté à la reconstruction à l'emplacement adéquat, multiplié par le contraste moyen au rang de décharge correspondant. (D) Pourcentage de réponses obtenues selon le nombre de cellules qui ont déchargé pour reconstruire une version donnée de l'image initiale. Quatre sujets ont participé au protocole comportemental, dans lequel 800 images cible et 200 images distracteur sont présentés au premier stade de reconstruction. Les sujets font défiler les images par quantité croissante de neurones qui déchargent jusqu'à répondre quand ils voient un animal. Lorsqu'ils jugent qu'il n'y a pas d'animal, les sujets font défiler toute la séquence de reconstruction avant de passer à l'essai suivant. Les barres donnent la répartition des réponses selon le pourcentage de neurones, et la ligne indique le cumul de ces réponses. Avec moins de 2% de neurones, la quasi-totalité des images est correctement catégorisée. Données adaptées de Delorme, Richard & Fabre Thorpe, 1999; Guyonneau, 2006; VanRullen, Gautrais, Delorme & Thorpe, 1998; VanRullen & Thorpe, 2001a; VanRullen & Thorpe, 2001b.

Afin de tester si les interférences ont lieu au niveau de la rétine/CGL ou après, il serait d'ailleurs intéressant d'envisager une nouvelle expérience dans laquelle les images cibles sont présentées à un œil et le masque est présenté soit à ce même œil soit à l'autre œil. En comparant les courbes de SOA dans ces deux conditions, il devrait être possible de dire si les interférences rétinien ont un rôle important, puisque qu'elles ne peuvent avoir lieu que dans le cas où la cible et le masque sont présentés au même œil... (Affaire à suivre !)

Il faut cependant rajouter que d'autres mécanismes peuvent être envisagés comme ceux définis par Breitmeyer entre le signal magnocellulaire du masque et le signal parvocellulaire de la cible (Breitmeyer, 1984; Breitmeyer & Ganz, 1976), bien que de précédents travaux utilisant cette tâche de catégorisation proposent un rôle majeur du canal magnocellulaire dans le transfert rapide d'informations grossières sur l'image naturelle mais suffisantes pour répondre dans la tâche (Delorme, Richard & Fabre Thorpe, 1999; Macé, Thorpe & Fabre-Thorpe, 2005). Enfin, il faut aussi penser que certains mécanismes itératifs impliqués dans les processus de segmentation (Hupe et al., 1998; Lamme, Super & Spekreijse, 1998), sont probablement rendus difficiles par la présence des informations du masque à la place de celles du stimulus sur les aires de projection.

2.4.4 Transmission continue ou discrète des informations ?

L'augmentation très progressive de la performance avec l'augmentation du SOA révèle une fonction de masquage particulièrement intéressante. L'absence de seuil sur la fonction de masquage suggère que les informations non seulement s'accumulent progressivement, mais de manière continue tout au long de la fonction SOA. Cette remarque fait référence à la manière dont sont transmises les informations tout au long du parcours sensori-moteur. En effet, plusieurs modèles de transmission de l'information peuvent être envisagés (Bullier & Nowak, 1995; Eriksen & Schultz, 1979; McClelland, 1979; Ratcliff, 1988). D'un côté, le modèle sériel soutient l'idée que les informations sont traitées de manière complètement discrète le long du parcours cortical, ce qui implique qu'à chaque étape, les traitements doivent être terminés avant d'envoyer les informations aux aires suivantes. Le versant opposé de ce fonctionnement prône un modèle continu dans lequel les informations incomplètes sont transmises d'une aire à l'autre très rapidement, et sont complétées au fur et à mesure que les traitements s'effectuent. Cette opposition très forte entre les deux modèles n'a pas fini de faire couler de l'encre. Pendant des années, les données expérimentales ont appuyé le modèle sériel à travers différents types de protocoles comportementaux, comme la méthode de soustraction de Donders ou la méthode des facteurs additifs de Sternberg (Donders, 1969; Sternberg, 1969)(voir aussi figure 1.5). Dans ces protocoles, on considère que si la manipulation de deux facteurs a un effet additif sur le temps de réaction, ces facteurs interviennent à des étapes différentes de traitement. Si au contraire l'augmentation des temps de réaction pour ces deux facteurs est inférieure à l'augmentation des temps de réaction pour chacun des deux facteurs pris isolément, alors les étapes de traitement régissant ces deux facteurs sont en partie communes. Même si ces travaux ont permis d'avancer considérablement nos connaissances sur les différentes étapes qui peuvent constituer la boucle sensori-motrice, l'approche comportementale reste néanmoins limitée à cause des hypothèses qu'elle pose au départ, qui justement ne permettent pas d'envisager l'existence d'une transmission autre que sérielle. Parallèlement, les études sur la physiologie des systèmes sensoriels et moteurs ont mis en évidence l'incroyable dynamique de l'architecture corticale et la possibilité d'une transmission purement discrète de l'information est apparue incompatible avec ce type d'organisation cérébrale (Miller, 1988). Dès la fin des années 70, Eriksen et Schultz avaient déjà posé un cadre théorique dit de « flux continu » pour expliquer les données obtenues dans une tâche de focalisation attentionnelle (Eriksen & Eriksen, 1974; Eriksen & Schultz, 1979). Dans cette tâche, les sujets devaient répondre rapidement (parmi deux réponses possibles) à un stimulus présenté au centre de l'écran et entouré de distracteurs dont les propriétés différaient ou non de la cible. Les TR enregistrés étaient plus courts lorsque les distracteurs étaient similaires à la cible (condition « congruente ») que lorsqu'ils appelaient la réponse alternative (condition « incongruente »). D'autre part, la diminution de la distance cible – distracteur augmentait d'autant l'effet perturbateur. Eriksen et Schultz ont conclu de cette série de travaux que les distracteurs activent automatiquement une réponse qui, dans la condition incongruente, entrent en compétition avec la réponse correcte. Ils ont alors développé un modèle dans lequel les informations sont transmises des étapes sensorielles aux étapes motrices de manière unidirectionnelle.

L'information accumulée dans le système sensoriel est transmise en continu au système moteur, qui au départ active le réseau de préparation motrice sur l'ensemble des réponses possibles, puis au fur et à mesure que l'information sensorielle se précise, engage des processus de sélection plus précis. Dans ce modèle, une réponse est émise lorsque le niveau d'activation de la région motrice atteint un seuil. Il est étonnant de voir à quel point ce modèle se recadre bien dans les travaux électrophysiologiques cités ci-dessus. Comme nous l'avons vu précédemment avec les études de Newsome, Shadlen et Schall, de nombreuses données expérimentales récentes suggèrent que les différentes étapes de traitement travaillent en parallèle, permettant ainsi aux structures impliquées dans les processus décisionnels d'accumuler de l'information au fur et à mesure que les structures sensorielles réunissent des indices pertinents sur une cible (Newsome, Britten & Movshon, 1989; Schall, 2001; Shadlen, Britten, Newsome & Movshon, 1996; Shadlen & Newsome, 2001). Les modèles sous-jacents à ces travaux électrophysiologiques ne sont pas compatibles avec une transmission strictement sérielle de l'information.

Alors, doit-on pencher en faveur d'une transmission discrète ou d'une transmission continue des informations dans la boucle sensori-motrice ? L'organisation fonctionnelle des différentes structures corticales –au moins entre les parties sensorielle et motrice – implique obligatoirement que l'information doit d'abord passer par une étape avant d'atteindre l'autre. En contrepartie, il paraît maintenant évident que les structures motrices commencent à travailler avant que les traitements perceptuels ne soient achevés, justifiant une organisation en pipeline⁵. Le modèle en cascade de McClelland intègre bien ces différentes contraintes, puisqu'il suppose que les flux d'informations se recouvrent dans le système, et lorsque l'activation à une étape donnée dépasse un certain seuil, alors l'information est transmise aux étapes suivantes (Ashby, 1982; McClelland, 1979). Finalement, tout porte sur la notion de granularité de l'information qui est transmise d'une étape à l'autre (Hasbroucq et al., 2002; Miller, 1988; Mouret & Hasbroucq, 2000). Alors que dans les modèles sériels et strictement discrets, la taille du grain représenterait la totalité de l'information qui peut être intégrée à une étape donnée, les modèles parallèles et strictement continus supposent que la taille des grains est infinitésimale. Dans les deux cas, on est confrontée à la réalité biologique des structures neuronales qui semble imposer un fonctionnement à la fois continu et sériel. Si le recouvrement important des gammes de latence d'une étape à l'autre confirme un fonctionnement en pipeline, l'intégration des informations et la nature des transmissions synaptiques requièrent un délai dédié pour chaque étape de traitement (Bullier & Nowak, 1995). Cependant dans ce dernier cas, des travaux comme ceux de Rolls et Tovée montrant que la majorité de l'information est contenue dans les 30 premières millisecondes de la décharge neuronale conforte l'idée que les premiers grains d'information, même s'ils sont petits, permet de véhiculer des indices qualitativement importants.

⁵ J'emploie le mot pipeline au sens oléoduc du terme, et non pas dans le sens informatique utilisé pour les processeurs pipeline. Explicitement, les étapes de traitement fonctionnent en pipeline lorsqu'une partie de l'information est envoyée de l'étape A à l'étape B, alors que l'étape A reçoit et traite toujours des informations nouvelles.

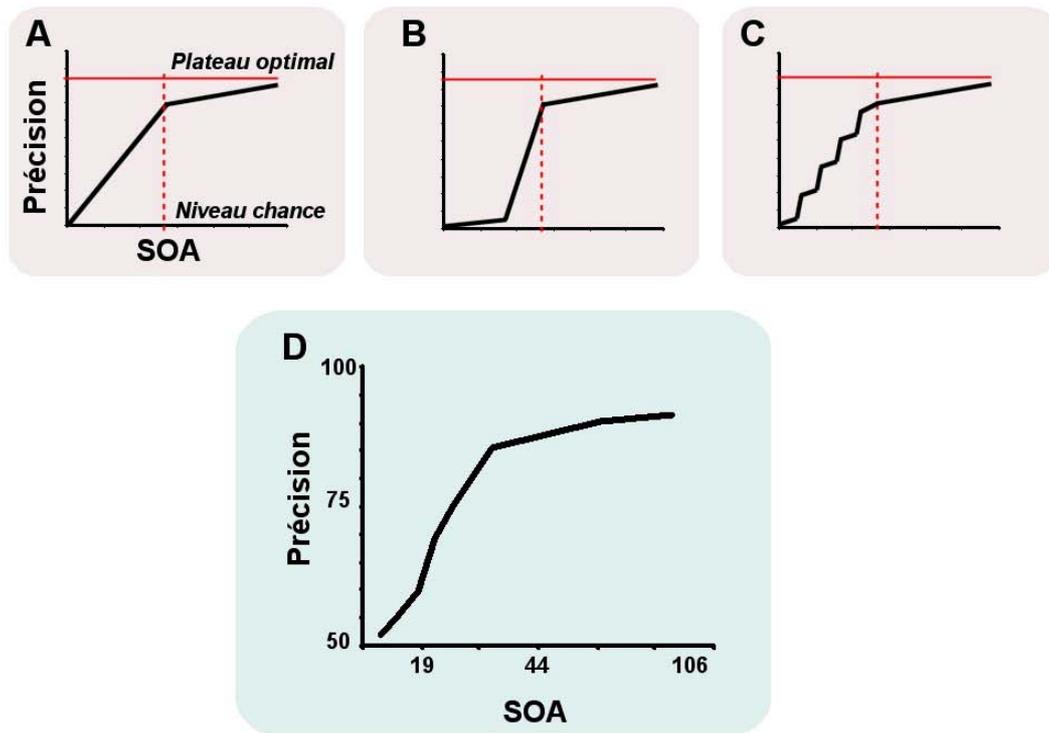


Figure 2.5 : Courbes de précision en fonction du SOA, selon différents modèles de transmission de l'information. (A) Dans un modèle strictement continu, la précision doit augmenter linéairement avec l'augmentation de la quantité d'information jusqu'à obtenir des performances optimales. (B) et (C) Dans les modèles discrets, les courbes de précision doivent marquer des plateaux transitoires selon le nombre d'étapes limitantes dans le parcours des informations. (D) La fonction de masquage obtenue dans l'expérience présentée montre une augmentation presque linéaire de la précision avec le SOA. La légère inflexion observée pour le deuxième pas de SOA à 19 ms, pourrait néanmoins indiquer une légère transition discrète.

Comment interpréter nos données par rapport au mode de transmission des informations dans le système ? L'aspect très régulier de la courbe de précision en fonction du SOA dans notre expérience de catégorisation d'images masquées confirmerait un fonctionnement parallèle et continu (Figure 2.5). Cependant, la légère inflexion observée pour le SOA inférieurs à 20 ms pourrait suggérer l'existence d'une phase de traitement demandant un minimum de processus itératif. Comme nous le verrons dans le chapitre suivant, cette inflexion est encore plus marquée dans une expérience utilisant un mode de réponse un peu différent, où les sujets doivent répondre à la fois lorsqu'ils pensent avoir vu une cible et lorsqu'il n'en ont pas vu (réponses de type yes/no et non pas go/no-go). Même si cette phase légèrement limitante peut se localiser à différents endroits du parcours des informations dans le système visuel, il peut être intéressant de rappeler que justement, les premières réponses qui sortent de la rétine ne se font pas avant 15-20 ms, ce qui rejoint l'idée développée dans la section précédents, suggérant que les interférences de masquage pourraient être majeures dans cette région, du à des mécanismes de compression importants.

Cependant, il est possible que le moyennage de la précision sur la totalité des essais et des sujets puissent cacher des transitions de type discret (Miller, 1988 #92, merci à Thierry Hasbroucq d'avoir soulevé ce point). Finalement, si nos données ne permettent pas de trancher en faveur d'un modèle strictement continu, les performances remarquables obtenues dans des conditions de perception limitée appuient fortement l'idée que l'information est véhiculée d'une étape à l'autre selon une architecture en pipeline; résultat qui se confirmera d'autant plus dans le chapitre suivant.

Conclusion

Cette première expérience de masquage a permis de confirmer à quel point le système visuel est efficace pour effectuer une tâche complexe de catégorisation. Malgré des interférences massives intervenant relativement tôt dans la hiérarchie visuelle (avant V4 ?), une quarantaine de millisecondes suffisent pour intégrer suffisamment d'informations et répondre correctement à la tâche. D'autre part avec ce paradigme, nous disposons d'un outil efficace pour étudier la dynamique temporelle des traitements, nous permettant d'aborder des questions complémentaires concernant l'organisation spatio-temporelle du transfert de l'information.

Catégorisation visuelle : Traitements sensoriels, décisionnels et moteurs

3

3.1 Tâches manuelle et saccadique

Nous avons vu dans le chapitre précédent que la tâche de catégorisation devait reposer sur une intégration rapide de l'information au niveau de l'analyse sensorielle puisque dès 40 ms d'intervalle entre l'image et le masque, les performances atteignent un plateau. Ces résultats sont donc en accord avec les données électrophysiologiques chez le singe montrant que les 30-40 premières millisecondes de la réponse neuronale dans les aires visuelles supérieures sont suffisantes pour intégrer la majorité de l'information. Cependant, les performances élevées au dessus de 40 ms de SOA n'indiquent pas seulement que l'analyse sensorielle a été épargnée par les effets perturbateurs du masquage, elles indiquent aussi que les étapes ultérieures des traitements ont pu s'effectuer de manière optimisée, en favorisant une prise de décision efficace et des traitements moteurs adaptés. Même s'il paraît de plus en plus clair que l'enchaînement de ces différentes étapes ne se fait pas de manière strictement sérielle mais fonctionne d'avantage sur un mode parallèle et asynchrone (Humphreys, Riddoch & Price, 1997; McClelland, 1979; McClelland & Rogers, 2003; Rumelhart, Hinton & McClelland, 1986), il apparaît néanmoins que les traitements perceptuels, décisionnels et moteurs doivent avoir lieu dans des structures anatomiques différentes. Connaissant les contraintes temporelles dans ce type de tâche, une question se pose alors sur la manière dont les informations sont véhiculées tout au long de ce parcours et sur l'impact que peut avoir la réduction de l'information disponible à l'entrée de la boucle sur les traitements ultérieurs.

Ces questions ont eu d'autant plus d'importance après avoir observé les résultats d'une nouvelle tâche de reconnaissance d'images naturelles, développée par l'équipe et en particulier par Holle Kirchner dès 2004. L'idée de départ était d'observer la faisabilité des traitements de catégorisation ultra-rapide en utilisant des réponses saccadiques à la place des réponses

manuelles. L'effecteur oculaire est connu pour être plus rapide que l'effecteur manuel, entre autres grâce aux distances neuronales plus courtes entre les systèmes sensoriel et moteur (Busettoni, Masson & Miles, 1997; Masson, Rybarczyk, Castet & Mestre, 2000). Or la rapidité des réponses comportementales manuelles suppose que les traitements visuels sont optimisés, et il n'était pas évident que l'utilisation d'un effecteur moteur plus rapide garantisse le même niveau de précision. Le protocole de catégorisation a donc été adapté pour que les réponses puissent être générées avec les yeux ce qui a impliqué quelques changements dans la nature même de la tâche. En effet, les réponses saccadiques sont biologiquement adaptées pour effectuer des réponses de sélection entre plusieurs choix et d'autant plus efficaces que le déplacement du regard se fait vers une cible. La tâche de catégorisation manuelle est de type go/no-go (le sujet doit soulever la main lorsque l'image flashée contient un animal : go, et ne doit pas bouger pas dans le cas contraire : no-go). Or il est particulièrement difficile d'inhiber une réponse oculaire dans le cas d'une non-réponse (condition no-go), ce qui se traduit en général par une augmentation très nette des temps de réaction sur la condition go (Juttner & Wolf, 1992; Kirchner & Colonius, 2005). D'ailleurs, les expériences pilotes ont même montré que les sujets, et notamment ceux qui sont experts dans la tâche manuelle, n'arrivaient quasiment pas à effectuer la tâche car leur regard partait sur le stimulus de manière presque automatique, indépendamment du statut cible ou distracteur. Le nouveau protocole a donc nécessité l'affichage de deux images en périphérie de part et d'autre de la croix de fixation, l'une contenant une cible et l'autre non. Rappelons que Guillaume Rousselet avait montré dans une tâche go/no-go que les performances étaient similaires qu'il y ait une ou deux images présentées sur l'écran (Rousselet, Fabre-Thorpe & Thorpe, 2002). Dans le cas de cette nouvelle tâche oculaire, le sujet devait ainsi effectuer une saccade du côté où se trouvait la cible. Il est important de noter que les traitements visuels engagés dans les deux tâches ne sont pas strictement identiques et dans le cas de la réponse manuelle on parlera d'une tâche de go/no-go, alors que dans le cas de la réponse saccadique, il est plus juste de parler de tâche de choix forcé.

En ce qui concerne les résultats expérimentaux (Figure 3.1), non seulement la précision reste très bonne dans la tâche de choix saccadique (90,1% dans la tâche de choix saccadique vs. 95,3% dans la tâche go/no-go manuelle), mais les temps de réaction saccadiques sont presque deux fois plus courts (valeurs médianes : 228 ms vs. 419 ms), avec les premières réponses significativement correctes apparaissant dès 120 ms (Kirchner & Thorpe, 2005). L'observation de ces résultats a soulevé un certain nombre de questions. Si l'on considère qu'un délai de 20-25 ms est nécessaire pour déclencher le mouvement moteur au niveau du circuit du tronc cérébral, il reste 95-100 ms de latence pour effectuer les traitements purement sensoriels et décisionnels. Un délai aussi court pose des contraintes strictes, à la fois sur :

- la nature des informations qui ont pu être intégrées en un temps aussi court,
- le niveau de traitement sensoriel qu'il est nécessaire d'accomplir pour répondre correctement à la tâche,
- les circuits neuronaux utilisés entre systèmes sensoriel et moteur,
- la manière optimisée dont s'enchaînent les différentes étapes de traitement : analyse sensorielle, processus décisionnels et déclenchement de la réponse motrice.

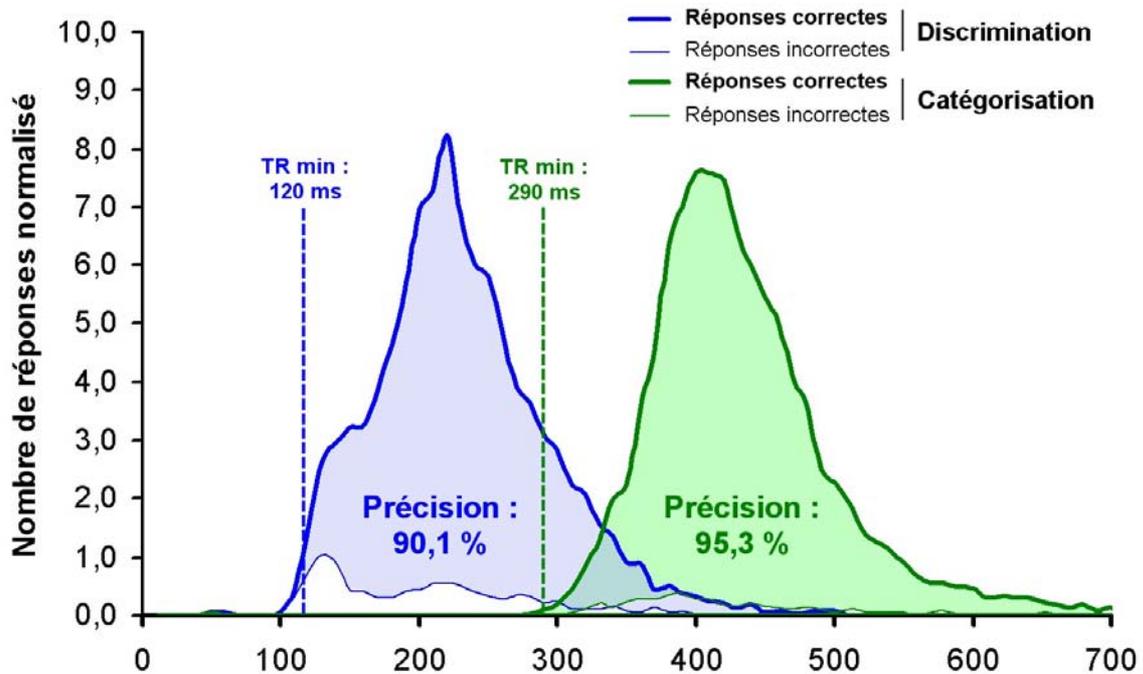


Figure 3.1 : Distribution des temps de réaction par intervalle de 10 ms dans les tâches de catégorisation manuelle et de discrimination saccadique. Les résultats de la tâche de catégorisation portent sur 12 sujets (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001) et ceux de la tâche de discrimination portent sur 15 sujets (Kirchner & Thorpe, 2005). Dans les deux tâches, les images étaient présentées une vingtaine de millisecondes, et les cibles et distracteurs étaient équilibrés dans chaque série. Dans la tâche de discrimination, les deux images étaient présentées à 6° de part et d'autre de la croix de fixation. Les temps de réaction minimaux (TR min) sont déterminés par un test du χ^2 afin d'évaluer la latence minimale à partir de laquelle les réponses sur les cibles sont significativement plus nombreuses que les réponses sur les distracteurs.

Les traitements visuels dans la tâche de choix saccadique reposent-ils (uniquement) sur des indices bas-niveau ?

Dans la tâche de choix saccadique, bien que les images aient été équilibrées en luminance moyenne, il n'était pas exclu que les sujets puissent s'appuyer sur un biais d'indices bas-niveau entre les catégories d'images cible et distracteur. En effet il apparaît que le calcul de certains indices statistiques du premier et second ordre semblent différer significativement entre les deux catégories d'images (Figure 3.2 et voir Kirchner & Thorpe, 2005). Cependant lorsque l'on enlève du set complet 1 ou 2 images présentant des valeurs extrêmes pour un test donné, aucun indice isolé ni aucune combinaison d'indices ne semblent pouvoir dissocier les groupes d'images cible et distracteur en terme de performance.

Figure 3.2 (page suivante): Tests statistiques des indices bas-niveau calculés sur les groupes d'images cibles et distracteurs. Les statistiques de premier ordre sont basées sur la distribution des valeurs en niveau de gris des images. Les statistiques de second ordre sont basées sur les matrices de co-occurrences des valeurs en niveau de gris des images pour des pixels adjacents dans les quatre directions de l'image (0°, 45°, 90° et 135°). Enfin les indices d'amplitude spectrale sont calculés grâce à des courbes d'ajustement sur les composantes horizontale et verticale de la transformée de Fourier de l'image. La première colonne donne la valeur moyenne pour chaque catégorie d'images, suivie par la déviation standard à la moyenne (s.e.m.). Le calcul du t-test, de ses degrés de liberté (df) et sa probabilité exacte ($p < 0,05$) indique si les groupes d'images cibles et distracteurs diffèrent significativement. Si c'est le cas, les N images aux valeurs extrêmes sont enlevées de chaque catégorie, et les temps de réaction médian (ms) et précision (%) sont calculés sur le set d'images restant. (adapté de KIRCHNER, H & THORPE, S.J., 2005).

	Moyenne	s.e.m.	t-test	df	p<0,05	N	TR median	Précision
Premier ordre								
Moyenne	91.20	1.42	-0.29	38.0	n.s.			
Distracteur	90.60	1.49						
Variance	2239.42	221.32	2.83	32.7	0.008	2	230	90.5
Distracteur	3382.62	338.39						
Signal/Bruit	2.06	0.10	-2.04	38.0	0.048	1	229	90.8
Distracteur	1.73	0.12						
Ecart-type	0.51	0.03	2.64	38.0	0.012	2	229	90.5
Distracteur	0.63	0.04						
Skewness	0.36	0.11	0.61	38.0	n.s.			
Distracteur	0.46	0.12						
Kurtosis	3.07	0.19	-0.64	38.0	n.s.			
Distracteur	2.90	0.20						
Second ordre								
Energie	0.0006	0.0006	2.68	28.9	0.012	2	229	91.0
Distracteur	0.0013	0.0011						
Contraste	321.57	176.89	3.12	29.1	0.004	3	230	90.9
Distracteur	582.70	329.66						
Entropie	8.48	0.54	0.16	38.0	n.s.			
Distracteur	8.51	0.66						
Homogénéité	0.17	0.06	1.41	38.0	n.s.			
Distracteur	0.21	0.08						
Corrélation	0.49	0.13	2.01	38.0	0.052	1	229	90.8
Distracteur	0.59	0.16						
Amplitude spectrale directionnelle								
Horizontale	1.43	0.03	2.85	38.0	0.007	2	229	90.8
Distracteur	1.27	0.05						
Verticale	1.46	0.04	2.9	38.0	0.006	2	229	91.1
Distracteur	1.31	0.04				10	230	90.2

Peut-on parler de niveau de traitement sensoriel minimal ?

Dans la tâche de choix saccadique, nous venons de voir que les informations sur l'image permettant de dissocier une cible d'un distracteur ne semblent pas reposer sur des indices simples. Cependant, un TR minimal aussi court que 120 ms doit obligatoirement impliquer des traitements sensoriels limités. D'autre part, il paraîtrait logique de conclure qu'une diminution si nette des temps de réaction entre la tâche go/no-go manuelle et la tâche de choix saccadique suppose que le niveau d'analyse doit être moins élevé dans la deuxième tâche, et que l'information extraite à un niveau précoce de la hiérarchie corticale court-circuite le parcours nerveux pour déclencher des réponses rapides. Cependant, cette hypothèse se place dans une vision sérielle de l'architecture du système visuel, et en plus des données de la littérature qui ont montré les limites du modèle sériel, l'expérience de masquage décrite dans le chapitre 2 suggère plutôt que les traitements impliqués dans la tâche de go/no-go manuelle sont effectués en parallèle et transférés selon un mode pipeline. Finalement, plutôt que de parler de traitement sensoriel minimal, il paraît plus juste de se demander comment les informations sensorielles les plus saillantes sur le stimulus permettent de déclencher une réponse motrice, et où se situent les différences entre les deux protocoles décrits précédemment. En d'autres termes, il s'agit de se demander comment se fait le lien entre l'analyse sensorielle et l'action dans les deux types de protocole, et plus précisément comment s'intègrent les différents processus liés à la prise de décision.

Quels circuits neuronaux peuvent être raisonnablement envisagés dans les deux types de tâches ?

Le TR minimal de 120 ms dans la tâche de choix saccadique pose des contraintes sévères sur le parcours emprunté par les informations visuelles tout au long de la hiérarchie corticale. La grande différence de temps de réaction entre les deux tâches doit probablement pouvoir être expliqué, du moins en partie, par des circuits neuronaux différents (Figure 3.3). Comme nous l'avons vu dans le chapitre 1, les réponses oculaires font intervenir des structures comme le FEF et le CS, qui sont eux-même directement connectés au circuit oculaire du tronc cérébral. En contrepartie, les informations permettant de déclencher les réponses manuelles doivent emprunter un parcours plus long, faisant intervenir le cortex prémoteur et moteur pour finalement atteindre les muscles du bras et de la main via la moelle épinière. Dans les deux tâches, les chemins parcourus semblent être optimisés puisque les temps de réaction enregistrés lors de la présentation d'images nouvelles ne diffèrent pas des temps de réaction obtenus sur des images familières (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001; Kirchner & Thorpe, 2005). On peut donc penser que cette différence importante de temps de réaction entre les deux tâches est exclusivement due à un parcours plus long des informations pour déclencher les réponses manuelles. Or il est important de préciser que les tâches diffèrent aussi dans la nature même de la question posée aux sujets. Dans le cas du go/no-go manuel, le sujet doit déterminer si oui ou non l'image présentée contient un animal (la question est donc « Est-ce qu'il y a un animal ? »), alors que pour le choix forcé saccadique, les sujets savent à l'avance qu'un animal est présent dans une des deux images mais il doit indiquer dans quelle scène l'animal apparaît (la question est donc « Est-ce que l'animal est à gauche ou à droite ? »). On ne peut donc pas exclure que des différences apparaissent aussi au niveau de

l'analyse visuelle ou de la prise de décision. Une question centrale porte sur la manière dont les informations visuelles sont transmises aux régions motrices, et en particulier sur les structures qui sont impliquées pour ces différents traitements.

La tâche de choix saccadique nécessite implicitement des processus de localisation, puisque la saccade doit être effectuée à gauche ou à droite vers l'image cible. Or il apparaît très clairement dans la littérature une dissociation des voies visuelles empruntées selon que la tâche requière une reconnaissance d'objet (voie ventrale) ou une action dirigée vers un objet (voie dorsale)(Goodale & Milner, 1992; Ungerleider, 1985; Ungerleider & Haxby, 1994). Les réponses neuronales de la voie dorsale sont aussi connues pour avoir des latences très courtes, bien plus que celles de la voie ventrale, ou les premières réponses apparaissent après 70 ms dans IT. La rapidité des réponses dans la tâche de choix oculaire peut-elle s'expliquer par l'utilisation de la voie dorsale (Figure 3.3 A) ?

Comme nous l'avons vu précédemment, l'absence de corrélation entre la performance et les indices statistiques simples sur les images suggère que l'analyse visuelle permettant d'intégrer suffisamment d'informations pertinentes pour répondre à la tâche va au-delà des informations disponibles dans V1. La tâche de choix saccadique n'engage pas seulement des processus de localisation mais semble requérir des traitements de reconnaissance d'images relativement avancés. L'utilisation de la voie dorsale ne pourrait donc se faire qu'à partir des informations intégrées dans certaines aires extrastriées ventrales, probablement jusque V4 au moins. En revanche le transfert des informations au delà de cette aire reste d'avantage hypothétique.

Il faut souligner que les cellules du FEF peuvent avoir des latences de décharge particulièrement courtes, parfois dès 40 ms en réponse à des flashes lumineux (Nowak & Bullier, 1997; Schmolesky et al., 1998; Thompson, Hanes, Bichot & Schall, 1996). Il semble aussi que certaines cellules de cette région peuvent avoir une composante visuelle (Bichot, Thompson, Chenthal Rao & Schall, 2001; Schall, 2002), mais il paraît improbable que les réponses précoces de ce groupe de neurones puissent engager les traitements d'une tâche aussi complexe que du choix forcé sur des scènes naturelles qui sont des stimuli beaucoup plus complexes que ceux utilisés dans les expériences citées (Figure 3.3B). Une autre voie possible pour transmettre l'information visuelle au FEF implique les circuits sous-corticaux de l'amygdale, connu pour avoir des réponses visuelles aux stimuli portant une valeur émotionnelle (Morris, Ohman & Dolan, 1998; Whalen et al., 1998). Or les scènes naturelles ont peu de valeur émotionnelle, mis à part quelques images particulières comme celles d'araignées et de serpents ... Encore une fois, même si on ne peut pas en exclure une participation indirecte, cette voie n'est pas à privilégier pour expliquer le transfert et l'intégration de propriétés complexes comme ceux engagés dans la tâche de choix saccadique.

De la même manière pour la tâche go/no-go manuelle, le parcours suivi par les informations visuelles pour déclencher la réponse motrice n'est pas encore bien défini. Dans l'article de 2001 publié dans Science (Thorpe & Fabre-Thorpe, 2001), il était proposé que les informations extraites jusque IT sont transmises aux aires prémotrices et motrices par l'intermédiaire du cortex préfrontal (Figure 3.3C). Cette voie est d'autant plus plausible que le cortex préfrontal est connu pour avoir un rôle majeur pour les processus impliqués dans la prise de décision ou dans

l'apprentissage de règles et d'associations (Freedman, Riesenhuber, Poggio & Miller, 2001; Huettel, Song & McCarthy, 2005; Kim & Shadlen, 1999; White & Wise, 1999). Or les travaux de connectivité peuvent laisser envisager une autre possibilité : le système visuel dorsal envoient des projections vers le cortex prémoteur via le cortex pariétal (médial et latéral); projections qui sont bien plus nombreuses que celles qui existent entre les cortex préfrontal et prémoteur (Boussaoud, di Pellegrino & Wise, 1995; He, Dum & Strick, 1995; Lu, Preston & Strick, 1994). Cette voie parallèle est probablement à privilégier pour des traitements visuels impliquant une coordination visuo-motrice (« vision pour l'action ») mais il reste intéressant d'envisager la possibilité d'un transfert de l'information des aires visuelles ventrales aux aires dorsales pour pouvoir rejoindre le cortex prémoteur. Cette issue serait d'autant plus envisageable si l'on montrait que les informations n'ont pas besoin d'être intégrées jusque IT pour les besoins de la tâche...

Dans les travaux qui seront présentés à la suite, nous ne pourrons de toute façon pas répondre entièrement à ces questions, mais il paraissait utile de résumer ici une problématique anatomique réelle, qui demandera irrémédiablement l'utilisation d'outils d'imagerie complémentaires, comme l'IRM fonctionnelle ou la TMS.

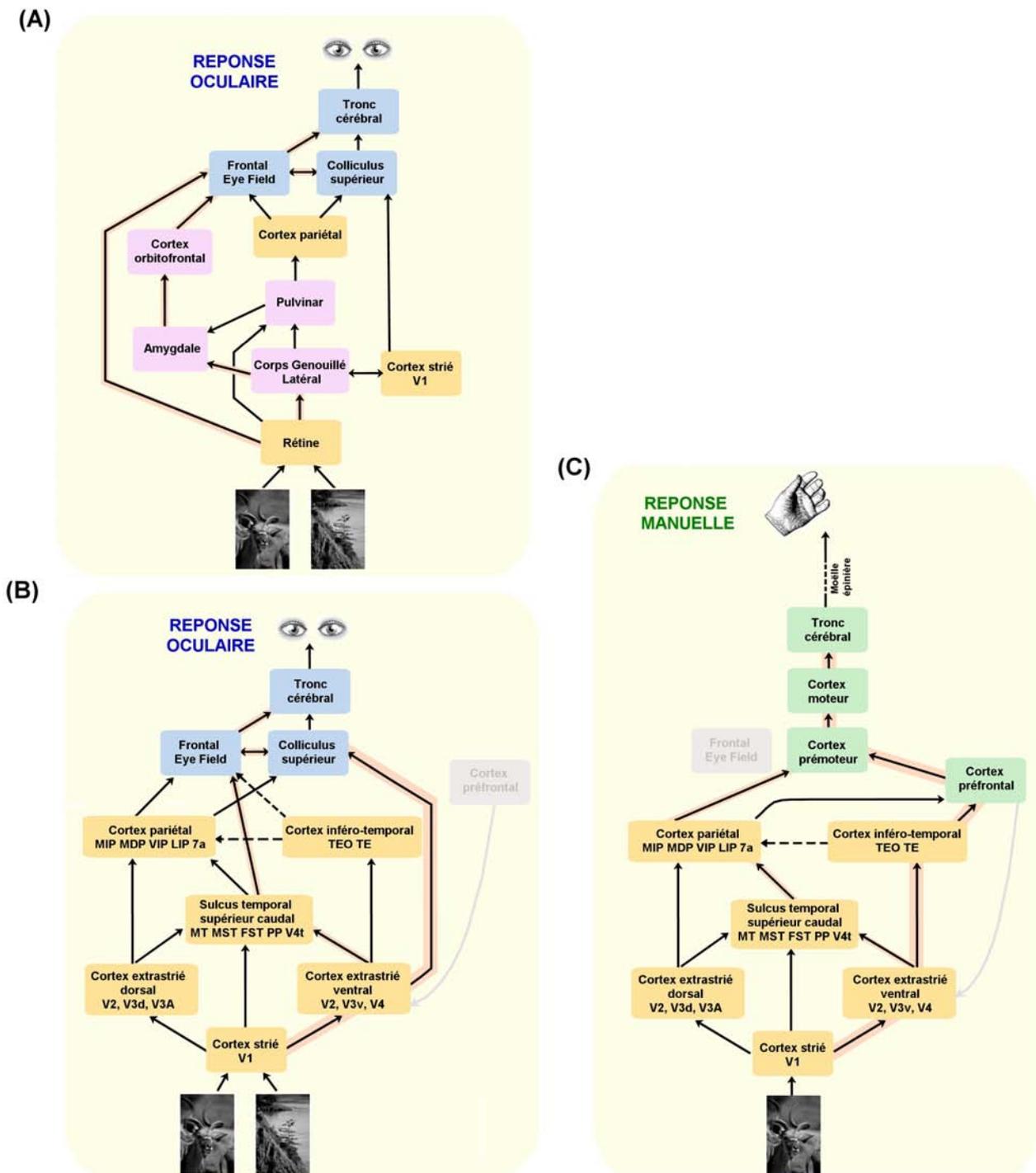


Figure 3.3 : Schémas des connexions entre les aires sensorielles et les aires motrices pour des réponses saccadiques ou manuelles. (A) Schémas des connexions principalement sous-corticales permettant de relier les entrées visuelles rétinienne aux structures impliquées dans le contrôle oculaire. (B) Schéma des connexions entre les différentes structures corticales du système visuel et les aires motrices oculaires. (C) Schéma des connexions entre les différentes structures corticales du système visuel et les aires motrices manuelles. Les flèches pointillées indiquent que les connexions sont peu nombreuses. Les flèches soulignées d'un trait orange indiquent un parcours possible (pour les flèches épaisses, probable) des informations entre les aires sensorielles et motrices. Pour les deux modes de réponses, un problème majeur porte sur l'interaction hypothétique entre les voies ventrales et dorsales pour transmettre l'information le plus rapidement possible vers les aires motrices.

3.2 Objectifs

L'objectif de départ de cette série d'expériences était de pouvoir comparer plus directement les tâches go/no-go manuelle et de choix saccadique en manipulant les paramètres qui les différencient : entrée visuelle, nature de la tâche, mode de réponse et nombre d'effecteurs. L'utilisation du protocole masquage devait permettre de décomposer les traitements au cours du temps afin d'obtenir des profils de masquage à comparer dans les quatre tâches. Cette fois, le phénomène de masquage n'est pas le centre de l'étude mais doit être d'avantage considéré comme un outil psychophysique permettant la comparaison des traitements au cours du temps.

Il faut aussi se replacer dans un contexte où les résultats obtenus dans la tâche de choix saccadique soulevaient beaucoup de questions et laissaient de nombreuses personnes sceptiques sur les conclusions apportées. Cette nouvelle série d'expériences devait permettre d'essayer de répondre à certaines de ces questions, voire de valider un protocole expérimental nouveau, particulièrement efficace pour étudier la reconnaissance d'objet.

3.3 Protocole expérimental

Comme nous l'avons vu dans la première partie, les deux protocoles expérimentaux diffèrent sur plusieurs points :

- A l'entrée de la chaîne de traitement, les tâches diffèrent sur le type d'informations qui entrent dans le système visuel. Dans la tâche go/no-go manuelle, une seule image est flashée au centre de l'écran, alors que deux images périphériques sont présentées dans la tâche de choix saccadique.
- La question à laquelle doivent répondre les sujets doit déterminer l'objectif de l'analyse sensorielle et des traitements décisionnels. Dans la tâche go/no-go, il s'agit de se demander si l'image contient un animal, alors que dans la tâche de choix saccadique, les sujets savent qu'il y a une cible dans une des deux images présentées, mais doivent décider si elle se trouve à gauche ou à droite du point de fixation. Pour cette raison, dans la suite de cette présentation, on parlera de tâche de catégorisation dans le premier cas, et de tâche de discrimination (entre deux images) lorsqu'il s'agit de déterminer laquelle de deux images contient une cible.
- Enfin la sortie motrice diffère évidemment de part la nature de l'effecteur (la main ou les yeux).

Afin de pouvoir comparer ces deux types de tâche, nous les avons reproduit en condition de masquage tout en introduisant deux tâches supplémentaires présentant des caractéristiques intermédiaires (Table 3.1 et Figure 3.4) :

Expérience 1 : Tâche de catégorisation avec des réponses manuelles de type go/no-go. Les caractéristiques de cette tâche sont similaires à celle de la tâche de catégorisation manuelle décrite précédemment. Les données de cette expérience ont été publiées dans Vision Research et développées dans le Chapitre 2.

Expérience 2 : Tâche de catégorisation avec des réponses manuelles de type yes/no. Les caractéristiques sont identiques à la tâche précédente, mais les sujets doivent cette fois répondre avec les index des deux mains appuyés sur les touches Ctrl d'un clavier, en indiquant du côté de leur main dominante si l'image contient un animal, et en répondant de l'autre côté si elle n'en contient pas.

Expérience 3 : Tâche de discrimination avec des réponses saccadiques. Les caractéristiques sont similaires à celles décrites précédemment dans la tâche de discrimination saccadique.

Expérience 4 : Tâche de discrimination avec des réponses manuelles. Les sujets effectuent la même tâche que dans l'expérience 3, mais au lieu d'indiquer par une saccade si l'image contenant un animal est à gauche ou à droite de la croix de fixation, les sujets répondent en levant le doigt des boutons Ctrl d'un clavier, du côté où se trouve la cible.

Les détails expérimentaux sont explicités dans l'article qui suit, en révision dans le journal *Journal of Experimental Psychology: Human Perception and Performance*.

	Expérience 1	Expérience 2	Expérience 3	Expérience 4
Entrée visuelle	1 image centrale	1 image centrale	2 images périphériques	2 images périphériques
Nature de la tâche	Catégorisation	Catégorisation	Discrimination	Discrimination
Type de décision	Go/no-go	Yes/no	Choix forcé	Choix forcé
Type d'effecteur	Main	Mains	Yeux	Mains
Nombre d'effecteurs	1	2	1	2

Table 3.1 : Descriptions des 4 expériences en fonction de leurs caractéristiques visuelles, décisionnelles et motrices.

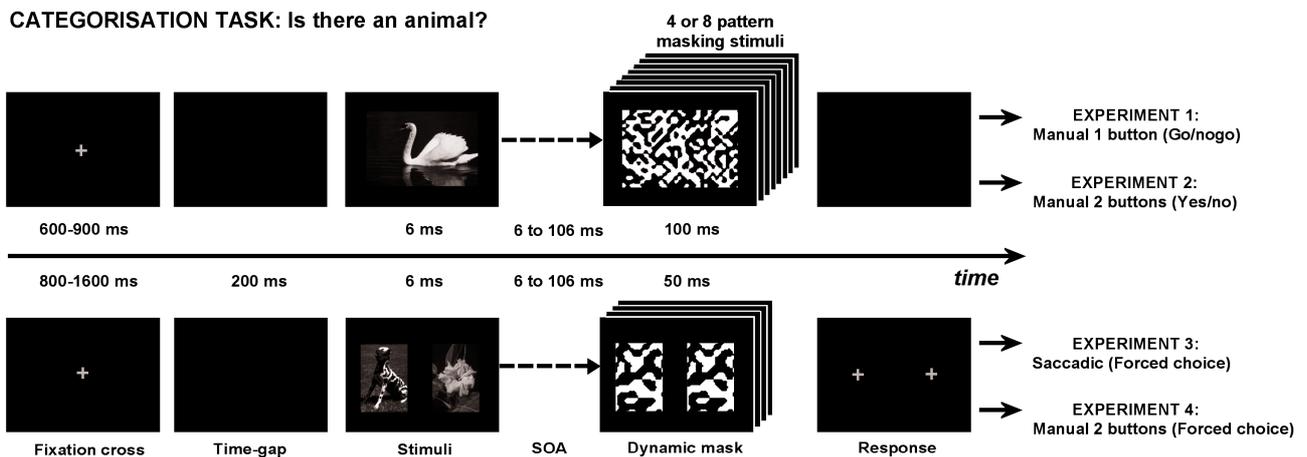


Figure 3.4 : Protocoles expérimentaux dans les 4 expériences. Chaque essai commence par l'affichage d'une croix de fixation pendant une durée variable. Dans trois expériences (excepté Exp. 1), nous avons introduit un intervalle fixe de 200 ms entre la disparition de la croix de fixation et l'affichage de la cible (« gap interval »). Dans les tâches de catégorisation, une image centrale était flashée et suivie d'un masque composé de 8 images contrastées à 100%. Dans les tâches de discrimination, deux images latérales étaient présentées, l'une contenant un animal. En fonction de l'expérience, les sujets devaient répondre manuellement sur un ou deux boutons ou par des saccades oculaires. Dans toutes les tâches, les sujets avaient pour consigne d'être le plus rapide et le plus précis possible. Les mêmes sets d'images naturelles ont été utilisés dans les expériences 2, 3 et 4.

3.4 Résultats

3.4.1 Pourcentage correct

Les analyses ANOVA montrent que le pourcentage correct moyenné sur toutes les conditions de SOA est significativement supérieur dans les tâches de discrimination (Exp. 3 : 76,1% et Exp. 4 : 77%) par rapport aux tâches de catégorisation (Exp. 1 : 72,8% et Exp. 2 : 73,4%). En revanche la précision est indépendante du mode de réponse, qu'elles soient manuelles ou saccadiques.

Le pourcentage correct augmente naturellement avec le SOA. Dans les 4 expériences, il commence au niveau chance pour le premier pas de SOA (6 ms), atteint un plateau autour de 44 ms de SOA, et termine à des valeurs comparables aux données recueillies dans des expériences sans masquage pour les SOA les plus longs (Delorme, Richard & Fabre-Thorpe, 2000).

Dans les tâches de catégorisation, les sujets avaient pour consigne d'essayer de se forcer à répondre positivement sur la moitié des essais de chaque série, puisque les essais cibles et distracteurs étaient équiprobables. Malgré cette consigne, le taux de réponse est fortement inhibé dans les conditions difficiles de SOA, avec seulement 20% de réponses dans la condition 6 ms. De manière intéressante, c'est plus le taux de réponses Missed (réponses manquées sur les cibles) qui est affecté par le masquage que le taux de réponses False Alarms (réponses incorrectes sur les distracteurs). Ce résultat révèle une faiblesse du paramètre « pourcentage correct » pour comparer des tâches qui ne reposent pas sur le même mode de validation de la réponse. Dans les tâches de catégorisation, à la fois une augmentation de False Alarm et de Missed peuvent expliquer une baisse de performance, alors que dans les tâches de discrimination, une diminution de performance s'explique uniquement par une augmentation de False Alarms. Nous avons donc calculé le d' , plus adapté pour la comparaison de ce type de tâches (Green & Swets, 1966; Macmillan & Creelman, 1991).

3.4.2 d'

Dans les tâches de catégorisation, la mesure de la sensibilité est déterminée par le calcul $Z_H - Z_{FA}$, où H est le taux de réponses Hits (réponses correctes sur les cibles) et FA le taux de False Alarms (Macmillan & Creelman, 1991). Dans les tâches de choix-forcé (discrimination), ce calcul doit prendre en compte la somme des variabilités des deux distributions (Macmillan, 2002), ce qui conduit à ajouter un facteur de correction tel que le d' se mesure par $(1/\sqrt{2})(Z_H - Z_{FA})$.

De la même manière qu'avec le calcul du pourcentage correct, la mesure du d' indique que les quatre expériences ne diffèrent pas selon le mode de réponse, mais selon la nature de la tâche, avec une valeur moyenne de d' inférieure pour les tâches de discrimination par rapport aux tâches de catégorisation. Mais il est surtout intéressant de voir que ces comparaisons de sensibilité observent deux phases en fonction du SOA : pour les SOA courts en dessous de 31 ms, les mesures de d' ne sont pas significativement différentes entre les quatre expériences. En revanche au-delà de ce SOA, des différences apparaissent : les tâches de catégorisation manuelle (Expériences 1 et 2) produisent des valeurs significativement meilleures que celles obtenues dans les tâches de discrimination (Expérience 3 et 4). D'autre part, il est intéressant d'observer que

pour ces deux dernières tâches, la mesure du d' ne diffère sur aucun point de SOA, que les réponses soient données manuellement ou par des saccades oculaires.

3.4.3 Temps de réaction

A la fois le mode de réponse (manuel ou oculaire) et la nature de la tâche ont un impact important sur les temps de réaction. Dans les trois tâches manuelles, le TR moyen dépasse les 400 ms, ce qui est largement supérieur au TR moyen enregistré dans la tâche oculaire, de seulement 250 ms. D'autre part, les tâches de discrimination permettent de générer des réponses plus courtes (TR moyen Exp. 3 : 250 ms, Exp.4 : 404 ms) que les tâches de catégorisation (Exp. 1 : 422 ms, Exp. 2 : 477 ms).

En revanche les temps de réaction ne varient pas beaucoup avec le SOA, malgré la grande variabilité inter-sujet. Une légère augmentation de TR peut tout de même être observée dans les tâches de catégorisation entre les conditions de SOA les plus faciles et les plus difficiles.

3.4.4 Compromis précision-vitesse

Il est courant d'observer que les réponses rapides induisent une précision moins élevée que celles obtenues avec des temps de réaction plus longs (Fitts, 1954; Meyer, Osman, Irwin & Yantis, 1988). Nous avons déjà observé ce phénomène de compromis rapidité-précision (speed-accuracy trade-off) dans différentes tâches de l'équipe (Kirchner & Thorpe, 2005; Thorpe, Fize & Marlot, 1996). Cependant, avec une procédure de masquage, on peut se demander si le compromis rapidité-précision est toujours présent, puisqu'il peut paraître inutile de retarder le déclenchement de la réponse motrice lorsque les informations disponibles sur le stimulus sont bloquées par le masquage.

Nous avons donc calculé la corrélation entre les temps de réaction et la précision dans les quatre expériences, en considérant la moyenne de ces paramètres pour chaque sujet. Dans les tâches de discrimination, un compromis rapidité-précision est observé pour des SOA longs, supérieurs à 44 ms, mais rien n'est visible avant cette valeur. Dans la tâche de catégorisation yes/no, la même tendance est observée, bien que le compromis soit moins marqué après 44 ms. En revanche dans la tâche de catégorisation go/no-go, aucun compromis rapidité-précision n'est observé, quel que soit le SOA. Ces résultats suggèrent que lorsque les sujets ont à choisir entre deux alternatives de réponses ('gauche' ou 'droite', 'oui' ou 'non'), le système visuel peut augmenter la précision en retardant le déclenchement de la réponse, mais ceci reste vrai à condition que le masque interrompe les traitements relativement tard, autorisant ainsi le système à travailler sur une quantité d'informations suffisante. Pour des SOA plus courts, il ne semble pas utile de retarder la réponse, probablement puisque toute l'information disponible a déjà été intégrée.

ARTICLE 2

Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms

Bacon-Macé N, Kirchner H, Fabre-Thorpe M & Thorpe SJ
Journal of Experimental Psychology: Human Perception and Performance
(in revision)

Effects of Task Requirements on Rapid Natural Scene Processing: From Common Sensory Encoding to Distinct Decisional Mechanisms

**Nadège M. Bacon-Macé, Holle Kirchner,
Michèle Fabre-Thorpe and Simon J. Thorpe**

Using manual responses, human subjects are remarkably fast and accurate at deciding whether a natural scene contains an animal, but recent data shows that they are even faster to indicate with saccadic eye movements which of two scenes contains an animal. Such speed could represent a considerable advantage for studying the time course of object recognition. The understanding of the underlying processing in the two paradigms requires a comparison between the mechanisms underlying scene categorisation and those underlying scene discrimination. Here we used a masking procedure with four different tasks in which we manipulated sensory, decisional and motor constraints. With Stimulus Onset Asynchronies below 30-40 ms, sensitivity was similar for all experiments suggesting that the initial part of the underlying sensory processing is common across tasks, whether subjects have to categorise a target or to locate it. In contrast, the variations observed in reaction times and sensitivity measures for SOAs above 40 ms are discussed in relation to feedback modulations, decisional processes and motor output delays.

A large number of experiments have now been performed on rapid scene categorisation and all have demonstrated that humans and monkeys are very efficient at detecting animals or other target categories in natural scenes (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Rousselet, Mace, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a). The general paradigm developed by Thorpe et al. (Thorpe et al., 1996) is based on a go/no-go task that involves the very brief presentation of a single image, to which subjects have to respond manually by releasing a button if they detect a target. Human subjects generally score over 90% correct, with mean reaction times ranging from 400 to 430 ms, and hits start to significantly outnumber false alarms at around 250-280 ms. More recently, a choice saccade task has been introduced in which two photographs are flashed simultaneously in both hemifields, and participants have to make a saccade towards the photograph that contains an animal (Kirchner & Thorpe, 2005). Whereas accuracy remains high at around 90% correct, saccadic reaction times are much shorter, with a mean value of 230 ms, and correct responses outnumber errors at latencies as short as 120 ms. Furthermore, an analysis of the low-level differences between target and distractor images have shown that they were unable to account for

these very fast responses (Kirchner & Thorpe, 2005). The comparison of the above paradigms raises a number of questions such as the nature of the pathways involved from perception to action and the temporal dynamics of visual processing. How can we explain that with similar accuracy in target detection, reaction times can be so different from one task to another? Moreover, while the choice saccade task seems to be a promising paradigm for the study of visual perception, we still need to understand where and how the localisation process takes place in the whole stream of object recognition.

The comparison of reaction time in go/no-go and forced-choice tasks relates to the extensive literature in experimental psychology concerning mental chronometry and its theoretical analysis (Meyer, Osman, Irwin, & Yantis, 1988). According to these theories, performance depends on a sequence of processes that includes sensory encoding of the stimuli, storage of information in long term memory, decision making and programming of an appropriate motor response. Currently, a growing literature on electrophysiology and brain imaging focuses on the neural processes that take place between perception and action, corresponding to the mechanisms of neural selection and decision (Heekeren, Marrett, Bandettini, & Ungerleider,

2004; Schall, 2001, 2004; Seidemann, Zohary, & Newsome, 1998; VanRullen & Thorpe, 2001b; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). A consequence is that the perception-decision-action stream is no longer considered as a simple sequential system but rather as a continuous flow in which information accumulates over time until reaching a decisional threshold that will trigger the motor output (Gold & Shadlen, 2000; Hanes & Schall, 1996; Perrett, Oram, & Ashbridge, 1998; Romo & Salinas, 2003; Shadlen & Newsome, 1996). At the functional and anatomical level, current models are now closer to a "distributed" or "parallel and asynchronous" system (Humphreys, Riddoch, & Price, 1997; McClelland, 1979; McClelland & Rogers, 2003; Rumelhart, Hinton, & McClelland, 1986). This suggests that the different requirements in the two tasks described above may have independent effects on performance. These differences can be listed in three points.

Firstly, at the beginning of the processing course, tasks differ in terms of the physical data that enter the visual system (one central vs. two peripheral images). Previous categorisation studies have shown that performance is virtually identical when presenting one or two images in the near periphery (Rousselet, Fabre-Thorpe, & Thorpe, 2002), despite the increased amount of information that needs to be processed when two images are presented simultaneously. But even a parallel system cannot explain by itself the better performance obtained in the choice saccade task. Rather, the peripheral presentation should induce a slight drop of performance, as has previously been observed (Fize, Fabre-Thorpe, Richard, Doyon, & Thorpe, 2005).

Second, at the end of the processing course, the anatomical pathways involved in motor preparation are clearly dissociated in both tasks: saccadic eye movements will presumably rely on a network involving structures like the frontal eye fields (FEF), the lateral intra-parietal area (LIP) and the superior colliculus (SC), all of which have close connections with the oculomotor plant in the brain stem. On the other hand, manual responses will need to take a longer route, involving motor and premotor cortical areas to finally reach the hand effectors through the spinal cord. It should be noted that there is evidence that these pathways are used almost optimally in the go/no-go task since training on specific images failed to increase the speed at which they could be processed in the go/no-go manual categorisation task (Fabre-Thorpe et al., 2001). Similarly, in the choice saccade task, although saccadic reaction times decreased during the course of the study, this improvement of general motor skill did not interact with the type of image present (novel vs. repeated), again indicating that information flow from

perception to action was close to optimal (Kirchner & Thorpe, 2005). Thus, the long reaction times seen in the categorisation task could potentially be entirely due to the longer neural path length required for the manual responses. However we cannot exclude the possibility that other differences at the level of visual analysis or decisional processes may also be involved.

In fact, there are also obvious differences in the very nature of the questions asked by the two tasks. In the standard go/no-go task, subjects have to determine whether or not the image contains an animal (i.e. the question is: "Is there an animal?"), whereas, in the choice saccade task, participants know in advance that an animal will be present in one of two images but have to indicate in which of the two scenes the animal appears (answering the question "Is the animal on the left or on the right?"). To keep in mind these different requirements, we will refer to "categorisation tasks" in the first case and to "discrimination tasks" in the second case. To what extent could this difference between the tasks explain the advantage of the saccadic discrimination task? There is evidence in the literature that different cortical streams are involved when the visual task requires object recognition (ventral "what" stream) or when it involves controlling goal directed action (dorsal "where" stream) (Goodale & Milner, 1992; Ungerleider, 1985; Ungerleider & Haxby, 1994). Neurons in the dorsal stream are also known to have very short latency responses, considerably shorter than in ventral stream areas such as the infero-temporal cortex. But the question "where is the animal?" is not trivial to answer and probably needs both ventral and dorsal processing. However, it has been shown that some dorsal stream areas contain shape selective neurons (Sereno & Maunsell, 1998). Could it be that the much faster responses seen in the saccadic discrimination task only rely on dorsal stream processing?

Another emergent idea suggests that many processes involved in recognition do not have to be considered as sequential but rather as simultaneous and parallel. For example, scene segmentation is usually thought to occur before recognition but a recent challenging study has shown that the presence of an object and its category label become available simultaneously (Grill-Spector & Kanwisher, 2005), suggesting that object segmentation and categorisation do not occur in sequence.

To disentangle these various explanations, we decided to run a set of experiments that systematically varied both the type of motor output (oculomotor vs manual) and the type of decision: Presence or absence of an animal (using either a go-no or a yes/no task), or choice of the scene that contains an animal (left vs right discrimination

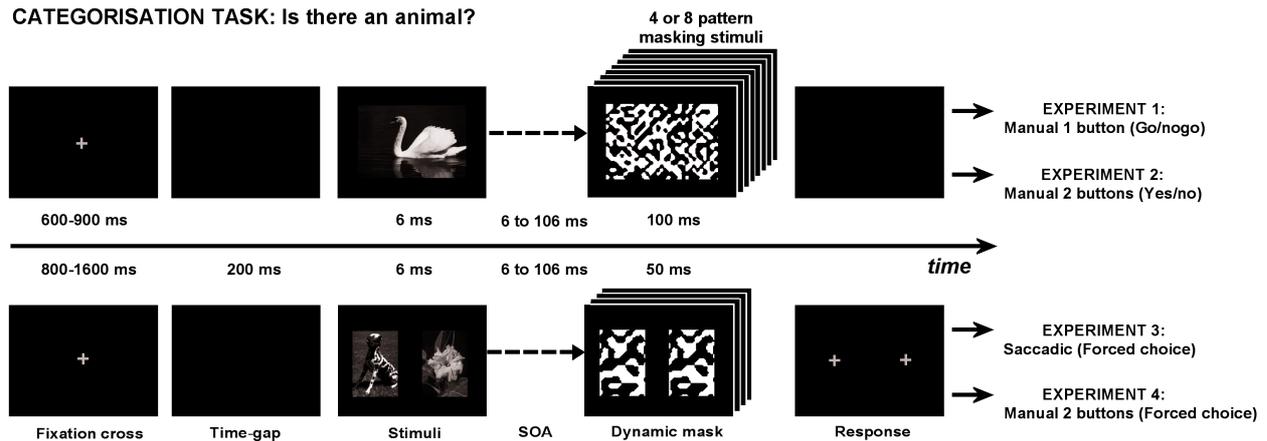
CATEGORISATION TASK: Is there an animal?**DISCRIMINATION TASK: Where is the animal?**

Figure 1: Experimental design in the four experiments. Each trial begins with the display of a fixation cross for a variable delay. In three experiments (except experiment 1), we introduced a gap interval of 200 ms. In the categorisation tasks, one central image was flashed and followed by a highly contrasted pattern mask after a variable SOA. In the discrimination tasks, two lateral images were presented, one of which was the target. Depending on the experiment, subjects had to respond manually with one 1 or 2 buttons, or with a saccadic eye movement. In all tasks, subjects were told to be as accurate and fast as possible.

task). To get a more detailed picture of the temporal dynamics, we used a masking procedure

Method

The general procedure for the four experiments is shown in Figure 1. In experiment 1 and 2, subjects had to perform a task in which one central image was briefly flashed on the screen and followed by a masking stimulus. The question for the subject was thus “Is there an animal in the image?”. We will refer to these tasks as categorisation tasks. In experiment 3 and 4, two images were presented, one in each hemifield, and followed by a mask. In all trials, one of the two images contained the target and the question was thus “Which of the two scenes contains an animal?”. These experiments are referred to as discrimination tasks.

Participants

A total of 31 subjects took part in the experiments (15 females, 16 males, with a mean age of 27.5 years), with normal or corrected-to-normal vision. They all volunteered for the study and gave their written informed consent. Several subjects took part in more than one experiment, but in each task, at least one third of the subjects were totally naive with regard to the experimental procedure.

Stimuli

Natural scenes.

In the four tasks, we used a large variety of achromatic natural scenes from the Corel database.

to interrupt processing at different levels of the recognition process.

Target photographs could contain mammals, birds, reptiles or fish, and distractors were also varied, including landscapes, indoor or outdoor scenes, and man-made objects. Each image was seen only once in a given task. Images were different in the first task (horizontal) and second task (vertical), but the same set of images was used for the last three tasks.

Mask.

The masks were the same as those used in a previous study (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005). They were produced by filtering a white noise pattern at 4 different spatial scales, and increasing the contrast of the resulting images to 100%. Each mask was rotated and mirrored, so that in total, a pool of 16 masks was available. In the first task, the mask was composed of a sequence of 8 masks, with each spatial scale appearing twice in the sequence. However, since the masking effects were demonstrated to be as strong with 4 masks as with 8 (Bacon-Mace et al., 2005), we used only 4 masks in the three other tasks with each spatial frequency displayed once. Each mask was flashed for two frames resulting in a total mask duration of around 100 ms (Exp 1) and 50 ms (Exp 2-3-4).

Experimental Design

Participants were seated in a dark room, in front of a screen adjusted to a 800 x 600 pixel resolution

with a 160 Hz refresh rate. Photographs of natural scenes were flashed for one single frame, that is 6.25 ms, and were followed by the mask with a variable Stimulus Onset Asynchrony (SOA). In each series, there were as many targets as distractors and all SOA conditions (6.25, 12.50, 18.75, 31.25, 43.75, 81.25 and 106.25 ms) were randomly intermixed, but in the same proportion. Furthermore in the four tasks, we added one special condition in which only the mask was presented, without any scene being flashed before. This condition was unknown by the subjects and occurred on roughly one out of ten trials in each series.

Experiment 1: Categorisation tasks with go/no-go manual responses.

This protocol is similar to that used in a large number of earlier studies. A trial began with the display of a fixation cross in the centre of a black screen for 600 to 900 ms at random. Then a target or distractor image (15.3°H x 10.4°V) was presented, followed by the mask. Subjects had to release a button as fast as possible if the image contained an animal or to maintain contact otherwise. Moreover, they were asked to try to release the button on 50% of the trials, whatever the masking condition. Each of the 16 participants (8 females, 8 males, mean age 29.4) performed 16 series of 90 trials, and were previously trained with two series. Unlike the three other experiments of this study, Experiment 1 contained an extra SOA condition of 25 ms. Detailed procedure and full data on this task have already been published (Bacon-Mace et al., 2005).

Experiment 2: Categorisation task with yes-no manual responses.

This task was comparable to the first experiment, as subjects had to decide whether or not there was an animal in the centrally presented masked image. But contrary to the first experiment, they had to provide a response on each trial: They maintained their index fingers on the two 'Ctrl' buttons of a keyboard throughout the series, and released the button on the side of their dominant hand to indicate that they had detected the presence of an animal in the scene, or release the button on the opposite side if they thought it was a distractor. Motor responses were thus very comparable to the fourth experiment (see later). Contrary to the first experiment, photographs were presented vertically (7.7°H x 11.5°V, 305 x 445 pixels) and were also used in the third and fourth experiment. Twelve subjects (5 females, 7 males, mean age 26.5) performed 12 series of 80 trials, after 1 block of training trials.

Experiment 3: Forced-choice discrimination with saccadic responses.

A trial began with the display of the fixation cross at the middle of the screen for a random interval between 800 to 1600 ms. To facilitate saccade initiation, the fixation cross disappeared 200 ms before the presentation of the images (Fischer & Weber, 1993; Kirchner & Thorpe, 2005). Two vertical natural scenes were flashed in the left and right hemifields with the centre of the image at 6°. They were followed by two lateral masks, covering the surface of the images. Masks on both sides were identical. Each subject, whose head was stabilised by a forehead and chin rest, had to make a saccade towards the side where an animal had appeared. Possible saccade endpoints were indicated by two fixation points that were displayed for 1 s after the masks at the centre of the images. Eight participants (4 females, 4 males, mean age 27.8) performed 12 series of 64 trials. For training they performed two series without masking and a third series with masking.

Experiment 4: Forced-choice discrimination with manual responses.

The experimental design was similar to the third task, with the same set of images but a different response mode. Throughout the series, subjects had to maintain each of their index fingers on the two 'Ctrl' buttons of a keyboard (Hasbroucq, Mouret, Seal, & Akamatsu, 1995). They had to indicate the side where the animal had appeared by releasing the 'Ctrl' button on the appropriate side. Subjects were told to be as fast and accurate as possible. Eight subjects (3 females, 5 males, mean age 26) performed 12 series of 64 trials, after having performed 1 training series.

Globally, the number of trials in the different experiments was settled to be equivalent for statistical analysis, except the first experiment which presented a higher number of trials because of the associated EEG recordings (Bacon-Mace et al., 2005). The additional training series in Experiment 3 were performed to habituate non-expert subjects to making saccadic eye movement responses.

Response Recording

In the first task, manual responses were detected using a button with an infra-red diode system, and this signal was directly recorded by the stimulation software. In the third task, eye position was recorded by two EOG electrodes (NuAmps, Neuroscan Inc.) with a sampling rate of 1 kHz, a notch at 50 Hz and a low-pass filter at 90 Hz. Before analysing the signal, a baseline correction was applied on a range starting from 400 ms preceding the stimulus onset to the beginning of image presentation. A saccadic response was validated for a trial if the subtracted signal

between left and right EOG electrodes reached an amplitude of at least $\pm 35 \mu\text{V}$. The nearest signal inflection preceding this point was then considered as the onset of the response. In the second and fourth tasks, manual responses were detected with the two Ctrl buttons of a keyboard connected to the stimulation software (Presentation software-Neurobehavioral Systems).

Analysis

Effects on performance within and between tasks were assessed by using an ANOVA with a Greenhouse-Geisser correction for non-sphericity. If necessary, post-hoc analyses were performed using t-tests with a Bonferroni correction or Wilcoxon tests.

Results

Accuracy

In this section, the accuracy of the subjects is analysed for the four tasks as a function of SOA. We first determined the accuracy in terms of total percentage correct. But the results were also analysed by calculating d' , which is known to be a better parameter for comparing tasks that involve both one (go/no-go, yes-no) and two alternative (forced-choice) detection paradigms (Green & Swets, 1966; Kroll, Yonelinas, Dobbins, & Frederick, 2002).

Percentage correct.

In all tasks, there is one chance in two to perform a correct answer on a given trial. Chance level is thus calculated at 50% and significance determined by a χ^2 test.

Anova analysis showed that accuracy in the discrimination tasks was higher than accuracy in the categorization tasks ($F(3,307) = 9.28$, $p < 0.001$), and interestingly did not depend on the particular motor response to be produced. When all SOA conditions were pooled together, accuracy in the discrimination tasks averaged 76.95% with manual responses (individual range [67.41-79.61]) and 76.14% with saccadic responses ([72.29-82.78]). In contrast, in the categorisation tasks, subjects scored an average 72.78% in the go/no-go task ([64.82-79.20]), which was not significantly different from the accuracy in the yes-no task, with 73.41% ([66.63-77.02]).

As expected, accuracy varied with the SOA conditions (Figure 2A). In all experiments the percentage correct begins at chance level with 6 ms of SOA which reflects that visual processing was totally blocked, but rapidly increases until reaching a plateau above 44 ms of SOA. Subjects were already above chance level ($p < 0.001$) with 12 ms SOA in three of the tasks, and only the manual yes-no experiment failed to reach statistical

significance in this SOA condition. The maximum value is obtained with the two last SOA conditions (81 ms and 106 ms), in which accuracy comes close to or exceeds 90% correct. This accuracy is comparable to the accuracy obtained in a previous go/no-go experiment using grey-level stimuli without masking (Delorme, Richard, & Fabre-Thorpe, 2000). Interestingly, the four tasks had a common global accuracy profile with increasing SOA, although post-hoc analyses revealed some significant differences between Experiment 1 (categorization with manual go/no-go) and the three other experiments in the 19 ms SOA condition ($F(3,43) = 12.58$, $p < 0.001$).

In the two discrimination experiments, targets were presented equally often in the left or right side of the display and subjects, regardless of their preferred hand (left or right), did not show any significant response bias towards one or the other side (percentage of responses on the left and right side: 48.73% vs. 51.27% in the saccadic responses, 51.10% vs. 48.90% in the manual responses). Moreover, in the condition where no target image was presented before the mask, the response rate for each hemifield was close to 50%, indicating that subjects responded randomly regardless of the side (49.61% vs. 50.39% in the saccadic responses, 48.54% vs. 51.46% in the manual responses).

In the two categorisation tasks, subjects were told that there was about the same number of target and distractor trials in a series, and they were encouraged to respond on half the trials. Despite this instruction, the number of 'go' or 'yes' responses (Hits) varied considerably with the difficulty of the task (manual go/no-go $F(6,111)=239.62$, $p < 0.001$); manual yes-no $F(6,83)=30.59$, $p < 0.001$), as shown in Table 1. In both experiments, response rates were stable for the three longest SOA conditions (44, 81 and 106 ms), but strongly decreased in all other more challenging SOA conditions ($p < 0.001$, Wilcoxon tests with Bonferroni correction). Furthermore, in the more difficult SOA condition of 6 ms, response rates (either Hits or False-Alarms) were comparable to the response percentage performed when no target image was presented before the mask (response percentage in SOA 6 ms / no-target condition: 22.3% vs. 19.6% in the go/no-go experiment, 31.1% vs. 25.1% in the yes-no experiment).

Globally, in the categorisation tasks, subjects showed strong inhibition when they had to initiate a response with limited information about the stimulus. In fact, two factors can explain a decrease in performance: either an increase in the number of false-alarms or a decrease in the number of hits. Although response latencies did not increase much (see below), the number of hit responses strongly decreased at short SOAs. In contrast, in the discrimination tasks, only an

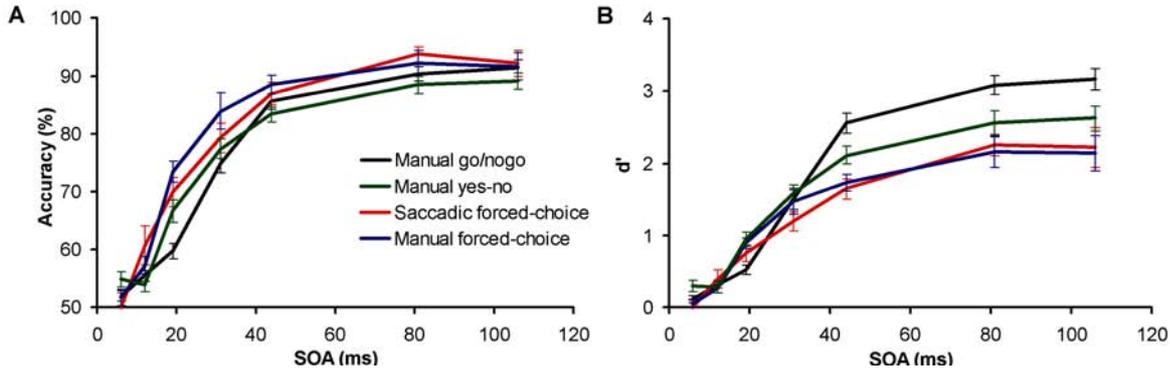


Figure 2: Mean performance as a function of SOA, in the four experiments. (A) Mean accuracy, with chance-level at 50%. In the categorisation tasks, the percentage was calculated over all correct responses, and thus reflects both Hits to targets and correct rejections with distractors. (B) Mean d' in the go/no-go and the yes-no tasks, and mean corrected d' in the forced-choice tasks. Error bars represent the mean standard errors.

increase in the number of false-alarms can produce a decrease in the percentage of correct responses. Thus accuracy measures, although very instructive, cannot be considered on their own to analyse performance variation as a function of task difficulty. In the next section, we will provide details of the d' analysis, which is known to be a bias-free measure of sensitivity (Green & Swets, 1966; Macmillan & Creelman, 1991).

d' analysis.

In the one-alternative paradigm, the detection sensitivity measure d' is calculated as $Z_H - Z_{FA}$, where H is the hit rate and FA the false alarm rate (Macmillan & Creelman, 1991). In the two-alternative paradigm, the equation needs to take into account the summation of the variability for two distributions (Macmillan, 2002). As a result, the d' measure must consider a correction factor and is calculated as $(1/\sqrt{2})(Z_H - Z_{FA})$. These values were calculated in the four experiments, separately for each SOA condition and for each individual subject. For some subjects, rates were 1 (in particular some subjects scored 100% correct in the easiest conditions), which would lead to an infinite value for the z-score. In such cases, we used a rate of 0.99 instead.

Figure 2B shows the d' values as a function of SOA for all four experiments. Across all SOA conditions, d' values were significantly different between experiments ($F(18,307) = 4.528$, $p < 0.001$), but post-hoc analyses indicated that the paradigms that used the same number of alternatives (either categorisation or discrimination) did not differ statistically. Thus, discrimination tasks d' were not different between the manual (on average 1.24 [0.68-1.5]) and saccadic response paradigms (1.21 [0.97-1.68]). Similarly in the categorisation tasks, the yes-no d' did not differ from the d' calculated for the go/no-go task (1.48 [0.96-1.76] vs. 1.61 [1.02-2.12]).

More interestingly, Figure 2B clearly shows two separate phases as d' increases with SOA. First, for

SOAs up to and including 31 ms, the four curves are quite close, and globally differences between experiments did not reach significance. The only exception was the 19 ms SOA condition where, as described above for accuracy, the manual go/no-go task had a slightly inferior d' value compared to both the saccadic discrimination ($p < 0.03$) and the manual yes-no paradigms ($p < 0.002$). However with SOAs above 31 ms, d' values varied considerably between experiments (44 ms: $F(3,43) = 10.867$, $p < 0.001$; 81 ms: $F(3,43) = 7.78$, $p < 0.001$; 106 ms: $F(3,43) = 6.857$, $p < 0.001$), and it is interesting to note that the highest d' values were seen in the manual go/no-go task. The d' in the yes-no task was intermediate but both discrimination tasks had values that were substantially lower. The comparison between the two discrimination tasks and the go/no-go task reached significance from an SOA of 44 ms onwards ($p < 0.006$).

Therefore, whereas the categorisation paradigms had a higher sensitivity than the discrimination paradigms with SOAs equal or longer than 44 ms, d' showed very similar values in all four experiments for SOAs of 31 ms and below, suggesting that early visual processing might be very similar in the four tasks.

Reaction Times

The response mode (eye or hand) and the type of paradigm (categorisation or discrimination) both had a strong impact on reaction times (RT). Mean RTs were significantly different across the four tasks, whatever the SOA ($F(3;307) = 147.03$, $p < 0.001$). In the three manual tasks, mean RTs were all above 400 ms (see table 1). By contrast, saccadic reaction times (SRT) were much shorter with a mean value of only 250 ms (individual range [193-339], SOA range [245-255]). Moreover, the speed of response also depended on whether the task involved a discrimination or a categorisation paradigm. Mean RTs were shorter in both discrimination tasks regardless of the

effector (saccadic: 250 ms, manual: 404 ms [307-572] [391-426]) than they were in the two categorisation tasks, either with go/no-go (422 ms [362-509] [399-453]) or yes-no responses (477 ms [367-626] [470-486]).

Furthermore in the two discrimination tasks, mean RT did not fluctuate much with SOA, although there were large differences between subjects (see Table 1 and Figure 3A). In contrast, in the categorisation tasks we observed an increase in RT for the shortest SOAs: In the go/no-go experiment the difference in mean RT between short (below 44ms) and long (equal or longer than 44ms) SOAs reached significance ($p < 0.01$) with a maximum difference of 54 ms. In the yes-no experiment, the same decreasing RT profile was observed on correct yes-no responses at increasing SOAs, and although it did not reach significance, the difference reached a maximum of 40 ms.

Figure 3B shows the RT distributions in the four experiments. Minimal reaction times were calculated by χ^2 tests as the first 10 ms time bin for which correct responses significantly exceed errors (see colour code in the Figure caption). This value is an estimation of the minimal processing duration from visual input to motor output. Globally, comparing minimal reaction times in the four

experiments led to two main conclusions. First, processing speed critically depends on the effector, as saccadic discrimination responses can be initiated more than twice as fast as manual discrimination responses (120 ms vs. 270 ms). Second, the task requirement is also critical: discrimination tasks (saccadic or manual) lead to shorter RTs than categorisation tasks (go/no-go or yes-no).

Speed-Accuracy Trade-Off

In the previous sections, we analysed performance separately in terms of accuracy and response speed, but these parameters are often closely correlated because of speed-accuracy trade-offs. Usually, a shift to shorter reaction times results in an increase in errors, whereas focusing on accuracy leads to longer reaction times (Fitts, 1954; Meyer, Irwin, Osman, & Kounios, 1988). Such trade-offs have been found in several of our previous experiments with natural scenes (Kirchner & Thorpe, 2005; Thorpe et al., 1996). However, by limiting the amount of available information with a masking procedure, one can ask whether the compromise between speed and accuracy is still present. It is possible that if all the available information has already been processed, there would be little to gain from delaying the motor response.

We thus calculated the correlation between RT and accuracy in the four tasks, on the basis of the mean data of each subject. Figure 4 shows the speed-accuracy correlation index as a function of SOA, separately for each task. In the manual go/no-go task, there was no speed-accuracy trade-off, whatever the SOA. In contrast, in the discrimination tasks a correlation was present for SOAs longer than 44 ms, whereas it was reduced or even absent with SOAs shorter than 31 ms. Finally, in the yes-no task, a speed-accuracy trade-off was present although less marked than in the discrimination tasks. This suggests that when subjects have to choose between two possible response alternatives ('left' or 'right', 'yes' or 'no'), the visual system can improve accuracy by slowing down the response. But this only seems to be true when the mask disrupts visual processing relatively late, allowing the visual system to process further the input of available information. In contrast, with short SOAs, it may not be useful to allow time for more information processing if all the available information has already been extracted.

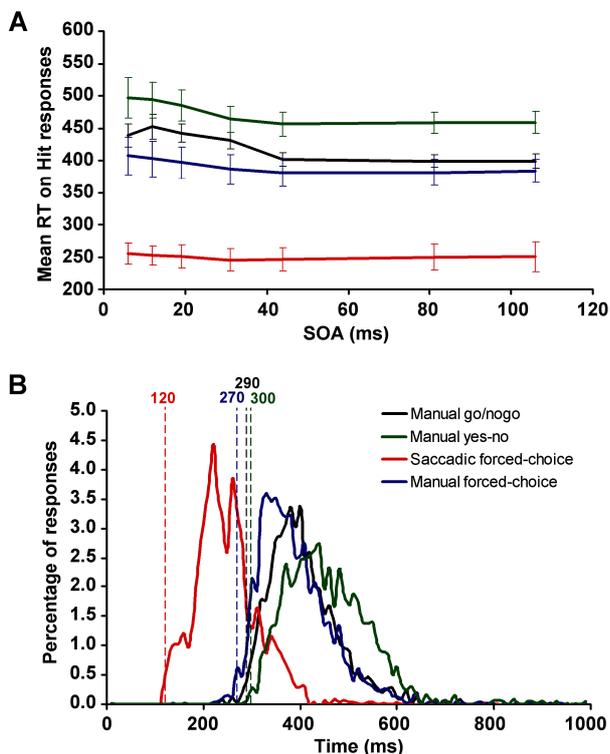


Figure 3: Reaction times in the four experiments. Only values for Hits were considered in the Yes-no task. (A) Mean reaction times as a function of SOA (+/- SEM). (B) Overall reaction time distributions, calculated as the difference between percentages of correct and incorrect responses for each time bin (10 ms bin width), in the four experiments. Dotted lines indicate the minimum time (first significant bin) for each distribution.

Discussion

The present study analysed human performance in four different tasks in order to study rapid visual processing of natural scenes under high temporal constraints that result from a dynamic mask. In

order to compare the four paradigms, it is important to identify clearly the differences that characterise them, which brings out three major dissociations as schematised in Figure 5. The first dissociation can be seen at the **perceptual stage**, and directly concerns the processing that can be performed on the raw input to the visual system: In the paradigms used in experiments 1 and 2, only one image was flashed centrally on the screen, whereas the paradigms used in experiments 3 and 4 required the parallel processing of two different images. Second, the tasks differed at the **decisional level**. In experiments 1 and 2, subjects were asked to indicate whether or not an animal was present, the judgment thus was categorical in nature. In experiments 3 and 4, on the other hand, subjects were able to perform a discrimination between two scenes: they could rely on the assumption that an animal was always present but had to determine on which side the animal appeared. Thus, categorisation tasks (experiments 1 and 2) require a decision that is based on the analysis of a single stimulus whereas discrimination tasks (experiments 3 and 4) involve the processing (and possibly the comparison) of two simultaneously presented stimuli in order to decide which one contains the animal. Finally at the very end of the perception/action stream, we varied the **response mode**, which either involved a

saccadic eye movement (experiment 3) or manual responses (experiments 1, 2, and 4). The motor stage thus showed large differences depending on which effectors were used (saccadic eye movement or manual response) and whether the response involved one effector and one response (go/no-go categorisation task with manual response), one effector but two different responses (saccadic discrimination task), or finally two effectors and one response (manual responses with the two hands in the discrimination task and in the yes-no categorisation task). Although these differences might be thought to have a critical impact on sensory-motor processing, we show here that performance measures were comparable in all tasks, in particular with SOAs of 31 ms or less.

Early Sensory Processing Is Independent Of The Task

When the mask is presented shortly after the stimulus, visual processing is strongly affected as demonstrated by low performance. A logical hypothesis that has been proposed previously is that the amount of information available from low-level properties of the stimulus is limited because of interference between the stimulus and the mask that takes place early in the cortical pathway, probably before area V4 (Bacon-Mace et al., 2005). The d' analysis indicated similar sensitivity

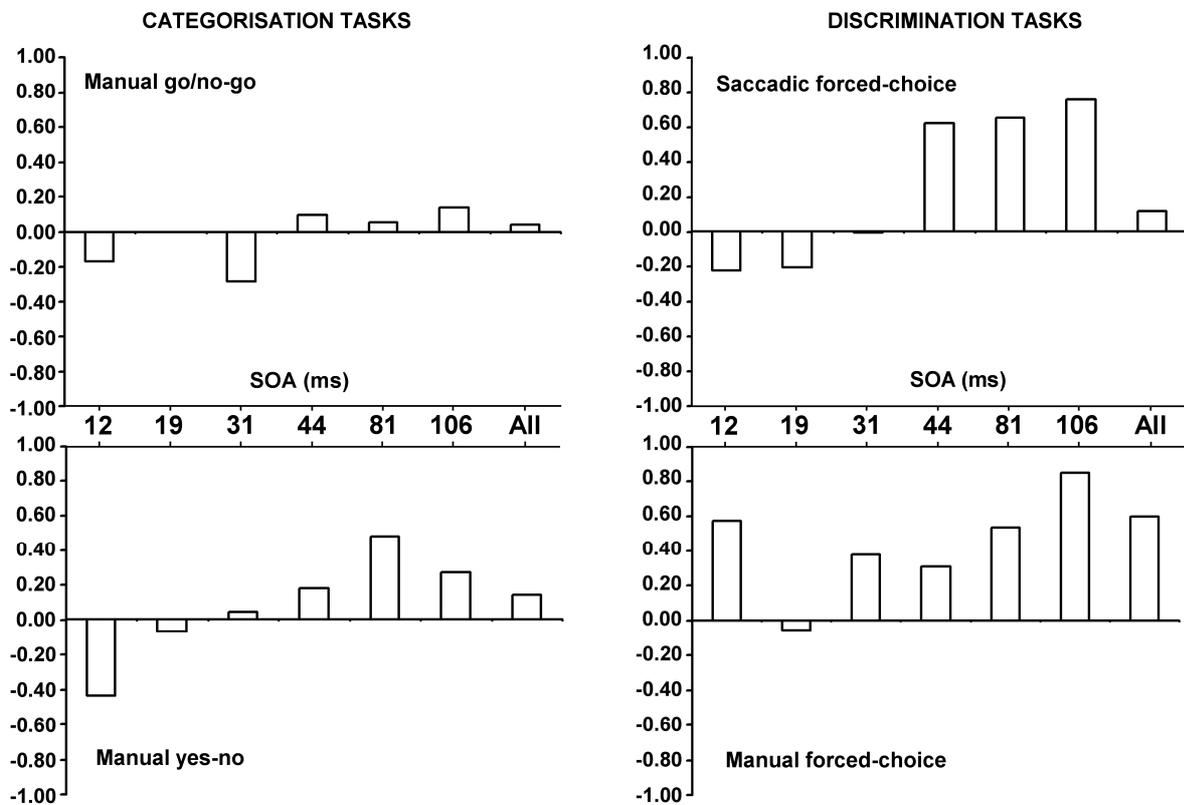


Figure 4: Correlation index (r) as a function of SOA (from 12 to 106 ms) and for all SOAs grouped together (bar in the right side). For each SOA condition, the correlation is calculated between the mean accuracy and the mean reaction time of each subject, with a Pearson test. Each bar thus represents a correlation index that takes into account as many points as there are subjects in a given experiment.

in all experiments when the mask was presented shortly after the stimulus, with an SOA of 31 ms or less, suggesting that the amount of information that can be extracted from the stimuli at these SOAs may be similar between experiments. Whatever the task and the response mode, the visual system seems to integrate stimulus information with the same efficiency, and disrupting the processing leads to the same limitations. When the mask is close to the image, information about the stimulus enters the visual system but is wiped out by the information about the mask as it goes along the visual pathway. Only information that can be extracted very rapidly can escape interference from the mask and be used further down-stream (Bacon-Mace et al., 2005; Romo & Salinas, 2003; VanRullen & Thorpe, 2002). We therefore propose that in all four experiments, maximum performance is determined by the amount of visual analysis that could be achieved before masking occurs at early sensory stages. This idea is reinforced by the absence of a speed-accuracy trade-off for SOA values shorter than 31 ms, which suggests that the visual system is working in a "just-in-time" mode and cannot manipulate the available information to refine potential interpretations of the stimulus.

Depending On task And Response Mode, Decisional And Motor Processes Are Dissociated With Longer SOAs

With SOAs of 44 ms and longer, accuracy improves and reaches a plateau in the four experiments, above 83% correct. The interference

between the stimulus and the mask information occurs later and the visual system has more time to process the stimulus. Interestingly, from this point onwards, d' measures show that sensitivity begins to differ between the experiments. Specifically, the discrimination experiments had a lower sensitivity than the categorisation tasks (Figure 2B). This effect was not visible in the raw accuracy measures when no correction was applied in the discrimination tasks to compensate for the processing of two images at the same time (Macmillan, 2002).

Categorisation tasks have a higher d' compared to discrimination tasks.

Several non-mutually exclusive hypotheses can explain the d' reduction in the discrimination tasks compared to the categorisation tasks (Figure 5). A first point concerns the raw data entering the system, because at least some drop in performance could be expected with the image eccentricity in the discrimination tasks (6°). In fact, the impact of eccentricity on performance has already been determined in previous categorisation studies in which, at eccentricities of 4-13° from the centre, a decrease of 3-4% was accompanied by an increase in mean reaction times of about 15 ms (Fize et al., 2005; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001).

Second, the analysis of mean reaction times showed that the latencies in the discrimination experiments were remarkably stable from one condition to another, even with the shortest SOAs,

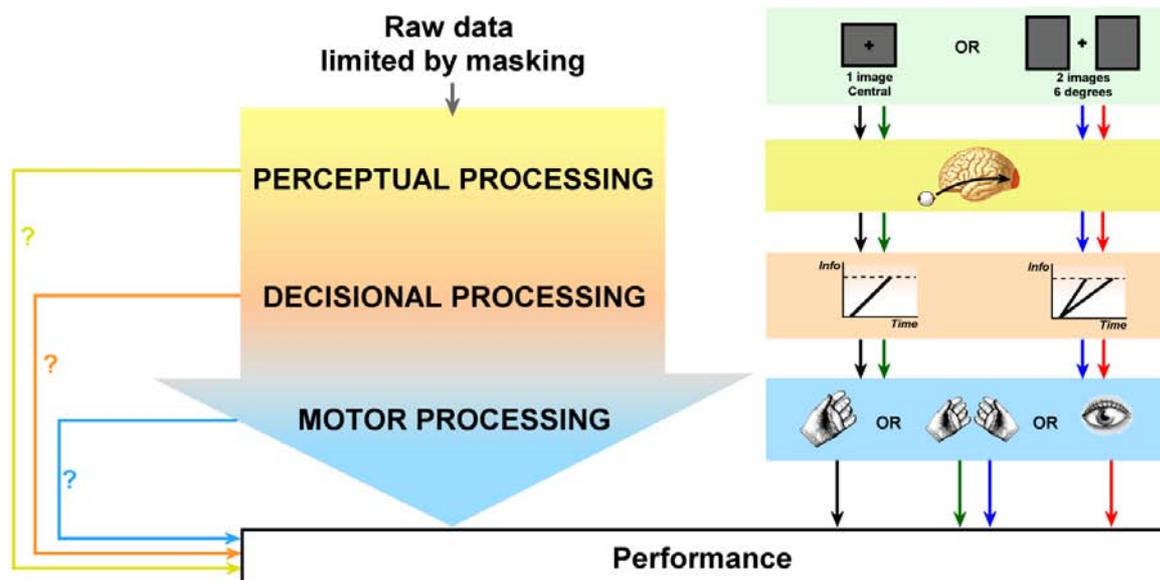


Figure 5: Schematic representation of the different stages within the visual sensory-motor processing stream. Differences that characterise the four experiments are indicated for each stage. Colour code of the arrows corresponds to the data presentation in figures 2 and 3.

contrary to the categorisation experiments in which reaction times increased with decreasing SOA (Figure 3A and Table 1). Thus, not only are reaction times shorter in the discrimination tasks, they also seem to be unrelated to the task difficulty. Typically, in the discrimination paradigms, whether manual or ocular, it could be that the final decision is based on more automatic response triggering (Romo & Salinas, 2003; Schall, 2003), with motor preparation sufficiently advanced when accumulated information about the stimuli has activated a coherent representation of the decision so that the appropriate response has only to be released (Kirchner & Thorpe, 2005). This could constitute one explanation for the short reaction times in discrimination as compared to categorisation tasks. However, this hypothesis might also account for the lower sensitivity in discrimination tasks, because the visuo-motor system may sometimes produce a slightly premature response relative to the selection process, leading to greater inaccuracy.

This hypothesis was further supported by the analysis of speed-accuracy trade offs (Figure 4) that indicated that in the discrimination experiments and for SOAs longer than 44 ms, the slowest participants tended to be more accurate than the fastest subjects. This suggests that when information is sufficiently preserved, a response strategy can be applied to modulate the compromise between speed and accuracy. In contrast in the categorisation experiments, this speed-accuracy trade-off was either reduced (yes-no task) or totally absent (go/no-go task) even with the longest SOAs. How can we explain that there is a speed-accuracy trade-off in the discrimination but not in the categorisation tasks and that it only appears at longer SOAs? In both categorisation tasks, subjects had a low response rate with short SOAs, and despite the instruction to respond on half of the trials in each series, they were strongly inhibited in the more difficult conditions (the response rate decreased to around 20% with an SOA of 6 ms in both tasks). A first explanation for the absence of speed-accuracy trade-offs in these tasks would be that the subjects were very concerned by the instruction given to force the rate of manual responses, which could prevent them from applying other strategies such as a modulation of speed and accuracy. The existence of such an inhibitory effect is particularly plausible, given that speed-accuracy trade-offs have already been observed in previous experiments using the go/no-go task but without masking (Delorme et al., 2000; Thorpe et al., 1996). In the discriminations tasks on the other hand, subjects were asked to respond in each trial because an animal was always present. In order to explain speed-accuracy trade-offs in such choice tasks, psychological models of decision-making

have proposed that these trade-offs should be related to moving the decision boundaries which are located at the end of a sensory-decisional accumulation process closer together so that responses will be made faster, but would be more prone to errors (Ratcliff & Rouder, 1998; Smith & Ratcliff, 2004). Our data is in agreement with this prediction as reaction times in the discrimination tasks were faster than in the categorisation paradigms. However, according to these models we were expecting a correlation between reaction times and accuracy (Palmer, Huk, & Shadlen, 2005), whereas RTs were constant for all SOA values, despite the fact that accuracy ranged from chance level all the way to 90% correct. Therefore, it is all the more remarkable that we observed speed-accuracy trade-offs within the long SOA conditions. It thus seems that decision boundaries were low so that accuracy was lower than in the categorisation tasks, but at the same time flexible so that they allowed some variability when more and more sensory information was available in long SOAs.

Finally, an interesting secondary issue remains as to whether a decision is reached after independent and parallel processing of each image or whether the decision is based on a comparison process. Given the response speed, it has previously been proposed that the two images are processed independently, each image in one ventral system, until one of these processes reaches the decision boundary so that a response to the opposite side can be released (Kirchner & Thorpe, 2005). However, an alternative hypothesis supposes that the visual system takes advantage of the presentation of the two images by directly comparing the outcome of the processing results in the two hemi-fields. This comparison might involve higher-level areas such as (dorsolateral) prefrontal cortex, and in the case of eye movements, the frontal eye fields. With regard to the difference in accuracy between discrimination and categorisation tasks, the integration hypothesis would predict that the rate of information accumulation on two simultaneous but integrated stimuli may be greater, leading to shorter reaction times, than in case of a single image task (categorisation) when subjects are forced to compare the target with an internal representation. A disadvantage of such comparison mechanisms would be that when the visual system has to process noisy information, as would be the case with masking, the sensory information reaching the boundary might be less reliable so that responses would be more likely to be made in error.

Discrimination tasks, whether manual or ocular, have similar sensitivities.

Another interesting point comes from the similar sensitivities observed in the two discrimination tasks, despite the fact that one uses the manual and the other the ocular motor system. Remarkably, although accuracy and d' are both similar, mean RTs are very different as saccadic responses can be triggered almost 150 ms before manual responses (250 ms vs. 391 ms). It is difficult to define exactly which anatomical pathways are involved in these experiments, but the contribution of the ventral pathway, which is known to be implicated in object recognition (Felleman & Van Essen, 1991; Goodale & Milner, 1992; Tanaka, 1993) seems highly likely and comparable in both tasks, especially at the sensory level. But in view of the extremely short reaction times one can wonder which processing steps could be bypassed to allow the first meaningful response to be released so fast. A plausible cortical route for manual response preparation might go from the ventral pathway to the prefrontal areas, and reach the spinal cord via the premotor and motor cortex (Boussaoud, di Pellegrino, & Wise, 1995; Foxe & Simpson, 2002; Rizzolatti, Luppino, & Matelli, 1998). The 150 ms difference in reaction times between the two tasks can partly be justified by the numerous stages from visual to motor areas and longer transmission delays from one stage to another, together with the additional delays required to go through the spinal cord to the hand. However, the similar sensitivity in the manual and saccadic tasks at each SOA step suggests that the amount of sensory information is similar, and that the paradigm alone (discrimination here) determines the level of information processing necessary to make a response. Whether the mode of response is manual or ocular may constitute a factor that operates at a relatively late stage, leaving initial processing relatively unchanged.

Categorisation tasks show a small difference of sensitivity.

Figure 2B shows a difference between the go/no-go and the yes-no experiments, for SOAs longer than 44 ms. Not only has the go/no-go task a better d' at this SOA range, but reaction times are also shorter than in the yes-no task. Differences between these two sorts of tasks have already been reported in the context of lexical decision experiments (Hino & Lupker, 2000; Perea, Rosa, & Gomez, 2002; Siakaluk, Buchanan, & Westbury, 2003). They are usually justified by the presence of more competitive stages related to the decisional and motor mechanisms in the yes-no task. Indeed, in the yes-no experiment, subjects are required to release one of two buttons, which implies that two distinct motor programs have to be prepared and initiated, whereas the response

selection process is apparently simpler in the go/no-go task in which subjects respond with one button, and only in the case of an affirmative response.

d' validity

Finally, it is important to note that previous studies that have aimed at comparing discrimination and categorisation paradigms have typically failed to find any difference in d' after applying the correction for the processing of two simultaneous stimuli in a forced choice task (Green & Swets, 1966; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Yonelinas, Hockley, & Murdock, 1992). We were thus surprised to observe such marked differences between these tasks at the longest SOAs, when perception was at its best. Indeed, traditional signal detection theory has been the subject of debate since it does not always provide a reliable measure of the accuracy in memory tests (Kroll et al., 2002). Other models involving modifications of the classical signal detection theory have been recently proposed, and were found to provide a better fit for the observed Receiver Operating Characteristics (ROC). They are based on additional measures of performance such as confidence judgements, and thus could not be applied here. However, it is noteworthy that these studies mainly used recognition memory tests, which are quite different from our analysis of sensory processing in the visual system. Moreover, the debate mostly focuses on the number of sources of information needed to explain competitive responses, which is not representative for our experiments where the source of information is probably unique, coming from the sensory processing in the ventral pathways (Banks, 2000; Dunn, 2004; Smith & Duncan, 2004).

Conclusion

The comparison of four experimental paradigms shows a clear dissociation in the effect of masking on performance depending on how early processing is disrupted, with a critical value around 40 ms. When information processing is disrupted early on, sensitivity is similar for all four experimental protocols, whatever the paradigm or the response mode. This is of considerable interest as it suggests that information integration processes are the same in the initial phase of sensory encoding. Such results also support the feed-forward model of information transmission, as previously proposed to explain the remarkable speed with which subjects can respond in such categorisation and discrimination tasks (Bacon-Mace et al., 2005; Delorme, Rousselet, Mace, & Fabre-Thorpe, 2004; Fabre-Thorpe et al., 2001; Kirchner & Thorpe, 2005; VanRullen & Koch, 2003). Modulatory effects on performance that would involve feedback processing are not

observed, either because they induce identical behavioural effects in the four tasks, or because they intervene later, after the first 30-40 ms of processing have been completed. Indeed, the variations observed in the measures of sensitivity and reaction times when processing is disrupted after longer delays may be explained by several processes, including feedback modulations, decisional processing or/and motor selection, preparation or triggering. The important fact is that all these operations intervene relatively late, after a common phase of sensory processing has been completed.

Finally, one of the main results of the present study is the clear validation of paradigms relying on eye-movement responses to analyse early perceptual processing. We have shown sensitivity to be similar whether the task involves manual or ocular responses, at least in the first 40 ms. But the fact that saccadic reaction times were almost twice as fast allows temporal constraints to be further refined and support the idea that early processing of visual information relies on an automatic mode despite the apparent complexity of processing required on natural scenes.

TABLE 1

	6	12	19	31	44	81	106	All SOAs
Number of responses								
EXPERIMENT 1: manual go/no-go categorisation								
Hits	310	443	636	1070	1232	1249	1250	6190
False-Alarm	262	303	387	431	320	214	189	2106
Miss	970	838	644	210	48	31	30	2771
Correct Rejection	1018	976	893	849	960	1066	1091	6853
EXPERIMENT 2: manual yes-no categorisation								
Hits	259	296	433	571	638	640	648	3485
False-Alarm	188	237	193	183	156	89	83	1129
Miss	462	425	287	142	83	77	73	1549
Correct Rejection	529	480	527	543	563	634	631	3907
EXPERIMENT 3: saccadic discrimination								
Hits	355	425	490	572	627	671	657	3797
False-Alarm	356	281	212	151	94	44	56	1194
EXPERIMENT 4: manual discrimination								
Hits	392	434	558	638	673	701	695	4091
False-Alarm	368	326	202	122	87	59	63	1227
Reaction time (ms)								
EXPERIMENT 1: manual go/no-go categorisation								
Mean	439	453	442	421	401	399	399	422
S.D.	68	75	53	46	41	39	41	48
Min	340	347	368	350	326	332	322	362
Max	535	575	545	501	474	466	481	509
EXPERIMENT 2: manual yes-no categorisation								
Mean	497	494	485	464	457	458	459	473
S.D.	104	92	80	68	62	54	56	72
Min	376	362	374	374	367	373	378	372
Max	750	673	644	592	546	527	541	611
EXPERIMENT 3: saccadic discrimination								
Mean	255	252	250	245	246	250	250	250
S.D.	43	39	47	45	47	55	60	48
Min	209	211	200	189	192	173	173	193
Max	339	327	338	326	338	347	357	339
EXPERIMENT 4: manual discrimination								
Mean	407	402	396	386	381	381	384	391
S.D.	78	73	64	61	55	55	48	61
Min	304	294	305	304	305	317	318	307
Max	528	489	474	463	458	457	455	472

Table 1: Number of responses and reaction times on Hit responses in the four experiments, as a function of SOA. The number of responses are presented depending on their type: Hit (H), False alarm (FA), Miss (M) and Correct rejection (CR) responses.

REFERENCES

- Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res*, *45*(11), 1459-1469.
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychol Sci*, *11*(4), 267-273.
- Boussaoud, D., di Pellegrino, G., & Wise, S. P. (1995). Frontal lobe mechanisms subserving vision-for-action versus vision-for-perception. *Behav Brain Res*, *72*(1-2), 1-15.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res*, *40*(16), 2187-2200.
- Delorme, A., Rousset, G. A., Mace, M. J., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res*, *19*(2), 103-113.
- Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychol Rev*, *111*(2), 524-542.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci*, *13*(2), 171-180.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, *1*(1), 1-47.
- Fischer, B., & Weber, H. (1993). Express Saccades and Visual Attention. *Behav Brain Sci*, *16*, 553-567.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol*, *47*(6), 381-391.
- Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain Cogn*, *accepted*.
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Exp Brain Res*, *142*(1), 139-150.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390-394.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci*, *15*(1), 20-25.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger Publishing Co.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition. *Psychol Sci*, *16*(2), 152-160.
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427-430.
- Hasbroucq, T., Mouret, I., Seal, J., & Akamatsu, M. (1995). Finger Pairings in Two-Choice Reaction Time Tasks: Does the Between-Hands Advantage Reflect Response Preparation? *J Mot Behav*, *27*(3), 251-262.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859-862.
- Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *J Exp Psychol Hum Percept Perform*, *26*(1), 166-183.
- Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. *Philos Trans R Soc Lond B Biol Sci*, *352*(1358), 1275-1282.
- Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*, *38*(10), 1333-1341.
- Kirchner, H., & Thorpe, S. J. (2005). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res*, *in press*.
- Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: implications of comparisons of yes-no and forced-choice tests for models and measures of recognition memory. *J Exp Psychol Gen*, *131*(2), 241-254.
- Macmillan, N. A. (2002). Signal detection theory. In H. Pashler (Ed.), *Steven's Handbook of Experimental Psychology* (third ed., Vol. 4: Methodology in Experimental Psychology, pp. 43-90). New York: John Wiley & Sons, Inc.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychol Rev*, 86(4), 287-330.
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat Rev Neurosci*, 4(4), 310-322.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: mental processes inferred from speed-accuracy decomposition. *Psychol Rev*, 95(2), 183-237.
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biol Psychol*, 26(1-3), 3-67.
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5(5), 376-404.
- Perea, M., Rosa, E., & Gomez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Mem Cognit*, 30(1), 34-45.
- Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, 67(1-2), 111-145.
- Ratcliff, R., & Rouder, J. F. (1998). Modeling response times for two-choice decisions. *Psychol Sci*, 9, 347-356.
- Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalogr Clin Neurophysiol*, 106(4), 283-296.
- Romo, R., & Salinas, E. (2003). COGNITIVE NEUROSCIENCE: Flutter Discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci*, 4(3), 203-218.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci*, 5(7), 629-630.
- Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J Vis*, 3(6), 440-455.
- Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. Cambridge, MA, USA: MIT Press.
- Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nat Rev Neurosci*, 2(1), 33-42.
- Schall, J. D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Curr Opin Neurobiol*, 13(2), 182-186.
- Schall, J. D. (2004). On building a bridge between brain and behavior. *Annu Rev Psychol*, 55, 23-50.
- Seidemann, E., Zohary, E., & Newsome, W. T. (1998). Temporal gating of neural signals during performance of a visual discrimination task. *Nature*, 394(6688), 72-75.
- Sereno, A. B., & Maunsell, J. H. (1998). Shape selectivity in primate lateral intraparietal cortex [see comments]. *Nature*, 395(6701), 500-503.
- Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proc Natl Acad Sci U S A*, 93(2), 628-633.
- Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Mem Cognit*, 31(1), 100-113.
- Smith, D. G., & Duncan, M. J. (2004). Testing theories of recognition memory by predicting performance across paradigms. *J Exp Psychol Learn Mem Cogn*, 30(3), 615-625.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci*, 27(3), 161-168.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262(5134), 685-688.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520-522.
- Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *Eur J Neurosci*, 14(5), 869-876.
- Ungerleider, L. G. (1985). The corticocortical pathways for object recognition and spatial perception. In C. Chagas, R. Gattass & C. Gross (Eds.), *Pattern recognition mechanisms* (Vol. 54, pp. 21-37). Vatican City: Pontifical Academy of Sciences.
- Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol*, 4(2), 157-165.
- VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *J Cogn Neurosci*, 15(2), 209-217.
- VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30(6), 655-668.
- VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci*, 13(4), 454-461.

“Effects of task requirements on rapid natural scene processing ”

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Res*, *42*(23), 2593-2615.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proc Natl Acad Sci U S A*, *100*(10), 6275-6280.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *J Exp Psychol Learn Mem Cogn*, *18*, 345-355.

3.5 Analyses complémentaires

3.5.1 Effet de séquence

Au vu des temps de réaction particulièrement courts dans les tâches de discrimination, il est intéressant de regarder si les essais rapides ne sont pas influencés par un effet de répétition de la cible dans un hémichamp donné. De la même manière pour les tâches de catégorisation, il s'agit de savoir si les présentations successives d'essais de type go favorisent une diminution de temps de réaction, comme il a déjà été observé dans l'équipe (Delorme, 2000). Un effet de répétition semble pouvoir trouver son origine à différents niveaux, que ce soit par un effet d'amorçage du système visuel lorsque plusieurs cibles se suivent (Campbell & Proctor, 1993; Laming, 1968), par une modulation du seuil décisionnel à partir duquel la réponse peut être sélectionnée (Notebaert & Soetens, 2003; Rabbitt, 1968), ou encore par une étape plus avancée de préparation motrice (Bertelson, 1961; Dorris, Pare & Munoz, 2000; Gore, Dorris & Munoz, 2002).

Cependant, la figure 3.5 montre qu'aucun effet de séquence ne peut être mesuré pour les 4 expériences, même si une légère tendance non significative (trop de bruit avec le masquage, pas assez d'essais ?) est observée pour les tâches de discrimination, dans lesquelles la précision augmente lorsque la cible est présentée séquentiellement dans un hémichamp donné.

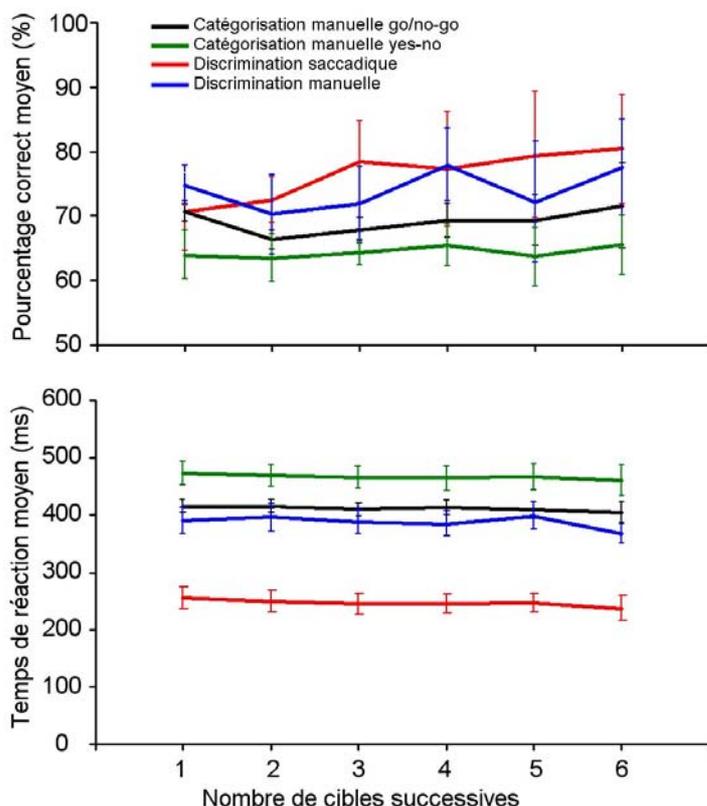


Figure 3.5 : Absence d'effet de séquence dans les quatre expériences. La précision et les temps de réaction ont été moyennés à chaque fois qu'un nombre donné d'essais successifs contenaient une cible. Pour les tâches de discrimination, on a regardé les essais successifs contenant une cible du même côté de la croix de fixation. Les valeurs au delà de 6 essais successifs ne sont pas indiqués, à cause de la trop grande variabilité des données due à leur faible occurrence. Ni la précision ni le temps de réaction ne sont sensibles à l'effet de répétition.

3.5.2 Le cas particulier des saccades express

Dans la tâche de discrimination saccadique, il est intéressant d'observer la distribution des TR pour les différents sujets. Rappelons que la valeur moyenne des temps de réaction est de 250 ms dans cette tâche, avec les premières réponses significativement correctes apparaissant dès 120 ms. Lorsque l'on regarde la distribution des temps de réaction prenant en compte les réponses de tous les sujets (Figure 3.7A), on observe un premier « décrochement » autour de 150 ms avant une montée plus franche de la courbe vers les valeurs moyennes. Or il apparaît que ce premier « décrochement » est principalement dû à des réponses fournies par deux sujets en particulier, dont les distributions présentent une forme bimodale avec un premier pic apparaissant autour de 150 ms puis autour de 210 ms (Figure 3.7B et C). Il est à noter que ces deux sujets peuvent reproduire ce profil de distributions, que nous avons retrouvé en analysant *a posteriori* les données enregistrées dans l'expérience initiale publiée dans Vision Research (Kirchner & Thorpe, 2005), ainsi que dans une autre expérience utilisant des réponses saccadiques, dont les données sont soumises à *Journal of Vision* (Guyonneau, Kirchner & Thorpe, soumis). Il est possible que ces données soient réunies dans un article prochainement.

Ces réponses précoces sont à relier aux saccades dites « express », décrites chez le singe puis chez l'homme dans des tâches utilisant un intervalle de temps fixe entre l'apparition du point de fixation et du stimulus (procédure de « gap interval », voir (Fischer & Boch, 1983; Fischer & Ramsperger, 1984). L'observation de distributions bimodales a permis de considérer ces saccades rapides comme une catégorie différente des réponses plus tardives dites « standards ». Cependant, les conditions de production de saccades express ne sont pas bien définies, et dépendent de nombreux facteurs expérimentaux.

D'abord, et même chez l'animal entraîné, la probabilité de produire des saccades express varie énormément entre sujets : certains singes ne produiront jamais de saccades express même après un entraînement intensif alors que d'autres peuvent générer jusqu'à un tiers de leurs réponses dans ces gammes de latences même sans aucun « gap interval » (Schiller, Sandell & Maunsell, 1987). Cette variabilité rejoint en cela les observations que nous avons faites sur notre groupe de sujet puisque seulement 2 sujets sur 8 présentent une bimodalité. Cette remarque permet aussi de conclure que l'utilisation du « gap interval » n'est pas indispensable pour générer des saccades express (Boch & Fischer, 1986; Schiller, Sandell & Maunsell, 1987).

D'autre part, la fréquence de ces réponses express augmente avec l'excentricité de la cible (Weber, Aiple, Fischer & Latanov, 1992) mais diminue avec le nombre de positions possibles de la cible (Pare & Munoz, 1996). De la même manière, aucune saccade express n'est produite si la cible est présentée en même temps qu'un set de distracteurs (McPeck & Schiller, 1994; Weber & Fischer, 1994), ce qui laisse penser qu'elles ne peuvent pas être générées si la tâche consiste à la fois à localiser la cible dans l'espace et à la sélectionner parmi des distracteurs. Notre tâche de discrimination consiste justement à localiser une cible sélectionnée en défaveur d'un distracteur, mais contrairement aux travaux ci-dessus, le set de distracteur est réduit à un seul stimulus.

La latence moyenne du pic de saccades express enregistrées chez nos deux sujets est plutôt tardive en comparaison des latences de la littérature sur le sujet, qui détermine ce type de réponses précoces dès 100 ms chez l'homme et 70 ms chez le singe. Or il semblerait que les

latences des saccades express soient particulièrement liées aux paramètres physiques des stimuli, que ce soit le contraste, la taille ou l'excentricité (Boch, Fischer & Ramsperger, 1984; Fischer & Weber, 1993; Schiller, Slocum, Carvey & Tolia, 2004). Les scènes naturelles sont des stimuli dont les paramètres physiques varient beaucoup entre item et que nous n'avons pas cherché à optimiser pour obtenir des réponses express, il n'est donc pas surprenant d'observer chez nos deux sujets des latences de réponses un peu plus tardives que celles reportées classiquement.

Enfin dans une étude récente (Schiller, Slocum, Carvey & Tolia, 2004), Schiller a montré qu'il est possible d'observer des saccades express dans des conditions de vision naturelle (en présentant une image naturelle pendant une dizaine de secondes à l'écran, ou en laissant les singes libres de regarder le laboratoire autour d'eux, ou encore en présentant une vidéo). Mais la distribution observée, bien qu'elle démarre tôt en latence, suit une forme unimodale contrairement à la condition où les animaux doivent localiser une cible pouvant apparaître à 4 endroits différents sur l'écran et précédée d'un « gap interval » de 100 ms. La distribution des temps de réaction saccadiques est dans ce dernier cas clairement bimodale. Les auteurs concluent que la bimodalité ne peut être observée qu'en condition de laboratoire, car les traitements reposent sur des contextes de vision différents : dans le cas de vision naturelle, la bimodalité de la distribution peut être masquée par une plus grande variabilité des temps de réaction saccadiques alors que le sujet doit non seulement analyser le stimulus dans son champ de vision central mais aussi décider où diriger la saccade suivante. En revanche, dans une tâche de sélection oculaire, le stimulus central -c'est à dire le point de fixation- est connu et reste le même dans tous les essais, générant un degré d'analyse limité.

Finalement les conditions expérimentales d'apparition des saccades express commencent à se définir plus précisément, mais il reste la question centrale de la nature de ces saccades. Sont-elles la conséquence d'une anticipation, ou d'un état d'alerte permettant au système prémoteur de se préparer à répondre en même temps que le système visuel intègre ses informations ? L'hypothèse de réponses anticipées semble maintenant exclue, d'une part parce que la précision obtenue avec des réponses précoces est largement au dessus du niveau chance, d'autre part parce que les latences des saccades express varient avec les paramètres physiques des stimuli, ce qui paraît incompatible avec un mode anticipatoire. De plus, ils semblerait que les réponses anticipées soient isolées dans la distribution, apparaissant en général avant 85 ms (Bronstein & Kennard, 1987; Fischer & Weber, 1992; Weber, Aiple, Fischer & Latanov, 1992). Enfin, il a été observé que les patients avec des lésions du lobe frontal peuvent toujours générer des saccades express mais sont incapables de fournir des réponses anticipées (Braun, Weber, Mergner & Schulte-Monting, 1992). Toutes ces données suggèrent que les réponses express sont dirigées par une composante visuelle, c'est à dire qu'elles ont besoin d'un trigger visuel pour pouvoir être initiées contrairement aux réponses anticipées. Il semblerait que la disparition du point de fixation fournit un signal d'alerte qui apporte une information temporelle sur le moment auquel va apparaître la cible (Ross & Ross, 1980; Ross & Ross, 1981). Certains auteurs font l'hypothèse que les saccades express sont le résultat de mécanismes de facilitation dans la programmation du signal prémoteur au niveau du colliculus supérieur (Reuter-Lorenz & Fendrich, 1992). Différents modes de facilitation peuvent être envisagés, faisant intervenir le désengagement de l'attention du point de

fixation au moment où celui-ci disparaît (Cavegn & Biscaldi, 1996; Cavegn & d'Ydewalle, 1996; Fischer & Weber, 1993) ou une augmentation de l'activité dans le réseau des neurones de mouvement (Cavegn & Biscaldi, 1996; Cavegn & d'Ydewalle, 1996; Findlay & Walker, 1999).

Malgré toutes ces données, il reste difficile d'expliquer une telle variabilité inter-sujets, pourquoi certains sujets fournissent majoritairement des saccades express alors que d'autres n'ont aucune réponse en dessous de 200 ms comme c'est le cas dans notre expérience ? Le niveau d'entraînement dans la tâche oculaire était équivalent sur tous les sujets, mis à part peut-être 2 sujets relativement plus entraînés que les autres, et dont les distributions sont cependant parfaitement unimodales (Figure 3.7C).

Dans tous les cas, les réponses enregistrées sur les sujets MLA et LRE confirment que l'ensemble des traitements visuels, décisionnels et moteurs permettant de déterminer laquelle des deux images contient un animal peut être accompli dans un délai aussi court que 120 ms.

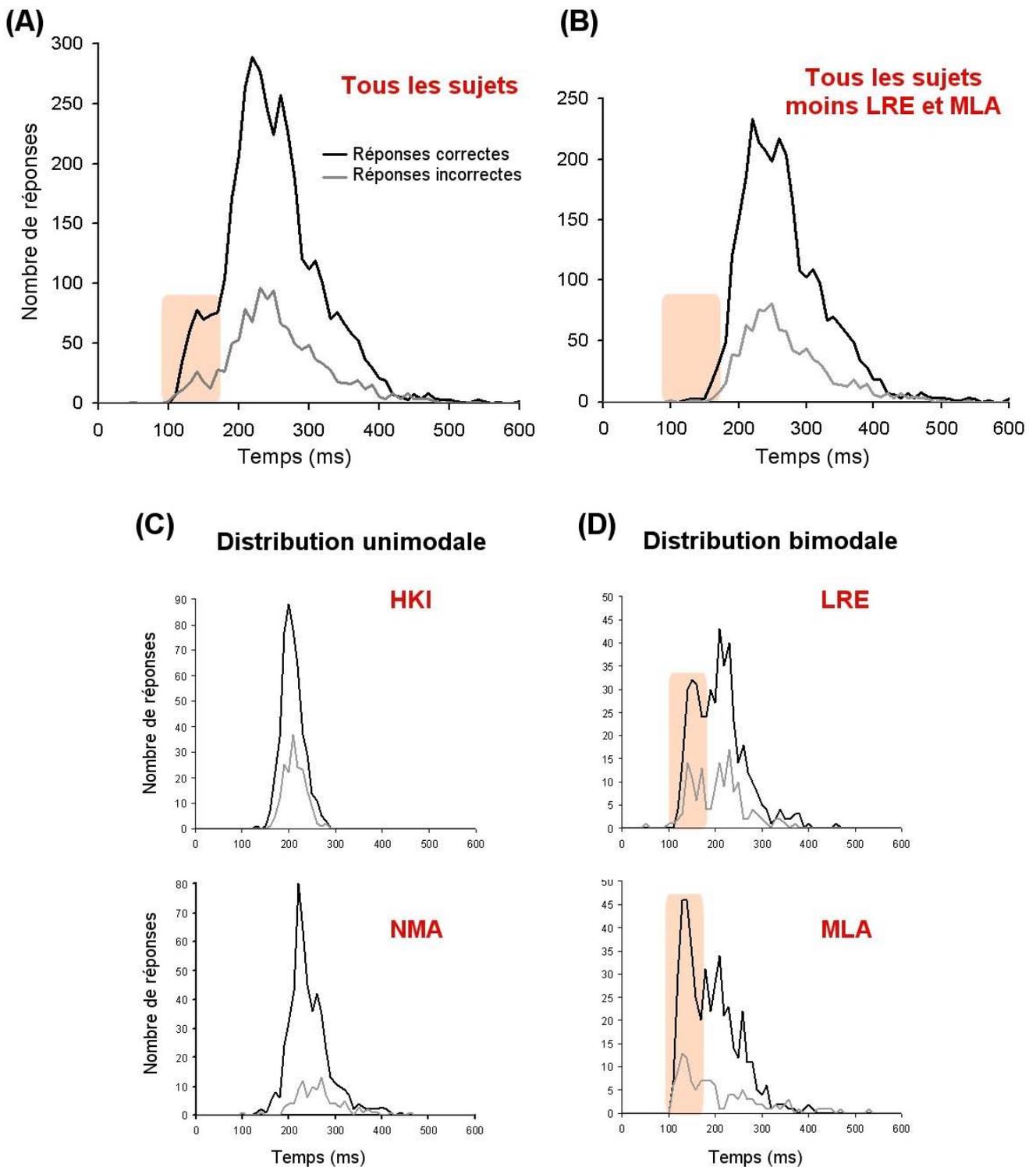


Figure 3.6 : Distributions des temps de réaction pour les réponses correctes et incorrectes, sur l'ensemble des conditions de SOA. (A) Distributions des temps de réaction pour les 8 sujets. (B) Distributions des temps de réaction pour tous les sujets, après avoir enlevé les données des sujets LRE et MLA, présentant une distribution bimodale. (C) Distributions des temps de réaction pour deux sujets caractéristiques du mode unimodal. (D) Distributions des temps de réaction pour les deux sujets montrant de la bimodalité. Ces deux sujets ont des temps de réaction moyens particulièrement courts (211 ms et 195 ms). Malgré le nombre important de réponses incorrectes générées dans la première partie de la distribution, le pas de TR minimal est inférieur à celui des autres sujets (130 et 120 ms, $p < 0.05$). La partie colorée en rouge clair met en évidence la zone de la distribution concernée par les réponses rapides de la première phase de réponse bimodale.

3.6 Discussion

3.6.1 Quelle que soit la tâche, les traitements visuels commencent par une phase commune d'analyse sensorielle

Lorsque le masque est présenté juste après l'image, les traitements visuels sont fortement affectés comme le montrent les mauvaises performances. D'après ce que nous avons vu dans le Chapitre 2, l'information disponible sur les propriétés bas-niveau de l'image est limitée à cause des interférences entre l'image et le masque, supposées avoir lieu assez tôt dans la hiérarchie corticale, probablement avant V4 (Bacon-Macé, Mace, Fabre-Thorpe & Thorpe, 2005), voire peut-être même au niveau de la rétine. Seule l'information qui a pu « échapper » à cette interférence peut être utilisée ultérieurement dans le système. Cette idée est renforcée par l'absence de compromis rapidité-précision pour des SOA inférieurs à 31 ms, ce qui implique que le système n'arrive pas à manipuler l'information pour extraire d'autres indices permettant de préciser une possible interprétation du stimulus.

Or ce qui est particulièrement intéressant dans cette deuxième étude, c'est que la sensibilité au masquage est identique dans les 4 expériences pour des SOA inférieurs à 31 ms, ce qui suggère que la quantité d'information extraite sur les stimuli semble similaire. Quelle que soit la nature de la tâche et le mode de réponse, le système visuel semble intégrer les informations avec la même efficacité, et la perturbation des traitements par un masque conduit aux mêmes limitations. Cette observation va dans le sens de nos hypothèses de départ sur le parcours suivi par les informations entre les voies sensorielles et motrices. Si les 30-40 premières millisecondes de SOA indiquent une sensibilité similaire entre les tâches de catégorisation et de discrimination, il y a de fortes chances pour que les traitements sensoriels soient communs aux deux tâches dans cette première phase, confirmant une intégration commune des informations peut-être jusque V4.

3.6.2 En fonction de la tâche et du mode de réponse, les composantes décisionnelles et motrices se dissocient dans le temps

Pour des SOA supérieurs à 44 ms, la précision augmente et atteint un plateau dans les 4 expériences, au dessus de 80 %. Cette amélioration s'explique facilement puisque l'interférence entre l'image et le masque intervient plus tard, laissant au système visuel plus de temps pour intégrer l'information sur le stimulus. Mais il est intéressant d'observer qu'à partir de ce SOA, les mesures de d' indiquent une sensibilité différente entre les 4 expériences, qui peuvent être résumées ainsi :

- Les tâches de discrimination ont des valeurs de d' inférieures à celles des tâches de catégorisation,
- Les tâches de discrimination, qu'elles soient saccadique ou manuelle, suivent un profil de d' parfaitement superposé pour tous les SOA,
- Dans les tâches de catégorisation, le d' est légèrement meilleur pour le go/no-go que pour le yes/no.

Impact de la nature de la tâche : comparaison des tâches de catégorisation et de discrimination

Au moins deux hypothèses peuvent expliquer un d' inférieur dans les tâches de discrimination en comparaison aux tâches de catégorisation. D'une part, les données entrant dans le système visuel diffèrent dans les deux types de tâche, puisque dans le cas de la catégorisation, une image centrale est présentée alors que dans le cas de la discrimination, deux images sont présentées à 6° du point de fixation. Or l'impact de l'excentricité de la cible a déjà été mesurée dans ce type de tâche, avec une baisse de 3-4% de précision et une augmentation d'environ 15 ms du TR lorsqu'une scène naturelle est présentée à $4-13^\circ$ du centre (Fize et al., 2005; Thorpe, Gegenfurtner, Fabre-Thorpe & Bulthoff, 2001).

Une deuxième explication porte sur le caractère peut-être plus « automatique » de la tâche de discrimination. Les TR dans les tâches de discrimination sont particulièrement stables quel que soit le SOA, contrairement aux tâches de catégorisation. Non seulement les latences sont plus courtes, mais elles semblent également insensibles aux variations de difficultés entre les différentes conditions de la tâche. Il semblerait que les réponses fournies en condition de choix forcé se basent sur un mode de déclenchement de la réponse plus automatique (Romo & Salinas, 2003; Salinas & Romo, 1998; Schall, 2003), c'est à dire que la préparation motrice serait déjà bien avancée au moment où l'analyse sensorielle a permis de construire une représentation suffisamment cohérente du stimulus pour initier une réponse. Ce mode de fonctionnement présente l'avantage d'autoriser des temps de réaction courts, mais il peut aussi entraîner une précision légèrement inférieure à cause des processus de sélection de la réponse un peu précipités. Prenons le cas des réponses saccadiques en particulier, dont les circuits pré-moteurs et moteurs sont mieux connus sur le terrain électrophysiologique de la prise de décision. Dans cette tâche, non seulement les TR sont stables malgré la difficulté mais ces TR sont en plus particulièrement courts. On peut alors poser l'hypothèse qu'il existerait une sorte de « générateur de temps de réaction », qui ordonnerait au système moteur de déclencher une réponse à des latences données, que le traitement de l'information sensorielle ait déjà donné des résultats pertinents ou non. Dans le cas où les conditions de perception sont optimales, la réponse générée sera probablement correcte, mais dans le cas où la quantité d'information sensorielle est réduite, la probabilité de générer une réponse juste sera bien moindre.

Comme nous l'avons vu dans le chapitre précédent, le FEF (mais il semble que ce soit aussi vrai dans le CS ou dans la région pariétale LIP) semble contenir deux types de neurones participant tous les deux aux processus de transformation du signal sensoriel en signal moteur (Figure 3.7A). Le premier type de neurones répond dans un premier temps de manière uniforme à tout stimulus qui peut potentiellement être une cible pour la tâche, puis le signal diffère après un certain temps selon le statut du stimulus : si c'est une cible, le signal est maintenu à un niveau de décharge important alors que si c'est un distracteur l'activité est fortement réduite (Schall, 2002; Schall & Bichot, 1998). Lorsque la cible et le distracteur se ressemblent, le délai de divergence significative entre les signaux se rallonge. Cependant ces neurones, qui semblent participer à la transformation du signal sensoriel en signal décisionnel, sont distincts des neurones qui interprètent plus spécifiquement la décision et déterminent la réponse motrice. Alors que les

premiers ont un profil d'activité constant même lorsque la réponse motrice n'a pas lieu et n'ont pas de latences de décharge corrélées avec les temps de réaction, le second groupe de neurones en revanche a un taux de décharge croissant avec une pente liée à la vitesse à laquelle les informations sensorielles s'accumulent dans le système : lorsque les stimuli cibles et distracteurs se ressemblent, l'activité croît avec une pente moins forte qu'avec des stimuli plus dissociables sur le plan de leurs indices diagnostiques. Ainsi la saccade est initiée seulement lorsque l'activité atteint un certain seuil d'activité -plus ou moins consciemment sous l'influence du critère de décision du sujet- et le délai pour atteindre ce seuil semble être corrélé aux temps de réaction comportementaux. L'observation de compromis rapidité-précision peut donc tout à fait s'expliquer par une variation du critère de décision du sujet lorsque les conditions de perception sont correctes. Dans les conditions les plus dégradées, s'imposer des latences de réponses plus tardives en augmentant le seuil de réponse est beaucoup moins utile puisque de toute façon, l'information sensorielle disponible est limitée. Ceci explique l'absence de compromis rapidité-précision pour les SOA les plus courts.

Cependant, il a aussi été remarqué qu'une partie de la variabilité des TR ne peut pas s'expliquer uniquement par la pente de décharge neuronale plus ou moins importante selon la quantité d'informations sensorielle disponible (Carpenter, 1999; Ratcliff, Carpenter & Reddi, 2001). La source de cette variabilité supplémentaire dans les processus de sélection de la réponse ne semble pas bien définie, mais certains auteurs relient l'introduction de ce facteur aléatoire avec la Théorie des Jeux : un système imprédictible fonctionnant avec une certaine proportion de paris pour sélectionner sa réponse est un système écologiquement efficace (Carpenter, 1999) et voir le site tutoriel sur le modèle LATER <http://www.cudos.ac.uk/later.html>). Cette variabilité des TR est tout à fait compatible avec l'idée qu'il existerait un générateur indépendant de latences, dont la variance de la distribution expliquerait la variabilité des TR non corrélée au bruit sensoriel (Figure 3.7B). L'origine de ce générateur peut être envisagée au niveau de structures que l'on sait impliquées dans la régulation des rythmes biologiques comme la formation réticulée du tronc cérébral par exemple. Une autre hypothèse encore plus simple impliquerait les structures pré-motrices et motrices elles-mêmes, par l'intermédiaire du circuit qui relie les aires visuelles précoces (voire sous-thalamiques) aux structures motrices (Figure 3.3). Par exemple, il semble que les premières réponses enregistrées dans le FEF après la présentation d'un flash lumineux ont des latences aussi courtes que 40 ms (Nowak & Bullier, 1997; Schmolesky et al., 1998; Thompson, Hanes, Bichot & Schall, 1996). Même si ces réponses ne sont pas sélectives aux propriétés du stimulus visuel, il est possible d'imaginer qu'elles constituent un signal de départ pour préparer le système moteur à générer une réponse après un intervalle de temps relativement stable. C'est cette stabilité que l'on pourrait retrouver dans notre tâche de discrimination saccadique, même lorsque le SOA -et donc la quantité d'information sensorielle disponible- est faible.

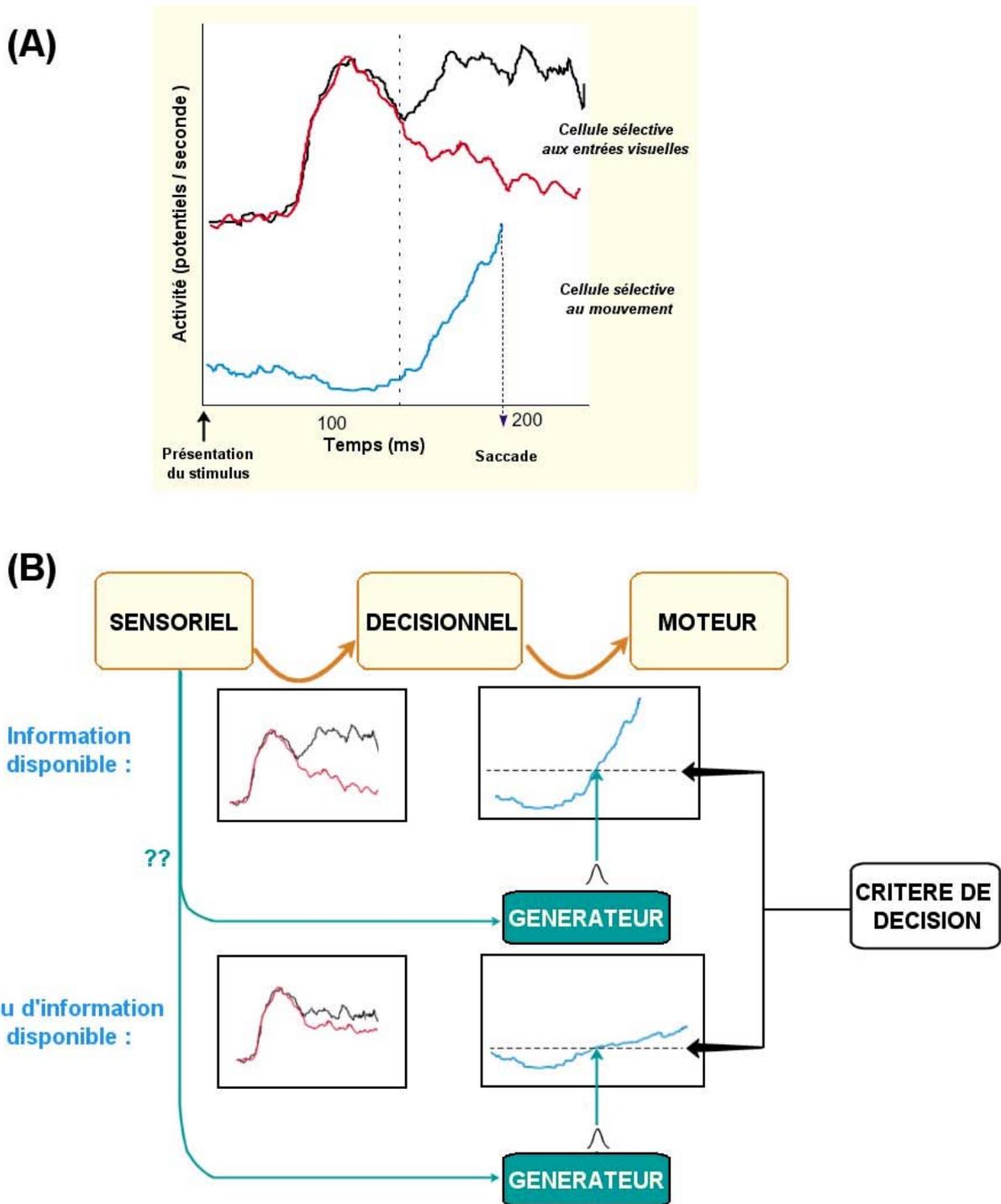


Figure 3.7 : Transformation du signal sensoriel en signal moteur. (A) Activité d'un neurone du FEF lorsqu'une cible (courbe noire) ou un distracteur (courbe rouge) apparaît. La partie initiale de la réponse est commune aux deux signaux qui divergent après un certain temps. L'activité pour la cible est maintenue alors que celle pour le distracteur décroît. Ces neurones participent à la transformation du signal sensoriel en signal décisionnel. Un deuxième groupe de neurones en revanche a des réponses dont la latence est corrélée aux TR saccadiques et semble refléter la transformation du signal décisionnel en signal moteur (courbe verte). Adapté de Reddi, 2001 (B) La pente de décharge des neurones sélectifs au mouvement dépend de la quantité d'information sensorielle disponible. Il est possible d'envisager l'existence d'un générateur de temps de réaction, sous l'influence des données sensorielles précoces à l'entrée du système, qui permettrait de déclencher la réponse motrice avec des latences relativement uniformes, ce qui pourrait expliquer que les TR obtenus dans notre tâche de discrimination saccadique soient si stables même pour les

SOA les plus courts. Le seuil de réponse est alors sous l'influence de deux paramètres : la variabilité dans ce générateur ainsi que la variabilité imposée plus ou moins consciemment par le sujet selon son critère de décision.

Il est intéressant de se demander si ce générateur est propre au système oculaire ou si la préparation motrice engagée pour les réponses manuelles peut aussi reposer sur un fonctionnement similaire. Tout d'abord, il faut rappeler que les temps de réaction manuels sont beaucoup plus longs que les réponses saccadiques. D'autre part d'après nos résultats, les TR manuels augmentent légèrement avec la difficulté dans les tâches de catégorisation. On ne retrouve donc pas dans ces tâches ni la stabilité des TR que nous trouvons pour les tâches de discrimination, ni de compromis rapidité-précision et ce, quel que soit le SOA. Cependant, les contraintes « écologiques » imposées pour des réponses manuelles sont beaucoup plus strictes que celles imposées sur les réponses oculaires : si nous générons des saccades trois fois par seconde sans jamais mettre en péril nos yeux, les mouvements manuels en revanche ont plutôt intérêt à être précis pour éviter de placer la main dans une position risquée (exemple classique : poser la main sur un serpent !), même si cette précision se paye en terme de délai de réponse. Il n'est donc pas étonnant d'observer une automaticité des traitements beaucoup moins importante avec des réponses manuelles et une capacité de modulation par des compromis rapidité-précision réduite. Cependant, l'existence d'un générateur paraît encore envisageable, à partir du moment où l'on considère que la réponse sera de toute façon relativement tardive. Lorsque la quantité d'information est suffisante, les TR seront donc longs, mais la précision (en terme de d' dans nos expériences) sera meilleure que pour la tâche oculaire. Lorsque la quantité d'information est réduite, pour des SOA courts par exemple, la précision sera comparable puisque les traitements supplémentaires n'auront pas permis d'extraire d'avantage d'indices diagnostiques sur le stimulus.

Une question sous-jacente à cette comparaison entre tâches de catégorisation et de discrimination porte sur la possibilité que les tâches de choix forcé puissent bénéficier des mécanismes de comparaison des informations issues des deux stimuli. Contrairement aux tâches de catégorisation, dans les tâches de discrimination, deux sources d'informations sont présentées et le système doit décider laquelle des deux sources est la plus pertinente pour représenter la cible. Est-ce que l'information s'accumule en parallèle et de manière indépendante autour de chaque source d'information jusqu'à ce que l'une d'elle atteigne le seuil décisionnel, ou est-ce que la décision est prise grâce à un processus de comparaison des informations issues des deux sources ? Autrement dit, la quantité d'information accumulée doit-elle dépasser un seuil donné sur l'une ou l'autre des sources ou doit-elle dépasser celle de l'autre source d'un écart donné ? Ces questions font référence aux modèles de la psychologie expérimentale développés pour analyser les paramètres intervenant dans la prise de décision (Carpenter & Williams, 1995; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004). Certains auteurs ont essayé de concilier les modèles statistiques développés depuis une vingtaine d'années pour prédire les réponses comportementales avec les données expérimentales qui ont permis d'enregistrer l'activité neuronale dans des structures majoritairement impliquées dans les processus décisionnels (Glimcher, 2001; Glimcher, 2003; Hanes & Schall, 1996; Shadlen, Britten, Newsome & Movshon, 1996; Smith & Ratcliff, 2004;

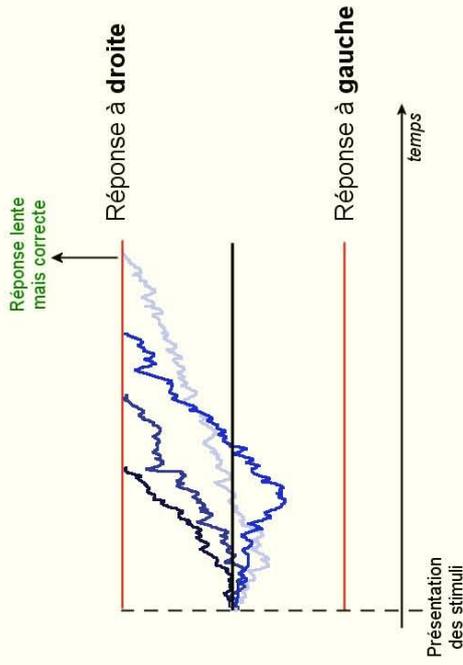
Sugrue, Corrado & Newsome, 2005). La figure 3.8 présente les deux hypothèses au travers de schémas simplifiés. Pour chaque hypothèse, deux situations sont envisagées selon que les seuils décisionnels sont fixés haut (auquel cas la réponse est lente mais précise) ou plus bas (auquel cas la réponse est rapide mais moins précise). Les deux types de modèles, qui fonctionnent sur une analyse soit indépendante soit comparée des deux sources d'information, sont en fait capables d'expliquer le phénomène du compromis rapidité-précision, et nous ne pourrions pas utiliser cet indice pour valider l'un ou l'autre des modèles (Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004). D'autres expériences de choix forcé seraient nécessaires, sans forcément utiliser un protocole de masquage, mais en manipulant par exemple les niveaux des seuils à travers une variation de la fréquence des cibles entre séries. En général, il est observé que si la fréquence de cible est supérieure à 50 % dans une série donnée, les seuils d'activation à atteindre sont abaissés (Ratcliff & Rouder, 1998). Il serait intéressant d'observer lequel des deux modèles valide le mieux les réponses comportementales observées avec ce type de manipulation.

De manière générale, la comparaison des tâches de catégorisation et de discrimination soulève des questions fondamentales sur la manière dont les informations visuelles sont évaluées et transformées en signal moteur. Même si certains points restent encore en suspens, il paraît clair que la manipulation de certains paramètres de ces protocoles permettra probablement d'apporter dans le futur des réponses fiables à ce qui représente maintenant un thème de recherche essentiel : la prise de décision.

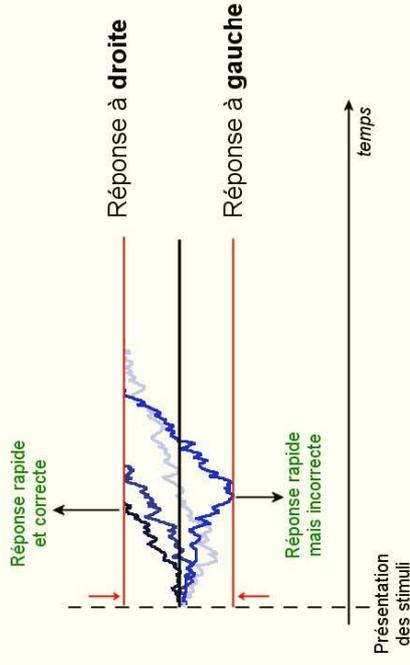
Figure 3.8 (page suivante): *Deux modèles d'accumulation de l'information dans le contexte d'une tâche de choix forcé, alors que le système doit choisir entre deux possibilités : réponse à droite ou à gauche.* (A) Schémas représentatif d'un modèle de diffusion, pour lequel les informations relatives à une source d'information sont autant d'indices ne s'accumulant pas pour l'autre source. (B) Schémas basés sur les modèles de compétition, pour lesquels les deux sources d'informations s'accumulent indépendamment et en parallèle. Chaque tracé bleu représente l'accumulation des indices permettant de décider si la réponse a lieu du côté droit ou gauche dans un essai donné. Les fluctuations du signal reflète le bruit dans le processus de décision. Les différentes courbes bleues, avec des pentes variables, se réfèrent au fait que la vitesse avec laquelle les informations s'accumulent dépend des propriétés du stimulus. Plus les informations sur la cible sont pertinentes (par exemple en terme de formes), plus les informations s'accumulent vite et la réponse déclenchée sera rapide et correcte. L'accumulation des indices se fait jusqu'à atteindre un seuil (lignes rouges). Mais selon le contexte de la tâche, le niveau des seuils à atteindre fluctue. Différents paramètres peuvent expliquer ces variations, comme la consigne donnée au sujet d'être précis (dans ce cas les seuils sont élevés, c'est à dire qu'il faut avoir une quantité importante d'informations pour déclencher la réponse) ou rapide (dans ce cas, les seuils sont abaissés, ce qui explique un plus grand nombre d'erreurs). Ainsi les deux modèles sont capables d'expliquer le phénomène du compromis rapidité-précision.

**Analyses
COMPAREES
des deux sources
d'information**

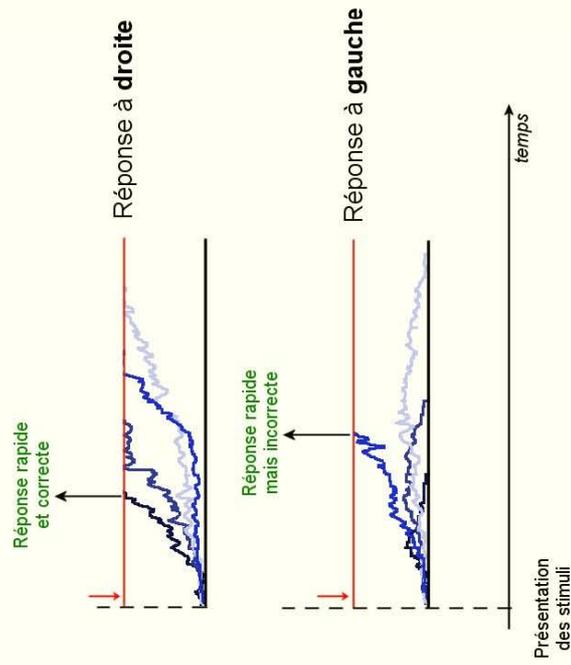
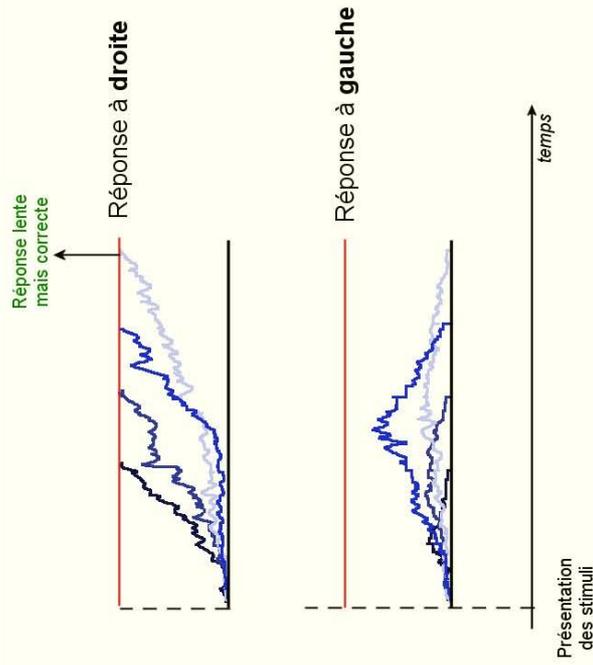
Seuils élevés
(par exemple,
consigne de précision)



Seuils bas
(par exemple,
consigne de rapidité)



**Analyses
INDEPENDANTES
des deux sources
d'information**



Impact du mode de réponse : comparaison des deux tâches de discrimination

Quel que soit le SOA, les valeurs de d' sont identiques entre les deux tâches de discrimination, que les réponses soient saccadiques ou manuelles. Pourtant, les temps de réaction diffèrent énormément avec 250 et 404 ms de TR moyens sur l'ensemble des SOA. Mais encore une fois, il faut noter que dans les deux tâches, le TR reste particulièrement stable quelles que soient les conditions de difficulté. Ces observations laissent penser qu'une grande partie des traitements est commune aux deux tâches, puisque le masquage affecte les traitements de la même manière dans les deux cas. Même si nous avons vu qu'il est difficile de poser des hypothèses sur les circuits neuronaux impliqués pour réaliser ce type de tâches, il est raisonnable de penser que les circuits doivent différer principalement à la fin de la boucle sensori-motrice, au moment de l'encodage moteur des informations intégrées. La différence de 150 ms dans les temps de réaction doit probablement pouvoir être expliquée par un trajet beaucoup plus long des informations lorsque la réponse est manuelle, alors que l'effecteur moteur est plus éloigné et les étapes motrices plus nombreuses. En revanche, une sensibilité similaire entre les deux tâches quelque soit le SOA suggère que la quantité d'information est comparable, et que seule la nature de la tâche -ici une discrimination- détermine le niveau d'analyse sensorielle nécessaire pour déclencher une réponse. Que le mode de réponse soit oculaire ou manuel, il s'agit certainement d'un facteur important, mais intervenant relativement tard dans la chaîne des processus et n'affectant pas les traitements précoces.

Des résultats récents viennent d'ailleurs confirmer cette observation. Dans un article de 2004, Brenner et Smeets ont développé un protocole original dans lequel les sujets devaient effectuer un mouvement du bras pour pointer le plus rapidement possible une cible, un carré de couleur rouge, présentée dans l'un des deux hémichamps sur l'écran (Brenner & Smeets, 2004). Dans la moitié des essais, un carré vert apparaissait dans l'autre hémichamp. Dans certains cas, dès que le mouvement était déclenché, la position des deux stimuli était inversée à l'écran, obligeant le sujet à modifier la trajectoire de son mouvement pour pointer correctement la position de la cible. Les auteurs concluent que 120 ms sont suffisantes pour ajuster le mouvement, ce qui suppose que les traitements de reconnaissance de la couleur et de modification de la trajectoire du mouvement ont pu être exécutés en un temps aussi court. Or dans une étude toute récente, ces mêmes résultats ont été retrouvés, non pas avec des carrés de couleur, mais avec des scènes naturelles contenant ou non un animal ! Autrement dit, les 120 ms enregistrées dans cette tâche de pointage ainsi que dans nos tâches de choix forcé supposent que les traitements de discrimination entre deux stimuli peuvent être produits extrêmement rapidement, indépendamment du mode de réponse (manuel ou oculaire).

Impact du nombre d'effecteurs : comparaison des deux tâches de catégorisation

Pour les SOA les plus longs, les mesures de d' indiquent une meilleure sensibilité de la tâche de catégorisation go/no-go en comparaison à la tâche yes/no. Dans les deux expériences, les sujets se posaient la question « Y'a t-il un animal dans l'image ? ». Mais dans le premier cas, ils devaient répondre avec une seule main, en levant le doigt uniquement si la réponse était positive.

Dans le deuxième cas, ils devaient répondre avec leurs deux mains, à la fois pour une réponse positive ou négative. Non seulement la sensibilité est plus affectée par le masquage dans l'expérience yes/no, mais les temps de réaction sont aussi augmentés, quels que soient les SOA, avec une différence moyenne de plus de 50 ms. Des différences entre ces deux types de tâches ont déjà été reportées dans le contexte d'expériences de décision lexicale (Hino & Lupker, 2000; Perea, Rosa & Gomez, 2002; Siakaluk, Buchanan & Westbury, 2003), elles sont en général attribuées à une forme de compétition intervenant au niveau des étapes décisionnelles et motrices dans la tâche yes/no. En effet, alors que les sujets ont à répondre sur l'un des deux boutons, deux programmes moteurs doivent être préparés et un seul initié. D'autre part au niveau décisionnel, s'il paraît naturel de répondre alors qu'un certain nombre d'indices se sont accumulés en faveur de la présence d'une cible, il est beaucoup plus difficile de décider que le stimulus ne contient pas cette cible, ce qui justifie un augmentation considérable des temps de réaction.

3.6.3 Validation du protocole oculaire pour l'étude de la reconnaissance d'objet

Une des conclusions qui est apparue au fur et à mesure de l'analyse de cette série de données, c'est que la tâche de discrimination oculaire paraît particulièrement bien adaptée pour analyser les traitements précoces de la reconnaissance d'objet. La sensibilité identique dans la première phase de SOA entre les deux types de tâches -catégorisation et discrimination- suggère qu'une partie considérable des traitements sensoriels est similaire. Or les temps de réaction sont presque deux fois plus courts ! Des plages de temps de réaction comprises entre 120 et 250 ms posent d'avantage de contraintes sur le système visuel pour extraire et intégrer l'information pertinente dans la tâche. D'autre part, comme nous l'avons vu dans la section précédente, les résultats obtenus dans les tâches de choix forcé suggèrent que l'enchaînement des traitements allant de l'analyse sensorielle au déclenchement de la réponse motrice semble particulièrement optimisé. Par ailleurs, les erreurs générées par ce mode de traitement à caractère automatique pourraient être une source d'analyse supplémentaire. En effet, en plus des questions concernant le mode de transfert de l'information et le parcours suivi tout au long de la hiérarchie corticale, ce type de protocole doit permettre d'étudier, à travers l'analyse des réponses incorrectes ou à travers la manipulation des espaces de recherche entre groupe d'image cible et distracteur, les éléments diagnostiques servant à la reconnaissance des scènes naturelles.

3.6.4 Vision subliminale et conscience

Dans la tâche de discrimination saccadique simple (sans masquage), un phénomène surprenant avait été évoqué par les sujets (phénomène que je peux bien décrire, pour l'avoir ressenti moi-même). Les sujets décrivaient une impression de pouvoir déclencher la saccade vers une image sans trop savoir pourquoi, pour se rendre compte une fois que leur regard était posé sur l'image périphérique qu'il s'agissait bien de la cible, comme si il y avait un décalage dans le temps entre leur réponse motrice et leur compréhension de la scène. Certains sujets avaient même l'impression de déclencher leur réponse avant que les images n'apparaissent (ce qui se révélait faux bien sûr à l'analyse des données). Cette sensation était d'ailleurs assez grisante puisque qu'elle donnait l'impression de pouvoir prédire où la cible allait apparaître... Or avec le masquage, le phénomène semblait beaucoup moins marqué chez ces mêmes sujets, voire même assez frustrant car il nous manquait un retour « conscient » sur notre réponse. D'autre part, dans toutes les tâches de masquage, les sujets étaient très pessimistes sur leurs résultats, et tout le monde était sûr d'avoir détecté peu de cibles. Or quand on leur montrait leurs performances réelles 15 minutes après, certains se disaient étonnés d'avoir finalement « pas si mal réussi »³.

Les temps de réaction particulièrement courts, notamment dans la tâche de discrimination où les premières réponses significativement correctes apparaissent dès 120 ms, laissent penser que tous les traitements sensoriels d'analyse de la scène n'ont pas pu être effectués. Certes, des indices diagnostiques ont permis de décider que la scène contenait bien la cible cherchée (en tout cas, plus que celle d'à côté), mais cela ne veut pas dire que la scène a été perçue et intégrée dans sa totalité. Une question fondamentale se pose sur le niveau de traitement sensoriel minimal qui est nécessaire pour déclencher une réponse motrice... A-t-on besoin de développer une interprétation aboutie et consciente de la scène pour déclencher une réponse motrice ? Le fait de relever sa main subitement après l'avoir posé à côté d'une branche que l'on confond avec un serpent laisse penser que non ... On ne peut certainement pas écrire une thèse complète sur des expériences de photographies masquées sans parler de vision subliminale et de la conscience biologique. Les études sur la construction d'une représentation consciente des scènes visuelles se sont multipliées cette dernière décennie, et on comprend en effet que cela puisse constituer une source de questionnement fondamentale pour la compréhension des capacités cognitives humaines.

³ Ce phénomène avait déjà été observé auparavant dans l'équipe dans une tâche où les cibles étaient présentées en périphérie du champ visuel (Thorpe et al., 2001). Toutes les conditions de présentation étaient mélangées dans les séries, ce qui fait que la cible pouvait apparaître entre $-70,5^\circ$ et $+70,5^\circ$ du point de fixation central. La tâche était de catégoriser les animaux en répondant selon le mode go/no-go. Or dans les conditions les plus difficiles d'excentricité, certains sujets ne répondaient quasiment jamais car ils se disaient sûrs de ne pas pouvoir dissocier une cible d'un distracteur. Les expérimentateurs ont dû insister sur le fait qu'ils souhaitaient obtenir un taux de réponse proche de 50%, même dans ces conditions difficiles de perception (un peu comme dans un choix forcé). Avec cette contrainte de réponse, la performance atteignait quand même 60% de pourcentage correct, significativement au-dessus du niveau chance, au grand étonnement des sujets eux-mêmes !

Qu'est ce que la vision subliminale ?

En général, on dit d'une image qu'elle est subliminale lorsqu'elle est présentée tellement brièvement ou avec un contraste tellement faible qu'elle n'est pas perçue consciemment. Dans la publicité (ou en politique, voir Figure 3.9), ces images seraient utilisées dans le but d'influencer le comportement des consommateurs (ou des électeurs) grâce au peu de sens que le système visuel a pu extraire du stimulus. Si aucune étude bien contrôlée ne semble avoir confirmé que des images présentées subliminalement peuvent modifier les attitudes, il est certain qu'elles ont un impact pour la perception immédiate du monde visuel, c'est d'ailleurs un principe de base dans les expériences utilisant la technique de l'amorçage. Comment définit-on la vision subliminale et quels sont les niveaux de traitement qu'elle engage ?



Figure 3.8 : Utilisation d'image subliminale pendant la campagne américaine pour les présidentielles de septembre 2000. Durant la campagne, G.W. Bush a fait diffuser un spot télévisé dans lequel une image contenant le mot RATS est insérée juste après une photo de son adversaire politique Al Gore.

Il semble en fait que de nombreux traitements sensoriels puisse être accomplis sans qu'ils soient nécessairement accessibles à la conscience. Les preuves sont fournies par un ensemble d'études comportementales ou d'imagerie cérébrale, à la fois sur des sujets sains et sur les patients souffrant de différentes formes de pathologies neuropsychologiques. Dans les années 70, Pöppel et collaborateurs montrent que des patients en partie aveugles à cause d'un scotome hémianoptique sont capables de diriger leur regard vers une cible située dans leur champ aveugle alors qu'ils se disent incapables de la détecter (Poppel, Held & Frost, 1973). De la même manière, il a été observé à de nombreuses reprises que la précision comportementale dans des tâches de choix forcé est nettement meilleure que le report conscient de la scène visuelle chez les patients souffrant d'agnosies visuelles (défiance dans la reconnaissance des formes), de prosopagnosie (impossibilité de reconnaître des visages), d'achromatopsie (défiance de la vision des couleurs), de cerveau dédoublé (patients épileptiques dit « split-brain », dont on a séparé chirurgicalement les deux hémisphères) ou encore d'héminégligence (une partie d'un hémisphère, en général droit, est lésé, ce qui rend le patient comme « aveugle » sur une moitié de son champ visuel) (Driver & Vuilleumier, 2001; Milner et al., 1991; Schacter, Buckner & Koutstaal, 1998; Weiskrantz, 1996). Les patients sont parfois capables de performances aussi bonnes que des sujets sains lorsqu'il s'agit de donner une réponse motrice, alors que le report conscient de la scène visuelle ou d'un stimulus en particulier est complètement dégradé. Il est d'ailleurs intéressant d'observer que les

stimuli non perçus consciemment sont analysés non seulement à un niveau perceptuel mais aussi à un niveau sémantique. Par exemple dans une expérience d'amorçage de mots et d'images chez des patients hémipariétaux, lorsque deux stimuli sont présentés simultanément dans les deux hémichamps, le report des sujets pour l'hémichamp contralatéral à la lésion est évidemment au niveau chance (McGlinchey-Berroth, 1997). Mais lorsqu'ils doivent effectuer une tâche de décision lexicale sur un mot présenté au centre de l'écran à la suite des deux stimuli, qui est relié ou non aux images, ils montrent le même degré d'amorçage sémantique quel que soit l'hémichamp, ce qui suppose que l'image présentée dans l'hémichamp aveugle a pu être traitée au moins au niveau sémantique.

De la même manière chez le sujet sain, il existe maintenant de nombreux travaux montrant que les stimuli masqués sont en partie traités par le système visuel, facilitant des traitements ultérieurs au niveau perceptuel, sémantique ou moteur (Bar & Biederman, 1998; Dehaene et al., 2001; Dehaene et al., 1998; Rossetti, 1998). L'idée que des réponses motrices rapides et stéréotypées puissent être générées à partir d'une analyse sensorielle non aboutie a mené certains auteurs à qualifier cet état de mode « zombie » (Campbell, 1970; Crick & Koch, 2003; Dehaene & Naccache, 2001; Dennett, 2001). Ce mode subliminal, qui a un avantage écologique certain, fonctionnerait sur un transfert rapide et feed-forward des informations, à durée de vie réduite si l'étape ultérieure de consolidation neuronale est limitée par la présentation successive d'un masque par exemple (Bullier, 2001; Lamme, 2000 #318; VanRullen & Koch, 2003a). En revanche, la construction d'une représentation consciente semble requérir l'activation d'un réseau de structures, réseau qui reste à déterminer mais qui impliquerait probablement les cortex visuel, pariétal et frontal (Baars, 2002; Bar & Biederman, 1999; Lamme & Roelfsema, 2000; Milner, 1998; Rees, Kreiman & Koch, 2002; Sergent & Dehaene, 2004).

Fonction intégrative de la représentation consciente

Comme le montrent les expériences électrophysiologiques sur le masquage et la rivalité binoculaire, la perception consciente n'est accessible que si la décharge liée au stimulus est forte et prolongée dans le temps, permettant ainsi d'activer des connexions longue distance reliant un ensemble de structures à la fois perceptuelles et motrices (Dehaene & Naccache, 2001; Edelman, 1999; Koch, 2004). Ce fonctionnement implique obligatoirement une sélection drastique des informations qui entrent massivement dans le système sensoriel.

« The central nervous system, like so many people in today's hyperconnected world, suffers from information overload. So much data about the constantly changing environment is streaming in along the sensory pathways that the brain is unable to process all of it in real time. [...] Natural selection pursued a strategy that amounts to summarizing most of the pertinent facts about the world compactly and sending this description to the planning stages to consider the organism's optimal course of action. [...] This few items, labeled with qualia, are sent off to the planning stages of the brain to help decide a future course of action. For example, you may see a dog in front of you, baring its teeth and growling, and an open door to your right. At that moment, everything else is irrelevant.» (Koch, 2004, p233)

Une des fonctions premières de la perception consciente serait donc de faire la différence entre les signaux importants du monde sensoriel et le bruit interne du système ou l'information non pertinente. Encore une fois, cette idée suppose qu'il existe un seuil à atteindre pour que la perception devienne consciente, qui doit certainement être plus haut que le seuil temporel à partir duquel l'activité neuronale devient significativement sélective au stimulus (Rolls, 2004). Finalement, il semblerait qu'il existe au moins trois seuils dans le traitement de l'information sensorielle : un seuil qui correspond à l'activation minimale du système permettant de dissocier l'information sur le stimulus du bruit du système, un seuil permettant de déclencher une réponse motrice adaptée de manière stéréotypée, et un dernier seuil qui permet d'accéder à une représentation consciente du monde sensoriel. Entre le premier et le deuxième seuil, l'information qui n'est pas assez pertinente est rapidement ignorée alors qu'entre le deuxième et le troisième seuil, le système fonctionne en mode subliminal non conscient (zombie) et peut produire des réponses motrices. Ca n'est qu'au dessus du troisième seuil que pourrait émerger une représentation consciente de ces informations. Ces différents processus implicites de sélection ont une importance fondamentale pour les traitements décisionnels, car la représentation compacte du monde sensoriel pendant un intervalle de temps prolongé permet au cerveau de choisir entre différentes alternatives d'action, et il est d'ailleurs intéressant de se demander dans quelle mesure elle participe à encoder une trace mnésique de ce qui a été perçu par le système sensoriel (Sergent, Baillet & Dehaene, 2005). Mais l'augmentation de l'efficacité du système décisionnel se paie par un coût temporel, et il semble qu'il faille attendre plusieurs centaines de millisecondes pour pouvoir construire une représentation consciente (Di Lollo, Enns & Rensink, 2000; Koch, 2004; Lamme, 2000).

Mesure non objective du phénomène conscient

La majorité des études qui cherchent à déterminer par des techniques d'imagerie le rôle éventuel de telle ou telle structure dans la construction d'une représentation consciente, essaie de corréler l'activité enregistrée dans une région donnée au report subjectif des sujets (Bar & Biederman, 1999; Bar et al., 2001; Kanwisher, 2001; Rees, 2001; Rees, Kreiman & Koch, 2002; Tong, Nakayama, Vaughan & Kanwisher, 1998; Watson, Valenstein, Day & Heilman, 1994). Or un grand problème soulevé par les études sur la conscience porte sur la mesure non objective du phénomène conscient. La conscience est un phénomène introspectif, et le report subjectif des sujets pose souvent un problème de précision, voire de fiabilité (Gazzaniga, LeDoux & Wilson, 1977; Merikle, Smilek & Eastwood, 2001).

La table 3.2 résume différentes mesures d'évaluation du phénomène conscient dans quelques études de masquage. Les méthodes utilisées sont très variables selon les études mais on peut globalement distinguer trois niveaux d'analyse qui portent soit sur :

- la présence ou non de la cible,
- une propriété en particulier de la cible,
- une sélection de son identité parmi un nombre limité de possibilités ou son identification libre.

S'il n'existe certainement pas de mesure parfaite pour reporter le phénomène conscient, il est important de bien regarder la méthode de report qui a été choisie lorsque l'étude cherche en plus à

en déterminer les corrélats anatomiques. Dans tous les cas, cette échelle graduelle dans l'analyse de la scène visuelle suggère que le phénomène conscient n'obéit peut-être pas à une loi du tout-ou-rien comme cela a été proposé, mais que son établissement se fait de manière plutôt progressive (Baars, 2002; Bar et al., 2001; Koch, 2004). Nous reviendrons sur cette idée dans le chapitre suivant, à travers la présentation d'un ensemble de travaux en RSVP.

Etude	Stimuli	Mesure du phénomène conscient
Bar & Biederman, 1998; Bar & Biederman, 1999	Dessins d'objets masqués	Dénomination libre des objets ou en choix forcé parmi 4 noms de catégorie d'objets
Dehaene et al., 1998	Nombre masqué écrit en chiffre ou en lettres (stimulus d'amorçage)	- Dénomination (lecture) - Stimulus absent / présent ? - Stimulus chiffre / lettre ?
Bar et al., 2001	Dessins d'objets masqués	4 niveaux de reconnaissance : - a reconnu l'identité de l'objet - a une idée sur la forme mais pas sur l'identité - a vu une image - n'a même pas vu d'image
VanRullen & Koch, 2003a	Photographies de 10 objets agencés dans une seule scène masquée	Dénomination des objets reconnus et mémorisés puis dénomination en choix forcé de tous les objets parmi 20 noms de catégorie d'objets
VanRullen & Koch, 2003b	Photographies de scènes naturelles masquées contenant ou non des animaux	- Catégorisation animal/non-animal - Catégorisation animal/non-animal selon une échelle de confiance allant de 0 à 2.
Rolls, 2004	Photographies de 5 visages connues et masquées	- Photographies normales ou modifiées ? - Identification
Marois, Yi & Chun, 2004	Photographie de scènes d'intérieur ou d'extérieur (position T2 dans un flux RSVP – protocole d'attentional blink)	Choix forcé : pas de scène / scène d'intérieur / scène d'extérieur ?
Grill-Spector & Kanwisher, 2005	Photographies d'objets ou de visages	Catégorisation et identification parmi un nombre limité de catégories et sous-catégories

Table 3.2 : Résumé de différentes méthodes d'évaluation du phénomène conscient dans des tâches de masquage.

Si l'on revient aux commentaires qu'ont fait les sujets après avoir passé les expériences de discrimination saccadique, il semble que ce protocole offre des perspectives intéressantes pour l'étude du phénomène conscient. Une éventuelle dissociation entre les modes subliminal et conscient a déjà été reportée par certains sujets sans que ce soit le centre de l'étude. L'utilisation d'un protocole de masquage pourrait permettre d'analyser au cours du temps le niveau d'analyse qu'il est possible de réaliser en mode subliminal d'une part, ainsi que l'évolution parallèle de la représentation consciente des stimuli. Dans une première tentative tout à fait informelle, j'avais essayé de mesurer qualitativement la dégradation du percept conscient dans une tâche de discrimination d'images masquées, où les différentes conditions de SOA n'étaient pas mélangées aléatoirement dans chaque série mais présentées par bloc. Le sujet (moi-même) devait effectuer la tâche de discrimination saccadique tout en comptant le nombre d'essais sur lequel il était sûr de pouvoir identifier la cible (capable de les dénommer mentalement). Ainsi une comparaison des performances réelles et de cette estimation quantitative du percept révélait que pour les SOA très courts et les SOA très longs, le percept était proche de l'estimation (soit je n'avais presque rien vu dans les blocs difficiles, soit au contraire j'avais très bien réussi et dans les deux cas, mes prédictions étaient justes). En revanche, pour les SOA intermédiaires (entre 19 et 44 ms), il y avait une différence très nette entre ce que je *pensais* avoir correctement identifié et ce que j'avais *réellement* discriminé. Mes réponses correctes étaient bien supérieures à mes prédictions, jusqu'à 25 % pour la condition 25 ms de SOA. Il est évident que ce protocole peine sur de nombreux points, à la fois sur la mesure utilisée pour évaluer le percept conscient (c'est d'ailleurs pour cela que je tenais à en parler dans le paragraphe précédent) et sur tout un tas d'effets perturbateurs annexes, comme une perte de contrôle sur les seuils décisionnels alors que les conditions de SOA sont regroupées par bloc. Cet exemple est tout à fait anecdotique, mais je reste vraiment persuadée que le protocole de discrimination saccadique offre de nombreuses possibilités pour étudier au cours du temps le développement du phénomène conscient.

3.7 Conclusion

La comparaison de ces quatre protocoles expérimentaux a montré une dissociation très nette de l'effet de masquage selon le moment à partir duquel les traitements sont perturbés. 30 ms d'intervalle image-masque semble être une latence critique. En dessous de cette valeur, la sensibilité est la même pour tous les protocoles, quelle que soit la tâche et quel que soit le mode de réponse. Nous proposons donc que les processus d'intégration de l'information sont similaires dans cette phase initiale de traitement sensoriel. Cette hypothèse va d'ailleurs dans le sens d'un modèle de transmission feedforward de l'information, comme cela a été suggéré à de nombreuses reprises dans l'équipe pour expliquer la grande rapidité avec laquelle les sujets peuvent répondre à des tâches complexes de catégorisation et de discrimination (Bacon-Macé, Mace, Fabre-Thorpe & Thorpe, 2005; Delorme, Rousselet, Mace & Fabre-Thorpe, 2004; Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001; Kirchner & Thorpe, 2005; VanRullen & Koch, 2003a). Cependant, ces résultats ne sont pas (encore) suffisants pour exclure l'intervention de mécanismes feedback, car leurs effets peuvent être similaires dans les quatre expériences, les rendant invisibles dans ces données. Une autre possibilité serait que les effets interviennent après 30 ms de SOA, au moment où les courbes de sensibilité commencent à différer. Comme nous l'avons vu dans la discussion, les écarts de sensibilité obtenus après cette latence entre les quatre expériences peuvent aussi s'expliquer par des différences en termes de processus décisionnels, mais aussi de sélection ou de préparation de la réponse motrice. Ce qu'il est important de souligner, c'est que ces mécanismes interviennent relativement tard, après une phase commune de traitement sensoriel.

4.1 Que se passe-t-il quand le masque a un sens ?

Nous avons vu dans les chapitres précédents que le protocole de masquage rétrograde offre des conditions de perception adaptées pour étudier le décours temporel des traitements visuels impliqués dans la reconnaissance des scènes naturelles. Les résultats des expériences précédentes ont confirmé que le système visuel est particulièrement efficace pour intégrer rapidement l'information pertinente dans la tâche, même lorsqu'un masque puissant interfère fortement lors de l'extraction des propriétés de bas-niveau de la cible. L'originalité de nos travaux a porté en particulier sur le fait que nous avons travaillé avec des scènes naturelles et les expériences de masquage de scènes naturelles sont relativement rares dans la littérature, malgré le fait qu'elles se rapprochent d'avantage des conditions de vision naturelle dans laquelle le système est confronté en permanence à de nouvelles informations. Cependant pour se rapprocher encore d'avantage des conditions de vision naturelle, au moins deux modifications majeures doivent être apportées à notre protocole : d'une part, il faut que la cible soit non seulement suivie mais aussi précédée par d'autres images à intégrer et d'autre part, ces images perturbatrices doivent elles-mêmes être constituées de scènes naturelles. Nous nous retrouvons alors typiquement avec les conditions d'un protocole RSVP (Rapid Serial Visual Presentation) tel qu'il a été défini par Potter dans les années 70 (Potter & Levy, 1969). Comme nous l'avons déjà souligné dans le chapitre 1, les travaux de Molly Potter sont précurseurs pour notre compréhension de la dynamique temporelle des traitements visuels. En présentant des scènes successives à des fréquences allant jusqu'à 10 images par seconde, Potter a pu observer que les performances de détection d'une image particulière dans la séquence restent très bonnes (Potter & Levy, 1969). En revanche, lorsqu'il s'agit de mémoriser les images de la séquence afin d'indiquer si une image test présentée après le flux appartenait ou non à la séquence, les performances tombent très rapidement au niveau-

chance (Potter, 1975; Potter, 1976). Ces expériences l'ont conduit à distinguer deux phénomènes de masquage : un masquage *perceptuel*, qui interfère avec les propriétés bas-niveau de la cible et un masquage *conceptuel* qui perturbe l'encodage mnésique des informations visuelles (Potter, 1976). Subramanian a récemment proposé une interprétation neuronale du masquage conceptuel, en suggérant que si la majorité de l'information peut être extraite dans les 50 premières millisecondes de la décharge neuronale, la consolidation mnésique requiert au moins 350 ms supplémentaires (Rolls, Tovee & Panzeri, 1999; Subramanian, Biederman & Madigan, 2000; Tovee & Rolls, 1995). En fait, la majorité des auteurs utilisant le paradigme RSVP se sont intéressés à analyser la dégradation des traitements post-perceptuels, à travers l'étude de certains phénomènes tardifs comme l'Attentional Blink ('cécité attentionnelle' (Chun & Potter, 1995; Raymond, Shapiro & Arnell, 1992) et le Repetition Blindness ('cécité à la répétition' (Bavelier, 1994; Kanwisher, 1987). Mais finalement, très peu de travaux ont été publiés sur les mécanismes du masquage perceptuel dans la tâche RSVP, alors que justement ce protocole offre l'occasion de tisser de nombreux liens entre la littérature de la psychophysique et de l'électrophysiologie sur le masquage rétrograde et la littérature sur les opérations cognitives plus avancées impliquant l'intégration des propriétés sémantiques ou lexicales, attentionnelles et mnésiques. D'autre part, si la problématique principale des expériences RSVP doit permettre d'analyser précisément le déroulement des traitements, la manipulation des stimuli cibles et distracteurs offre de plus la possibilité d'isoler la nature des propriétés intervenant dans les interférences mises en jeu (physiques, intermédiaires, sémantiques, lexicales ? / globales, locales ?) (Bachmann & Hommuk, 2005; Evans & Treisman, 2005).

Keysers et collaborateurs ont présenté des scènes naturelles dans des séquences RSVP à des singes macaques et ont été les premiers à décrire l'impact de l'augmentation de la fréquence de présentation sur la décharge neuronale (Keysers & Perrett, 2002; Keysers, Xiao, Foldiak & Perrett, 2001). Grâce aux propriétés de persistance neuronale présente tout au long de la voie visuelle, les auteurs argumentent que la représentation des stimulus successifs dans une séquence RSVP se recouvre dans le temps au niveau de chaque aire, ce qui engendre une dégradation de la réponse neuronale codant les stimuli en corrélation avec la performance comportementale. Parallèlement, ils démontrent que si la réponse neuronale est dégradée, la sélectivité reste néanmoins préservée pour la majorité des cellules, même lorsque les stimuli sont présentés à une fréquence de 72 images par seconde, ce qui laisse seulement 14 ms de délai pour que l'information spécifique à chaque image ait eu le temps d'être intégrée. A cette latence, la performance dans une tâche de détection est d'ailleurs toujours au-dessus du niveau chance. Si ces délais sont compatibles avec ceux que nous avons obtenus dans l'expérience de masquage rétrograde, ce type d'expérience soulève la question des mécanismes de masquage mis en jeu. En quoi les scènes naturelles utilisées en tant que masque diffèrent-elles des simples masques pattern ?

Dans un premier temps, nous allons présenter brièvement le protocole RSVP que nous avons adapté à notre tâche de catégorisation ainsi que les résultats de différentes expériences pilotes qui nous ont permis d'affiner la problématique. Puis nous présenterons les résultats d'une expérience de masquage dans laquelle nous avons manipulé différents paramètres, comme le type de masque

et la position de la cible dans la séquence RSVP. Enfin nous terminerons ce chapitre par la description d'une expérience RSVP dans laquelle les sujets devaient non seulement effectuer une réponse manuelle mais devaient aussi identifier verbalement le contenu des images cibles détectées.

A travers ces travaux, nous avons cherché à analyser explicitement les mécanismes de masquage impliqués dans ce type de protocole où les scènes naturelles constituent elles-mêmes des masques puissants. De manière plus implicite, nous avons aussi comparé les performances dans deux tâches de catégorisation: la catégorisation d'animaux et la catégorisation de visages célèbres.

4.2 Catégorisation d'images naturelles dans un flux RSVP : expériences préliminaires

4.2.1 Catégorisation animal / non-animal

Protocole expérimental

Le protocole expérimental est similaire à celui utilisé dans la tâche de catégorisation d'images masquées décrite dans le Chapitre 2 (Figure 4.1). Le sujet est assis dans une pièce sombre à 1 mètre d'un écran réglé à 100 Hz. Un essai commence par l'apparition d'une croix de fixation, suivi après un délai aléatoire de 800 à 1500 ms par une séquence de 20 images en niveau de gris (taille : environ 12° x 9°). Dans la moitié des essais de chaque série, les 20 images de la séquence sont des distracteurs. Dans l'autre moitié des essais, la séquence contient 19 distracteurs et 1 image cible qui pouvait être située en 6^{ème}, 10^{ème} ou 14^{ème} position dans le flux. Pour chaque essai, les images de la séquence sont présentées soit 10 soit 20 ms. L'intervalle entre le début de présentation de chaque image dans le flux (que nous appellerons SOA) varie entre 20 et 140 ms (8 conditions de SOA en tout : 20, 40, 50, 60, 80, 100, 120, 140 ms qui peuvent aussi être présentées en terme de fréquence d'images dans la séquence : 50, 25, 20, 17, 12, 10, 8 et 7 images par seconde). Toutes ces conditions sont aléatoirement mélangées au sein des séries, en prenant soin toutefois d'afficher autant d'essais cibles que de distracteurs.

Le sujet avait pour consigne de lever le doigt du bouton réponse pour tous les essais où il détectait une cible dans la séquence. Il ne lui était pas demandé de répondre rapidement, et pouvait déclencher sa réponse pendant le flux ou après, à condition de répondre dans la seconde qui suivait la fin de la séquence. 16 sujets ont participé à cette expérience, en suivant chacun 2 sessions de 12 séries de 48 essais. Les deux sessions n'étaient jamais séparées de plus de deux jours. Chaque condition représentait ainsi sur l'ensemble des sujets 192 essais cible et 192 essais distracteur (pour une condition de SOA, de durée d'image et de position donnée). Tous les essais présentant des erreurs dans la latence d'affichage des images ont été supprimés de l'analyse¹ (moins de 1%).

¹ Presentation est un logiciel de présentation de stimuli développé par Neurobehavioral Systems. Après avoir fait une série de tests avec une photodiode placée sur l'écran de stimulation, nous pouvons dire Presentation est un des rares programmes particulièrement fiables pour afficher avec précision les images

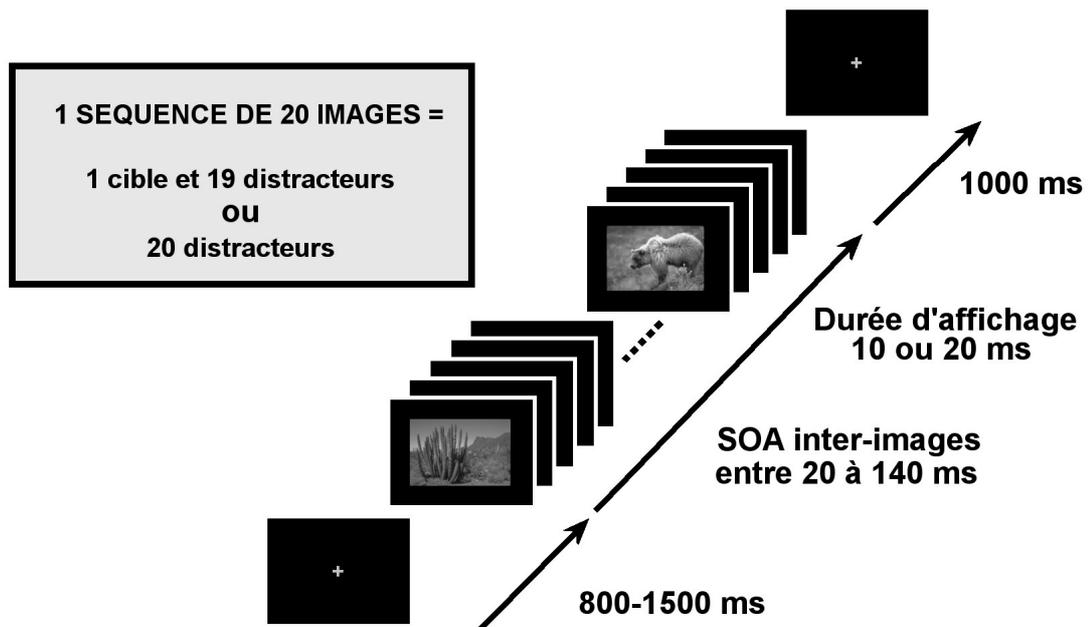


Figure 4.1 : Protocole expérimental dans la tâche RSVP.

En tout, plus de 11000 scènes naturelles ont été utilisées pour cette expérience, dont 576 cibles. Tous les stimuli étaient en niveau de gris, et de taille 600 x 400 pixels. Encore une fois, un effort a été fait pour utiliser des images aussi variables que possible, pour les cibles comme pour les distracteurs. D'autre part, aucune des images distracteur ne contenait d'humains même dans les scènes urbaines.

Résultats

Les performances dans la tâche RSVP sont basses, malgré des valeurs de SOA plus longues par rapport à l'expérience de masquage rétrograde (Figure 4.2A). En moyenne sur l'ensemble des SOA, les sujets réalisent 71,9% (Valeurs interindividuelles min-max [66,73 – 76,68]) de réponses correctes. Il est étonnant d'observer que la précision est encore au niveau chance à 20 ms de SOA, et augmente difficilement même pour des SOA beaucoup plus longs. Ainsi le dernier pas de SOA à 140 ms ne permet de réaliser que 87,5% ([76,43-94.24]) correct et la forme de la courbe laisse supposer que les sujets n'ont pas encore atteint leur performance maximale. Ces résultats sont bien moins bons que ceux obtenus dans l'expérience de masquage rétrograde, alors que dans cette dernière tâche, la consigne de rapidité était une contrainte supplémentaire pour certaines conditions de SOA. A 19 ms de SOA, la précision était déjà au-dessus du niveau chance

aux latences voulues. Cependant dans des expériences comme le RSVP qui demandent beaucoup de ressources mémoire pour stocker les images de la séquence avant leur présentation, il arrivent parfois que les délais ne soient pas respectés (en général, le programme se décale d'une ou plusieurs périodes correspondant à la fréquence de rafraîchissement d'écran). Heureusement, le logiciel Presentation indique dans son fichier de sortie la latence absolue à laquelle il présente réellement les stimuli, il est donc facile de déterminer les essais pour lesquels il y a eu une erreur.

avec près de 60% de réponses correctes, et les 85% correct étaient atteints dès 44 ms de SOA. Nous avons alors cherché à comprendre si cette expérience pilote de RSVP permettait d'isoler certains paramètres pouvant augmenter la difficulté de la tâche. Or ni la position de la cible dans la séquence (présentée en 6^{ème}, 10^{ème} et 14^{ème} position), ni la durée d'affichage des images dans la séquence (10 ou 20 ms) ne semble influencer le pourcentage correct, quel que soit le SOA. Ce dernier point est d'ailleurs intéressant car il semble que la majorité des études de RSVP ont utilisé des images avec une durée de présentation égale au SOA. Keysers et collaborateurs ont testé cette différence en électrophysiologie (données non publiées Keysers, 2000). Il apparaît que lorsqu'on garde un SOA constant mais que l'on varie la durée d'affichage des images de la séquence RSVP, les réponses neuronales enregistrées dans le STS sont inchangées. Encore une fois, ça n'est pas cette durée d'affichage de l'image qui détermine la performance, mais le temps qu'on laisse au système pour intégrer l'information. En revanche dans notre tâche, les sujets réalisaient en moyenne de meilleures performances lors de la 2^{ème} session (Figure 4.2D). Même si cette différence n'était pas significative, nous avons cherché à exploiter cette observation. Comme nous étions réellement surpris par les résultats obtenus, nous avons mené une deuxième expérience dans laquelle 3 sujets se sont entraînés à la tâche de manière intensive.

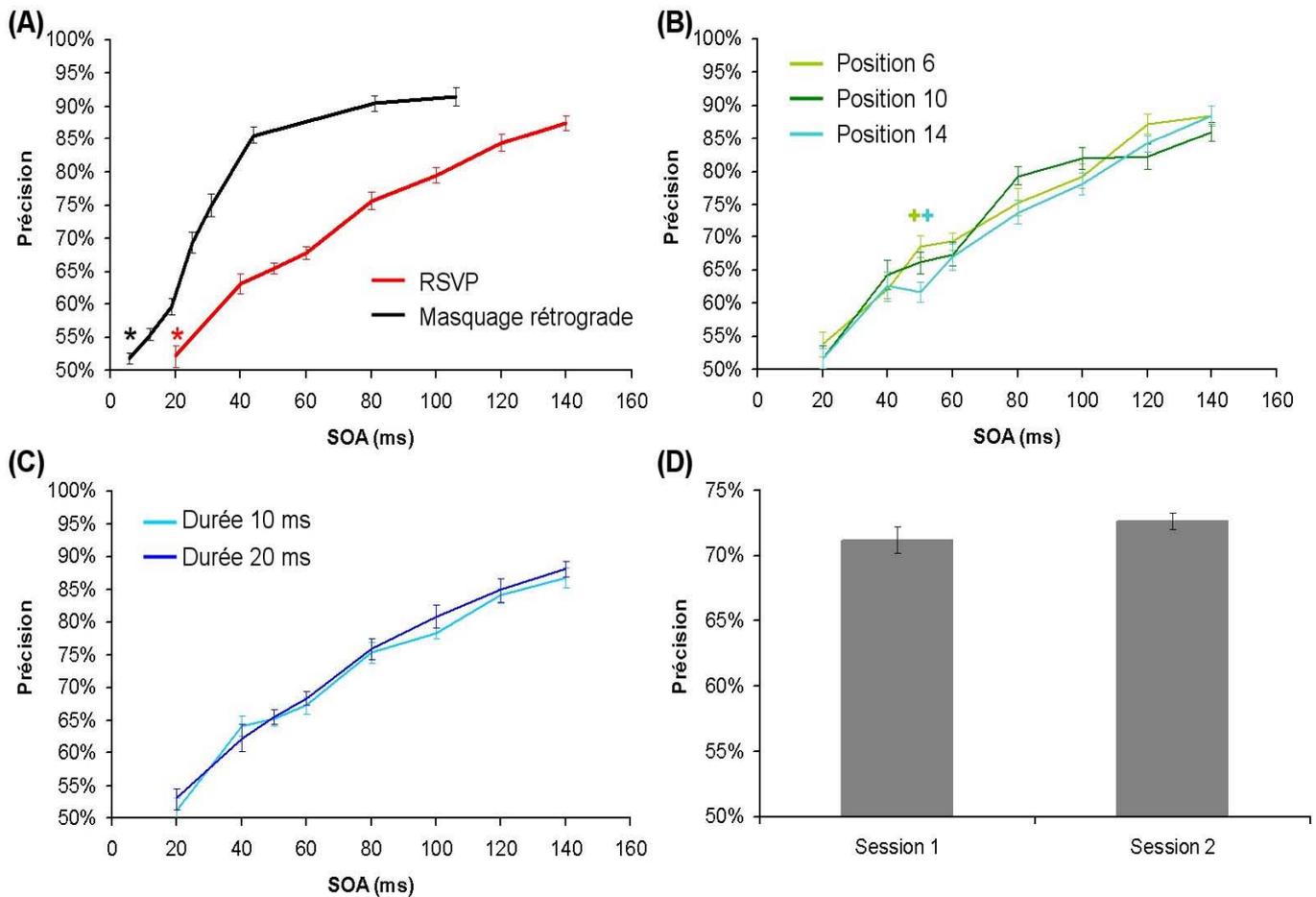


Figure 4.2 : Performance dans la tâche RSVP, alors que les sujets doivent détecter des animaux parmi des distracteurs neutres (+/- sem). La séquence d'images comprenait 20 scènes naturelles, et sur la moitié des essais, l'une de ces scènes contenait une image d'animal (A) Précision en fonction du SOA dans cette tâche en comparaison à la tâche de masquage rétrograde décrite dans le chapitre 2 et publiée dans Vision Research. La performance est beaucoup moins bonne dans la tâche RSVP, contrairement à nos attentes. Les étoiles indiquent que la précision est au niveau chance pour la condition de SOA ($p < 0.01$). (B) Précision en fonction du SOA pour les différentes positions de la cible dans la séquence. Mis à part la condition 50 ms, la précision ne change pas avec la position de l'image. Les symboles indiquent que la précision diffère significativement lorsque la cible est présentée en 6^{ème} et 14^{ème} position, à 50 ms de SOA ($p < 0.02$) (C) Précision en fonction du SOA selon que les images sont présentées 10 ou 20 ms. La durée d'affichage de l'image n'a pas d'impact sur la performance. (D) Précision moyenne sur toutes les conditions de SOA, séparément pour les sessions 1 et 2. La légère augmentation de précision d'une session à l'autre (71,1 % vs. 72,6) n'est pas significative.

4.2.2 Effet de l'entraînement dans la catégorisation animal / non-animal

Protocole expérimental

Le protocole expérimental est le même que dans l'expérience précédente. Les mêmes séries et images ont été utilisées, en ayant pris soin d'effectuer un mélange aléatoire des essais entre série, tout en respectant l'équiprobabilité des conditions de SOA, position et durée d'affichage.

Trois sujets ont réalisé 10 sessions de 576 essais du lundi au vendredi sur 2 semaines. Deux de ces sujets peuvent être considérés comme expert dans les tâches de catégorisation, le troisième étant complètement naïf.

Résultats

Comme le montre la figure 4.3, la précision dans la tâche de RSVP n'augmente pas, même avec un entraînement intensif. Même si l'on peut observer une légère hausse de la performance au cours de la première semaine, qui se poursuit encore après le week-end, ces différences ne sont pas significatives.

Nous pouvons donc conclure que cette tâche est réellement difficile pour les sujets, et beaucoup plus que la tâche de masquage rétrograde présentée précédemment. Or les effets de masquage dans la tâche RSVP sont générés par les images de scènes naturelles elles-mêmes, qui sont beaucoup moins contrastées que le masque noir et blanc utilisé dans le chapitre 2. Nous avons justement posé l'hypothèse que le masque contrasté à 100 % permettait des effets perturbateurs maximaux, en saturant le système visuel au moment de l'extraction des propriétés bas-niveau de la cible. Il faut donc conclure que les scènes naturelles peuvent avoir des effets perturbateurs différents, n'influençant pas seulement les étapes précoces des traitements. Les sujets nous ont d'ailleurs fait de nombreuses remarques intéressantes à ce sujet et en particulier certains justifiaient leur difficulté à détecter la cible par le fait que leur attention restait « bloquée » sur un distracteur précis dans la séquence, une maison victorienne pour les uns, un avion de chasse qu'ils ont d'abord confondu avec un oiseau pour les autres ou encore un ballon dirigeable bien centré dans l'image. Ces remarques nous ont laissé entendre que le contenu sémantique des images pouvaient en partie expliquer la différence de performance entre les tâches de masquage et de RSVP. Nous avons donc réalisé une autre série d'expérience RSVP dans laquelle nous avons varié le contenu des catégories cibles et distracteurs.

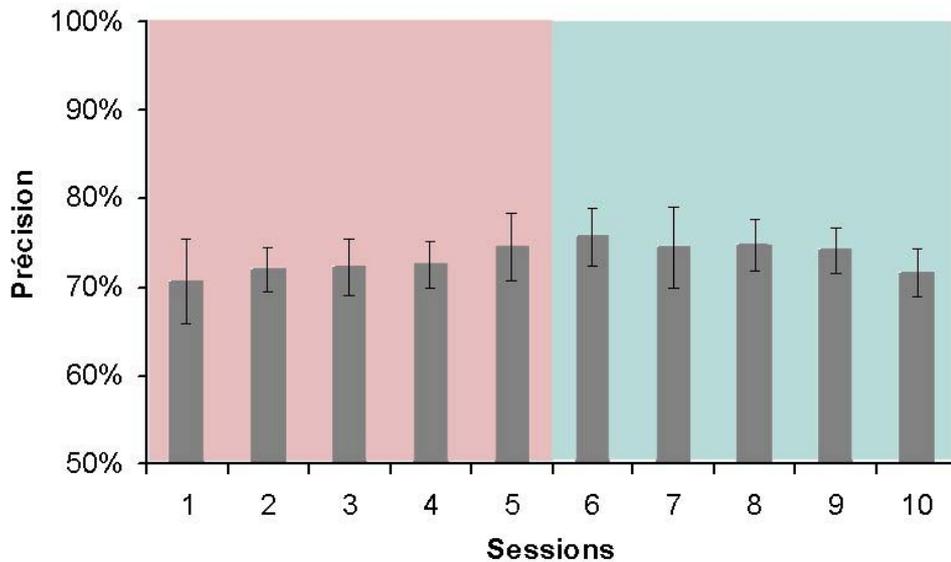


Figure 4.3 : Précision des trois sujets dans la tâche RSVP, au cours de 10 sessions (+/- sem).

4.2.3 Changement de catégories : catégorisation animal–humain–visages

Protocole expérimental

Le protocole expérimental reste très proche de celui décrit ci-dessus, avec l’affichage pour chaque essai d’une séquence de 20 images, dont l’une contient ou non une cible à détecter. Cependant, comme la durée d’affichage des images ne semble pas affecter les performances, toutes les images de la séquence ont été présentées 10 ms. Les huit conditions de SOA entre 20 et 140 ms ont été inchangées. De la même manière, la cible pouvait se situer en 6^{ème}, 10^{ème} et 14^{ème} position. 8 sujets ont effectué 15 séries de 48 essais, dont la moitié contenait des cibles (Figure 4.4). Dans 5 séries, les sujets devaient catégoriser des animaux alors que les autres images de la séquence contenaient des distracteurs neutres (Condition Animal). Dans les 10 autres séries, les sujets devaient catégoriser des humains parmi des animaux. La moitié des sujets a commencé l’expérience par la catégorisation animal, l’autre moitié par la catégorisation d’humains. Tous les sujets effectuaient une série d’entraînement avant une nouvelle catégorisation et chaque partie était séparée par une pause d’au moins 10 minutes.

Dans 5 des 10 séries de catégorisation d’humains, les images cibles contenaient des humains à différentes échelles et en nombre variable (Condition Humain). Dans ces séries, les distracteurs étaient les images d’animaux catégorisées dans la condition Animal, variables aussi en distance, taille et position. Dans les 5 autres séries de catégorisation d’humains, les images contenaient des humains vus de près de telle manière que le visage représentait une grande partie de la surface du stimulus. 3 des 5 séries contenaient des visages inconnus (Condition Visages Inconnus) et les deux autres contenaient des visages de personnages célèbres (Condition Visages Célèbres). Dans ces 5 séries, les distracteurs étaient des têtes d’animaux. Les séries visages et humains étaient

alternées une à une. Les sujets n'ont pas été informés de ces différentes conditions et avaient pour unique consigne de détecter les humains. La plupart des sujets ne se sont d'ailleurs pas rendus compte des différences entre les séries.

Stimuli

Les images d'animaux et de distracteurs neutres font partie du set d'images utilisé dans les expériences pilotes précédentes. Les distracteurs ne contenaient donc pas d'humains. Les images d'humains et de visages non célèbres sont issues des librairies Corel et variaient en taille, situation et position de la cible dans la scène. Les images de célébrités ont été sélectionnées sur Internet, et contenaient des photographies de chanteurs, acteurs, politiciens, sportifs ... Un petit test effectué sur huit sujets n'effectuant pas l'expérience a montré que 98% des images de célébrités choisies sont connues par les sujets et 78% correctement identifiées.

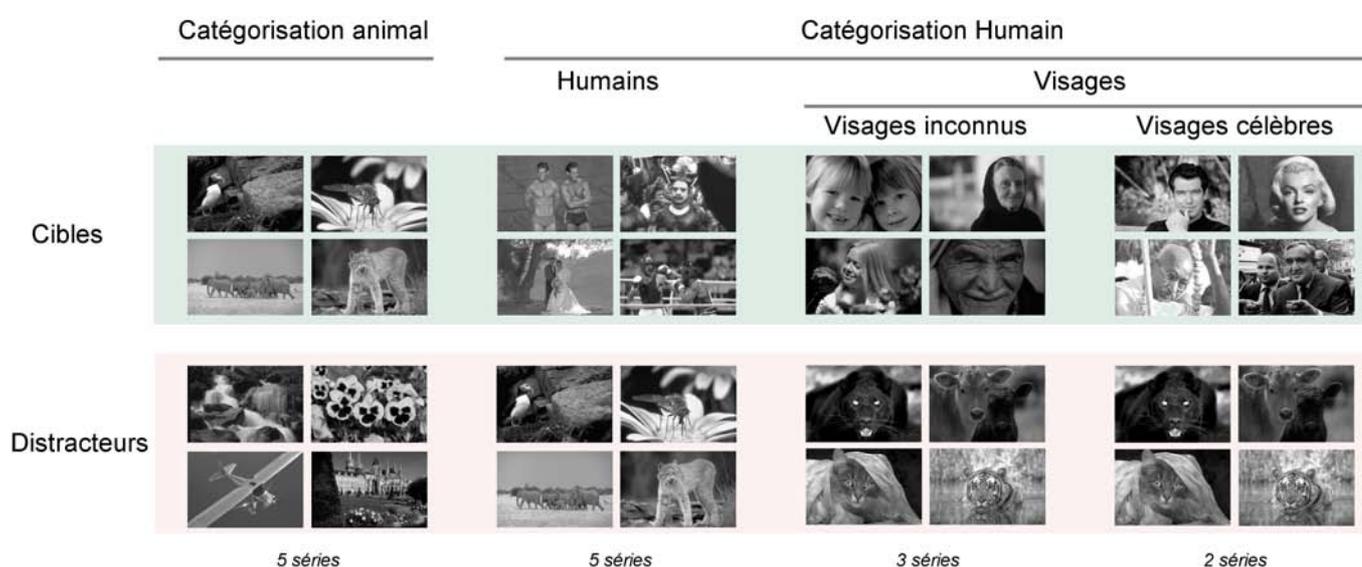


Figure 4.4 : Exemples d'images utilisées dans l'expérience RSVP, en fonction des séries.

Résultats

Encore une fois, la précision comportementale reste très basse lorsqu'il s'agit de détecter un animal dans un flux contenant des distracteurs neutres (Figure 4.5). La précision commence au niveau chance et atteint seulement 80% de réponse correcte à 80 ms de SOA. En revanche, les résultats sont très différents pour la catégorisation d'humains ($p < 0.001$). Dès le deuxième pas de SOA à 40 ms, la précision atteint 80% dans toutes les conditions, que les humains soient vus à différentes échelles ou de près, inconnus et célèbres. Il est même particulièrement surprenant d'observer que la précision est déjà proche de 75% à 40 ms de SOA pour les visages célèbres et qu'elle plafonne à 100% de réponses correctes dès 60 ms de SOA.

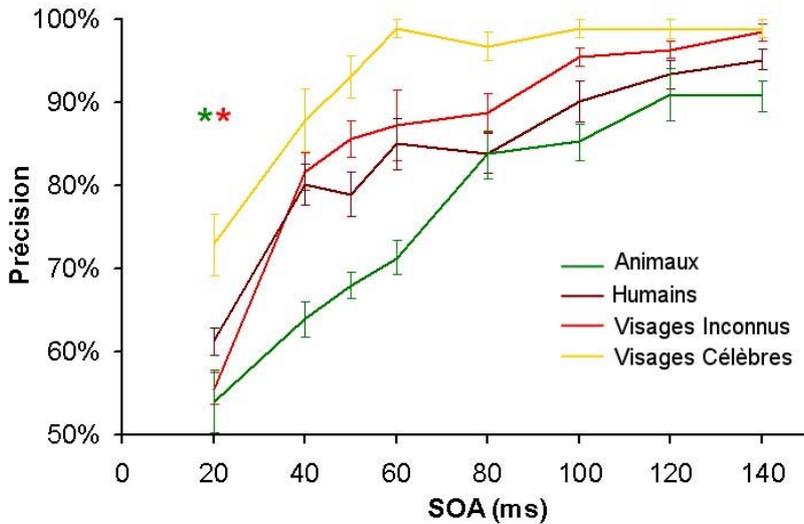


Figure 4.5 : Précision en fonction du SOA, pour les différentes conditions de catégorisation (+/- sem). Les étoiles de couleur indiquent les conditions pour lesquelles la précision est au niveau chance. La condition Animal est significativement différente des trois autres pour toutes les conditions de SOA groupées ($p < 0.004$). La condition Visages Célèbres diffère des deux autres catégorisations d'humains - Visages Inconnus et Humains- ($p < 0.02$)

Discussion : trop de limitations expérimentales

La variabilité de ces résultats selon les catégories soulève un nombre important de questions. D'une part, il est clair que la catégorisation d'humains est bien plus efficace dans cette tâche que la catégorisation animale. Doit-on attribuer la meilleure performance obtenue sur les visages cibles au fait que ces stimuli ont des propriétés plus résistantes au masquage, ou au fait que les distracteurs utilisés dans ces catégories sont moins efficaces ? S'il paraît tentant de conclure que les photographies d'humains - et *a fortiori* de visages d'humains - ont une spécificité biologique qui permet des traitements optimisés en vitesse, il n'est pas possible de répondre car les stimuli utilisés en tant que distracteurs diffèrent dans les deux types de catégorisation, soulevant ainsi une première limitation expérimentale. Dans le cas de la catégorisation animale, les distracteurs sont des scènes naturelles dites « neutres » car elles ne contiennent ni humain ni animal, alors que dans le cas de la catégorisation d'humains, les distracteurs sont des animaux. Il est possible d'imaginer que les distracteurs neutres, qui ont une variabilité plus grande que les distracteurs de têtes d'animaux, rassemblent des propriétés physiques plus perturbatrices pour les traitements visuels bas-niveau des cibles animales.

D'autre part, même si l'on observe que les visages célèbres sont encore mieux détectés dans le flux que les visages inconnus, il est encore difficile de conclure que la valeur sémantique du stimulus influence sa capacité à résister au masquage, car les sets d'images sont différents. Les photographies de célébrités ont été sélectionnées sur Internet et sont globalement beaucoup plus contrastées que les photographies Corel de visages inconnus (Figure 4.6). Or les propriétés de contraste de luminance sont d'une importance fondamentale dans ce type de tâche où l'on pose l'hypothèse que les mécanismes de masquage reposent principalement sur une perturbation des indices bas-niveau du stimulus cible.

Enfin, il faut rappeler que la capacité à détecter une image d'animal est beaucoup plus affectée par le masquage lorsque la cible est présentée dans une séquence RSVP que lorsqu'elle est suivie par un masque rétrograde n'ayant aucun contenu sémantique. Encore une fois, il serait tentant de conclure qu'en plus du masquage bas-niveau perturbant l'extraction des propriétés de la cible,

une autre forme de masquage intervient faisant référence à la valeur sémantique ou attentionnelle des stimuli distracteurs. Or il faut rappeler qu'une autre différence majeure dissocie le protocole RSVP du protocole de masquage rétrograde : lorsque les distracteurs sont présentés avant la cible, les traitements de celle-ci peuvent être affectés par du masquage antérograde. Il n'est donc pas possible avec ces données de savoir si les faibles performances obtenues en RSVP pour catégoriser des animaux, sont dues majoritairement au masquage antérograde qui vient s'ajouter au masquage rétrograde ou si une autre forme de masquage, que l'on pourrait qualifier de sémantique, interviendrait aussi.

En conclusion, cette expérience RSVP apporte des résultats surprenants puisqu'il apparaît clairement que le type de catégorie d'images à détecter influence directement la précision comportementale. Globalement, ces trois expériences pilotes ont permis d'observer que les scènes naturelles peuvent elles-mêmes constituer des masques puissants pour étudier le décours temporel des traitements dans une tâche de catégorisation. Cependant, il reste difficile d'expliquer les mécanismes de ce masquage car certains choix effectués au moment de la mise en place du protocole expérimental présentent des limitations. Dans la section suivante, nous proposons de tester les différentes hypothèses soulevées dans cette discussion à travers un set de trois expériences faisant intervenir des conditions de masquage plus contrôlées.

4.3 Catégorisation d'images naturelles dans un flux RSVP : comparaison des effets perturbateurs de trois masques

4.2.1 Objectifs

Dans trois expériences RSVP, nous avons demandé aux sujets d'effectuer des tâches de catégorisation de scènes naturelles contenant des animaux ou des visages de célébrités. Les masques de la séquence RSVP pouvaient être selon l'expérience : des patterns, des scènes naturelles ou des scènes naturelles « scrambled ». D'autre part, les stimuli cible ou distracteur pouvaient se positionner de la première à la dernière position dans la séquence de 20 images.

A travers cette série de trois expériences, nous avons deux objectifs principaux :

- Comparer les effets perturbateurs de différents types de masque ayant un contenu sémantique ou des caractéristiques physiques particulières. L'utilisation de masque pattern devait permettre de faire le lien entre les expériences RSVP et l'expérience de masquage rétrograde présentée dans le chapitre 2. D'autre part, alors que les masques constitués de scènes naturelles ont un contenu sémantique, les pattern et surtout les scènes naturelles « scrambled » n'en ont pas.
- Comparer les effets perturbateurs des masquages antérograde et rétrograde et évaluer l'impact de la combinaison de ces deux masquages sur les performances. Nous avons isolé le masquage rétrograde en présentant la cible en première position, le masquage antérograde avec la cible en dernière position, et les deux formes de masquage combinées avec une cible présentée à l'intérieur de la séquence.

4.3.2 Protocole expérimental

Sujets

Chacune des trois expériences a impliqué la participation de 12 sujets, dont au moins la moitié était totalement naïf dans ce type de protocole expérimental. Au total 25 sujets ont participé à ce set de trois expériences (moyenne d'âge 26 ans, 12 femmes et 13 hommes), plusieurs sujets ayant participé à plus d'une expérience.

Tâches

Le protocole expérimental est le même dans les trois expériences (Figure 4.6). Les sujets sont assis dans une pièce sombre, à un mètre d'un écran réglé à 160 Hz. Un essai commence avec l'apparition d'une croix de fixation au centre de l'écran pour une durée variable de 800 à 1500 ms. Une séquence de 20 images ($12.3^\circ \times 8.3^\circ$) est alors présentée : 19 sont des masques et 1 image est une scène naturelle, soit cible soit distracteur. Chaque image est présentée le temps d'un rafraîchissement d'écran, c'est à dire 6,25 ms, mais le délai entre deux images de la séquence (SOA) varie entre 6,25 et 118,75 ms (avec 6 pas : 6,25 – 12,5 – 25 – 43,75 – 81,25 – 118,75 ms). L'image stimulus, cible ou distracteur, peut se trouver à n'importe quelle position de la séquence : dans un quart des essais, elle est présentée en première position, dans un autre quart des essais en dernière position (20^{ème}) et dans la moitié des essais, aléatoirement entre la 2^{ème} et la 19^{ème} position. Dans toutes les séries, les conditions de SOA et de position du stimulus dans la séquence sont contrebalancées et le nombre d'essais cibles et distracteurs équiprobables.

Les réponses sont données selon un mode go/no-go, les sujets devant relâcher leur doigt d'un bouton réponse lorsqu'ils pensent avoir détecté une cible dans la séquence. Les sujets n'ont pas de consigne de rapidité et peuvent répondre pendant ou après la séquence.

Chaque expérience est divisée en deux parties de 6 séries de 96 essais chacune. La tâche de catégorisation change d'une partie à l'autre, la moitié des sujets commence par effectuer une tâche de catégorisation Animal puis effectue une catégorisation Visage. L'autre moitié des sujets réalise ces tâches dans le sens inverse. Sur l'ensemble des sujets, chaque condition de SOA a cumulé 576 essais pour chacune des tâches de catégorisation (soit 48 essais par sujet). Tous les essais présentant des erreurs dans la latence d'affichage des images ont été supprimés de l'analyse (moins de 5%).

Stimuli

La figure 4.6 montre des exemples de cibles et de distracteurs, ainsi que de masques utilisés pour chacune des trois expériences.

Cibles et distracteurs

Au total, 576 photographies ont été utilisées comme cibles. La moitié d'entre elles sont des scènes naturelles sélectionnées dans les librairies Corel et contiennent un set varié d'animaux incluant des mammifères, des oiseaux, des poissons, des reptiles, des insectes... L'autre moitié contient 4 photographies de visages de 96 célébrités (acteurs, chanteurs, politiciens, sportifs etc....) sélectionnés sur des magazines et Internet. Comme le montre une expérience de dénomination sur

ces mêmes images, décrite dans la deuxième partie de ce chapitre, les sujets reconnaissaient en moyenne 91% de ces visages et pouvaient en identifier correctement 84 %.

576 distracteurs ont aussi été sélectionnés dans les librairies Corel, avec des contenus aussi variables que ceux utilisés dans l'expérience de masquage rétrograde, mais ne contiennent ni humain ni animal.

Masques

Expérience 1 : *Masques Pattern*

Nous avons utilisé les mêmes masques que dans l'expérience de masquage rétrograde, composés de formes aléatoires noires et blanches à différentes fréquences spatiales et contrastés à 100% (16 masques au total). Dans chaque essai, les quatre masques précédant la cible ainsi que les quatre masques suivant la cible comprennent des échelles spatiales différentes. Tous les masques séparés d'au moins 5 positions du stimulus cible ou distracteur sont ensuite aléatoirement choisis dans le set de 16 images de masque.

Expérience 2 : *Masques Scènes*

Dans cette expérience, les masques sont choisis dans le set de 288 images composant la catégorisation inverse : lorsque les sujets effectuent une tâche de catégorisation Animal, les masques sont des photographies de visages de célébrités, et lorsque les sujets effectuent une tâche de catégorisation Visage, les masques sont des scènes naturelles contenant des animaux.

Expérience 3 : *Masques Scènes « scrambled »*

Comme dans l'expérience 2, les masques sont des images appartenant à l'autre catégorie. Cependant, ces images ont été modifiées par un algorithme Matlab, chaque image étant découpée selon un motif de Penrose et les 230 morceaux découpés ont été mélangés pour former une nouvelle image (principe du taquin mais avec un motif de découpage plus complexe). Le but de cette manipulation est de conserver un maximum de propriétés bas-niveau sur l'image, tout en rendant l'identification de la scène impossible afin d'en dégrader son contenu sémantique.

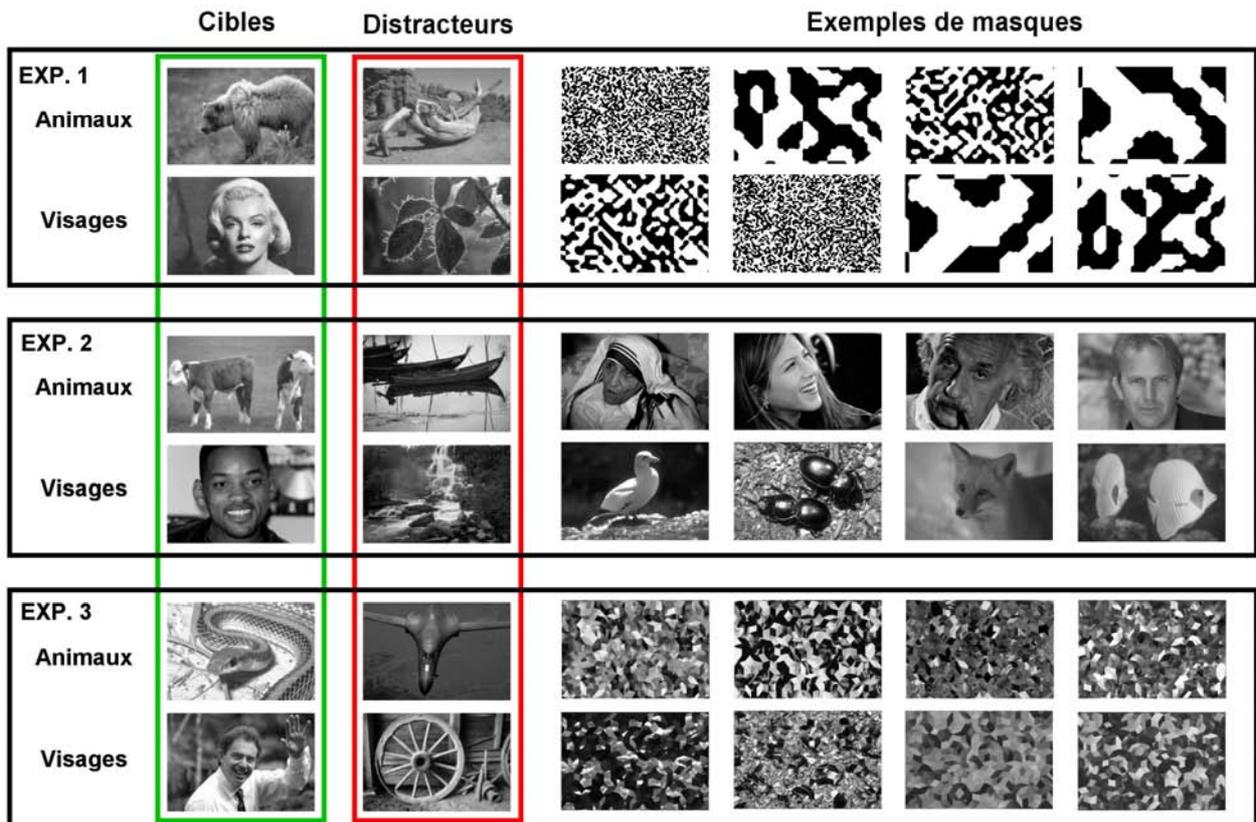


Figure 4.6 : Exemples d'images utilisées dans l'expérience RSVP, en fonction des expériences et des tâches à réaliser. Chaque flux RSVP était composé de 19 masques et d'une scène naturelle intacte cible ou distracteur. Dans chacune des trois expériences, les sujets devaient effectuer une catégorisation Animal sur la moitié des séries et une catégorisation Visages (de célébrités) sur l'autre moitié. Dans l'expérience 1, les masques sont des patterns noirs et blancs, dans l'expérience 2, ce sont des images de la catégorie inverse et dans l'expérience 3, ce sont aussi des images de la catégorie inverse mais dans une version « scrambled ».

4.3.3 Résultats

Globalement dans les trois expériences, les performances sont plutôt basses, mais on observe une différence importante selon la tâche effectuée : les sujets réalisent 70 % de réponses correctes dans la tâche de catégorisation Animal, et 83,3% dans la tâche de catégorisation Visage. Ces résultats sont en accord avec les expériences pilotes décrites ci-dessus.

Différents paramètres ont été manipulés dans cette nouvelle série d'expérience et dans la section suivante, nous présenterons les résultats de performance en fonction du SOA, de la position du stimulus dans la séquence et en fonction de l'expérience, séparément pour chacune des tâches réalisées (Animal/Visage).

Précision en fonction du SOA

Sans surprise, la performance varie significativement avec le SOA. Dans la tâche de catégorisation Animal, le premier pas de SOA conduit à une précision au niveau chance, quelle que soit l'expérience, autrement dit quel que soit le masque utilisé (Figure 4.7A). Puis la

précision augmente progressivement pour finalement dépasser les 90 % dans la condition la plus facile de SOA à 119 ms.

En revanche dans la tâche de catégorisation Visage (Figure 4.7E), la performance est bien meilleure quel que soit le SOA et la précision est déjà bien au-dessus du niveau chance dans les expériences 2 et 3 lorsque les masques sont des scènes naturelles intactes et « scrambled ». A 44 ms de SOA, la précision atteint déjà un plateau et les sujets sont presque tous à 100% de réponses correctes pour le dernier pas de SOA à 119 ms.

Précision en fonction de la position du stimulus dans la séquence

L'augmentation de la performance en fonction du SOA dépend aussi de la position de la cible dans la séquence RSVP (Figure 4.8). Lorsque le stimulus est présenté en 1^{ère} position, seuls les masques suivant ce stimulus interfèrent avec lui, nous permettant ainsi d'isoler le phénomène de masquage rétrograde. Lorsque le stimulus est présenté en dernière position (20^{ème} position), nous isolons le masquage antérograde. Enfin, lorsque le stimulus est présenté entre la 2^{ème} et la 19^{ème} position dans la séquence, les effets simultanés du au masquage rétrograde et antérograde peuvent être observés.

La figure 4.8A montre l'augmentation des performances dans les tâches de catégorisation Animal et Visage en fonction du SOA et séparément pour les trois conditions de masquage. Les tracés montrent clairement que la combinaison des masquages rétrograde et antérograde affecte d'avantage les traitements visuels par rapport au cas où un seul des phénomènes de masquage intervient. Cet effet est d'ailleurs vrai aussi bien pour la catégorisation Animal ($F(2,107) = 75.5$, $p < 0.001$) que pour la catégorisation Visage ($F(2,107) = 87.8$, $p < 0.001$). Dans ce dernier cas, il est intéressant d'observer qu'il faut la combinaison des deux masquages pour obtenir une précision au niveau chance avec 6 ms de SOA.

Dans la tâche Animal (Figure 4.8A), le masquage rétrograde est significativement plus efficace que le masquage antérograde quel que soit le SOA ($p < 0.001$) et en moyenne la précision est 7% inférieure dans cette condition. Cette différence n'est pas observée du tout dans la catégorisation Visage où les deux courbes se superposent parfaitement (Figure 4.8B).

Précision en fonction du masque utilisé

Un des objectifs principaux de cette série d'expérience était de comparer les performances des sujets dans une tâche de catégorisation de stimuli masqués en fonction du contenu des masques, qui pouvaient être soit des patterns, des scènes naturelles ou des scènes naturelles « scrambled ». Deux conclusions principales peuvent être tirées.

Premièrement, dans la tâche de catégorisation Animal, pour les SOA inférieurs à 25 ms, la précision est moins bonne lorsque les masques sont composés de patterns noirs et blancs (Figure 4.7A). Au delà de cette valeur, les effets les plus perturbateurs sont obtenus lorsque les masques sont des scènes naturelles qui dans ce cas, sont des photographies de visages de célébrités (SOA 44 ms: $F(2,35) = 10.3$, $p < 0.001$, SOA 81 ms: $F(2,35) = 3.6$, $p < 0.04$).

Deuxièmement dans la tâche de catégorisation Visage (Figure 4.7E), les masques pattern sont beaucoup plus efficaces que les autres masques jusqu'à ce que le SOA permette d'obtenir des

performances optimales ($F(2,215) = 6.7, p < 0.001$). En revanche aucune différence n'est observée lorsque les masques sont des scènes naturelles intactes ou « scrambled ».

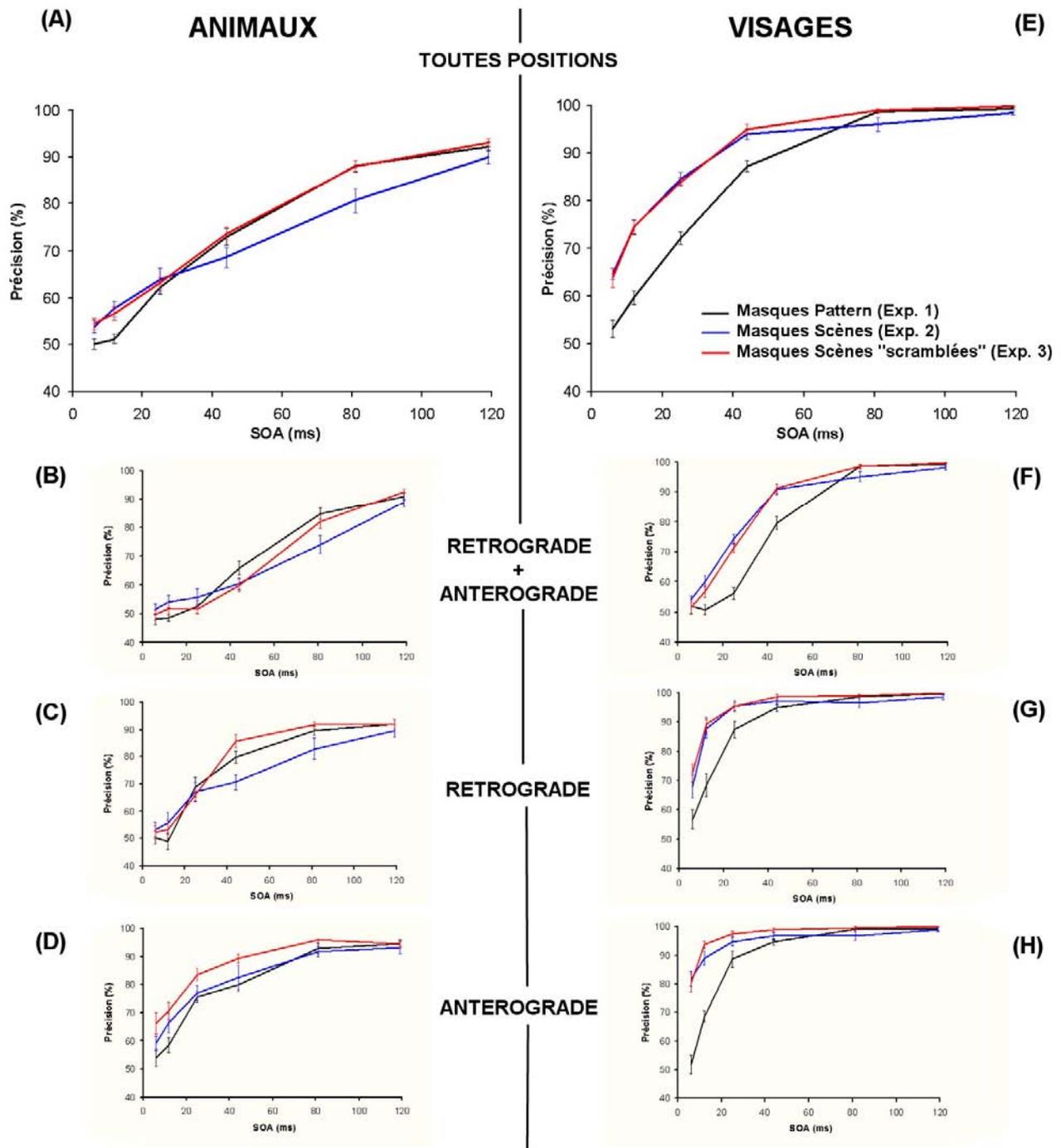


Figure 4.7 : Précision en fonction du SOA pour les différentes conditions expérimentales. (A) à (E) La précision augmente avec le SOA différemment selon le type de masque qui est utilisé dans la tâche, aussi bien dans la tâche de catégorisation Animal (A à D) que dans la tâche de catégorisation Visage (E à H). Les graphiques A et E montrent l'évolution de la précision en fonction du SOA, selon les types de masque, quelle que soit la position du stimulus dans la séquence RSVP. Les graphiques de petite taille indiquent plus précisément l'évolution de la précision en fonction du SOA pour les trois types de masque, selon la position du stimulus dans la séquence, qui génère soit du masquage rétrograde (B et F), soit du masquage antérograde (C et G) soit les deux phénomènes de masquage combinés (D et H).

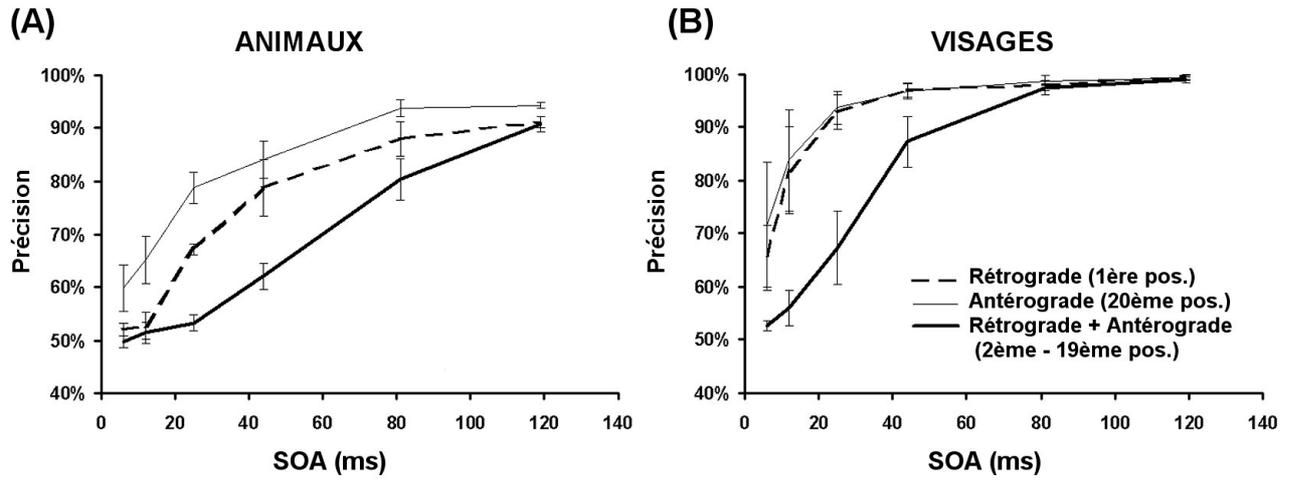


Figure 4.8 : Précision en fonction du SOA selon le phénomène de masquage enregistré, pour la tâche Animal (A) et pour la tâche Visage (B). Les données sont cette fois groupées sur les trois expériences.

ARTICLE 3

Investigating early masking effects using the RSVP paradigm

Bacon-Macé N & Thorpe SJ
En préparation

Investigating early masking effects using the RSVP paradigm

Nadège Bacon-Macé and Simon J. Thorpe

Centre de Recherche Cerveau et COgnition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil. 133, route de Narbonne. 31062 Toulouse, France.

ABSTRACT

Human observers are very good at deciding whether a briefly flashed natural scene contains an animal or a face and previous backward masking experiments have allowed us to study the dynamics of visual processing during this task (Bacon-Mace, Mace, Fabre-Thorpe & Thorpe, 2005). Here we used a Rapid Serial Visual Presentation (RSVP) paradigm to further analyse the nature of masking mechanisms when categorising animals or celebrity faces. Results show that the disrupting processing are operating relatively early in the visual pathway, by interfering with the low-level properties of the target. Moreover, the particular status of the famous faces pictures in the tasks hardly suggest that other kinds of higher-level attributes could also be in competition, and may be linked to the semantic content of the stimuli. The RSVP paradigm is a solid tool to investigate the dynamics of early visual processing, and is a potential way to make the link between the extensive literature on masking processing and on higher-level cognitive operations.

INTRODUCTION

Since the early 70's, the Rapid Serial Visual Presentation (RSVP) paradigm has provided a solid experimental background to study the degradation of perceptual processing. By presenting sequences of unrelated pictures at different rates, Potter observed a difference in detection and memory performance which led her to propose the existence of a very short-term memory system (Potter, 1976; Potter, 1993). She clearly distinguished between *perceptual masking*, which is supposed to degrade early forms of visual

processing and *conceptual masking*, implicated in her tasks by disturbing the storage of information into a mnemonic representation. Subramanian recently offered a neural interpretation of the conceptual masking described by Potter, proposing that while most of the information concerning the stimulus is contained in the first 50 ms of firing, an additional 350 ms is necessary for memory consolidation (Rolls, Tovee & Panzeri, 1999; Subramanian, Biederman & Madigan, 2000; Tové & Rolls, 1995). RSVP paradigms have

also been used to analyse other forms of post-perceptual impairments, including the Attentional Blink (Chun & Potter, 1995; Raymond, Shapiro & Arnell, 1992) and Repetition Blindness (Bavelier, 1994; Kanwisher, 1987). But finally, relatively few studies have been interested in using the RSVP procedures to study perceptual masking mechanisms, although many links can be made with the extensive psychophysical literature on forward and backward masking and its associated electrophysiological data. In a recent article, Evans and Treisman published the results of a complete set of experiments suggesting that feature analysis mediating detection of natural scenes could be very fast, whereas identification and localisation may require more time (Evans & Treisman, 2005). They used the RSVP paradigm to precisely isolate the processing required for detection and propose that the limits to performance could be the reflect of disruption effects in the first stages of perceptual processing. In their model, an initial rapid pass through the perceptual hierarchy could process the features achieving a crude representation of the target but sufficient for detection and to some extent for identification. Accurate and conscious identification must be mediated by further processing involving feature binding and the creation of an object file to represent the integrated object. Furthermore, they observed that feature overlaps between target and distractor categories strongly affect performance, but that the confusion may occur before the extraction of the semantic properties. An important challenge of this paper was precisely to understand the level of representation that can be built with the information integrated in the first pass, but one still needs to know how natural scenes are operating as disruptive masks. Keyzers and collaborators also used naturalistic pictures in RSVP sequences and were the first to describe the impact of increasing presentation rates on the neural responses (Keyzers & Perrett, 2002; Keyzers, Xiao, Foldiak & Perrett, 2001). Due to neural persistence that is present throughout the visual system, they postulated that the representations of the different stimuli in a RSVP sequence are overlapping in time within a single cortical area, which results in a degradation of the neural response coding the stimuli, in correlation with behavioural performance. However, they demonstrated that stimulus selectivity was still preserved for a majority of the cells recorded in the superior temporal sulcus, even when the sequences were presented at a rate of up to 72 images

per second. Questions remain concerning the nature of the information that could be sensitive to masking effects in early processing. Naturalistic masks probably affect the integration of low-level target properties but does the semantic content make them differ from the classical pattern or noise masks?

Here we propose three behavioural RSVP experiments in which we vary the categorisation task, the content of the masks and the position of the target in the sequence, in order to decompose distinct masking phenomena that may interfere in the early visual processing. The general paradigm is based on the Ultra-Rapid Visual Categorisation task developed by Thorpe and collaborators (Thorpe, Fize & Marlot, 1996), that have already proved to be particularly well adapted to study the temporal aspects of visual processing (Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005; VanRullen & Koch, 2003). A previous experiment has shown that categorisation processing can be very robust when masking the stimuli with a strong backward pattern mask (Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005). By varying the Stimulus Onset Asynchrony (SOA) between the stimulus and the mask, we observed that performance rapidly increases and was close to optimal above 40-50 ms, which is consistent with the electrophysiological studies showing that considerable information is available within the first 50 ms (Keyzers, Xiao, Foldiak & Perrett, 2001; Kovacs, Vogels & Orban, 1995; Rolls, Tovee & Panzeri, 1999). We proposed that those masking effects were the results of an interference between the stimulus and the mask, taking place early in the visual pathway while extracting the physical properties on the image. This phenomena was supposed to be maximal with the mask we used, composed by black and white patterns at different spatial frequencies, and we predicted weaker effects by using natural scenes as masking stimuli. Contrary to our first predictions, natural scenes appear to be much stronger masks than expected, as shown by the low performance at detecting a target in a natural scene.

METHOD

Experimental design

The experimental procedure was the same in the three experiments (Figure 1A). Subjects were seated in a dimly dark room, at one meter from a screen set at 160 Hz. A trial began with the display of a fixation

cross for a variable delay from 800 to 1500 ms, and was followed by a sequence of 20 pictures ($12.3^\circ \times 8.3^\circ$): 19 pictures were masking stimuli (noise patterns, natural scenes or scrambled images depending on the experiment) and 1 picture was an intact natural scene, either a target or a distractor. Each picture was displayed for one frame, that is 6.25 ms, but the delay between the onsets of two successive pictures (Stimulus Onset Asynchrony) varied from 6.25 to 118.75 ms (6.25, 12.5, 25, 43.75, 81.25 and 118.75 ms), leading to a complete trial duration of between 125 and 2 253 ms. The target/distractor picture could be presented at any position in the

sequence: in one out of four trials, it was presented at the first position, in another quarter at the last position, and on half of the trials its position ranged from the 2nd to the 19th position. In any one series, all these conditions were counterbalanced and targets and distractors trials were equiprobable.

Responses were given in a go/nogo mode, with subjects releasing their finger from a button only when they had detected a target in the stream. They were not pressed to answer quickly and could give their response during the sequence display or afterwards, but within one second before the next trial began.

Each experiment was divided in two parts of 6 series

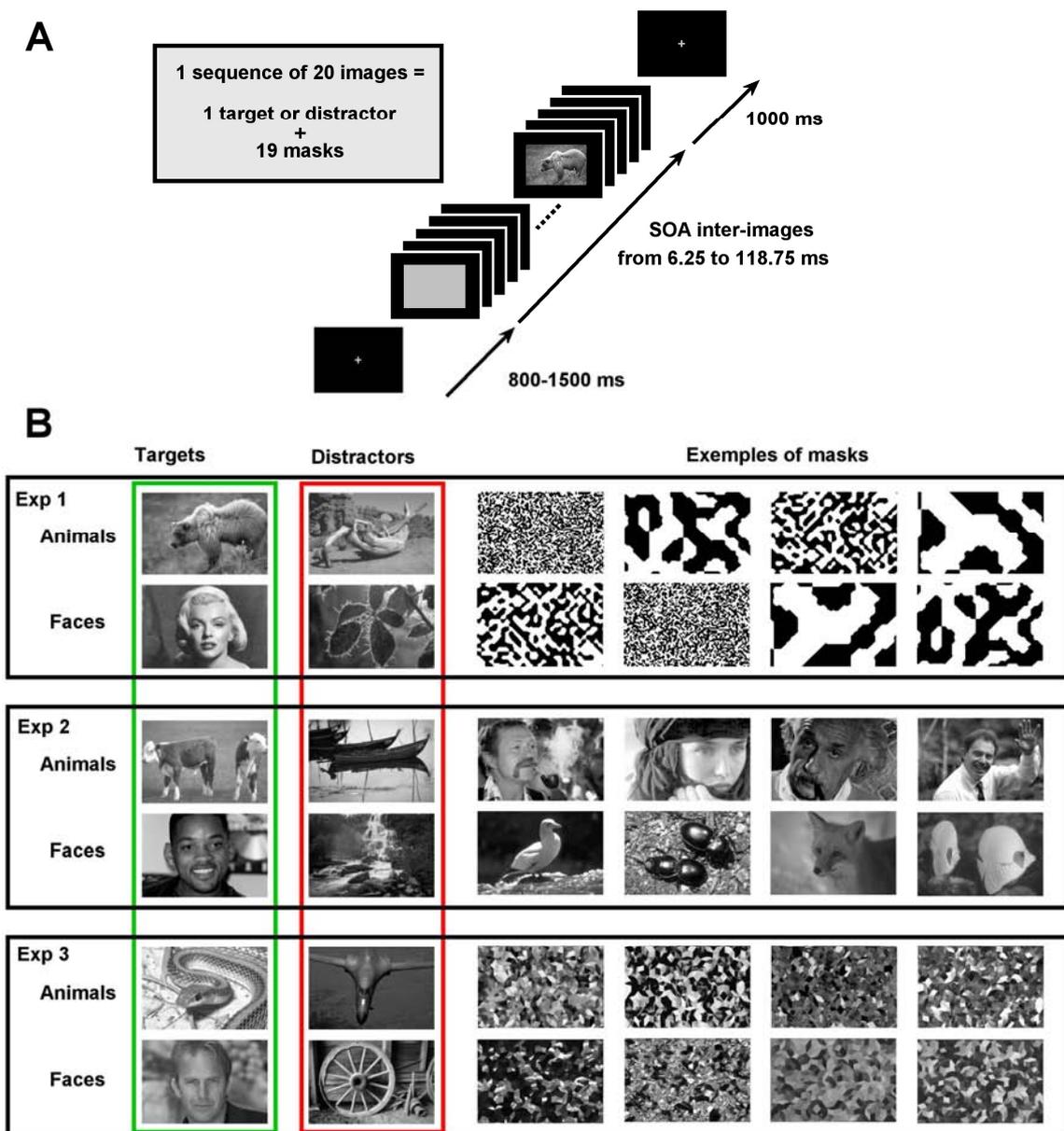


Figure 1: Experimental procedure in the three experiments. (A) Each trial was composed of a sequence of 20 pictures, containing 19 masking stimuli and one target/distractor. (B) In each of the three experiments, subjects had to perform two categorisation tasks in separated blocks: either an animal or a face categorisation. Distractor images were neutral natural scenes that did not contain animals nor faces. In Experiment 1, masking stimuli were composed by black and white patterns. In Experiment 2, the masks were the natural scenes of the opposite category and in Experiment 3, it was the scrambled versions of the opposite category images.

(of 96 trials), separated by a 10-minute-break. The categorisation task changed from one part to another, with half of the subjects detecting animals in the first part and faces in the second. The other subjects made the categorisation tasks in the inverse order. Across all the subjects, each SOA condition cumulated 576 trials for each categorisation task, but during the analysis, a checking measure excluded trials with a timing error (concerning less than 5% of the trials in total).

Stimuli

Figure 1B shows examples of target and distractors pictures, as well as the masks that were used in the three experiments.

Targets and distractors

A total of 576 grey-level pictures were used as targets. Half of them were natural scenes chosen from the Corel image database and contained a varied set of animals that included mammals, birds, fish, reptiles, insects... The other half were photographs of celebrity faces (actors, singers, politicians, sportsmen ...) selected from magazines and the web. A previous experiment (not published) revealed that around 91% of those faces stimuli were well known for 9 other subjects, and nearly 84% correctly named.

A further 576 grey-level natural scenes were chosen from the Corel database as distractors, and could be of any sort (man-made objects, landscapes, indoor, outdoor...) but contained neither animals nor faces.

Masking stimuli

Experiment 1: Pattern masks

The masking stimuli were black and white patterns, obtained from a white noise image filtered at 4 different spatial scales and contrasted to 100% (Figure 1B). Each picture was rotated and mirrored so that in total, a pool of 16 pictures was available (Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005). The four masking stimuli preceding and following the target/distractor were carefully selected in order to present each spatial scale once. Then all the masking stimuli separated from the target/distractor with 5 or more positions were randomly chosen from the pool of 16 masking stimuli.

Experiment 2: Scene masks

Here, the masking stimuli were chosen from the 288 target images of the opposite categorisation: when subjects had to categorise animals, masking stimuli were randomly picked from the pool of face targets. Similarly, animal targets constituted the masks in the face categorisation task.

Experiment 3: Scrambled masks

In this case, the masking stimuli were constituted by scrambled versions of the opposite categorisation images. As in Experiment 2, we used the pool of faces targets in the animal categorisation task and vice-versa, but after having scrambled them. The transformation was performed by a Matlab program that cut the pictures into about 230 pieces and shifted them according to an aperiodic Penrose tiling algorithm.

Subjects

In total, 25 different participants performed the experiments (mean age 26 [21-34], 12 females, 13 males) but several subjects participated in more than one experiment. Each task involved 12 subjects, at least half of whom were totally naive with respect to the experimental procedure. They all volunteered for the study, and gave their written informed consent.

RESULTS

Performance as a function of the SOA

Each trial was a sequence of pictures that contained 19 masking stimuli and one picture that on half of the trials was a target and on the other half, a neutral scene distractor. Accuracy is thus calculated as the percentage correct on target and distractor trials, with a chance level at 50%. Globally performance were quite low (Pattern [73.9%], Scenes [77,2%], Scrambled [78.8%]). Interestingly, the task was more difficult with pattern masks than the other masks at SOA below 44 ms ($p < 0.03$) but with longer SOA, the greatest disruptive effects were obtained with masks composed by natural scenes.

As expected, accuracy was strongly affected by the SOA (Figure 2A). When the stimulus was embedded within the flow, the first SOA step of 6 ms led to chance level whatever the mask used (Experiment 1: 50.1% individual range [38.2-56.3], Experiment 2:

52.9% [48.3-59.7], Experiment 3: 50.6% [43.5-55.4]). Then accuracy progressively increased and finally exceeded 90% with the last SOA condition at 119 ms (95.1% [90.2-97.9], 93.4% [85.1-97.9], 96.2% [92.7-97.9]).

Performance as a function of the target position

The increase of performance as a function of the SOA also depends on the target position in the sequence of 20 pictures. By presenting the target at the first position, we isolated the effects of backward masking. Conversely, when the stimulus was in the last (20th) position, only forward masking was present. Finally, the combined effect of both types of masking could be evaluated by presenting the target from the 2nd to the 19th position.

Figure 2B shows the increase in performance as a function of the SOA for the three masking conditions. As the results were comparable in the three experiments, data were grouped to one figure. The graph clearly shows that combining forward and backward masking results in far worse performance than when only one masking mechanism is present ($F(2,53) = 4.58, p < 0.015$). Performance was thus at chance level with 6 ms of SOA with both kind of masking. It is also noteworthy that backward masking is more disruptive than forward masking alone, whatever the SOA.

Performance as a function of the target category

Globally in the three experiments, subjects scored around 70% correct in the animal categorisation (Experiments 1 [69.5%], 2 [69.1%], 3 [71.5%]) and

83.3% in the face categorisation (Experiments 1 [78.4%], 2 [85.4%], 3 [86%]). Thus, detecting a famous face in the sequence of masking stimuli was much easier than detecting an animal. Whereas performance reached near 100% correct with SOA as short as 44 ms in the face categorisation, it had not yet reached its optimal value even with more than 100 ms of SOA in the animal categorisation.

An important part of the results emerges from the comparison of the disruption effects on the target categories depending on whether the masks were patterns, natural scenes or scrambled pictures. Two main observations could be inferred.

First, in the animal categorisation tasks, accuracy was lower when face photographs were used as masks than with scrambled versions of the same faces, at least for SOAs of 44 ms and above (Figure 3A). Furthermore, as shown in Figure 3B, this difference was maximised by using backward masking only (SOA 44 ms: $F(2,35) = 10.3, p < 0.001$, SOA 81 ms: $F(2,35) = 3.6, p < 0.04$). Although scrambling pictures can be expected to modify the spatial frequency spectrum, the scrambling procedure remains an efficient way of removing the semantic content of an image while retaining considerable information about the low-level properties. Importantly, the results show that face masks, which have a high semantic value, were more disruptive than the scrambled masks in the animal categorisation. Even the pattern masks have a lower impact on the animal categorisation than the faces masks, despite their high contrast.

Second, in the face categorisation task, pattern masks were much more disruptive than the other ones (All SOA: $F(2,215) = 6.7, p < 0.001$) before reaching the optimal accuracy. This was true whatever the position of the target in the sequence (Figure 3F, G and H). In

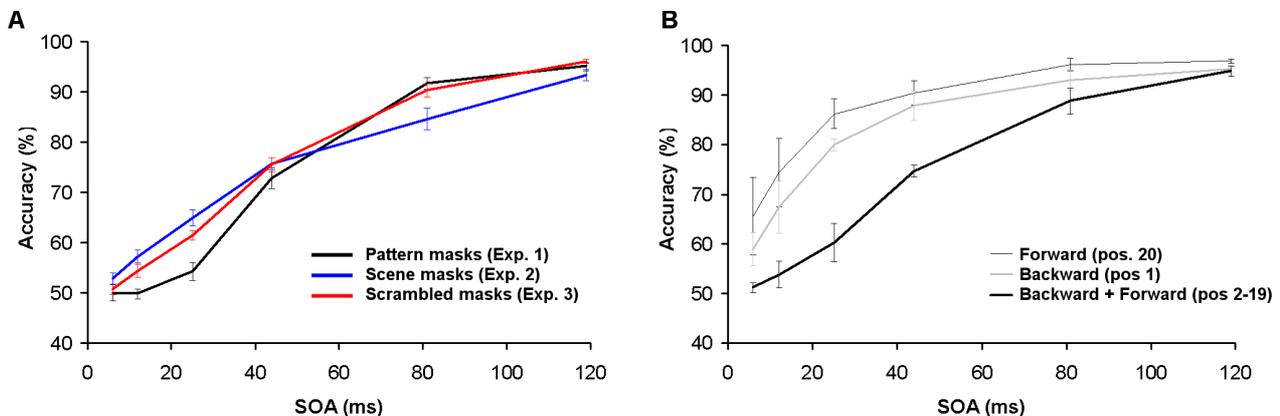


Figure 2: Accuracy as a function of the SOA in the different conditions. (A) Accuracy as a function of the SOA in the three experiments, according to the mask that can be a pattern (black curve), a scene (blue curve) or a scrambled scene (red curve). These data consider the trials when the stimulus is located from the 2nd to the 19th position in the sequence of pictures. (B) Accuracy as a function of the SOA, according to the position of the stimulus in the sequence. Processing are disrupted by backward masking when target is at the first position, (thin grey line), by forward masking when it is at the 20th position (thin black line), and by both masking mechanisms when it is ranged between the 2nd to the 19th position (bold line). Data were averaged across the three experiments.

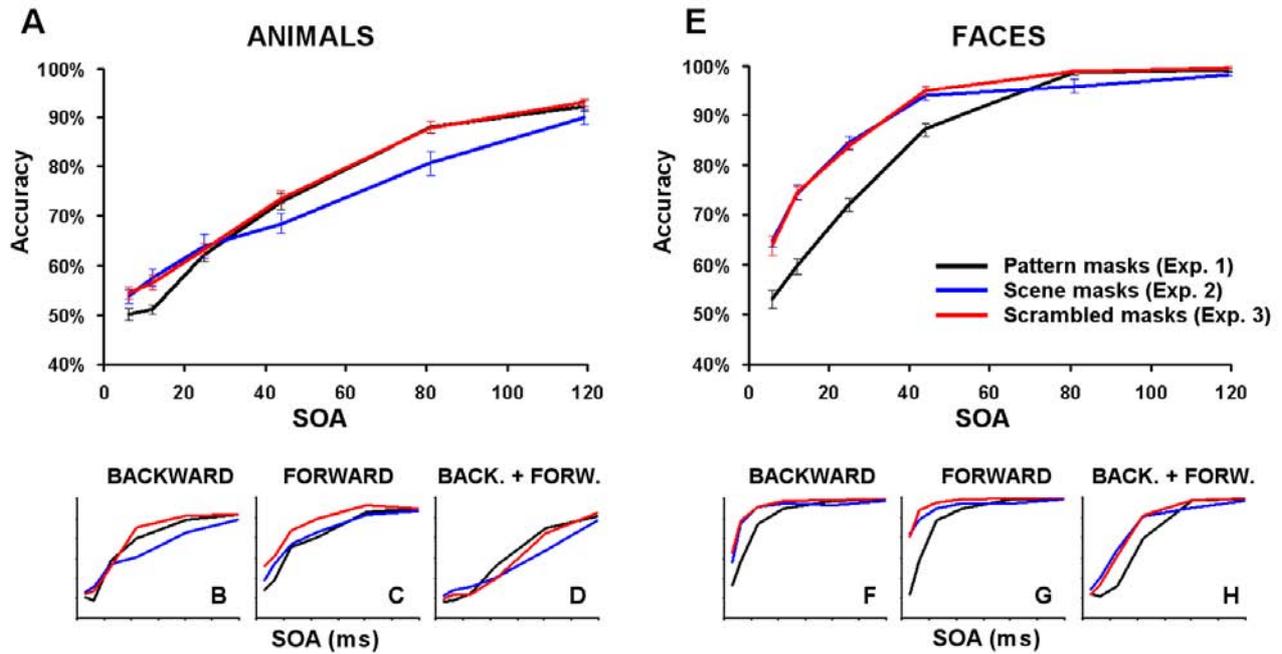


Figure 3: Accuracy as a function of the SOA, in the different experimental conditions. Data reported in the experiment that used pattern masks are in black curves (Exp. 1), scenes masks in blue curves (Exp. 2) scrambled masks in red curves (Exp. 3). Graphs (A) to (D) report the results in the animal categorisation, graphs (E) to (H) in the face categorisation. Graphs (A) and (E) averaged the data whatever the position of the target in the sequence of images. The bottom small graphs separate the different masking conditions: backward masking alone (graphs B and F), forward masking alone (graphs C and G) and both backward and forward combined (graphs D and H).

contrast with the previous animal task, there was absolutely no difference whether the masks were constituted by animal scenes or by scrambled versions of the same scenes.

DISCUSSION

Keyser et al. showed that STS neurons were still selective to their best naturalistic stimulus when presented within a sequence at a rate of 72 images per second, which leaves only 14 ms for any particular stage in the visual system to extract and integrate the information about the stimulus before the next one arrives (Keyesers, Xiao, Foldiak & Perrett, 2001). The associated behavioural results performed by human subjects (detection and memory tasks) were above chance level at this SOA, which confirmed that not all the information has been erased in the system. In our experiments, we used even faster presentation rates (up to 160 images/second with the 6 ms SOA condition), and found that, even with such short SOAs, categorisation performance could be well above chance level. However, the main point is that performance depends on two important factors: the nature of the target and of the masking stimuli. As an example, while famous faces can be easily detected within a sequence of animal scenes or scrambled

masks even at the shortest SOA, the use of pattern masks considerably reduces this performance (Figure 3E). In the same way, the animal categorisation seems to be more disrupted by faces masks than the other masks, in particular above 44 ms of SOA (Figure 3A). What do such findings imply for the masking mechanisms involved in the RSVP paradigm?

Low-level masking interference

Both the animal and face categorisation tasks were strongly affected by masking with 100% contrasted patterns, and the effects were even more persuasive in the face categorisation since face stimuli were particularly resistant to the other forms of mask (Figure 3A and E). In a previous study using this pattern masks, it was proposed that the amount of information available from the low-level properties of the stimulus is limited because of interference between the masks and the image that takes place early in the visual pathway, probably before V4 (Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005). This hypothesis is consistent with many studies on visual masking with noise or patterns, in which only the information that can escape the spatio-temporal interference between mask and stimulus can reach downstream processing stages. However, in the

previous published backward masking experiment, accuracy on animal categorisation was already close to optimal with SOAs above 40-50 ms, a result that can be linked with electrophysiological studies showing that many of the relevant physical properties on the stimulus can be extracted within the first 50 ms (Keysers, Xiao, Foldiak & Perrett, 2001; Kovacs, Vogels & Orban, 1995; Rolls, Tovee & Panzeri, 1999). In the present animal categorisation task, accuracy increased slowly with the SOA and even with more than 100 ms of SOA, performance had not yet reached its optimal value (Figure 3A) suggesting that a larger part of the stimulus integration process may be affected by masking. In fact, other masking mechanisms must be considered here, and firstly it should be noted that stimulus information must survive both forward and backward bottlenecks. The design of the experiments allowed us to analyse separately the different masking processes, and Figure 2B shows that the combination of forward and backward masking leads to much lower performance than when the mechanisms are isolated, in accordance with Keysers et al. (Keysers, Xiao, Foldiak & Perrett, 2001; Uttal, 1969). These authors also raised the possibility that the classical backward masking (targets are followed by the masking stimuli) could place the subject in an attentive state that may facilitate the processing of the first picture. However in our experiment, the three conditions (backward, forward and backward plus forward) were randomly presented in the series, and it is difficult to suppose that subjects were more attentive to the first image of the sequence while the target could appear at any position, all the more so that the best performance was obtained in the forward condition, with the target at the 20th position (Figure 2B). Combining backward and forward masking leads to lower performance and this should be interpreted rather by a stronger interference of the competing stimuli than by an attentive strategy of the subjects. The fact that backward masking was more disruptive than forward masking may appear surprising (Figure 3B). Little attention has been given to the neurophysiological impact of forward masking in the visual system, but psychophysical studies have reported that forward masking by noise or structured stimuli is typically weaker than backward masking, possibly because the target itself can produce a masking effect on the mask stimuli when it is presented in advance (Breitmeyer, 1984).

A face advantage

The advantage for face stimuli, both as stronger disruptive masks (Figure 3A) and as more resistant targets to masking effects (Figure 3E) can be interpreted in a number of ways.

(1) Faces and animals categories differ a lot in terms of basic features, which could have several consequences on the encoding of low-level properties. First, although we tried to use a set of pictures the more diverse as possible, the higher homogeneity of the face stimuli can explain why they are better recognised as targets in the sequences, since they are more uniform in shape, size, or even global configuration than in the case of the natural scenes containing animals (all condition mixed: 70% correct in the animal and 83.3% in the face categorisation). The visual system may have to pre-set and base its analysis on a smaller number of basic features in the faces compared to the animal task, which could explain the higher resistance to masking. Another important consequence of this homogeneity can arise from a possible tuning of specific spatial frequencies filters in the earlier areas towards the lower spatial frequency content of the face stimuli (Bachmann, Luiga & Poder, 2004; Fiser, Subramaniam & Biederman, 2001; Morrison & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002). This could explain that when faces are used as the masking stimuli, the tuning to the low spatial frequencies may impair the integration of the higher spatial frequency content of the animal images (Figure 3A). Nevertheless, it is interesting to note that in the animal categorisation task, the stronger masking effects by faces compared to the other masks does not appear from the first SOA conditions, but only from the fourth step, with an SOA of 44 ms. It may be difficult to justify such a shift by only low-level interference or pre-tuning, as they are supposed to take place at the very beginning of the processing and should be observable at the earliest SOA.

(2) Hence, other more high-level interpretations can be given to explain why the face category is so particular in this set of experiments. Faces are known to have a unique capacity to draw attention compared to other categories of objects. A number of studies have proposed that face processing may involve specialised units or areas within the brain, including studies with human fMRI (Kanwisher, McDermott & Chun, 1997), single cells in monkeys (Perrett, Rolls &

Caan, 1982) or event-related potentials (Bentin et al., 1996; Botzel, Schulze & Stodieck, 1995). Other researchers prefer to justify the special abilities of the visual system by other factors, like a greater learned expertise for face recognition (Gauthier, Curran, Curby & Collins, 2003; Tarr & Gauthier, 2000). Whatever the issue, the high performance in face recognition seems to involve some kind of automatic and mandatory processing (Boehm & Sommer, 2005; Farah, Wilson, Drain & Tanaka, 1995; Hershler & Hochstein, 2005; Lavie, Ro & Russell, 2003), and this face advantage over the other categories of objects appears to be particularly important in conditions of high perceptual load that involves competition among the stimuli (Ro, Russell & Lavie, 2001). In addition to the low-level masking effects described above, that were visible with all the masks used in the experiments (noise patterns, scenes or scrambled scenes), another type of interference may occur during the integration of the stimulus properties, particularly efficient when the masks contain a high semantic value, as it is the case with famous faces.

A semantic interference effect?

A number of studies support the idea that object identity is accessed automatically, without being required by task constraints (Boucart & Humphreys, 1997; James & Gauthier, 2004; Pins et al., 2004), a phenomenon that seems even more present for face stimuli (Boehm & Sommer, 2005; Farah, Wilson, Drain & Tanaka, 1995; Lavie, Ro & Russell, 2003). It has also been reported that implicit semantic information about unattended stimuli can affect the response to targets that require attention (Boucart & Humphreys, 1997; Gauthier, James, Curby & Tarr, 2003; James & Gauthier, 2004). Questions remain about when and where the visual and semantic sources of information integrated from different stimuli can interfere. Physical and semantic information processing are supposed to be carried out in different brain regions, even if they may all belong to the complex network involved in object recognition (Bar et al., 2001; Grill-Spector, Kushnir, Hendler & Malach, 2000; Malach et al., 1995). How could we explain the differential disruptive effect during the animal categorisation when the masks are constituted by famous faces, compared to the pattern and scrambled masks (Figure 3A)? Several hypotheses can be proposed, but one in particular would involve the

participation of disruptive mechanisms by the way of feedback connections, conveying information from areas that process the semantic content of the face masks. The arrival point of these connections may be localised at different levels of the visual cortical pathway, including the earlier areas as suggested by the differential effects that start early in the forward-masking only conditions (Figure 3C). In contrast, recurrent information about the semantic properties of masks that follow the target may intervene later, as suggested by the shift of the differential effect starting from 30-40 ms in the backward-masking only condition (Figure 3B). These recurrent semantic effects projecting early in the integration process could be useful under normal visual conditions for preparing the system to process a succession of semantically coherent scenes. It is important to note that what may appear to be close to the “conceptual masking” defined by Potter and collaborators nevertheless reflects different mechanisms (Chun & Potter, 1995; Potter, 1976), as it concerns the perceptual stage that precisely precedes the neural consolidation impairment described in the memory recognition tasks. In the present experiment we have analysed the masking effects that are occurring during the bottom-up processes that lead to target detection, which do not necessarily require the creation of a memory trace to postpone the recognition response. Rather, the semantic disruption defined here is supposed to interfere quite early in the integration process.

Interestingly, Evans and Treisman also raised the question of a potential semantic impact of the scene distractors in their set of RSVP experiments (Evans & Treisman, 2005). They actually observed that the presence of humans in the distractors severely impaired performance in an animal detection task but not in a vehicle detection task. However, when performing these same tasks with sequence of words instead of pictures, no such difference were found. They concluded that the disruptive effects of human pictures could be due to an important overlap of the visual features between humans and animals rather than to a semantic confusion. Nevertheless, the direct comparison of picture and word stimuli could potentially omit a number of intermediate processing, more closely related to the visual analysis of the objects. The visual and lexical systems required in each of the two tasks imply different anatomical structures and different time courses. Thus, detecting

animal among human pictures may require more processing, because of interference that could take place both at the level of basic feature extraction (since animals and humans share a lot of physical properties) and at the semantic level as humans are also animals.

CONCLUSION

Finally, through these different RSVP experiments, we could observe distinct masking mechanisms. By varying the SOA, the position of the target in the sequence, the content of target and masking stimuli, we could decompose through time several disrupting effects that may operate relatively early in the visual pathway. Therefore, we propose that low-level properties of the competing stimuli interfere with each

other, but that in addition, other kinds of higher-level attributes could also be in competition, and this may explain why famous faces pictures have this particular status in the tasks.

As already suggested by Keysers et collaborators, the RSVP paradigm is a powerful tool to investigate the dynamics of visual processing, as the analysis of a succession of pictures is close to what the visual system is lead to process through consecutive saccades. Furthermore, manipulating the complexity of the stimuli within the RSVP paradigm can allow to make the link between the extensive psychophysical literature on masking processing and on higher-level cognitive operations.

REFERENCES

- BACHMANN, T, LUIGA, I & PODER, E. (2004). Forward masking of faces by spatially quantized random and structured masks: on the roles of wholistic configuration, local features, and spatial-frequency spectra in perceptual identification. *Psychol Res*, 69:11-21.
- BACON-MACE, N, MACE, MJ, FABRE-THORPE, M & THORPE, SJ. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res*, 45:1459-69.
- BACON-MACÉ, N, MACÉ, MJ, FABRE-THORPE, M & THORPE, SJ. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res*, 45:1459-69.
- BAR, M, TOOTELL, RB, SCHACTER, DL, GREVE, DN, FISCHL, B, MENDOLA, JD, ROSEN, BR & DALE, AM. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, 29:529-35.
- BAVELIER, D. (1994). Repetition Blindness between visually different items: the case of pictures and words. *Cognition*, 51:199-236.
- BENTIN, S, ALLISON, T, PUCE, A, PEREZ, E & MCCARTHY, G. (1996). Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*, 8:551-565.
- BOEHM, SG & SOMMER, W. (2005). Neural correlates of intentional and incidental recognition of famous faces. *Brain Res Cogn Brain Res*, 23:153-63.
- BOTZEL, K, SCHULZE, S & STODIECK, SR. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res*, 104:135-43.
- BOUCART, M & HUMPHREYS, GW. (1997). Integration of physical and semantic information in object processing. *Perception*, 26:1197-209.
- BREITMEYER, BG. (1984). *Visual masking: an integrative approach (extracts)*. Oxford: Clarendon Press.
- CHUN, MM & POTTER, MC. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J Exp Psychol Hum Percept Perform*, 21:109-27.
- EVANS, KK & TREISMAN, A. (2005). Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform*, 31:1476-92.
- FARAH, MJ, WILSON, KD, DRAIN, HM & TANAKA, JR. (1995). The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. *Vision Res*, 35:2089-93.
- FISER, J, SUBRAMANIAM, S & BIEDERMAN, I. (2001). Size tuning in the absence of spatial frequency tuning in object recognition. *Vision Res*, 41:1931-50.
- GAUTHIER, I, CURRAN, T, CURBY, KM & COLLINS, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat Neurosci*, 6:428-32.
- GAUTHIER, I, JAMES, TW, CURBY, KM & TARR, MJ. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, 20:507-523.

- GRILL-SPECTOR, K, KUSHNIR, T, HENDLER, T & MALACH, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci*, 3:837-43.
- HERSHLER, O & HOCHSTEIN, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Res*, 45:1707-24.
- JAMES, TW & GAUTHIER, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Res*, 44:429-39.
- KANWISHER, N. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, 27:117-143.
- KANWISHER, N, MCDERMOTT, J & CHUN, MM. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17:4302-11.
- KEYSERS, C & PERRETT, DI. (2002). Visual masking and RSVP reveal neural competition. *Trends Cogn Sci*, 6:120-125.
- KEYSERS, C, XIAO, DK, FOLDIAK, P & PERRETT, DI. (2001). The speed of sight. *J Cogn Neurosci*, 13:90-101.
- KOVACS, G, VOGELS, R & ORBAN, GA. (1995). Cortical correlate of pattern backward masking. *Proc Natl Acad Sci U S A*, 92:5587-91.
- LAVIE, N, RO, T & RUSSELL, C. (2003). The role of perceptual load in processing distractor faces. *Psychol Sci*, 14:510-5.
- MALACH, R, REPPAS, JB, BENSON, RR, KWONG, KK, JIANG, H, KENNEDY, WA, LEDDEN, PJ, BRADY, TJ, ROSEN, BR & TOOTELL, RB. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A*, 92:8135-9.
- MORRISON, DJ & SCHYNS, PG. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon Bull Rev*, 8:454-69.
- PERRETT, DI, ROLLS, ET & CAAN, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res*, 47:329-42.
- PINS, D, MEYER, ME, FOUCHER, J, HUMPHREYS, G & BOUCART, M. (2004). Neural correlates of implicit object identification. *Neuropsychologia*, 42:1247-59.
- POTTER, MC. (1976). Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn]*, 2:509-22.
- POTTER, MC. (1993). Very short-term conceptual memory. *Mem Cognit*, 21:156-61.
- RAYMOND, JE, SHAPIRO, KL & ARNELL, KM. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform*, 18:849-60.
- RO, T, RUSSELL, C & LAVIE, N. (2001). Changing faces: a detection advantage in the flicker paradigm. *Psychol Sci*, 12:94-9.
- ROLLS, ET, TOVEE, MJ & PANZERI, S. (1999). The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci*, 11:300-11.
- SCHYNS, PG, BONNAR, L & GOSSELIN, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol Sci*, 13:402-9.
- SUBRAMANIAM, S, BIEDERMAN, I & MADIGAN, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cogn*, 7:511-535.
- TARR, MJ & GAUTHIER, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci*, 3:764-9.
- THORPE, S, FIZE, D & MARLOT, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520-2.
- TOVÉE, MJ & ROLLS, ET. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cogn*, 2:35-58.
- UTTAL, WR. (1969). The character in the hole experiment: Interaction of forward and backward masking of alphabetic character recognition by dynamic visual noise (DVN). *Percept Psychophys*, 6:177-181.
- VANRULLEN, R & KOCH, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *J Cogn Neurosci*, 15:209-17.

4.3.4 Résultats complémentaires

4.3.4.1 Moyennage des sets d'images

Une des grandes questions soulevées par cette série d'expériences porte sur les différences importantes de performance enregistrées dans les deux tâches de catégorisation. Il apparaît clairement que la détection des visages de célébrités dans un flux RSVP est beaucoup plus facile que la détection d'une image contenant un animal. Est-ce que ces écarts de performance peuvent s'expliquer par des différences bas-niveau entre les groupes d'images ? Il n'est pas simple de répondre à cette question, car si des différences existent, elles peuvent apparaître sous de nombreuses formes. Nous présentons ici le résultat d'un moyennage des images pixel par pixel séparément pour les photographies de visages et d'animaux. Comme le montre la Figure 4.9, le groupe d'images de visages de célébrités montre une homogénéité beaucoup plus grande que le groupe d'images contenant des animaux. Les photographies de visages ont été sélectionnées sur magazine et Internet, et un effort particulier a été fait pour que le groupe d'images soient aussi varié que possible : dans la position du visage, son orientation, sa taille relative dans l'image, les accessoires qui l'entourait (chapeau, cigarette ou pipe, vêtements) etc. ... Néanmoins, la forme globale du visage et du cou apparaît clairement sur l'image moyennée ainsi que le contour des cheveux et du menton. Comparativement, le schéma observé sur l'image moyennée à partir des images d'animaux est beaucoup plus diffus, et très peu d'informations de contraste local permettent de dégager l'impression d'une forme spécifique dans l'image. D'autre part, il est intéressant d'observer que l'image moyenne obtenue sur le groupe de distracteurs neutres (neutres car ce sont des scènes ne contenant ni animal ni visage) ressemble d'avantage à l'image moyenne du groupe Animal qu'à l'image moyenne du groupe Visage. L'image moyennée sur les distracteurs est plus diffuse, mais le haut de l'image est majoritairement constitué de luminance haute et le bas de l'image de luminance basse. Il n'est pas étonnant d'observer le même type de

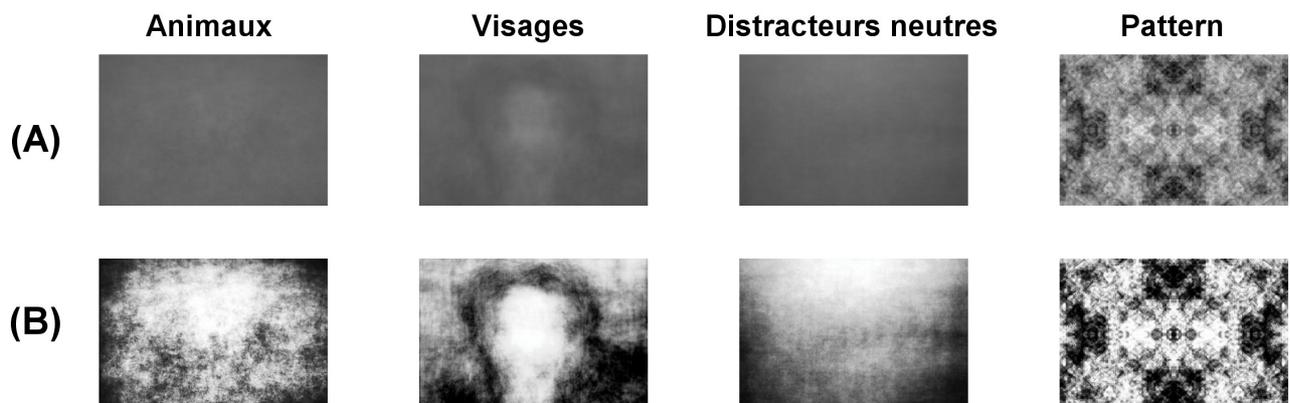


Figure 4.9 : Moyenne des images pour chaque catégorie de stimulus. (A) Les images sont moyennées pixel par pixel dans chaque groupe, ce qui donne un fond gris relativement uniforme. (B) Image moyenne égalisée. Les luminances sont redistribuées entre les valeurs des deux points correspondant aux luminances minimale et maximale.

schéma sur le groupe d'images contenant des animaux, car ces scènes naturelles sont souvent constituées de plusieurs plans horizontaux qui définissent le ciel, un plan éloigné et le premier plan situé au niveau du sol. Encore une fois, bien qu'un effort ait été fait pour sélectionner des images au contexte varié dans les deux groupes d'images, il est tout à fait possible d'affirmer qu'une image de distracteur neutre (dans un contexte naturel) ressemble d'avantage à une image d'animal (lui aussi souvent dans un contexte naturel) qu'à une photographie de visage. C'est ici un point fondamental : la meilleure performance obtenue lors de la catégorisation Visage est-elle due au fait que les images sont plus homogènes entre elles, ou au fait qu'elles sont plus distinguables des distracteurs comparativement aux images d'animaux ? Cette remarque est à relier aux espaces de recherche qui constitue les groupes d'images cibles et distracteurs et à la littérature plutôt limitée sur le sujet de la diagnosticité des cibles (Duncan & Humphreys, 1989; Hershler & Hochstein, 2005; Santhi & Reeves, 2004; Schyns, 1998; Treisman, 1991). Globalement, la difficulté d'une tâche de catégorisation est dépendante de plusieurs paramètres :

- La variabilité de l'espace des cibles (par exemple, les images d'animaux sont beaucoup plus variables et hétérogènes que les images de visages),
- La variabilité de l'espace des distracteurs,
- La distance entre les espaces cibles et distracteurs. Par exemple, chercher des chiens parmi des distracteurs neutres est plus facile que de chercher des chiens parmi des animaux.

Cette problématique a été abordée dans les travaux de thèse de Marc Macé avec lequel j'ai collaboré (article 4 en fin de chapitre) à travers une série d'expériences de catégorisation visuelle rapide dans lesquelles ces espaces ont été manipulés : les sujets devaient catégoriser des chiens (et d'autres animaux) parmi des distracteurs neutres (Condition 1), ou des chiens parmi des distracteurs dont la moitié était neutre et l'autre moitié contenait des animaux non chiens (Condition 2). Les temps de réaction augmentent considérablement dans la Condition 2 par rapport à la Condition 1, ce qui peut s'expliquer par la diminution de la distance entre les catégories cible et distracteur. Sachant que les cibles partagent un certain nombre de caractéristiques communes avec le groupe des distracteurs qui contient des animaux, il devient ainsi plus difficile de déterminer les indices diagnostiques spécifiques de la catégorie chien. Ces mêmes résultats ont d'ailleurs été retrouvés dans une tâche de catégorisation Oiseau / non-oiseau (Macé, Bacon-Mace, Nespoulous & Fabre-Thorpe, en préparation).

Nous ne reviendrons pas plus sur cet aspect de la diagnosticité des indices visuels associés aux catégories à détecter dans la suite de la thèse, mais je tenais à en parler, car ces notions d'espaces de recherche sont extrêmement importants pour notre compréhension générale des processus de catégorisation.

4.3.4.2 Analyse des images : taille relative des animaux dans l'image

Dans la même ligne de pensée que l'analyse complémentaire présentée ci-dessus, nous nous sommes demandés si la précision obtenue lors de la catégorisation Animal /non Animal pouvait s'améliorer avec certains types d'images. Effectivement, si les images de célébrités représentaient principalement les visages en gros plan, les images d'animaux en contrepartie étaient très variées dans la taille relative de la cible. Certains animaux étaient vus de loin, seuls ou en groupe, alors

que d'autres images portaient sur une partie d'un animal vue de près. Nous avons ainsi séparés deux groupes d'images, selon que la taille relative (subjective) des animaux était grande ou non (Figure 4.10). Nous avons alors recalculé la précision en fonction du SOA pour chaque groupe d'images. Il apparaît clairement que la précision est bien meilleure lorsque les animaux occupent une place importante dans l'image par rapport aux essais dans lesquels leur taille relative est réduite (Figure 4.10B). Cette différence n'est pas significative lorsque les masques sont des pattern mais dépasse les 12% en moyenne lorsque les masques sont des photographies de visages ($F(1,141)=8,4$ $p<0,004$), et 10% avec ces mêmes images constituant les masques « scrambled » ($F(1,141)=4,8$ $p<0,03$).

Si l'on regarde les images d'animaux sélectionnées pour leur petite taille relative (Figure 4.10A), il peut paraître normal que les sujets aient eu des difficultés à les détecter dans le flux d'images car la compréhension de ces images est réellement difficile. Ce qui est plus surprenant, c'est que cette différence n'est pas la même selon le masque utilisé, et l'écart est maximal lorsque les masques contiennent des visages. Finalement encore une fois, les images d'animaux pourraient pâtir de leur hétérogénéité en comparaison aux images de visages. Non seulement le système visuel doit baser son analyse sur un nombre plus important de caractéristiques bas-niveau lors de la catégorisation Animal, mais on peut même supposer que les masques de visages eux-même perturbent plus l'extraction des propriétés de bas-niveau que les masques d'animaux. On peut en effet imaginer un éventuel pré-réglage des canaux de fréquences spatiales vers les fréquences relativement basses lorsque des visages sont présents parmi les stimuli. Il est établi que les traitements visuels précoces filtrent les entrées visuelles avec un nombre limité de canaux, chacun réglé pour une bande de fréquence spatiale différente (Blakemore & Campbell, 1969; De Valois

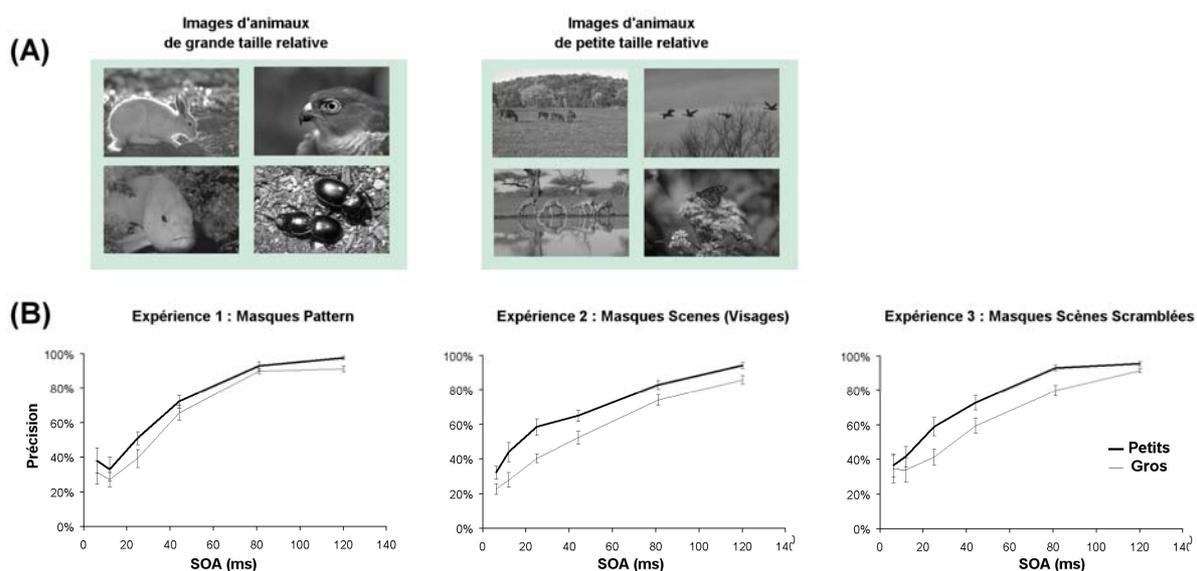


Figure 4.10 : Résultats de la tâche de catégorisation Animal dans les trois expériences pour des images cibles sélectionnées. (A) Exemples d'images dont la taille relative de l'animal dans l'image est grande ou petite. Deux groupes d'images comprenant respectivement 112 et 176 stimuli ont ainsi été sélectionnés. (B) Courbe de précision en fonction du SOA pour les trois expériences, selon le groupe d'images. Ces résultats ne prennent pas en compte les réponses sur les essais distracteurs. Quel que soit le masque, la précision est meilleure lorsque la cible contient un animal de grande taille relative.

& De Valois, 1980). De nombreux auteurs ont montré que l'analyse visuelle des visages porte majoritairement sur des gammes de fréquences spatiales relativement basses, conduisant à la construction d'une représentation configurale pertinente pour leur reconnaissance (Costen, Parker & Craw, 1996 ; Morrison & Schyns, 2001; Nasanen, 1999; Schyns, Bonnar & Gosselin, 2002). A l'inverse, la gamme de fréquences spatiales concernant des images d'animaux pourrait être plus large, puisque les images sont plus variées. Non seulement cette variété ne permet pas au système visuel de se pré-régler sur une gamme de fréquence particulière mais cela peut aussi expliquer pourquoi les images d'animaux sont tellement perturbées par les masques de visages. Ces dernières années, il a en effet été proposé que le système pré-règle ses gammes de fréquences selon un mode plutôt flexible (Bachmann, Luiga & Pöder, 2004; Fiser, Subramaniam & Biederman, 2001; Morrison & Schyns, 2001; Schyns & Oliva, 1999). Par exemple, Schyns et collaborateurs ont proposé que les éventuelles interactions entre les contraintes de la tâche et la nature de l'information visuelle pouvait moduler les besoins du système en terme de fréquence spatiale. De la même manière, Fiser et collaborateurs ont montré à travers un protocole RSVP que le système visuel ajuste son analyse sur la bande de fréquence majoritaire de l'image précédant la cible, avec un impact considérable sur les performances.

Dans nos expériences, nous pouvons supposer que lorsque les images de visages sont utilisées comme des masques, le système est réglé pour des fréquences spatiales relativement basses d'où sa difficulté à détecter les animaux qui peuvent être à n'importe quelle échelle spatiale. A l'inverse, les animaux utilisés en tant que masque ne perturbent pas ou peu les traitements sur les visages, puisque le système a eu besoin de se pré-régler sur une gamme de fréquences beaucoup plus vaste. Cette interprétation est peut être un peu poussée, et nos données sont plutôt limitées pour aboutir à de telles conclusions. Mais notre apparente facilité à détecter des visages dans des conditions de perception dégradée a suscité de vives discussions dans l'équipe, d'autant qu'elle contredit des résultats précédemment obtenus dans une tâche de catégorisation de photographies de visages et d'animaux, sans masquage, dans laquelle les performances étaient égales pour les deux catégories d'images (Rousselet, Mace & Fabre Thorpe, 2004). Cette série d'expériences RSVP n'avait pas pour objectif initial de comparer les performances pour ces deux catégories de stimuli, mais plutôt de comparer l'impact de différents types de masques sur l'analyse visuelle de chaque tâche de catégorisation prise indépendamment. Rien n'a donc été fait pour contrôler les propriétés bas-niveau des groupes d'images, d'où cette longue discussion ! Dans tous les cas, il est certain que les interprétations à donner doivent être multiples pour justifier à la fois les bonnes performances dans la tâche Visage et les effets perturbateurs importants des masques Visages. Si l'on ne peut pas exclure totalement la participation de facteurs de bas-niveau, il peut aussi être tout à fait possible d'envisager des effets plus cognitifs ... Nous reviendrons sur ce sujet dans la discussion qui suit.

4.3.5 Discussion

Dans leur article de 2000, Keysers et collaborateurs ont montré que les neurones du STS étaient encore sélectifs à leur stimulus préférentiel (scène naturelle) pour des fréquences de présentation de l'ordre de 72 images par seconde, ce qui laisse seulement 14 ms de délai au système visuel pour extraire et intégrer spécifiquement l'information de ce stimulus avant que les autres stimuli ne soient présentés à l'écran (Keysers, 2000). De plus, la performance comportementale réalisée par des sujets humains avec ces mêmes stimuli était nettement au-dessus du niveau chance pour des tâches de détection et de mémorisation, ce qui confirme que toute l'information n'a pas été complètement effacée par le masquage avec 14 ms de SOA. Dans notre expérience, nous avons utilisé des taux de présentation encore plus rapides, jusqu'à 160 images par seconde, et nous venons de voir que même avec ces fréquences de présentation très élevées, la performance dans une tâche de catégorisation reste relativement bonne, en particulier lorsqu'il s'agit de détecter des visages. L'objectif principal de cette série d'expériences était de montrer que la performance est modulée par au moins deux facteurs : la nature de la cible et la nature des stimuli utilisés en tant que masque.

4.3.5.1 Interférences de bas-niveau

Dans les deux tâches de catégorisation, les performances sont particulièrement affectées lorsque les masques sont constitués de patterns noir et blanc contrastés à 100%. Ce phénomène est d'autant plus marquant dans la catégorisation Visage pour laquelle les autres types de masque (Scènes et Scènes « scrambled ») ont un impact réduit. Comme nous l'avons vu dans le chapitre 2, nous avons proposé que les informations concernant les propriétés de bas-niveau du stimulus sont limitées à cause d'une interférence avec les informations du masque intervenant relativement tôt dans la chaîne des traitements visuels, probablement avant V4. Cependant, alors que la tâche de masquage rétrograde permettait de dépasser 80% de réponses correctes dès 40 ms de SOA, les différentes expériences de RSVP ont montré qu'il fallait attendre au moins 80 ms de SOA pour obtenir cette même performance dans la tâche de catégorisation Animal. En fait, d'autres mécanismes de masquage doivent être considérés dans les expériences RSVP et notamment l'intervention du masquage antérograde (les masques précédant la cible affectent aussi les traitements). La figure 4.8 montre en effet à quel point la performance chute avec l'utilisation combinée des deux masquages, rétrograde et antérograde (Keysers, 2000; Uttal, 1969). Lorsque nous avons commencé à mettre au point les expériences RSVP, nous étions étonné de voir que les performances étaient plutôt mauvaises dans l'ensemble par rapport aux expériences précédentes de masquage rétrograde. Or nous avons négligé un facteur important en considérant que les effets du masquage antérograde ne pouvaient être que minimes. Assez peu d'études se sont intéressées à l'impact neurophysiologique du masquage antérograde dans le système visuel, mais il semble que ce phénomène puisse s'expliquer par une « fatigue » des neurones lors de la présentation successive de plusieurs stimuli. Nelson a en effet reporté de l'inhibition au niveau de neurones du LGN chez le chat, inhibition qui peut durer jusqu'à 200 ms en présentant deux stimuli successifs avec des ISI relativement courts (inférieurs à 150

ms)(Nelson, 1991). Encore chez le chat, Stratford et collaborateurs ont montré que la stimulation des fibres afférentes au niveau du LGN produisait une baisse de l'amplitude de la réponse des neurones enregistrés dans la couche IV du cortex visuel primaire (Stratford et al., 1996). Cette baisse était d'autant plus marquée que l'intervalle de temps entre les deux stimulations diminuait en-dessous de 100 ms.

En revanche, il est étonnant d'observer que seulement dans la tâche de catégorisation Animal, le masquage rétrograde semble plus puissant que le masquage antérograde. Les résultats d'expériences psychophysiques ont conduit certains auteurs à proposer que le masquage antérograde par bruit ou par pattern est généralement plus faible parce que la cible elle-même exerce des interférences sur le traitement des masques qui la précède (Breitmeyer, 1984). Il reste toujours difficile d'expliquer pourquoi cette asymétrie n'est pas retrouvée dans la catégorisation des Visages, si ce n'est que les visages sont justement des stimuli particuliers, à la fois en termes de propriétés physiques et attentionnelles.

4.3.5.2 Supériorité du traitement des visages ?

Les trois expériences de RSVP ont montré que les stimuli de visages de célébrités sont à la fois des masques plus puissants et des stimuli plus résistants au masquage que les scènes contenant des animaux. Encore une fois, bien que cette comparaison n'était pas au départ l'objectif de cette série d'expériences, les différences importantes de performance doivent maintenant être interprétées.

Tout d'abord, comment peut-on expliquer que les images de visages sont plus résistantes au masquage que les images d'animaux ? Comme nous l'avons vu dans la description des analyses complémentaires, les images de visages et d'animaux diffèrent énormément sur le plan de leurs propriétés physiques. Le groupe d'images de visages est plus homogène que celui des scènes naturelles contenant des animaux ou que celui des distracteurs neutres, ce qui peut expliquer les meilleures performances dans la tâche de catégorisation puisque le système visuel peut se baser sur une quantité limitée d'indices diagnostiques.

Mais il est aussi possible d'envisager des interprétations plus cognitives de ce qui s'apparente à une « supériorité » du traitement des visages. Il est évident que les visages constituent une classe de stimuli privilégiés avec lesquels nous interagissons en permanence. La littérature sur la spécificité du traitement des visages est particulièrement vaste, et nous ne chercherons pas ici à présenter de revue exhaustive de ces travaux (pour de très bonnes revues, voir ces articles (Bentin et al., 1996; Carmel & Bentin, 2002; Farah, Wilson, Drain & Tanaka, 1998; Gauthier & Logothetis, 2000; Tarr & Cheng, 2003) ou encore ces thèses françaises : (George, 1997; Itier, 2002; Mouchetant-Rostaing, 2001))

Certaines données comportementales et électrophysiologiques soutiennent l'idée que les visages forment une catégorie de stimuli différente de toutes les autres catégories d'objets. Par exemple, les visages présentés à l'endroit sont généralement mieux reconnus que les autres catégories d'objets, tandis que les visages présentés à l'envers sont moins bien reconnus que les autres classes de stimuli (Diamond & Carey, 1986; Valentine, 1988; Yin, 1969). D'autre part, il existe des patients présentant des troubles spécifiques de la reconnaissance des visages et malgré le fait

qu'ils aient une vision normale, il leur est impossible de reconnaître le visage d'une personne, même familière (patients prosopagnosiques) (Farah, 1996; Moscovitch, Winocur & Behrmann, 1997; Renault et al., 1989; Sergent & Poncet, 1990). De nombreux auteurs ont ainsi proposé que l'analyse visuelle des visages implique des structures spécialisées, en se basant à la fois sur des données enregistrées en IRMf (Kanwisher, 2000; Kanwisher, McDermott & Chun, 1997; McCarthy, Puce, Gore & Allison, 1997), en enregistrements unitaires (Perrett, Rolls & Caan, 1982; Tsao, Freiwald, Tootell & Livingstone, 2006) ou en potentiels évoqués (Bentin et al., 1996; Botzel, Schulze & Stodieck, 1995). D'autres auteurs ont proposé que l'efficacité du système visuel pour reconnaître des visages peut s'expliquer par notre plus grande expertise pour cette catégorie de stimuli, acquise depuis la naissance (Gauthier, Curran, Curby & Collins, 2003; Tarr & Gauthier, 2000). Des études chez le primate ont en effet montré que les neurones de IT génèrent des réponses sélectives à la suite d'un processus d'apprentissage lié à la reconnaissance d'un pattern complexe, pattern qui était sans signification pour l'animal avant l'apprentissage (Miyashita, 1993; Thorpe, 1994). Que les visages soient traités par une région particulière du système visuel ou qu'ils bénéficient d'un effet d'expertise, il semble de toutes façons que les bonnes performances enregistrées dans les tâches de reconnaissance de visages impliquent l'existence de traitements rapides voire « automatiques » des visages (Boehm & Sommer, 2005; Farah, Wilson, Drain & Tanaka, 1995; Hershler & Hochstein, 2005; Lavie, Ro & Russell, 2003), qui semblent particulièrement utiles dans des conditions de surcharge attentionnelle impliquant une certaine forme de compétition entre les stimuli (Ro, Russell & Lavie, 2001). Encore une fois il est intéressant de revenir à cette idée que des groupes de neurones accumulent au cours du temps l'information sur les différents vagues d'entrées rétiniennes. Les enregistrements unitaires chez le singe suggèrent qu'au sein d'une même structure visuelle comme le cortex inféro-temporal, la vitesse des traitements dépend du taux d'accumulation de l'activité des neurones sélectifs aux stimuli (Perrett, Oram & Ashbridge, 1998). Et ce taux pourrait justement dépendre de la fréquence à laquelle ce stimulus a été vu auparavant. Plus un objet est familier comme c'est le cas des visages, plus les seuils de discrimination pourraient être atteints rapidement. Dans le cas de nos expériences de masquage, cette remarque est importante parce qu'elle pourrait expliquer en partie pourquoi les stimuli de visages résistent aussi bien aux mécanismes d'interférence. La courbe de performance en fonction du SOA atteint très vite un plateau, comme si la majorité de l'information pertinente avait pu « s'échapper » rapidement des étapes perturbatrices, certainement parce que les seuils de discrimination sont plus vite atteints que pour des images d'animaux.

D'autre part, il faut rappeler que nous avons choisi des visages de célébrités. Une des expériences pilotes présentée précédemment avait montré que les images de célébrités permettaient d'obtenir une précision encore supérieure à celle obtenue avec des visages inconnus. Même si cette comparaison manquait de contrôle du point de vue des propriétés physiques entre les groupes d'images, il est quand même possible d'argumenter que ces images de célébrités, parce qu'elles représentent des stimuli hautement familiers pour les sujets, permettent d'atteindre encore plus rapidement une activation sélective dans les aires concernées. Rappelons d'ailleurs que des enregistrements chez le sujet humain ont montré que la région temporale semble

présenter un haut degré d'invariance même pour des représentations particulièrement abstraites (Quian Quiroga et al., 2005). Par exemple certains neurones répondent spécifiquement à la présentation d'une célébrité en particulier (Jennifer Aniston), que le stimulus soit une photographie de cette célébrité, un dessin ou même seulement son nom. Ces réponses invariantes sont probablement issues de l'expérience que les sujets ont eu avec les images, les mots ou d'autres stimuli en association avec un individu ou un objet donné.

Globalement donc, les meilleures performances dans la tâche de catégorisation Visage peuvent s'expliquer à la fois grâce à leurs propriétés physiques (grande homogénéité, gamme de fréquences spatiales) et par des propriétés plus cognitives, qui font de cette classe d'images des stimuli connus, pour lesquels les sujets ont une grande expertise et pour lesquels les traitements visuels sont particulièrement bien optimisés. Mais cette remarque peut aussi s'appliquer aux stimuli de masques utilisés dans les trois expériences. Est ce que les propriétés bas-niveau des masques sont les seuls éléments à prendre en compte pour expliquer les mécanismes du masquage ou doit-on aussi considérer leur contenu informationnel ? Dans notre cas, dans les expériences 1 et 3, nous avons utilisé des masques qui n'avait aucun contenu sémantique pour les sujets (Patterns et Scènes « scrambled »), contrairement à l'expérience 2, dans laquelle les masques étaient constitués par les scènes naturelles cibles de la catégorisation opposée (Scènes). En particulier, il est intéressant de remarquer que les masques composés de visages semblent être particulièrement efficaces pour perturber les traitements dans la tâche de catégorisation Animal. Dès 44 ms de SOA, la précision atteinte dans cette tâche est moins élevée avec les masques composés de visages de célébrités qu'avec les autres masques, y compris les masques Pattern très contrastés. Ce décalage de 40 ms dans le SOA est intéressant car il permet de poser l'hypothèse de mécanismes de masquage un peu plus tardifs que ceux consistant à interférer avec les propriétés de bas-niveau de la cible, en faisant intervenir des propriétés plus complexes, voire le contenu sémantique des masques Visages...

4.3.5.3 Des mécanismes de compétition haut-niveau ?

Récemment, des enregistrements unitaires chez le singe ont montré que la région connue pour être particulièrement activée en IRMf pour des visages contenait en effet près de 97% des cellules sélectives à ces stimuli (Tsao, Freiwald, Tootell & Livingstone, 2006). Même si la présentation d'autres classes de stimuli, comme des animaux, doit aussi permettre d'activer cette région du STS, il est certain que la présentation de visages avec des intervalles de temps courts va mener à une certaine forme de compétition sous forme d'inhibition locale entre les représentations haut-niveau de ces stimuli (Keysers & Perrett, 2002). Cette compétition, qui se joue en faveur des stimuli de visages, est certainement un des facteurs les plus probables pour expliquer à la fois les bonnes performances pour catégoriser les visages en tant que cibles, et l'impact significatif de ces visages utilisés comme masques dans la tâche Animal.

D'autre part, un certain nombre d'études ont montré que l'identité des objets semble émerger de manière automatique, même lorsqu'elle n'est pas requise par les contraintes de la tâche (Boucart & Humphreys, 1997; James & Gauthier, 2004; Pina et al., 2004). Ce phénomène semble être

encore plus marqué pour les visages² (Boehm & Sommer, 2005; Farah, Wilson, Drain & Tanaka, 1995; Lavie, Ro & Russell, 2003; Rousselet, Macé, Thorpe & Fabre-Thorpe, soumis). Il a d'ailleurs été rapporté que l'accès implicite aux informations sémantiques de stimuli sur lesquels aucune attention particulière n'est portée pouvait influencer les réponses à des stimuli qui eux requiert de l'attention (Boucart & Humphreys, 1997; Gauthier, Curran, Curby & Collins, 2003; James & Gauthier, 2004). Ces interférences se retrouvent d'autant plus si le sujet doit inhiber le stimulus distracteur comme c'est le cas dans les expériences d'amorçage négatif (Tipper & Cranston, 1985; Tipper & Driver, 1988). Les processus d'inhibition exercés sur le stimulus d'amorçage interfèrent avec les traitements du stimulus cible, réduisant considérablement les performances si les deux stimuli sont sémantiquement liés. Il est donc tout à fait possible d'imaginer que les masques qui ont un contenu sémantique, et en particulier les masques composés de visages de célébrités pour lesquelles la « valeur » sémantique doit être élevée, exercent des interférences négatives sur l'encodage des propriétés perceptuelles du stimulus cible, d'autant que les deux catégories de stimuli sont sémantiquement proches. Cet argument pourrait expliquer pourquoi la performance dans la catégorisation Animal est d'avantage affectée par les masques Visages que par les autres masques.

Il reste cependant difficile de déterminer dans quelles aires cérébrales et dans quelle gamme de latence ces informations visuelles et sémantiques interfèrent entre les différents stimuli. En ce qui concerne les latences, est-il raisonnable de penser que des informations sémantiques ont déjà eu le temps d'être intégrées au moment où la représentation du stimulus cible entre en compétition avec celles des masques ? Dans ses expériences pionnières, Molly Potter comparait les performances de détection et de mémorisation d'un stimulus particulier en fonction du SOA. La condition de détection se déclinait en deux sous-conditions dans lesquelles soit l'image cible était présentée avant la séquence, soit cette image cible était décrite préalablement en quelques mots (« une femme assise sur un lit »). Lorsque l'image était présentée avant la séquence, le sujet pouvait se baser sur un ensemble de propriétés bas-niveau préactivées avant l'affichage des images du flux. En revanche, lorsque le stimulus cible était simplement décrit par un énoncé, le sujet devait avoir intégré en partie le sens de l'image pour le faire correspondre aux représentations plus abstraites générées par la description préalable de la cible. Cette condition est d'ailleurs très comparable à nos expériences de catégorisation. Or il est surprenant de remarquer que les courbes de performance restent très proches dans les deux sous-conditions, avec un décalage significatif mais inférieur à 10 points de pourcentage correct sur les essais cibles, alors

² Dans un article de l'équipe actuellement soumis, Rousselet et collaborateurs ont argumenté en faveur d'une automaticité des traitements pour la catégorie Visage, que l'on retrouve dans une moindre mesure pour les autres catégories comme Animal et Moyen de transport (Rousselet et al., soumis). En effet, les auteurs ont observé que l'activité différentielle enregistrée en potentiels évoqués, qui habituellement reflète le traitement différentiel entre essais cibles et distracteurs, est particulièrement faible et retardée pour la catégorie Visage. Cette observation contraste avec les performances comportementales qui montrent au contraire que les sujets sont particulièrement rapides et efficaces pour effectuer la tâche de catégorisation des visages. Les auteurs proposent donc que l'automaticité du traitement des visages est responsable de la diminution de cette activité, car ces stimuli sont traités indépendamment de leur statut cible ou distracteur. La soustraction des signaux moyennés séparément pour ces deux types d'essais ne permet donc pas de mettre en évidence l'existence de traitements différentiels.

que la performance perdait 50 à 70 % en comparaison dans la condition de mémorisation dans laquelle l'image cible était présentée après la séquence (Figure 1.6 dans le chapitre 1). Ces données confortent l'idée que les sujets ont le temps de « comprendre » une grande partie du sens des scènes, lorsqu'elles sont présentées jusqu'à des fréquences de 10 hertz..

D'autre part sur le plan anatomique, les propriétés visuelles et sémantiques semblent être intégrées dans des régions différentes (Gorno-Tempini, Cipolotti & Price, 2000; Mummery, Patterson, Hodges & Price, 1998), mais ces traitements appartiennent de toute façon au réseau complexe impliqué dans la reconnaissance d'objets (Bar & Biederman, 1999; Grill-Spector, Kushnir, Hendler & Malach, 2000; Humphreys, Price & Riddoch, 1999; Malach et al., 1995). Pour expliquer la différence obtenue dans notre expérience RSVP de catégorisation Animal entre les différents masques, nous pouvons supposer qu'il existe des mécanismes perturbateurs par le biais de connexions feedback, qui ré-injecteraient l'information des aires engagées dans les traitements sémantiques des masques Visages vers les aires sensorielles impliquées dans le traitement des cibles. Ces modulations feedback pourraient expliquer pourquoi les performances sont d'avantage affectées par les masques Visages comparativement aux autres masques dès les premiers pas de SOA dans la condition antérograde (Figure 4.7C), alors que la modulation intervient plus tardivement dans la condition rétrograde dans laquelle les masques sont présentés après la cible. Biologiquement, cette idée qu'il existe des interférences sémantiques « en retour » des premiers traitements pourrait paraître plausible dans les conditions de vision naturelle, afin de préparer le système à traiter une succession de scènes cohérentes entre elles.

Ce type d'interférence a aussi été retrouvé par d'autres auteurs dans des expériences RSVP utilisant des scènes naturelles (Evans & Treisman, 2005). Evans et Treisman ont observé que la catégorisation Animal était bien plus affectée par la présence d'humains dans les scènes constituant les masques que ne l'était la catégorisation de moyens de transport. Cependant ces auteurs préfèrent argumenter que les effets perturbateurs des images d'humains sont d'avantage causés par un important recouvrement des propriétés bas-niveau appartenant aux catégories Animal et Humain que par une éventuelle confusion sémantique. Elles justifient cet argument au moyen d'une expérience contrôle dans laquelle les sujets devaient reproduire le même type de tâches, non plus avec des scènes naturelles mais avec des mots. Dans ce cas, aucun effet perturbateur n'est obtenu sur le traitement des noms d'animaux lorsque les masques sont des mots faisant référence à des humains. Or la comparaison directe du traitement des images et des mots semble fragile ici car les systèmes visuels et lexicaux requis dans les deux types de tâches impliquent des régions anatomiques et une dynamique temporelle des traitements différentes. Même s'il est certain que les catégories Animal et Humain partagent un certain nombre de caractéristiques physiques communes, l'expérience contrôle ne permet pas non plus d'exclure la possibilité d'une confusion sémantique, d'autant que ... les humains sont aussi des animaux !

4.3.5.4 Nature des informations rapides

Il est inutile de souligner l'importance des travaux de Anne Treisman dans la littérature psychophysique et ses derniers travaux de RSVP réalisés avec son étudiante Karla Evans soulèvent un nombre important de questions sur la nature des indices diagnostiques utilisés dans des tâches de catégorisation alors que les conditions de perception sont limitées dans le temps. Dans une de leurs expériences en particulier, le sujet devait non seulement détecter une cible dans le flux mais aussi l'identifier. Si les sujets ne pouvaient pas donner de nom précis de l'animal ou du moyen de transport détecté, il leur était demandé de décrire tous les indices visuels qui leur avait permis de penser qu'une image cible était bien présente dans la séquence. Dans cette expérience, le SOA inter-image était de 75 ms, c'est à dire dans la partie supérieure de notre gamme de SOA. Les sujets atteignent en moyenne 80% de réponses correctes pour détecter des images d'animaux dans la séquence lorsque les masques ne contenaient pas d'humains ; résultat très proche de celui que nous avons obtenu dans l'expérience pilote décrite dans la section 4.2.3 (83,8% avec 81 ms de SOA). En revanche parmi ces détections correctes, à peine la moitié des cibles étaient correctement identifiées. Une analyse du report des sujets montre qu'ils étaient capables de fournir des informations approximatives mais relativement justes comme le nom de la catégorie superordonnée (« mammifère » pour une vache), le nom d'un animal très proche (« cerf »), ou des caractéristiques précises (« large et brun »). Globalement donc, dans plus de 90% des essais où la détection était correcte, les sujets pouvaient justifier leurs réponses en décrivant correctement une propriété de la cible. Les auteurs argumentent alors que les traitements perceptuels qui permettent de dégager des informations sur ces différentes propriétés sont suffisants pour effectuer rapidement une tâche de détection. En revanche l'identification de l'objet requiert obligatoirement des traitements supplémentaires consistant à « lier » ces informations entre elles pour générer un « fichier objet » accessible consciemment (Kahneman, Treisman & Gibbs, 1992). Plus particulièrement, les étapes perceptuelles pourraient être rapidement accomplies par un transfert rapide et feedforward de l'information le long de la hiérarchie visuelle afin de dégager les propriétés caractéristiques de la cible, suffisantes pour une tâche de détection. Puis si la scène est toujours disponible, l'attention est alors portée sur cette cible afin de lier ses différentes propriétés entre elles et autoriser la construction d'une représentation plus abstraite. Même si un éventuel désaccord peut porter sur la nature des propriétés dégagées rapidement (nous y reviendrons dans la section suivante), globalement, les travaux de Evans et Treisman sont tout à fait compatibles avec les hypothèses proposées par l'équipe ces dernières années (Delorme, Richard & Fabre Thorpe, 1999; Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001; VanRullen & Thorpe, 2001). D'ailleurs, nous avons déjà soulevé certains points dans le dernier chapitre, à travers la description du rapport des sujets qui suggérait une éventuelle dissociation entre leurs traitements perceptuels et conscients. La section suivante décrit une nouvelle (et dernière !) expérience RSVP dans laquelle nous avons aussi exploité le décalage entre ce qu'il est possible de détecter et ce que nous pouvons réellement dénommer lors d'une tâche de catégorisation.

4.4 Détection et identification d'images naturelles dans un flux RSVP : expérience de dénomination

Nous avons vu dans les expériences précédentes que les performances pour catégoriser un objet dans une séquence RSVP étaient particulièrement basses, ce que nous avons essayé d'expliquer en décrivant différents mécanismes de masquage. Afin de comprendre plus explicitement la difficulté qu'avaient les sujets dans ce type de tâche, nous avons réalisé une nouvelle expérience dans laquelle nous avons demandé aux sujets non seulement de *détecter* des stimuli d'animaux et de visages dans les séquences RSVP, mais aussi d'essayer de les *identifier*.

4.4.1 Protocole expérimental

Sujets

8 sujets ont participé à cette expérience (moyenne d'âge 29 ans, 3 femmes, 5 hommes).

Tâche

Comme dans les expériences précédentes de RSVP, les sujets ont réalisé une catégorisation Animal dans la moitié des séries puis une catégorisation Visage et l'ordre de ces deux tâches était contrebalancé selon les sujets. Cependant, le protocole diffère quelque peu de celui utilisé dans les expériences précédentes (Figure 4.11). Un essai commençait toujours par l'affichage d'une croix de fixation pendant 1200 à 1600 ms, suivie par la séquence de 20 images contenant 19 masques de scènes naturelles et une cible. Le sujet disposait d'une seconde pour appuyer sur la touche espace du clavier s'il pensait avoir vu la cible dans la séquence (Tâche Détection). Dans le cas de la catégorisation Visage, il devait alors donner une réponse verbale indiquant le genre du visage de la célébrité qu'il avait vu dans la séquence (Tâche Genre) ainsi que son identité (Tâche Identification). Si le sujet n'arrivait pas à retrouver rapidement le nom de cette célébrité, mais qu'il était capable de donner des informations précises sur cette personnalité (nom de film, nom du partenaire, titre de chanson...), la réponse était quand même évaluée correcte, même si ces essais sont restés peu fréquents en moyenne. De la même manière dans le cas de la catégorisation Animal, après avoir donné sa réponse positive de détection, le sujet devait indiquer verbalement le nom de l'animal perçu (Tâche Identification). Encore une fois, nous avons été indulgents dans l'évaluation des réponses verbales car les réponses étaient considérées correctes si l'espèce donnée était proche de l'espèce réelle (« chien » pour « renard » par exemple). Enfin après avoir donné sa réponse verbale, le sujet devait appuyer une nouvelle fois sur la touche espace du clavier pour faire apparaître l'image cible à l'écran aussi longtemps que souhaité afin de l'identifier sans contrainte de temps (Tâche Identification Libre).

La cible pouvait apparaître n'importe quand entre la 4^{ème} et la 16^{ème} position de la séquence. Le SOA inter-image variait entre 20 et 120 ms (4 conditions : 20, 50, 80 et 120 ms).

Chaque sujet réalisait deux sessions de 8 séries de 52 essais, sur deux jours consécutifs. Dans une session, il effectuait la catégorisation Animal et dans l'autre session la catégorisation Visage,

précédée à chaque fois d'une série d'entraînement. Dans chaque série, 48 essais comprenaient une cible, et 4 essais dits « catch trials » comprenaient 20 images de masques sans aucune cible pour vérifier la fiabilité des réponses du sujet dans les conditions où le SOA était le plus court. Les sujets n'étaient pas informés de cette condition. Globalement donc, chacune des conditions de SOA réunissait 768 essais pour l'ensemble des sujets dans une tâche donnée. Encore une fois, tous les essais avec une erreur de latence d'affichage ont été supprimés de l'analyse (environ 8%).

Stimuli

En tout, 416 images cible d'animaux ont été sélectionnées dans les librairies Corel (Figure 4.12), dont 32 ont été utilisées pour la deuxième présentation à l'écran des « catch trials » (les sujets n'étant pas informés de cette condition, il fallait tout de même présenter des stimuli pour la Tâche Identification Libre, même si aucune cible n'était présentée dans le flux RSVP). La moitié des images d'animaux comprenaient des têtes d'animaux en gros plan, l'autre moitié comprenait des animaux dans les différents plans de l'image. Les images ont été particulièrement choisies afin de représenter un nombre équilibré d'espèces connues pour appartenir au niveau de base dans la hiérarchie des catégories (Rosch et al., 1976). Ainsi 16 sous-catégories ont été isolées : canards, cerfs, chats, chevaux, chiens, éléphants, lapins, lions, loups, ours, renards, rhinocéros, serpents, singes, tigres, vaches.

De la même manière, 416 images de visages de célébrités ont été sélectionnées sur Internet et dans des magazines. En tout, ces images comprenaient 4 photographies de 96 célébrités différentes, les 32 images restantes utilisées pour les « catch trial » représentant les visages de 32 autres célébrités (chanteurs, acteurs, politiciens, sportifs etc...). Les images utilisées dans l'expérience présentée précédemment sont incluses dans cet ensemble de photographies.

Enfin, les masques étaient composés de scènes naturelles, ne comprenant ni animaux ni humains, choisis aléatoirement parmi un set de plus de 11000 images.

Analyse des réponses verbales

Les réponses verbales dans la tâche **Genre** sont simples à évaluer, puisqu'il s'agit en quelque sorte d'un choix forcé : le sujet répond soit 'Femme' soit 'Homme'.

Si le sujet a répondu qu'il a détecté une cible dans la tâche **Détection**, sa réponse verbale dans la tâche **Identification** peut entrer dans différentes catégories :

- (1) Identification correcte,
- (2) Identification proche, évaluée correcte : le sujet confond une oie avec un canard, ou indique qu'il a joué dans Terminator pour parler de Schwarzenegger,
- (3) Identification incorrecte : le sujet confond un éléphant avec une souris ou Schwarzenegger avec Jenifer Aniston
- (4) Aucune identification : le sujet ne peut rien dire sur la cible.

Nous avons considéré que les réponses étaient correctes dans les cas (1) et (2), et incorrectes dans les cas (3) et (4).

Les réponses verbales données dans la tâche **Identification Libre** respectent le même profil de réponses, avec une condition supplémentaire dans laquelle le sujet affirmait parfois qu'il

connaissait la célébrité présentée à l'écran mais ne pouvait pas donner suffisamment d'informations pour valider correctement la réponse. **Il est important de noter que les analyses présentées ci-dessous portent seulement sur les essais pour lesquels la tâche Identification Libre est correcte.**

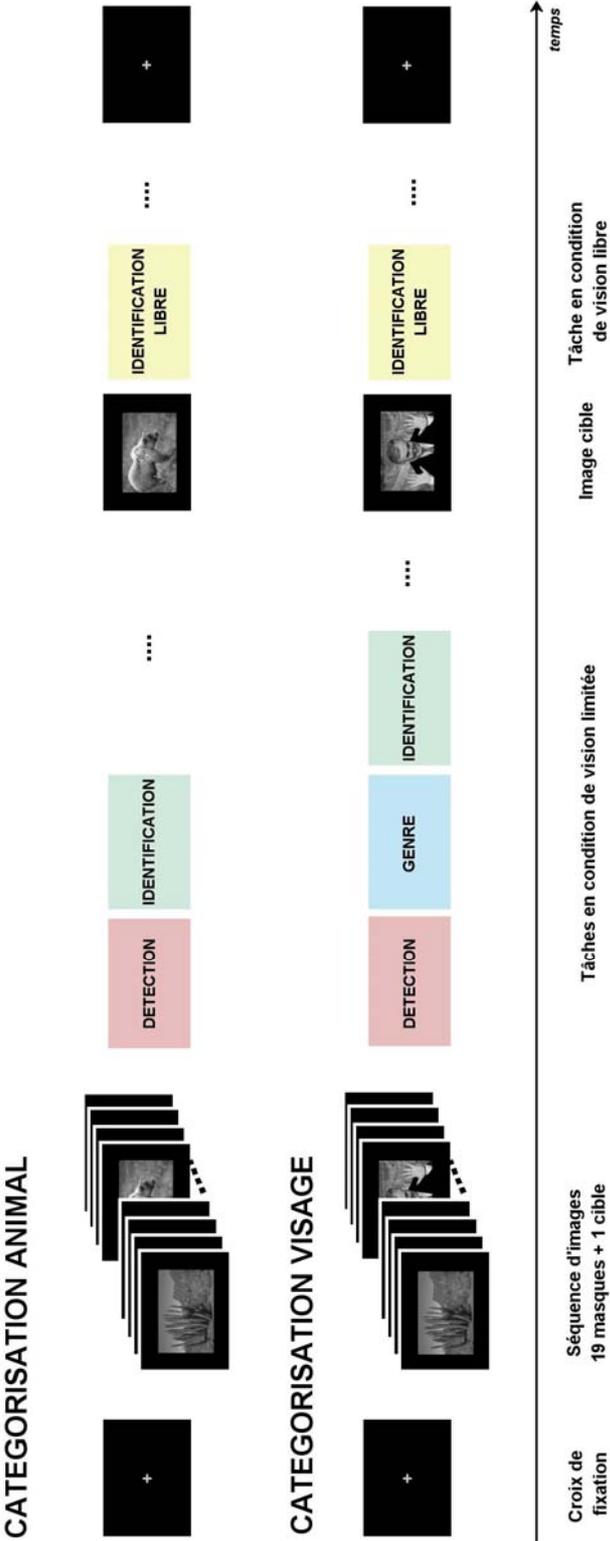


Figure 4.11: Protocole expérimental dans l'expérience RSVP faisant intervenir des tâches de détection et d'identification.



L. de Funès



T. Cruise



A. Boccelli



J-P. Foucault



J. Binoche



P. Collins



S. Connery



L. Fabius



E. Béart



Coluche



E. Murphy



H. Ford



S. Stallone



S. Marceau



M. Farmer



F. Mitterrand



K. Costner



S. Gainsbourg



I. Adjani



P. Anderson



Lorie



C. François



C. Aznavour



Madonna



F. Gall



J-C. Vandamme



G. Clooney



B. Bardot



B. Clinton



C. de Gaulle



Mère Thérèse



J-P. Belmondo



Dalida



C. Eastwood



B. Tapie



P. Laffont



F. Ardant



A. Delon



M. Gibson



B. Spears



T. Blair



P. Douste-Blazy



Y. Montand



C. Dion



J. Chirac



C. Chazal



J. Bové



N. Sarkozy

Figure 4.12 - Planche A



Figure 4.12 - Planche B



Figure 4.12 - Planche C

Figure 4.12 : Exemples de stimuli utilisés dans l'expérience RSVP de dénomination. (Planche A et B) En tout 4 photographies de 96 célébrités ont été sélectionnées, les planches n'en montrent qu'une seule par célébrité. (Planche C) La planche montre 2 images parmi les 16 qui ont été sélectionnées pour chacune des 16 sous-catégories d'animaux. (Planche D) Exemples de masques utilisés dans la tâche, composés de scènes naturelles ne contenant ni humains ni animaux.



Figure 4.12 - Planche D

4.4.2 Résultats

Performances globales

Le pourcentage d'identification correcte mesuré dans la tâche Identification Libre (les stimuli cibles étaient présentés aussi longtemps que le souhaitaient les sujets) montre que les images ont été correctement choisies : les sujets ont pu identifier la totalité des images d'animaux et près de 84% des images de célébrités (Figure 4.13). Dans cette dernière catégorie, à peine 9% des célébrités semblaient totalement inconnues des sujets (en moyenne), ce qui veut dire que 91% des visages leur paraissaient familiers.

Lorsque nous regardons les performances mesurées dans les tâches Détection et Identification, la précision est bien sûr moins bonne à cause des effets de masquage générés par la présentation des stimuli dans la séquence RSVP. Les sujets ont réalisé en moyenne 75,5% de réponses correctes dans la tâche Détection et seulement 46% dans la tâche Identification (Figure 4.13). Cependant le pourcentage correct obtenu dans la tâche de détection doit être considéré avec précaution car les essais « catch trials » insérés dans les séries indiquent que les sujets étaient particulièrement libéraux dans leurs réponses (les "catch trials" sont les essais pour lesquels les séquences RSVP contenaient 20 masques et aucune cible). Au total, les sujets ont répondu sur 3,5% des essais « catch trials » dans la tâche Visages, et sur 23% de ces essais dans la tâche Animal.

D'autre part il est intéressant de noter que si la performance de détection est moins élevée dans la tâche Animal que dans la tâche Visage (72,1% vs. 78,9%), ce qui est cohérent avec les résultats obtenus dans les expériences précédentes, le pourcentage d'identification est au contraire bien meilleur dans la tâche Animal que dans la tâche Visage (52,8% vs. 39,3%).

Aucun effet de l'ordre des sessions n'a été trouvé sur la performance de détection, sachant que les sujets effectuaient les deux tâches de catégorisation sur deux jours consécutifs. Encore une fois, ces résultats sont cohérents avec ceux que nous avons obtenus dans une expérience pilote qui montrait qu'un entraînement intensif dans la tâche n'améliore pas les performances.

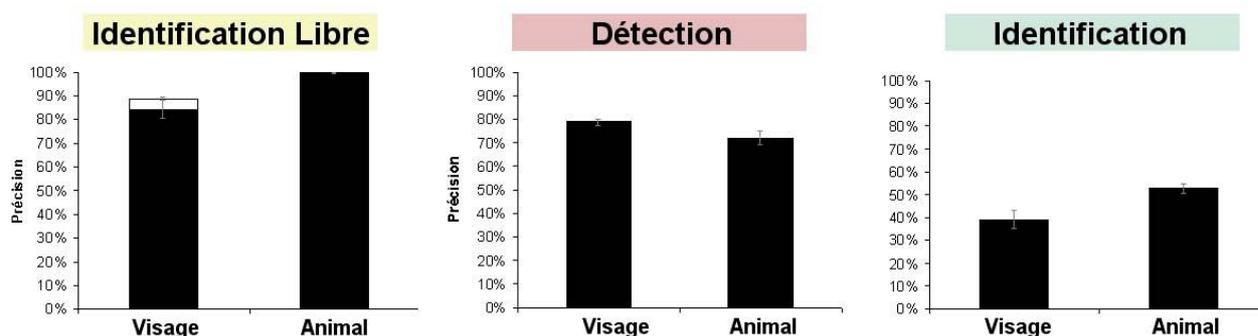


Figure 4.13 : Performances globales dans les différentes tâches pour toutes les conditions de SOA groupées (+/- sem). Les images de visages de célébrités et d'animaux ont été globalement bien choisies, puisque les sujets réalisaient des performances élevées dans la tâche Identification Libre pour laquelle l'image était affichée à l'écran aussi longtemps que le souhaitaient les sujets. La zone blanche au dessus de l'histogramme Visage indique le pourcentage des essais pour lesquels les sujets affirmaient connaître cette célébrité mais ne pouvait pas donner de nom. En revanche, les performances étaient bien moins bonnes dans les tâches Détection et Identification alors que les stimuli étaient présentées dans la séquence RSVP. Seules les essais où l'identification libre était correcte ont été gardés pour l'analyse des performances de Détection et d'Identification.

Performances en fonction du SOA

On retrouve un écart important de performance entre la tâche Détection et la tâche Identification quel que soit le SOA, pour les images d'animaux comme pour les visages. Dans de nombreux essais les sujets ont détecté une cible dans la séquence, mais ils ne peuvent pas indiquer précisément ce qu'ils ont détecté (Figure 4.14). Il est possible d'évaluer grossièrement le « coût » des traitements supplémentaires pour identifier les cibles en calculant le délai additionnel nécessaire pour atteindre le même niveau de précision dans les deux tâches. Au moins 30 ms supplémentaires dans le SOA seraient nécessaire pour identifier correctement un animal dans la séquence par rapport au moment où il a été détecté dans la séquence, et ce délai serait même supérieur pour identifier correctement les célébrités.

En revanche, il est étonnant d'observer que les sujets accèdent très rapidement aux informations sur le genre des visages, car les courbes de précision en fonction du SOA sont très proches entre les tâches Détection et Genre (Figure 4.14). La performance dans la tâche Genre diffère significativement de la tâche Identification ($p < 0.002$), mais pas de la tâche Détection. Ainsi, lorsque les sujets disposent d'un délai d'intégration tout juste suffisant pour détecter un visage, ils ne peuvent donner son identité mais savent si c'est un homme ou une femme.

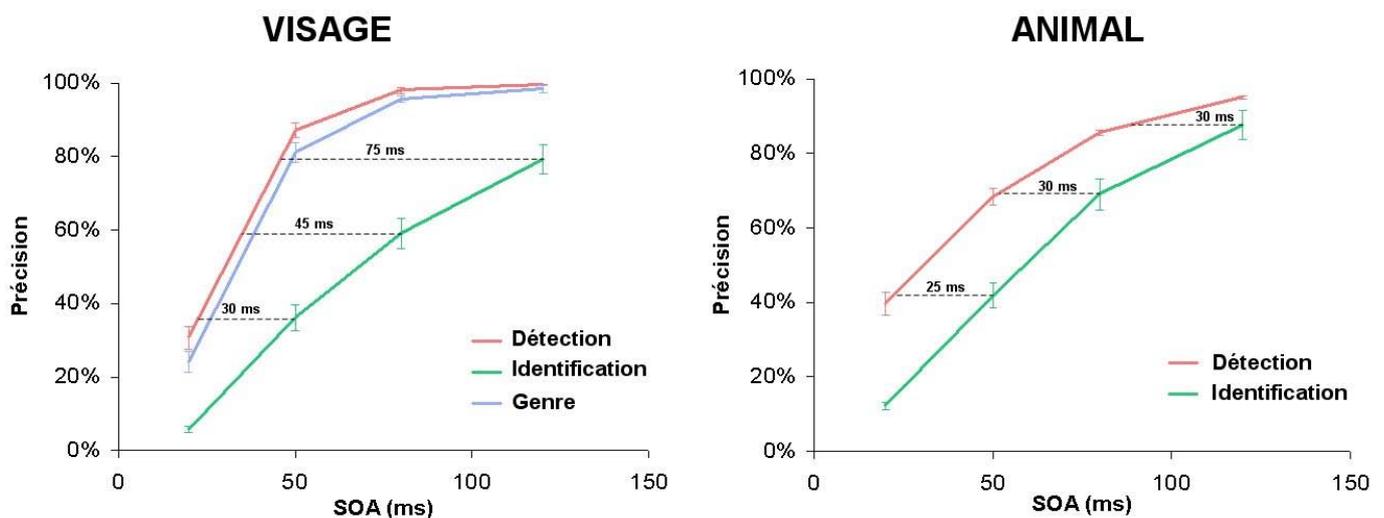


Figure 4.14 : Performance en fonction du SOA pour les différentes tâches de Détection, Identification et Genre. Sur l'ensemble des SOA, la précision est bien moins élevée dans la tâche Identification par rapport à la tâche Détection. L'écart de performance peut être évalué en mesurant le délai supplémentaire nécessaire pour atteindre le même niveau de précision dans les deux tâches (latences indiquées sur les graphiques). D'autre part, il est intéressant de remarquer que la courbe d'identification du genre des visages est très proche de celle consistant simplement à détecter les visages dans la séquence RSVP. Ces courbes de performance ont été calculées sur les essais cibles seulement et ne prennent pas en compte les essais « catch trial », qui sont peu nombreux en proportion (1/13).

4.4.3 Discussion

4.4.3.1 De la détection à l'identification

Les résultats de cette expérience semblent montrer que les processus visuels demandent moins de traitement pour catégoriser une cible que pour l'identifier. Cette observation se retrouve aussi bien pour des stimuli d'animaux que de visages. Autrement dit, même si le sujet a pu détecter une cible en se basant sur son appartenance à une catégorie, il ne peut pas pour autant donner de précisions sur celle-ci. Ces résultats rejoignent directement ceux de Grill-Spector et Kanwisher, qui ont montré dans une tâche de masquage que les sujets sont aussi efficaces pour détecter ou catégoriser un objet, mais qu'ils ont besoin de beaucoup plus de temps pour l'identifier (Grill-Spector & Kanwisher, 2005). Le décalage entre les deux courbes de précision en fonction du SOA enregistrées dans leurs tâches de Catégorisation et d'Identification est d'ailleurs comparable au nôtre, avec un écart d'un minimum de 20-30 ms de SOA à performance égale. Les auteurs infèrent de leurs résultats deux conclusions principales, qui se recadrent globalement bien avec les hypothèses générales proposées par notre équipe ces dernières années :

(1) *Les traitements visuels consistant à détecter un objet ou à le catégoriser reposent sur les mêmes processus.* Que les sujets aient à détecter une cible ou à la catégoriser, les deux tâches impliquent l'analyse de la scène afin de mettre en correspondance les informations extraites avec les représentations mnésiques codant pour une gamme d'indices spécifiques. Plus les caractéristiques de la cible sont prédictibles et dissociables des stimuli distracteurs, plus les seuils de discrimination sont atteints rapidement. Ainsi et contrairement à l'hypothèse de Grill-Spector et Kanwisher, nous pensons que des différences de performance peuvent apparaître entre les tâches de détection et de catégorisation selon le niveau de finesse requis dans l'analyse. Des différences de performance peuvent aussi apparaître entre les différents niveaux d'une catégorisation : les expériences menées par Marc Macé et Arnaud Delorme dans l'équipe ont montré que les performances sont significativement augmentées lorsque l'espace des cibles est réduit. Dans une série d'expérience manipulant les niveaux de catégorie, Macé et collaborateurs ont mis en évidence une augmentation des performances lorsque le sujet devait catégoriser des images de chiens parmi des distracteurs par rapport à la condition dans laquelle le sujet devait catégoriser des images d'animaux (catégorie beaucoup plus variable) parmi ces mêmes distracteurs (Macé, Joubert & Fabre Thorpe, en préparation). La version extrême de cette tâche consiste à réaliser une tâche simple de reconnaissance dans laquelle un seul stimulus préalablement appris doit être détecté (reconnu) parmi des distracteurs variés. Dans ce cas, la précision augmente parallèlement à la diminution de TR (Delorme, Rousselet, Mace & Fabre-Thorpe, 2004). Cependant, il en reste néanmoins que les processus impliqués dans la détection ou la catégorisation d'objet semblent reposer sur les mêmes chaînes de traitement, même si la manipulation de certains paramètres comme l'augmentation de la prédictibilité de la cible peut

rapidement conduire à des performances sensiblement meilleures lorsqu'il s'agit de détecter ou de reconnaître plutôt que de catégoriser un objet³.

(2) *L'identification du stimulus requiert plus de temps de traitement que ceux impliqués dans la catégorisation*, résultat que nous retrouvons largement dans notre expérience RSVP, aussi bien pour les stimuli d'animaux que de visages. Dans le cas de la catégorisation, le système visuel doit faire correspondre un objet à différents membres d'une classe donnée bien que ces objets aient des formes et des propriétés physiques différentes. Dans le cas d'une identification, le système visuel doit non seulement être capable de réaliser une catégorisation, mais il doit aussi faire la distinction entre les différents membres à l'intérieur même de cette catégorie (Palmeri & Gauthier, 2004). Nos données ne permettent cependant pas de savoir si les délais supplémentaires reflètent l'engagement de mécanismes différents ou si les mêmes mécanismes sont poursuivis plus longtemps. Cependant encore une fois, un certain nombre d'études laissent penser que les traitements sont similaires mais que l'information demandée -étant de plus en plus précise et spécifique- requière des délais plus importants pour générer une activité sélective lorsque le niveau de détail à atteindre est supérieur (Nosofsky, 1986; Palmeri & Gauthier, 2004; Riesenhuber & Poggio, 2000). Ces résultats vont tout à fait dans le sens des données électrophysiologiques enregistrées dans des protocoles de masquage ou de RSVP, montrant que la sélectivité des neurones à un stimulus dans une même structure augmente avec l'augmentation de la durée d'affichage de la scène. Si la partie initiale de la décharge neuronale apporte suffisamment d'informations pour détecter puis pour catégoriser des stimuli, il est possible d'imaginer que d'avantage d'activation soit nécessaire lorsqu'il s'agit de les identifier (Calder & Young, 2005; Liu, Harris & Kanwisher, 2002; Sergent & Signoret, 1992; Young & Yamane, 1993).

Finalement, nous nous retrouvons dans un modèle où le niveau de traitement requis dans différentes tâches (Reconnaissance – Détection – Catégorisation – Identification) semble former un continuum⁴ dans le niveau d'activation à atteindre pour répondre à une consigne demandant de

³ Afin de ne pas alourdir la discussion, je ne souhaitais pas rentrer dans le détail expérimental des travaux de Grill-Spector et Kanwisher. Cependant, il est possible de trouver une explication pour justifier un niveau de performance égal dans les tâches Détection et Catégorisation alors que nous proposons justement que les processus impliqués dans la catégorisation nécessitent d'être poursuivis plus longtemps. Il faut pour cela bien définir les termes faisant référence aux tâches effectuées. Dans la tâche de Détection, les sujets devaient répondre le plus rapidement possible si le stimulus masqué était une photographie d'objet. Dans ce cas, les distracteurs étaient composés de bruit ou de versions « scrambled » des objets. Dans la tâche de Catégorisation, les sujets devaient répondre le plus rapidement possible s'il pensaient avoir vu des images d'une catégorie donnée alors que les distracteurs étaient des photographies appartenant à une autre catégorie. Il est donc étonnant de voir que les performances restent à niveau égal malgré le fait que les cibles soient plus difficiles à discriminer des distracteurs dans la tâche Catégorisation par rapport à la tâche Détection. Or, il est important de remarquer que les indices diagnostiques des cibles dans la tâche de Catégorisation sont beaucoup plus spécifiques et réduits puisqu'ils appartiennent à une catégorie donnée, par rapport à la tâche de Détection dans laquelle le sujet peut être confronté à n'importe quel type d'objet. Il est possible d'imaginer que les seuils de discrimination à atteindre le sont d'autant plus facilement que les indices sur les cibles sont connus et spécifiques, comme dans le cas d'une catégorie d'objets limitée dans sa variabilité.

⁴ Comme son nom l'indique, le terme continuum se réfère d'avantage à une vision continue de la chaîne des traitements, ce qui semble tout à fait plausible biologiquement comme nous l'avons vu dans le

plus en plus de précision, qui se traduit par une augmentation du délai de traitement pour pouvoir atteindre ce niveau d'activation. Plus spécifiquement, cette idée fait référence à la hiérarchie des catégories et au recours croissant à des indices visuels précis lorsqu'il s'agit d'accéder à une description de plus en plus figurative des stimuli (Murphy & Smith, 1982; Rosch et al., 1976). Un objet est censé être reconnu mieux et plus rapidement au *niveau de base* (par exemple un chien) qu'au *niveau superordonné* (dans ce cas, animal) ou *subordonné* (caniche ou lévrier). Jolicoeur et Murphy ont remis en cause cette classification des catégories en introduisant le concept de *niveau d'entrée* qui permet au niveau subordonné d'être le niveau d'accès privilégié dans certains cas, en particulier pour expliquer les performances sur les objets atypiques (par exemple, un pingouin est catégorisé d'abord comme un pingouin avant de l'être comme un oiseau). Récemment, cette représentation des niveaux de catégories a de nouveau été remise en cause en faveur d'un système plus flexible faisant intervenir l'expertise des sujets pour une catégorie donnée et l'influence des contraintes de la tâche sur le niveau de traitement (Gauthier, Skudlarski, Gore & Anderson, 2000; Gauthier & Tarr, 1997; Johnson & Mervis, 1997; Schyns, 1998; Tanaka & Curran, 2001; Tarr & Cheng, 2003). Par exemple, si la majorité des sujets peu entraînés ont des performances meilleures pour catégoriser les objets au niveau de base par rapport aux autres niveaux de catégories, les sujets experts pour une catégorie donnée arrivent à effectuer une tâche de catégorisation aussi rapidement au niveau subordonné qu'au niveau de base. Pour donner une illustration, les sujets dits naïfs seront plus rapides pour accéder à la représentation des chiens qu'à celle des lévriers. En revanche, un dresseur ou un passionné de chiens réalisera la tâche aussi rapidement qu'il s'agisse de catégoriser des lévriers en tant que lévriers ou en tant que chiens. D'ailleurs lorsqu'on demande à ces sujets de dénommer les stimuli, les experts décriront presque systématiquement l'espèce du chien alors que les sujets naïfs auront d'avantage tendance à faire référence au nom prototypique « chien ». D'autre part, il semble que l'accès à ces différents niveaux de catégorie soit modulé par le mode de réponse dans la tâche. La majorité des études ayant décrit une supériorité du niveau de base par rapport aux autres niveaux de catégorie se sont appuyées sur des réponses verbales, impliquant obligatoirement l'intervention du système lexical. En revanche, et c'est ce qu'a proposé Marc Macé dans ses travaux de thèse (Macé, 2006; Macé, Bacon-Mace, Nespoulous & Fabre-Thorpe, en préparation), lorsque la réponse repose uniquement sur les traitements visuels, on se retrouve d'avantage dans un système où l'extraction rapide d'une représentation grossière de la scène permet de détecter dans l'image la présence d'un animal puis d'un chien, et que d'avantage d'informations et de traitements sont nécessaires pour déterminer la présence d'une espèce particulière de chien.

Chapitre 2. En revanche, il est important de noter que les différentes réponses fournies (Reconnaissance – Détection – Catégorisation – Identification) discrétisent l'information à l'issue des traitements, puisqu'elles représentent le résultat d'une consigne donnée. Rien ne s'oppose à l'idée qu'au moment de sa réponse, le sujet a pris sa décision en extrapolant, ou au contraire en ne considérant qu'une partie de l'information disponible. Les données enregistrées sur ces réponses discrétisées représentent l'ensemble de ces essais, et ne permet pas d'observer l'aspect potentiellement continu des traitements engagés dans la reconnaissance (au sens large) des objets.

Cette classification des catégories selon le niveau de représentation privilégié par les traitements visuels est fondamentale dans nos expériences, puisque l'identification des animaux et des visages de célébrités requiert un accès à une description plus figurative des stimuli par rapport à leur catégorisation, expliquant ainsi l'écart de performance entre les deux tâches. Un autre résultat majeur de cette série d'expériences semble montrer que les sujets sont aussi performants à identifier le genre des stimuli Visages qu'à les catégoriser (Figure 4.14). Où replacer ce niveau de catégorie Genre dans la classification des catégories et pourquoi les visages représentent-ils une catégorie de stimuli si particulière ?

4.4.3.2 « J'ai vu une femme aux cheveux longs ! » : propriétés de la catégorie Visage

Nous avons vu dans les résultats que la courbe de précision en fonction du SOA obtenue dans la tâche Genre se superposait presque parfaitement à la courbe de précision enregistrée dans la tâche Catégorisation (Figure 1.14). Autrement dit, lorsque les sujets ont repéré un visage dans la séquence, ils sont capables de dire si c'est un homme ou une femme. Ces résultats rejoignent ceux de Leila Reddy et collaborateurs qui ont montré que le genre des visages pouvait être traité sans attention, au même titre que la catégorisation simple des stimuli de visages (Reddy, Wilken & Koch, 2004). Il semblerait donc qu'il ne soit pas nécessaire d'accumuler d'avantage d'informations pour identifier le genre d'un visage à partir du moment où ce stimulus a été détecté dans le flux. Ces données suggèrent que le niveau d'entrée pour la catégorie des visages ne se base pas seulement sur l'intégration des indices permettant la détection d'un visage dans sa configuration globale mais il engage aussi (en parallèle ?) l'intégration des indices déterminant le genre. D'ailleurs, les premiers termes que l'on utilise pour décrire une scène ne sont pas « J'ai vu un visage » ou « J'ai vu un humain », mais plutôt « J'ai vu une femme » ou « J'ai vu un homme »... On imagine facilement que cette hypothèse est d'autant plus valide que les stimuli sont physiquement proches d'une scène naturelle, comprenant les cheveux, les vêtements et tous ces indices contextuels permettant d'orienter l'identification du genre.

D'autre part, il est important de noter que lorsqu'on demande aux sujets de donner le nom des célébrités détectées dans le flux RSVP (tâche Identification), le niveau figuratif, dans le sens de précision, est bien plus élevée que lorsqu'il s'agit d'identifier le nom de l'espèce des animaux. Une tâche équivalente avec des stimuli Animaux serait de demander aux sujets de donner l'identité des animaux, qui dans ce cas seraient connus pour être « Filou » le petit chat de la voisine ou « Lassie », le célèbre colley de la série télévisée ! Identifier les célébrités par leur nom demande aux sujets d'accéder au niveau de catégorie de l'exemplaire (exemplar en anglais), niveau où l'item a la propriété d'être unique. Il n'est donc pas étonnant d'observer que les performances dans la tâche Identification sont sensiblement moins bonnes avec les stimuli Visages qu'avec les stimuli Animaux (39,3% vs. 52,8%) puisque le niveau de catégorie requis n'est pas le même (Figure 4.14 et 4.15).

Finalement ces résultats montrent que la catégorie Visage est une catégorie particulièrement singulière et qu'il est important d'être prudent lorsqu'il s'agit de mettre en parallèle les performances obtenues en utilisant ces stimuli avec ceux obtenues pour d'autres catégories d'objets (Figure 4.15). Les études essayant de définir les corrélats à la fois comportementaux et

électrophysiologiques entre cette catégorie et d'autres catégories d'objets pour lesquelles nous avons une grande expertise doivent être poursuivies afin de déterminer si l'efficacité du traitement des visages est le résultat d'une optimisation du système au cours du temps pour des stimuli écologiquement importants (Gauthier, Skudlarski, Gore & Anderson, 2000; Tanaka, 2001; Tanaka & Curran, 2001; Tarr & Cheng, 2003). Une orientation possible de ces études comparatives pourrait impliquer la participation d'autres groupes de sujets, comme des enfants à différents âges ou des primates non-humains.

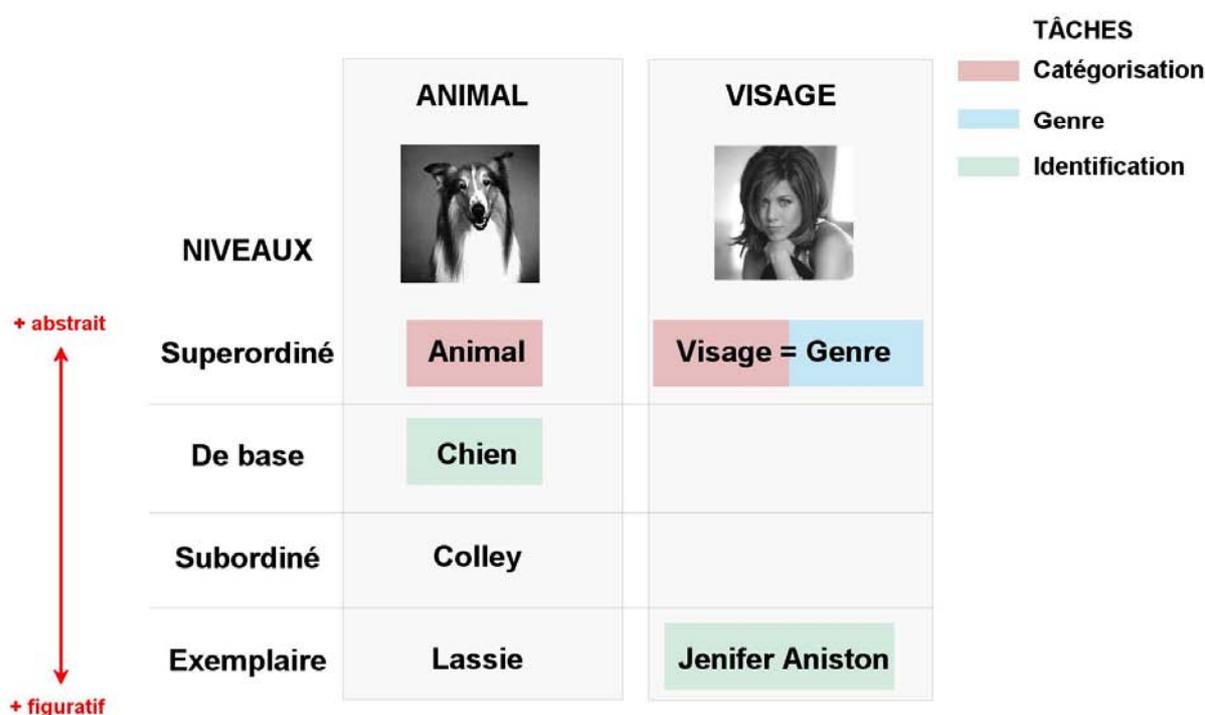


Figure 4.15 : Classification des catégories pour les stimuli Animal et Visage. Le niveau d'abstraction demandé dans les différentes tâches de notre expérience RSVP n'est pas identique pour ces deux catégories de stimuli. Cette mise en correspondance des classifications met l'accent sur la singularité de la catégorie Visages, qui semble difficilement comparable aux autres catégories d'objets.

4.4.3.3 Nature des indices diagnostiques

Nous terminerons cette partie par une question fondamentale et sujette à de nombreux débats dans la communauté des chercheurs s'intéressant à la reconnaissance d'objet : quels sont les indices diagnostiques qui permettent au système visuel de reconnaître efficacement une cible potentielle, même dans des conditions bruitées de perception ?

Dans leurs expériences RSVP, Evans et Treisman ont analysé la nature des réponses verbales fournies par les sujets. Si à peine 45 % des images cibles étaient correctement identifiées (prenons l'exemple d'une « vache »), les autres réponses des sujets donnaient cependant de bons indices pour affirmer que la perception des images étaient relativement avancée : en tenant compte des

réponses verbales données à un niveau de catégorisation supérieur (« mammifère »), des confusions sur l'espèce de l'animal (« cerf ») et des descriptions vagues mais justes des images (« animal gros et noir»), presque 95 % des réponses d'identification étaient correctes.

Une analyse plus poussée des réponses fournissant une description vague de l'image montre que sur 124 réponses :

- 58 donnent des indices sur la texture (fourrure, tâches, plumes ...),
- 55 donnent des indices sur une partie du corps (pattes, antennes, becs, ailes, yeux ...),
- 31 mentionnent la taille (petit ou gros),
- et seulement 10 font un report de la forme (pointu, rond, allongé, fin ...).

Dans notre expérience, je n'ai pas pris en compte les réponses des sujets décrivant globalement l'image qu'ils pensaient être la cible (exemple de « gros et noir»). Si ces réponses représentaient 16% des essais dans l'expérience de Treisman, elles étaient plutôt rares dans nos tâches, et c'est d'ailleurs pour cette raison que je ne les ai pas quantifié. Globalement, les sujets répondaient plutôt sur un mode du « tout ou rien » : soit ils étaient capables de donner un nom pour la cible, soit ils affirmaient clairement qu'ils « avaient vu quelque chose » mais ne savaient pas quoi. On peut supposer que les différences entre nos expériences tiennent au fait que nos séries comportaient toutes les conditions de SOA mélangées alors que les séquences d'images présentées dans les expériences d'Evans et Treisman étaient toutes présentées à 75 ms de SOA, ce qui peut conduire à des différences en terme de stratégie de réponse.

Toujours est-il que les auteurs s'appuient sur cette observation pour décrire un modèle dans lequel le système visuel s'appuie d'avantage sur des indices locaux de complexité intermédiaire, comme un bec ou une patte, que sur la configuration globale de la cible potentielle. Sans nous opposer complètement à cette idée, cette observation est un argument un peu fragile dans cette expérience puisque les sujets ont fourni d'avantage de réponses en décrivant la catégorie super-ordonnée (par exemple mammifère, 27% des réponses) qu'en décrivant des indices précis (moins de 16% des réponses). D'autre part, il faut aussi préciser que l'accès lexical à ces informations pénalisent la description d'une configuration globale, car s'il existe des mots précis pour décrire un bec, une patte ou de la fourrure, le vocabulaire est plutôt limité pour dire : « j'ai vu une forme, comme ça, qui s'étendait comme ça, avec un renflement de l'autre côté, et une pointe vers le haut, et j'avais l'impression que sur la partie inférieure droite, il y avait une sorte de bosse... ». Bref, les termes ne sont pas simples à trouver !

Il est important de rappeler que la taille des champs récepteurs des neurones (CR) augmente tout au long de la voie ventrale. Dans V1, les CR sont relativement petits (1-2° d'angle visuel), permettant d'établir une représentation fine de l'espace alors que dans IT, les CR couvrent l'ensemble du champ visuel, permettant ainsi d'appréhender de grands objets. Il a ainsi été montré que les neurones du cortex inféro-temporal peuvent coder pour des objets particulièrement complexes comme des visages, des animaux, des maisons ou des chaises (Gross, Bender & Rocha-Miranda, 1969; Logothetis & Sheinberg, 1996; Perrett, Rolls & Caan, 1982; Tanaka, 1996) (Sheinberg & Logothetis, 2001). Certaines études ont même rapporté que des neurones de IT sont sélectifs à des arrangements spatiaux d'éléments formant des visages ou divers types

d'objets (Logothetis & Sheinberg, 1996; Tanaka, 1996; Tanaka, 2003). Il est d'ailleurs intéressant de noter que cette capacité évolue au cours du traitement en lui-même, car si certains groupes de neurones sont plus sélectifs à la configuration normale d'un objet qu'à une configuration réarrangée, l'utilisation d'un protocole de masquage montre que cette différence est atténuée pour les SOA les plus courts (Rolls et al., 1994). On ne peut cependant pas oublier qu'une majorité de neurones dans ces aires hautement impliquées dans la reconnaissance codent pour des propriétés particulières des objets, et non pas spécifiquement pour l'objet dans son ensemble. Alors comment savoir sur quels mécanismes se base la reconnaissance des stimuli : en codant les composants de complexité intermédiaire qui caractérisent de manière prototypique l'objet (Evans & Treisman, 2005; Ullman, Vidal-Naquet & Sali, 2002) ou bien en utilisant la configuration globale de cet objet ? L'ensemble des résultats comportementaux et électrophysiologiques de l'équipe a plutôt favorisé la deuxième solution en décrivant un modèle général dans lequel la transmission de l'information se fait de manière massivement parallèle (sur le plan spatial) permettant de dégager rapidement une représentation rudimentaire de la scène suffisante pour catégoriser les objets. D'ailleurs, une expérience ayant manipulé le contraste des scènes naturelles montre que cette représentation rudimentaire pourrait être construite par les informations fournies par le système magnocellulaire dans la voie ventrale (Macé, Thorpe & Fabre-Thorpe, 2005). Le système parvocellulaire dont les informations arrivent plus tardivement permettrait alors d'affiner la nature de cette représentation de la scène visuelle. Cependant comme nous l'avons vu dans le chapitre 3, les temps de réaction mesurés dans la tâche de discrimination saccadique permettent de douter de l'implication des aires visuelles avancées comme le cortex inféro-temporal pour décider laquelle de deux images contient une cible donnée (Kirchner & Thorpe, 2005). L'idée que ce genre de tâche pourrait s'effectuer sur la base des seules informations intégrées dans V4, dont les champs récepteurs sont relativement petits, remettrait considérablement en cause la nature des représentations utilisées pour exécuter des réponses rapides adaptées.

Plutôt que de trancher pour un type d'indices plutôt qu'un autre (configuration globale vs. composants), il faut peut-être rappeler que de nombreuses études tendent maintenant à démontrer l'étonnante flexibilité du système visuel pour adapter au mieux ses traitements en fonction des besoins de la tâche (Biederman, Mezzanotte & Rabinowitz, 1982; Fiser, Subramaniam & Biederman, 2001; Henderson & Hollingworth, 1999; Morrison & Schyns, 2001; Schyns, 1998; Torralba, 2003).

« It was shown that the categorization itself could induce the creation of new, flexible building blocks that changed the immediate appearance and the featural representation of visual inputs. This revealed a limitation of the fixed feature framework : the idea that categorization processes operates on a 'perceived input' underestimated the influence that the categorization task exerts on the construction of real-world face, objects and scene percepts. » (Schyns, 1997)

Le système pourrait ainsi orienter et ajuster dynamiquement les traitements à effectuer sur les éléments diagnostiques de la scène, afin d'être le plus efficace possible dans une tâche donnée. Par exemple, il semblerait que lors de l'analyse des indices bas-niveau d'un stimulus de visage, les filtres de fréquence spatiale soient particulièrement orientés vers les échelles relativement

basses lorsqu'il s'agit de catégoriser un visage et vers des échelles plus hautes lorsqu'il s'agit de l'identifier ou de reconnaître son expression (Schyns, Bonnar & Gosselin, 2002; Schyns & Oliva, 1999). La représentation des formes complexes pourrait donc se baser sur une structuration des formes visuelles mettant en jeu à la fois des représentations fines et grossières (Bullier, 2001). Il est donc important de considérer la notion de saillance perceptive dans la scène, et il est tout à fait possible d'imaginer que si la segmentation figure-fond se fait facilement parce que le contraste de l'objet dans sa globalité est élevé par rapport au fond de la scène, le système pourra rapidement dégager une représentation holistique de l'objet. Au contraire, si une composante locale de cet objet est particulièrement saillante dans l'image, le système pourra privilégier le traitement de cette zone si elle correspond à une représentation prototypique et apprise de l'objet. La nature des informations diagnostiques repose donc sur un ensemble de paramètres à la fois contextuels à la tâche et inhérents aux stimuli qu'il faut (faudra) prendre en compte et analyser finement pour élaborer une théorie biologiquement plausible des mécanismes de la catégorisation.

ARTICLE 4

What's seen first : The animal or the bird?

Macé MJ-M, Bacon-Macé N
Nespoulous J-L & Fabre-Thorpe M
En préparation

What's seen first: The animal or the bird?

**Marc J.-M. Macé CA, Nadège M. Bacon-Macé,
Jean-Luc Nespoulous & Michèle Fabre-Thorpe**

Centre de Recherche Cerveau et COgnition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil. 133, route de Narbonne. 31062 Toulouse, France.

^{CA}Corresponding author : marc.mace@cerco.ups-tlse.fr

ABSTRACT

Since the principle work of Rosch and colleagues in 1976 it is commonly accepted that in perceptual tasks, basic level categories (dog, chair...) are accessed before superordinate level categories (animal, furniture...) and a large number of studies involving various experimental paradigms have since confirmed this results. But the basic level advantage has been recently challenged by the speed at which objects can be categorized at the superordinate level in a go/no-go paradigm using natural images (Thorpe et al., 1996; VanRullen & Thorpe, 2001b). Using the same protocol in this study, we specifically compared the speed of processing when subjects are performing a go/no-go visual categorization task at the basic or at the superordinate level. We found that categorizing at the basic level is 40-65 ms slower than at the superordinate level. The discrepancies of our results with previous experiments are explained by important paradigms differences such as lower language processing and higher visual system pre-activation. The faster access to a more general/abstract category than basic level could reflect a 2-stage process where the visual system can first analyze the image at a coarse level, with enough information to decide whether the object belongs to the superordinate category and then, with additional detailed processing, gain access to basic-level related information.

INTRODUCTION

In 1973, Rosch and colleagues proposed that the different levels of categorization, organized as a taxonomic system, present a favored access stage whatever the perceptual system used (Rosch *et al.*, 1976). This so-called *basic level* is defined as the most abstract level where objects still share a common shape. It has been proposed that this level

corresponds to an optimum in term of cognitive efficiency of categorization. This advantage of the basic level (ie, dog or chair) upon the *superordinate* (animal or furniture) and *subordinate levels* (shepherd or rocking chair) was confirmed in object naming and category membership verification experiments (Rosch *et al.*, 1976). Developmental

studies also observed that basic level categories are the first learned by children (Mervis & A., 1982) and include the most spontaneously used terms by adults to name objects (Rosch *et al.*, 1976). From these observations, it was inferred that the basic level should correspond to the stored mnemonic representation that is activated first when an object is perceived. Superordinate levels were then considered as abstract generalization of the basic level and subordinate levels as perceptually more inclusive categories. This idea has later been refined by Jolicoeur (Jolicoeur *et al.*, 1984a) and Murphy (Murphy & Brownell, 1985) who introduced the concept of *entry level category* to explain the reaction times advantages found for members belonging to atypical subordinate categories (a penguin is categorized faster as a penguin than as a bird, contrary to more common birds for which basic level is accessed first). Besides, other authors observed that expertise was also a factor that could modulate the entry level towards subordinate level and abolish in some specific cases the basic level advantage (Tanaka & Taylor, 1991; Johnson & Mervis, 1997). However, these observations concerned only the basic and subordinate levels, and this is only recently that the prevalence of basic level over superordinate level has also been challenged. The first element came from Murphy & Wisniewski (Murphy & Wisniewski, 1989) who reported that the RT advantage of basic over superordinate levels was reduced -but did not disappear- by presenting objects in full scenes instead of the usually isolated stimuli on a neutral background. More recently, experiments on rapid visual categorization at the superordinate level definitely challenged the traditional view. The surprising speed at which subjects can detect animals or vehicles in natural scenes in a simple go/no go task, raised many questions about the basic level dominance, at least in the visual modality (Thorpe *et al.*, 1996; VanRullen & Thorpe, 2001a). Cerebral activity associated with this task performance revealed that a differential signal between targets and non-targets appears as soon as 150 ms over the occipital electrodes, already challenging most models of object recognition. VanRullen *et al.*, first pointed the fact that it would be very difficult to expect the basic-level categorization to be even faster than this. Contrary to the original experiments that used a reduced set of drawing objects in isolation, these last studies involve a large variety of targets presented in natural scenes photographs. Another difference which could be fundamental is that subjects gave manual

responses, as opposed to the frequently used verbal responses from the previous experiments. In fact, all studies that reported a basic level advantage relied, at different degrees, on some lexical processing. Neurons that respond both to the picture of a celebrity and to its written name have been recorded in human medial temporal lobe, but their latency of response is usually around 300 ms at the earliest (Quiñero *et al.*, 2005). At which point does the semantic processing of a stimulus require an access to the lexical structure of categories? Experiments with macaque monkeys have raised many questions since it has been shown that they are nearly as accurate as human subjects to categorize objects like animals or food items, despite being largely faster (Fabre-Thorpe *et al.*, 1998). The lack of language in this specie does not prevent a high performance in the categorization task and suggests that a superordinate experiment strictly based on visual -and not lexical- processing can potentially lead to different results than previously obtained. Very recently, Large *et al.* (Large *et al.*, 2004) followed this idea and found an advantage for superordinate level in visual categorization of isolated objects in a yes/no task. However, the effect was relatively weak (less than 15 ms) and could have been due to a speed/accuracy trade-off, as subjects were 2% better in the slower basic level categorization task. As in the previous studies, they used a small number of isolated objects drawings and did not benefit from the contextual effect reported by Murphy *et al.* in 1989.

In the two present experiments, we wanted to optimize the visual processing of the stimulus and minimize the influence of language by using (1) categorization of full-scene natural images in separate blocks, either at the superordinate (animal/non animal) or the basic level (bird/non bird or dog/non dog) and (2) go/no-go motor responses to obtain fast responses and reduce higher level cognitive interactions during the task. It has been shown, using peripheral stimuli (Thorpe, 2001 ; Boucart *et al.*, Submitted) or dual-task paradigms (Li *et al.*, 2002), that this task involve a large portion of non-conscious processing. Here, the brief presentation of the images (<30 ms) and the instruction to respond as fast as possible (and within 1s) constitute temporal constraints that were likely to increase the amount of implicit processing in the task.

In the category-level studies mentioned above, subjects were generally tested on several categories and the non-targets were chosen among the other categories, some of them belonging to the same

higher level categories. This precaution is necessary to ensure that subjects categorize the images only at the requested level. We used the same constraint here and half of the non-targets in the basic level tasks were images from the same superordinate category (non-bird or non-dog animals).

We also recorded the EEG signal during task performance to obtain precise temporal information on the categorization process. To our knowledge, only two studies that investigated categorization levels hierarchy have used EEG. Tanaka's main result (Tanaka *et al.*, 1999) was that superordinate level categorization led to a greater negativity than basic level between 306 and 356 ms, probably reflecting more important semantic/cognitive processing when access to the superordinate level was required. But unfortunately, no comparison could be made with behavior. The second study (Large *et al.*, 2004 is a complete behavioral/EEG study with 3 different levels of categorization. They found that, in accordance with the observed behavioral RT decrease, the first EEG components associated with targets at the superordinate level had earlier latencies when compared to the other levels of categorisation. In the present paper, we will mainly look at the differences between target and non-target ERPs to calculate the precise moment when images are sufficiently processed by the visual system to allow their categorization.

MATERIALS AND METHODS

Participants

18 subjects (9 women and 9 men) were tested in each experiment (mean age: 32 in experiment 1 and 30 in experiment 2). Five of the subjects were tested in both experiments. They all volunteered and gave their written informed consent. They all had normal or corrected-to-normal vision.

Procedure

Subjects were seated at 1 meter from a computer screen in a dimly lighted room. They started the experiment by placing a finger over a response pad for at least one second. A fixation cross appeared for 300-900ms, immediately followed by a photograph of a natural scene, flashed at the centre of the screen for 26 ms (view angle: 20° x 13.5°). With non-target photographs, subjects had to keep their finger on the button (no-go response) and with a target images, they had to release the button as quickly and accurately as possible (go-response). They had 1 second to trigger their go response after which delay

their response was considered as a no-go. The inter-stimulus interval time (ISI) was random between 1.6 and 2.2 seconds (mean: 1.9s).

Two different experiments have been performed in distinct sessions. Within each of the experiment, the aim was to compare the performance in superordinate and basic level tasks. In both experiments, the superordinate level task was an animal/non-animal categorization. At the basic level, subjects were required to perform a bird/non-bird (Bird experiment, Fig 1A) or a dog/non-dog (Dog experiment, Fig 1B) categorization. Each subject completed 16 blocks of 96 trials: 10 at the basic level and 6 at the superordinate level. Each categorization task was preceded by a training block of 48 trials.

Half of the subjects began with the superordinate task, the other half started with the basic level task. In the superordinate animal/non-animal task, half of the animal targets were either birds (Bird experiment) or dogs (Dog experiment). Conversely, in the basic level (bird or dog) categorization task, half of the non-target were non-bird or non-dog animals, while the other half were non-animal (neutral non-targets) pictures. As each image was only seen once by a given subject, the main concern was to avoid any bias induced by the selection of natural photographs for the tasks. To prevent this bias, we counterbalanced images across conditions and subjects: all bird (or dog) photographs were seen by some subjects as targets in the animal task and by the others as targets in the bird (or dog) task. Similarly, all images of non-bird (or non-dog) animals were seen by some subjects as targets in the superordinate task and by the other subjects as non-targets in the basic level task. Non-animal (neutral) images were also seen by different subjects in the 2 tasks. With such protocol, the effect observed on performance can be confidently attributed to task requirements and not to an image selection bias.

Stimuli

We used a set of 1536 images in each task, chosen as varied as possible from a Corel Database (Fig 1). Images of birds and dogs contained a large panel of the species (birds of prey, parrots, sparrows, wading birds, gulls... or shepherds, poodles, mastiffs, spaniels, dachshunds...). Other animal images were also varied and could contain mammals, insects, fish, reptiles etc. Subjects had no *a priori* knowledge about the size, position or number of target(s) in the pictures. "Neutral" non-targets did not contain animals but were as varied, including plants,

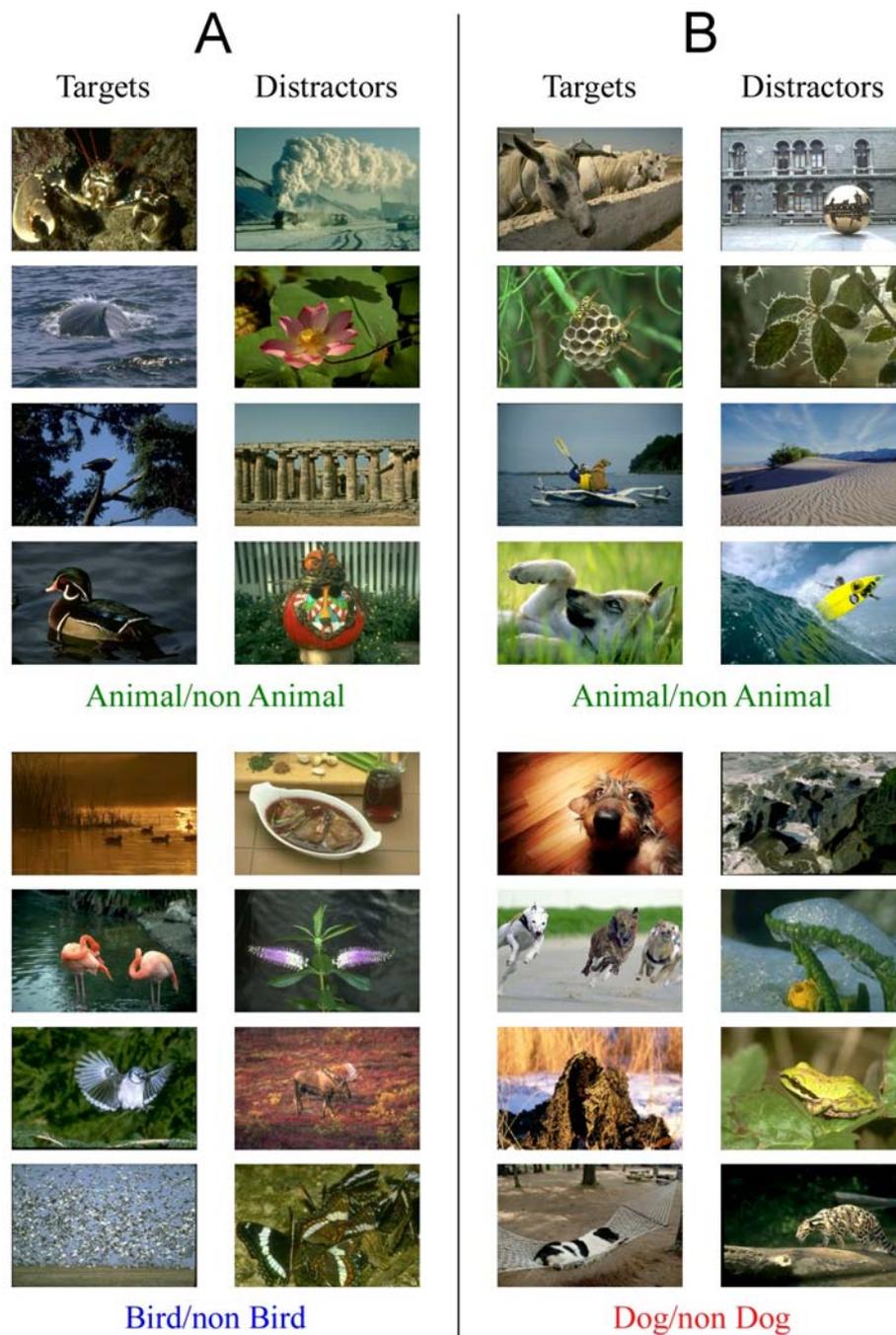


Figure 1: Examples of target and non-target images in the Bird experiment (A) and the Dog experiment (B) used either in the superordinate or in the basic level task. Note that in each experiment, half of the targets in the animal/non animal task are images of the corresponding basic level category (birds or dogs). In basic level tasks, half of the non-targets are images from the same superordinate category (non-bird or non-dog animals).

buildings, people, man-made objects, various landscapes...

EEG recording and analysis

Brain electrical activity was recorded from a 32 electrodes cap in accordance with the 10-20 system and completed by additional occipital electrodes connected to a Synamps amplifier system (Neuroscan Inc.). The ground electrode was placed along the midline, ahead of Fz. Impedances were kept below 5

k Ω . The signal was sampled at 1000 Hz and low-pass filtered at 100 Hz with a notch at 50Hz. Potentials were on-line referenced relative to electrode Cz and average re-referenced off-line. Baseline correction was performed using the 100 ms pre-stimulus interval. Two artifact rejections were applied over the [-100 ms; +400 ms] time period, first on frontal electrodes FP1 and FP2 with a criterion of [-50; +50 μ V] to reject trials with eye movements, and second on parietal electrodes Oz and Pz with a criterion of [-

30; +30 μ V] to remove trials with excessive alpha rhythms. Statistical tests were performed on raw data but a 30 Hz low-pass filter was applied for illustrations (Fig 4&5). Epochs were computed separately for correct target and non-target trials, for each categorization task. The differential activity was calculated by subtracting the averaged signal on correct non-target trials from the signal on correct

RESULTS

Performance in the different tasks was evaluated using both accuracy and go-response reaction times measurements. Targets and non-targets were equiprobable in each series, which set the chance level at 50%.

Behavior: superordinate level tasks (A/nA)

Accuracy: The control condition in both Bird and Dog experiments was an Animal/non-Animal task (A/nA) that has often been used in other studies (Thorpe *et al.*, 1996; Fabre-Thorpe *et al.*, 2001; VanRullen & Thorpe, 2001a). A particularity of the present experiment was that half of the targets were very varied photographs of different types of animals whereas the other half included only birds in the Bird experiment and dogs in the Dog experiment. These bird and dog photographs were also used as targets in the basic level categorization tasks (for different subjects to avoid any repetition bias), to allow analysis of the very same set of images when processed either at the superordinate or at the basic level. The global accuracy in the control A/nA task was similar between the two experimental series (95.8% and 95.5%). In control task of the Bird experiment, bird pictures were categorized as “animal” with an accuracy that was slightly higher than for the non-bird animals (accuracy on targets : 98.5 vs 94.8%; Khi^2 , $p < 0.05$). In the control task of the Dog experiment, subjects categorized the dog photographs as animal with an accuracy that was virtually the same than for all non-dog animal pictures (accuracy on targets : 96.4 and 96.8%, ns). In both control tasks, subjects were slightly better at responding on animal targets (hit responses) than at ignoring non-targets (correctly withhold responses): Bird experiment: 96.6 vs 95.0; Dog experiment: 96.6 vs 94.4; Khi^2 , $p < 0.05$ for both).

The important conclusion of these results is that the varied set of birds and dogs images chosen for the categorization at the basic level were categorized with the same accuracy (dogs) or even a better

target signal. Significant differences between the two conditions were assessed by paired t-tests at the $p < 0.01$ level, at each scalp location for each 1-ms time bin. The time bin for which a significant t-test value was reached and followed by at least 15 consecutive significant bins was taken as the onset latency of a differential activity. All values reported in the text met this criterion.

accuracy (birds) than the set including all other animal photographs.

Speed: Concerning the speed of responses, in both the Bird and Dog experiments, the animal control task was performed with comparable mean reaction time (mean RT : 394 ms in the Bird and 386 ms in the Dog experiment). Interestingly in both control tasks, the pictures of birds and dogs were categorized faster than the pictures of other animals. Birds were categorized as animal with a mean RT of 385 ms (402 ms for the other animals; t-test, $p < 0.01$) and dogs with a mean RT of 377 ms (394 ms for the other animals, t-test, $p < 0.01$). These differences are accounted for by some very long latency responses recorded on some non-bird or non-dog animal targets. Such long latency responses are due to specific photographs that need long processing times to be analyzed (Fabre-Thorpe *et al.*, 2001). In contrast, the very first responses appeared at the same latencies (fig 2). To evaluate the reaction time of the earliest responses that cannot be attributed to anticipation we use as index the minimal processing time (MinRT). This value corresponds to the first time bin in the RT distribution from which correct responses significantly outnumber false alarms. It reflects the shortest processing time necessary for the visual system to discriminate between targets and non-targets in a given task. In both control A/nA tasks the MinRT was the same for the bird and dog photographs compared to all other types of animals (Bird experiment: 270 ms for birds and other animals (250 ms when taken altogether); Dog experiment: 260 ms for dogs and other animals (also 250 ms altogether)).

This control task allows to conclude that the bird and dog photographs were analyzed as any other animal image in the control A/nA task. If anything, they might be slightly easier to process as they were categorized in average 10 ms faster, with fewer long latency responses and even with a higher percentage of correct responses (+3%) for the bird photographs.

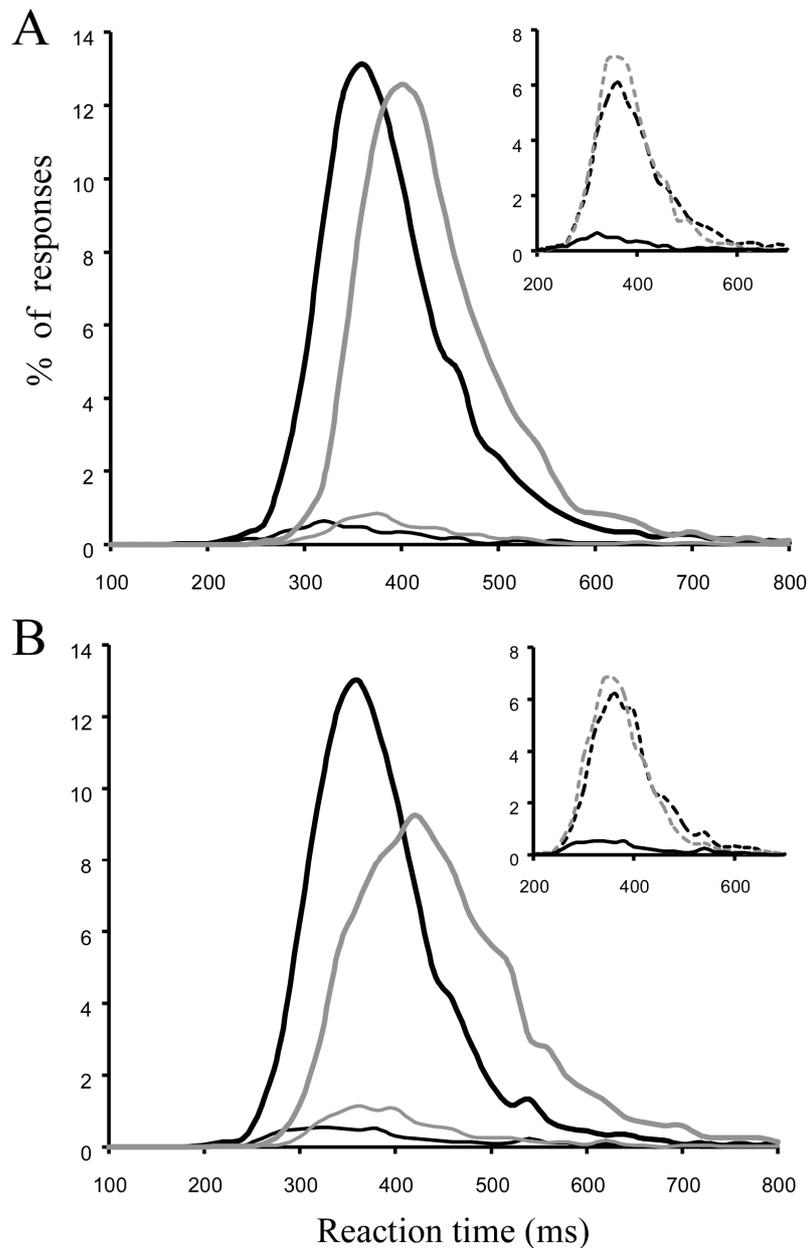


Figure 2: Reaction time distributions in the Bird (A) and in the Dog (B) experiments on correct (thick lines) and incorrect (thin lines) trials, calculated with 10 ms bin width. Reaction time distributions were computed separately in the superordinate level task (black curves) and the basic level task (gray curves). The inserts correspond to RT distributions within the animal/non animal task for birds (A) and dogs (B) in dotted gray lines and the other animals in dotted black lines.

Behavior: basic level tasks

Accuracy: In the basic level categorization task, subjects scored in average 95.6% correct in the Bird task and 92.6% in the Dog task. They reached accuracy scores that were slightly lower when categorizing the bird photographs at the basic level (97.2%) than at the superordinate level (98.5%; Khi^2 , $p < 0.05$). The same effect was observed with dogs categorized at the basic level compared to the superordinate level (94.7% vs. 96.8%; Khi^2 , $p < 0.05$). As in the superordinate tasks, subjects were better at responding on targets than at ignoring non-targets in

both basic level tasks (97.2% vs 93.9% in the bird task, 94.7% vs 90.5% in the dog task; Khi^2 , $p < 0.05$). As the non-targets pictures contained at the same time neutral images and non-target animals, it is important to specifically look at the performance on photograph of non-target animals. Indeed, the false alarms were in a very large majority elicited by non-target animals. In the Bird task, 90% of the false alarms were committed on non-bird animals; a proportion that was even higher in the Dog experiment with 95% of the false alarms on non-dog animals. This result also means that when subjects

performed the categorization task at the basic level, they were able to correctly ignore neutral non-animal photographs with a very high degree of accuracy (99% correct or over in the bird and the dog tasks). Interestingly, it seems that some animal categories were more likely to induce false alarms but strongly depends on the basic level categorization task considered. Figure 3 shows the difference of false alarms percentage between the Bird and the Dog task as a function of the animal non-target categories (% of FA in the Dog task minus % of FA in the Bird task). We see that some categories of animals had completely opposed patterns of errors in the 2 experiments. For example insects and sea animals elicited far more error in the Bird task than in the Dog task and this was exactly the opposite for bears and felines.

Speed: The mean reaction time at the basic level was 434 ms in the Bird task and 452 ms in the Dog task. These mean RT values are considerably longer than

those obtained for the same images of birds and dogs categorized at the superordinate level (control A/nA tasks). Globally, subjects were respectively 40 ms and 65 ms slower in the Bird and in the Dog experiments to categorize images at the basic level. This effect was not limited to the mean RT. As illustrated in figure 2, the whole RT distributions in the Bird and the Dog experiments were shifted towards longer latencies at the basic level. The shape of the RT distribution in the Dog task was also asymmetrical with a particularly large number of long latency responses.

This shift in latencies was also present on the earliest responses : this is reflected by the increased minimal processing time (MinRT) in the basic level tasks compare to the control tasks: 300 ms in the Bird task and 290 ms in the Dog task. Thus, MinRTs were respectively increased by 50 ms and 40 ms when compared to the control superordinate tasks (Fig 2).

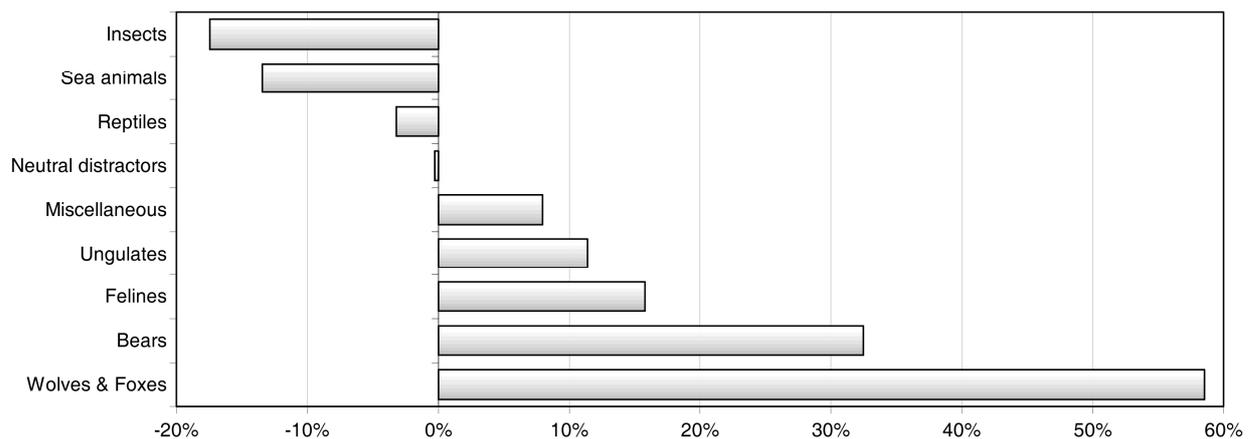


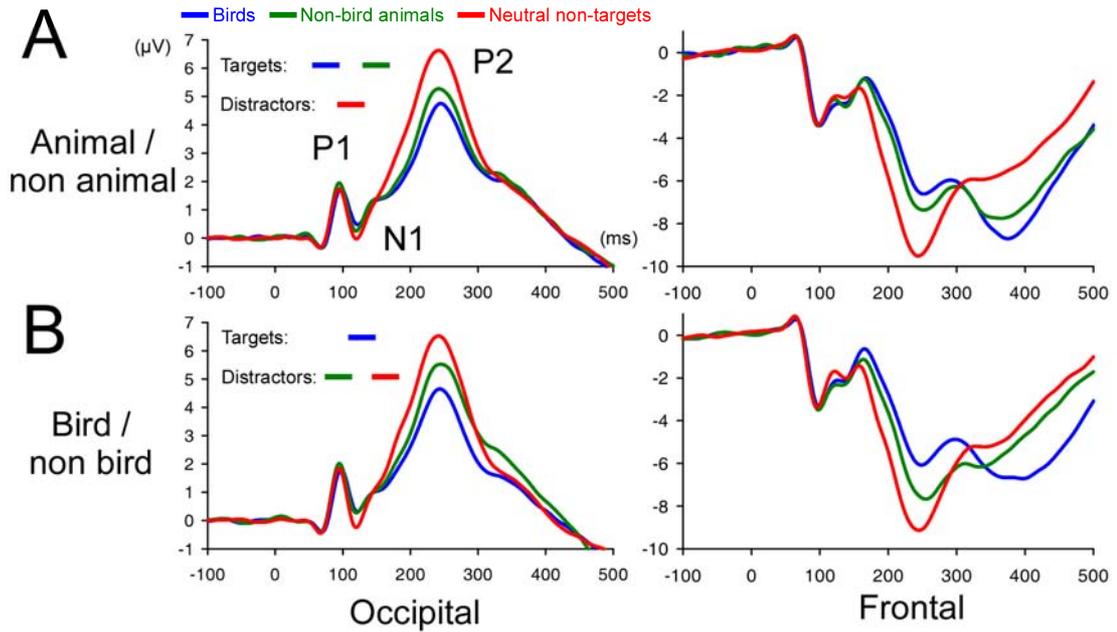
Figure 3: Bars represent the difference of false alarm rates between the Bird and the Dog tasks for several subgroups among animal non-targets. Negative values (insects, sea animals...) mean that subject had a higher false alarm rate for this category of animals when they were looking for birds compare to dogs. For example, the positive value of 32.5% for bears is the result of the subtraction between 38.5 (false alarm rate for bears in the Dog task) and 6.0 (false alarm rate for bears in the Bird task). These discrepancies in the nature of the committed errors could take its origin in the specific pre-activation state of the visual system in order to recognize rapidly its targets.

Electrophysiology: basic and superordinate level tasks

We averaged separately the signal on target and non-target trials. In the control tasks, we also separated dogs and birds signal from the other animals and in the basic level tasks, animal non-targets signal from the other non-targets. Latencies and amplitudes of the early components P1 and N1, supposed to reflect the physical encoding of the stimulus, were similar

between the different conditions within all the tasks (Fig 4 A-D). The first significant effects were found on the P2 wave (150-300 ms) with a larger amplitude on neutral non-targets compare to the other images in all tasks, both at occipital and frontal sites. In the control tasks, there was a slight difference in the amplitude of the P2 between non-bird animals and birds, but not between non-dog animals and dogs (respectively Fig 4A and 4C). In

Bird experiment



Dog experiment

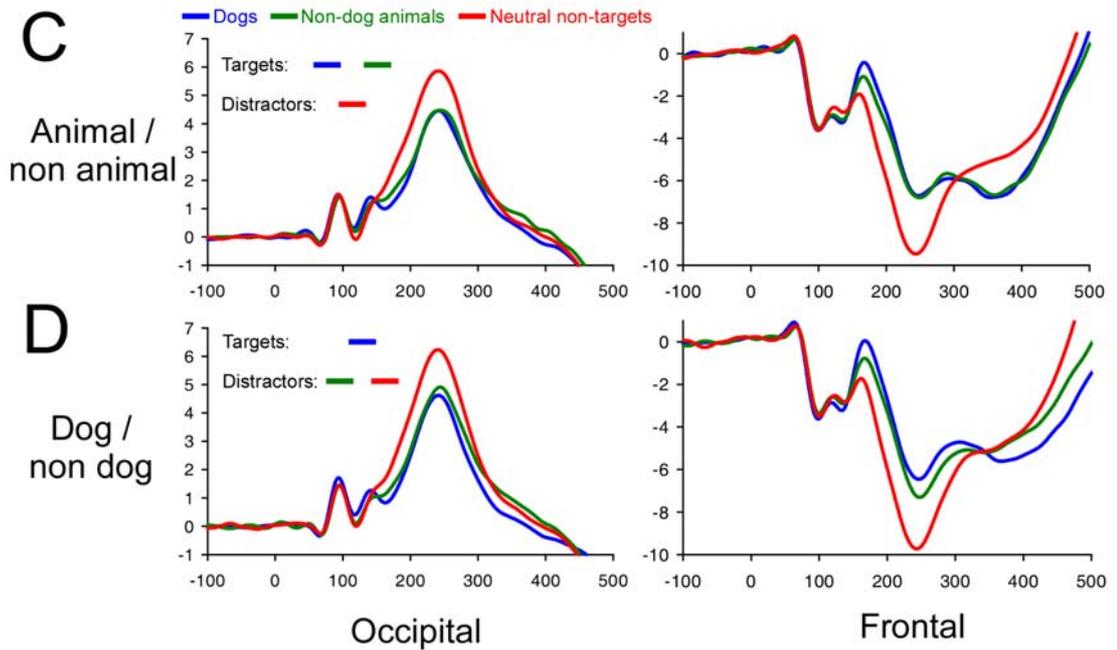


Figure 4: Event Related Potentials (ERP) recorded in Bird (A&B) and Dog experiment (C&D) on over 6 occipital (O1-O2-O9-O10-Iz-Oz) and frontal (FP1-FP2-F7-F8-F3-F4) electrodes. ERPs on birds (A&B) or dogs (C&D) are in blue, non-bird (A&B) or non-dog (C&D) animals in green and neutral non-targets in red. The images of birds/dogs and other animals were either categorized at the superordinate level (A&C) or at the basic level (B&D).

the basic level tasks, P2 amplitude difference between bird and non-bird animals was slightly increased and a small difference appears on P2

amplitude between non-dog animals and dogs (respectively Fig 4B and 4D).

To look closely at these differences, we computed the differential activity (DA) by subtracting the average signal on correct non-target trials from the average signal on correct target trials (Fig 5). A small early DA was found around 70 ms both in the Bird and in the Dog tasks (half of posterior electrodes were significantly different from the baseline) whereas no such early differences were seen in both control tasks. These very early differences have been shown to reflect physical differences between the target and non-target image sets (VanRullen & Thorpe, 2001b). This component could have been expected, as images of birds or dogs share a large amount of similar features within them, whereas the set of images containing all the other animals is more heterogeneous and thereby less distant from the non-target pictures. We won't discuss further these early ERP differences, as they may be related to low level physical encoding of the images rather than to an early processing mechanism that would be used to analyze the same stimulus but only in the basic-level categorization task.

In both experiments, we found the first large differential activity that probably reflects the advanced cognitive processing which allows the detection of the animal targets between 150 and 250 ms (Thorpe *et al.*, 1996; VanRullen & Thorpe, 2001b; Rousselet *et al.*, 2004). The average onset latency of this task-related DA was measured on 7 posterior electrodes (P7-P8-PO7-PO8-O9-O10-Oz). In the Bird experiment, it was found on average at 164 ms in the control task and 160 ms in the basic level task. In the Dog experiment, these latencies were respectively 159 ms and 161 ms. Thus the DA onset latencies in the basic level tasks are very similar to those observed in the superordinate tasks. If we look more specifically at the trials with bird and dog photographs within the control tasks, the onset latencies are nearly unchanged (166 and 161 ms respectively) but the amplitude of the DA is considerably higher when compared to the target

trials containing the other animal species (Fig 5 A&B: dark grey vs black).

The fact that we could not find any difference in the onset latency of the different conditions might be surprising in regard to the behavioral results. With MinRTs shifted by 40-50 ms towards longer latencies in the basic level categorization tasks compare to superordinate level tasks, one would expect a delayed onset of the differential activity. This is clearly not the case and an explanation may come from recent experiments (Bacon-Macé *et al.*, 2005a; Macé *et al.*, 2005a) which give us evidence that the difficulty of a task is more related to the amplitude of the DA than to its latency. This was the case in both basic level tasks which had lower DA amplitude when compared to superordinate tasks, together with lower behavioral results (Fig 5 A&B: light grey vs dark grey/black). However, this correlation does not hold here, as in the control task of the Dog experiment, DA was clearly larger on dogs than on the other animals while behavioral accuracy was at a similar level (Fig 5 A&B: dark grey vs black).

Although the differential activity elicited in go/no-go A/nA control tasks is clearly visible on EEG recordings, it seems from the present results that no clear differential activity could be recorded between non-target animals and target animals in the basic level tasks.

On figure 5, a difference between the two experiments in the profile of their basic level DA curves is visible. The DA for the dogs reached a lower amplitude than for the birds. The latency and the time course are quite similar up to a point where a marked plateau appears in the dog task. This effect could be due to the higher feature similarity between dogs and the other animals than between birds and the other animals, a characteristic that could explain both this lower DA signal and the lower behavioral performance.

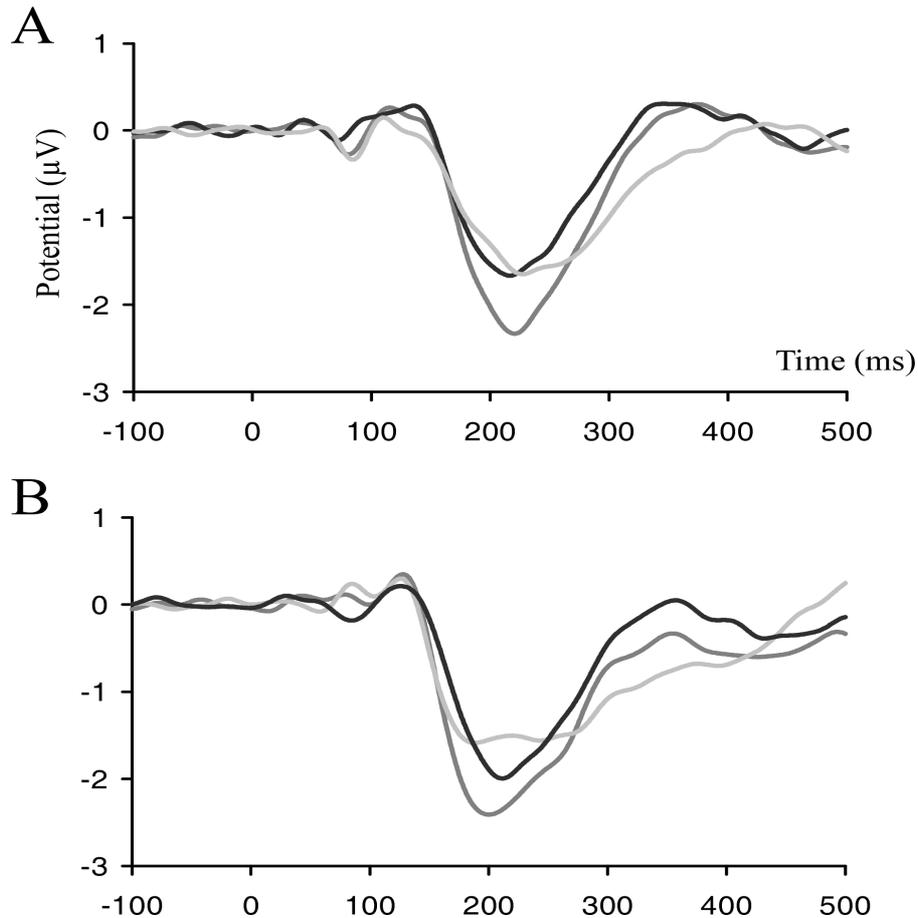


Figure 5: Differential activity (DA) signal calculated by subtracting the signal recorded on correct non-targets from correct targets on 7 posterior electrodes (P7-P8-PO7-PO8-O9-O10-Oz) in Bird (A) and Dog (B) experiments. The DA latencies were similar for all conditions in most of the electrodes and only the amplitude of the DA differs across the conditions. In the control tasks, the amplitude of Birds or Dogs minus non-targets (dark grey) was larger than non-bird or non-dog animals minus non-targets (black). In the basic level tasks, we calculated the DA between the targets and the non-targets (light grey).

DISCUSSION

The two experiments presented here strongly suggest that the accepted idea supporting that basic level is perceptually accessed before superordinate level must be challenged. In a go/no-go categorization task using complex natural scenes photographs, subjects are largely faster (and also better in Dog experiment) at finding animals than birds or dogs. This result is surprising in regard to the large set of previous experiments that have demonstrated a basic level

advantage both in terms of speed and accuracy. Part of our 40-65 ms reversed advantage could be explained by the stimulus we used here. Murphy et al., (Murphy & Wisniewski, 1989) showed that the display of detailed scenes reduced the basic level advantage measured on isolated objects. However, according to this study, the "full scene effect" should not be important enough on its own to completely revert the traditional speed advantage of the basic level. Murphy et al. proposed that this effect is due to

coherent interactions between the background, potential objects in the scene and the target. However, we propose a complementary hypothesis to explain this radical inversion in processing speed, by raising the differences in terms of lexical requirements between the experiments. The large majority of previous works on levels of categorization require a lexical access additionally to the visual processing of objects (category verification tasks, lexical priming, category naming, lexical input to switch target category at each trial, etc). The consequence could have been that the revealed architecture of category structure was more derived from the lexical category structure than actual visual processing. Although no doubt subsists regarding the superiority of basic level in language, the results found here could reflect a category structure more strictly linked to visual perception. Our experiment was designed to minimize the interactions with language, and monkey's high performance in this task constitutes in itself a good demonstration of this point. In the visual system, the superordinate level may not constitute an abstraction from the basic level as previously proposed (Rosch *et al.*, 1976; Jolicoeur *et al.*, 1984a), but rather represents the less detailed level at which some visual properties can be grouped to primarily construct distinct categories. This idea is very close to the hypothesis of an architecture that would process information from coarser to finer representations, as theoretically proposed by Schyns *et al.* (Schyns & Oliva, 1994). A number of experimental findings support this model: electrophysiological studies in monkeys (Sugase *et al.*, 1999), functional imaging in humans (Liu *et al.*, 2002) or psychophysical data (Sergent, 1985, Parker *et al.*, 1992). During the reconstruction of the scene along the visual pathway, objects contours are the highest saliency locations (strongest contrast), and should be processed first, conveyed by the low spatial frequency information. This information is available at the very beginning of the integration process and may be sufficient to infer the coarse representation of the scene necessary to perform a superordinate categorization task. In contrast, the basic level categorization may need to wait for more details on the stimulus that require additional processing. Recent RSVP (Rapid Serial Visual Presentation) experiments (Bacon-Macé, in preparation) showed that only coarse information is included in the very first visual representation, and that more time is

needed to access the details of the image. For example, in a task where subjects had to categorize animals, they were able to respond for the presence of an animal with great confidence while being unable to report whether this animal was a rabbit or a horse. So, in conditions where the information capture is severely limited by temporal constraints, the categorization task that shows best performance is at the superordinate level. At these extreme conditions, only the information that is extracted, conveyed and processed at first could be used to respond. A comparable result had been found in a masking experiment on natural images (Grill-Spector & Kanwisher, 2005). Grill-Spector *et al.* showed that a categorization at a level intermediate between basic and superordinate level can be done with the same accuracy as a simple detection task. These experiments provide more evidence that the superordinate level is accessible as soon as the very first information are available, contrary to what happen for more inclusive levels.

We calculated differential activities based on the EEG signal of correct trials. Contrary to the behavioral results, no clear difference was found between the latencies of differential activities onsets in the superordinate and the basic level tasks for the two experiments. If we suppose that an image of a bird or a dog is categorized first at the superordinate level and afterwards at the basic level, there is no reason why the onset latency between these two tasks should differ. What is certainly harder to discriminate are the basic level images from the images that belong to the same superordinate category. This case corresponds to the light gray curve in the Figure 5, and no differences are clearly visible in its DA onset latency compare to the black curve (superordinate task). This could be because some features that birds or dogs exclusively share among them allow some images to be distinguished early from the other animals and neutral non-targets. However, this DA has a reduced amplitude compare to the bird or dog DA (dark gray) in the superordinate task. The visual system seems uncomfortable at discriminating the images of birds and dogs from the other animals and incidentally, the number of false alarms on animals non-targets was 10 to 25 times greater than on neutral non-targets.

It is interesting to note that the behavioral results on the two experiments were not totally similar, the dog/non dog categorization seems to be harder to

perform compared to the bird/non bird task. This effect may originate from the high prototypicality of dogs relatively to the other animals. Dogs are very typical mammals, not only they occupy a special place in the animal kingdom for humans, but they also share more features with the other animals than the birds may do (like 4 legs and some fur). In contrast, the feathers, wings, beaks and aerodynamic shapes of birds can lead to certainly more peculiar and thus distinguishable views. The expected consequence is that more visual features are diagnostic for birds and they could be more easily distinguished from the other animals comparatively to dogs. As an illustration, typical errors in the bird/non bird task were made on insects and fishes (an aero/hydro-dynamic global shape effect ?) whereas errors in the dog task were committed on canines, bears, felines and ungulates (Fig 5). Besides, we also observed that the amplitude of the differential activity was more reduced on the dog/non-dog task compared to the animal/non-animal task, reflecting the greater difficulty for the visual system to separate the dog representations from the other animals. This degree of difficulty in the tasks raised the question of the choice of pictures in this experiment. We tried to use images as varied as possible and the set we used could be considered as relatively representative of the animal kingdom, a point that could be criticized in previous studies working on the categorization levels with a very small number of stimuli. Even if some differences exist between our two basic level tasks, performance on both bird and dog categorization were strongly affected compared to the animal categorization. Following Schyns and colleagues, (Schyns, 1998) we argue that finding the most diagnostic features that characterize the targets relatively to the non-targets images is the essential point to perform such categorization tasks. A logical sequel of this study would consist in a basic level categorization task where the proportion of “tricky” non-targets (same superordinate category as the targets) would be changed. By this way, we could measure the effects of varying the number of shared features between targets and non-targets. Using a

target category at a more inclusive level than the basic level (subordinate or exemplar level) could lead to even shorter reaction time if no images from the same higher order level are used. In such tasks, the visual system could be more and more confident at doing prediction regarding the target features and better pre-activation of the visual representations could shorten the RT. There could even be a shift towards simple detection of low-level feature if predictability becomes complete, as in a previous experiment where a single natural image was used as target (Delorme *et al.*, 2004a). Subject were around 40 ms shorter to detect this single image than to categorize at the superordinate level. This result also means that unconstrained categorization tasks at levels inferior to the superordinate level could at most be 20-40 ms faster.

Manipulating category structures and boundaries is not a new idea and has previously been performed to explain typicality and expertise effects (Tanaka & Taylor, 1991; Murphy & Brownell, 1985). The underlying hypothesis is that training on a particular set of stimuli can possibly modify the representations in the inner visual processing and facilitate recognition by increasing encoding speed. Regarding the atypical members of a category, the effect should be due to some better combination of diagnostic features that could help to boost their recognition. However, the results we present here are not limited modifications in the category structure with repeatedly presented stimuli or a peculiar set of features. They imply that the process of categorization inside the visual system is different from what had been thought for several decades. The representations involved in our task are linked to low level cortical structure organization, with a limited plasticity (no great differences in purely visual capabilities after millions of years of separate evolution between human beings and macaques). We have now to investigate how these different category structures are linked and cooperate to rapidly attach a semantic and a lexical content to any objects.

ACKNOWLEDGMENT

This work was supported by the Integrative and Computational Neuroscience ACI program of the CNRS. Financial support was provided to M.J.-M. Macé by a Ph.D. grant from the French government. The authors declare that they have no competing financial interests.

REFERENCES

- Bacon-Macé, N., Mace, M.J., Fabre-Thorpe, M. & Thorpe, S.J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res*, 45, 1459-1469.
- Boucart, M., Fatima, N., Desprez, P., Defoort-Dhelemmes, S., Macé, M.J.-M. & Fabre-Thorpe, M. (Submitted). Implicit but not explicit recognition at very large visual eccentricities. *J Exp Psychol Hum Percept Perform*.
- Delorme, A., Rousselet, G.A., Mace, M.J. & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res*, 19, 103-113.
- Fabre-Thorpe, M., Delorme, A., Marlot, C. & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci*, 13, 171-180.
- Fabre-Thorpe, M., Richard, G. & Thorpe, S.J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, 9, 303-308.
- Grill-Spector, K. & Kanwisher, N. (2005). Visual recognition. *Psychol Sci*, 16, 152-160.
- Johnson, K.E. & Mervis, C.B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal-of-experimental-psychology-General*, 126, 248-277.
- Jolicoeur, P., Gluck, M.A. & Kosslyn, S.M. (1984). Pictures and names: making the connection. *Cognit Psychol*, 16, 243-275.
- Large, M.E., Kiss, I. & McMullen, P.A. (2004). Electrophysiological correlates of object categorization: back to basics. *Brain Res Cogn Brain Res*, 20, 415-426.
- Li, F.F., VanRullen, R., Koch, C. & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A*, 99, 9596-9601.
- Liu, J., Harris, A. & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat Neurosci*, 5, 910-916.
- Macé, M.J., Thorpe, S.J. & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *Eur J Neurosci*, 21, 2007-2018.
- Mervis, C.B. & A., C.M. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, 53.
- Murphy, G.L. & Brownell, H.H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *J Exp Psychol Learn Mem Cogn*, 11, 70-84.
- Murphy, G.L. & Wisniewski, E.J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *J Exp Psychol Learn Mem Cogn*, 15, 572-586.
- Parker, D.M., Lishman, J.R. & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, 21, 147-160.
- Quiñero, R., Reddy, L., Kreiman, G., Koch, C. & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102-1107.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Rousselet, G.A., Thorpe, S.J. & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res*, 44, 877-894.
- Schyns, P.G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, 67, 147-179.
- Schyns, P.G. & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychol Sci*.
- Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *J Exp Psychol Hum Percept Perform*, 11, 846-861.
- Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400, 869-873.
- Tanaka, J., Luu, P., Weisbrod, M. & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *Neuroreport*, 10, 829-835.
- Tanaka, J.W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognit Psychol*, 23, 457-482.
- Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.
- Thorpe, S.J., Gegenfurtner, K. R., Fabre-Thorpe, M., Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *Eur J Neurosci*, 14, 869-876.
- VanRullen, R. & Thorpe, S.J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, 30, 655-668.
- VanRullen, R. & Thorpe, S.J. (2001b). The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci*, 13, 454-461.

Synthèse et perspectives

L'ensemble des résultats de cette thèse ont été obtenus grâce à une tâche de catégorisation dans laquelle les sujets devaient effectuer une réponse motrice rapide sur la base d'informations sensorielles incomplètes. Le protocole de masquage a non seulement permis de confirmer une nouvelle fois que les traitements visuels sont robustes et rapides (Chapitre 2), mais aussi que les informations parcellaires peuvent être transformées en signal décisionnel de manière optimisée, sous la dépendance du type d'effecteur (Chapitre 3) et du niveau de représentation requis par la tâche (Chapitre 4). Une quarantaine de millisecondes de SOA semblent être suffisantes pour extraire la majorité de l'information pertinente dans la tâche de catégorisation animal/non-animal, délai qui est encore plus court lorsqu'il s'agit de catégoriser des visages. Cette latence est tout à fait compatible avec l'idée que la reconnaissance d'objet est d'abord basée sur un transfert rapide des premières informations tout au long de la hiérarchie des aires corticales visuelles. De par les contraintes temporelles, ce transfert doit se faire de manière feedforward et massivement parallèle (Rousselet, Thorpe & Fabre-Thorpe, 2004a; Rousselet, Thorpe & Fabre-Thorpe, 2004b; VanRullen & Koch, 2003; VanRullen & Thorpe, 2001), et bénéficie peut-être de la vitesse de conduction plus rapide de la voie magnocellulaire (Macé, Thorpe & Fabre-Thorpe, 2005). La détermination des contraintes temporelles qui pèsent sur le système dans ce type de tâches écologiques (puisque notre monde visuel est structuré sous forme de catégories) permet aussi de contraindre les modèles développés pour reproduire le fonctionnement des processus visuels. Les données recueillies par l'équipe depuis 1996 ont d'ailleurs amené à remettre en question la nature du codage neuronal jusque-là accepté classiquement (Gautrais & Thorpe, 1998; Thorpe, Delorme & Van Rullen, 2001; VanRullen, 2000; VanRullen & Thorpe, 2002). Sachant qu'il existe une dizaine d'étapes synaptiques entre la rétine et IT et que les premières réponses sélectives à des stimuli complexes apparaissent avant 100 ms dans IT, le transfert entre deux étapes ne doit pas prendre plus de 10 ms en moyenne. Cette période semble bien trop courte pour qu'une fréquence de décharge puisse être calculée puisque très peu de neurones auront déchargé plus d'une fois durant cette période. En revanche, un codage par rang semble beaucoup plus adapté à ces contraintes temporelles (Gautrais & Thorpe, 1998; Thorpe, Delorme & Van Rullen, 2001). Dans ce cas, le code neuronal ne correspond pas à un taux de décharge mais à l'ordre dans lequel les neurones déchargent, en supposant que les informations sont transmises de manière spatialement asynchrone, les neurones codant les propriétés les plus saillantes du stimulus déchargeant le plus rapidement. Ce type de codage permet donc de déterminer rapidement la présence d'indices saillants dans la scène visuelle (par exemple les bords des objets), en convertissant l'information représentée au départ par une intensité en une latence relative. Le débat n'est pas clos entre les partisans des différents codages neuronaux, et le codage en taux de décharge a encore des arguments très convaincants permettant un fonctionnement avec des étapes de 20-30 ms sur population de cellules (Rolls, Treves & Tovee, 1997). Cependant, les travaux comportementaux comme électrophysiologiques s'intéressant aux contraintes temporelles des traitements dans les systèmes sensoriels permettront à terme de valider le ou les modèles de codage les plus

compatibles biologiquement, et l'enjeu est d'autant plus important que le codage neuronal est la base même des mécanismes d'intégration de l'information.

Un thème majeur développé dans cette thèse porte sur l'hypothèse que les informations s'accumulent progressivement dans le système sensoriel et le franchissement de seuils semble constituer un « filtre » indispensable pour valider le résultat d'une compétition marquée entre des sources multiples d'information. C'est justement cette accumulation que l'on a cherché à observer au cours du temps en perturbant les traitements à différentes étapes du parcours cortical des informations. Pour les SOA les plus courts, les interférences spatio-temporelles sont vraisemblablement maximales dans les aires visuelles précoces, probablement avant V4 et peut-être même dès la rétine. Pour des niveaux de représentations plus avancées, codant pour des propriétés plus complexes, la compétition entre les informations appartenant aux différents stimuli prend alors une autre forme, et l'utilisation de visages en tant que stimuli a été particulièrement informative de ce point de vue. En effet la reconnaissance des visages est au centre d'un débat entre les partisans d'un module dédié au traitement de ces stimuli, identifié anatomiquement dans une région du gyrus fusiforme, et les partisans de l'expertise, pour lesquels les visages sont des stimuli sur-appris mettant en jeu des structures capables d'effectuer des discriminations fines. Les arguments proposés par les deux camps sont souvent très convaincants, mais il est certainement possible de réconcilier leurs données en proposant que les taux d'accumulation des informations, qui se traduisent par le taux de décharge des neurones sélectifs aux stimuli, dépendent de la fréquence à laquelle ont été vus ces stimuli auparavant (Perrett, Oram & Ashbridge, 1998). Plus un objet est familier comme c'est le cas des visages, plus les seuils de discrimination vont être atteints rapidement. Encore une fois, il est intéressant de faire le lien entre ces travaux comportementaux et les hypothèses développées par les « modélisateurs » de notre équipe. Rudy Guyonneau et Timothée Masquelier ont poursuivi le développement du modèle fondé sur le codage par rang (Guyonneau, 2006; Guyonneau, Vanrullen & Thorpe, 2005; Masquelier & Thorpe, en préparation) en introduisant une règle d'apprentissage dans laquelle les synapses sont renforcées lorsque les stimuli sont présentés plusieurs fois (STDP Spike Timing-Dependant Plasticity). Cette règle d'apprentissage, basée sur des observations électrophysiologiques (Bi & Rubin, 2005; Markram, Lubke, Frotscher & Sakmann, 1997), permet de sélectionner les neurones qui ont servi à activer une représentation donnée, de telle manière qu'une synapse excitatrice renforce toute connexion qui a émis une décharge avant que la décharge post-synaptique ne soit elle-même émise et affaiblit les connexions qui lui ont transmis une décharge après la décharge post-synaptique. Les modifications sont d'autant plus importantes que les intervalles de temps qui séparent les deux évènements sont courts. Cet algorithme d'apprentissage non-supervisé a ainsi prouvé qu'il était possible de « reconnaître » des objets de complexité intermédiaire dans des scènes naturelles (comme un œil, une bouche. Voir la vidéo « demo_STDP_face.avi » sur la page Cerco de Timothée Masquelier : <http://cerco.ups-tlse.fr/~masqueli/>), avec une grande précision et une grande robustesse au bruit, à la rotation, au changement d'échelle ou aux variations de contraste. L'idée que des stimuli que nous voyons aussi fréquemment que les visages bénéficient d'un réseau de connexions particulier avec une

sélectivité amplifiée paraît donc tout à fait plausible, d'autant que les propriétés bas-niveau des visages sont relativement homogènes et prototypiques.

Le protocole de choix saccadique offre de belles perspectives, à la fois parce qu'il permet de travailler sur des gammes de latences particulièrement courtes, mais aussi parce que le système oculomoteur est assez bien connu. Un autre avantage incontestable de ce protocole est qu'il permet d'analyser directement la nature des indices utilisés pour la reconnaissance d'objet, en opposant directement dans un même essai des images cible et distracteur. Pas de doute que le mystère « Visage » va susciter de nombreuses expériences dans les années qui viennent.

D'autre part, un des thèmes majeurs à développer dans le futur doit concerner la manière dont les informations sensorielles sont transformées en signal moteur, et plus spécifiquement les mécanismes et structures engagés dans la prise de décision.

« Precisely because the processes leading up to these decisions are hidden from us, we enjoy a robust, confident picture of the world that belies the fact that sensory signals are uncertain, the consequences of actions unpredictable. We think that we see, but actually we guess. » (Carpenter, 1999)

Un grand problème soulevé par les travaux de cette thèse, c'est qu'il est difficile avec un protocole purement comportemental d'identifier les structures impliquées dans les différents processus. Il est alors possible d'envisager de remplacer le protocole de masquage par un autre outil pour perturber localement des traitements : la TMS (Stimulation Magnétique Transcranienne). Des aires comme V1 sont relativement faciles à atteindre avec ces stimulations magnétiques, ce qui devrait permettre de confirmer son implication majeure dans les interférences observées dans nos tâches de masquage. Encore plus intéressant, la stimulation de régions comme le FEF à des intervalles de temps variables, pourrait aussi permettre de vérifier certaines hypothèses comme son éventuelle rôle dans la génération automatique de temps de réaction.

Enfin, même si nous n'avons pas fini de soulever de nouveaux points de recherche autour d'un protocole aussi riche que celui de la catégorisation de scènes naturelles, rappelons que si les premières vagues de traitement permettent de dégager une représentation grossière de la scène, suffisante pour déclencher des réponses motrices, d'avantage de traitements sont nécessaires pour affiner cette représentation et probablement pour en avoir un accès conscient. Comme nous l'avons déjà évoqué dans le chapitre 3, le protocole de choix saccadique paraît encore une fois particulièrement bien adapté pour étudier l'émergence potentiellement graduelle du phénomène conscient.



DEPART SYNCHRONE
(des informations à l'entrée
de la rétine)



PASSAGE DE SEUIL
(au dessus du bruit
dans le système sensoriel)



PARCOURS ASYNCHRONE
(des informations le long
de la voie visuelle)



COMPETITION
(spatio-temporelle entre
les représentations
des différents stimuli)



PASSAGE DE SEUIL
(décisionnel pour la
représentation gagnante)



LIAGE
(des informations
entre elles)

Références

- ADOLPHS, R. (2003). Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci*, 4:165-178.
- ALBRECHT, DG, GEISLER, WS, FRAZOR, RA & CRANE, AM. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *J Neurophysiol*, 88:888-913.
- ALLISON, T, PUCE, A, SPENCER, DD & MCCARTHY, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex*, 9:415-30.
- AMARASINGHAM, A, CHEN, TL, GEMAN, S, HARRISON, MT & SHEINBERG, DL. (2006). Spike count reliability and the Poisson hypothesis. *J Neurosci*, 26:801-9.
- ANBAR, S & ANBAR, D. (1982). Visual masking: a unified approach. *Perception*, 11:427-39.
- ARIKUNI, T, WATANABE, K & KUBOTA, K. (1988). Connections of area 8 with area 6 in the brain of the macaque monkey. *J Comp Neurol*, 277:21-40.
- ASHBY, FG. (1982). Deriving exact predictions from the cascade model. *Psychol Rev*, 89:599-607.
- BAARS, BJ. (2002). The conscious access hypothesis: origins and recent evidence. *Trends Cogn Sci*, 6:47-52.
- BACHMANN, T. (1984). The process of perceptual retouch: nonspecific afferent activation dynamics in explaining visual masking. *Percept Psychophys*, 35:69-84.
- BACHMANN, T. (1997). Visibility of brief images: the dual-process approach. *Conscious Cogn*, 6:491-518.
- BACHMANN, T & HOMMUK, K. (2005). How backward masking becomes attentional blink. *Psychol Sci*, 16:740-2.
- BACHMANN, T, LUIGA, I & PODER, E. (2004). Forward masking of faces by spatially quantized random and structured masks: on the roles of wholistic configuration, local features, and spatial-frequency spectra in perceptual identification. *Psychol Res*, 69:11-21.
- BACON-MACÉ, N, MACE, MJ, FABRE-THORPE, M & THORPE, SJ. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Res*, 45:1459-69.
- BAIZER, JS, UNGERLEIDER, LG & DESIMONE, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J Neurosci*, 11:168-90.
- BAR, M. (2004). Visual objects in context. *Nat Rev Neurosci*, 5:617-29.
- BAR, M & BIEDERMAN, I. (1998). Subliminal visual priming. *Psychological Science*, 9:464-469.
- BAR, M & BIEDERMAN, I. (1999). Localizing the cortical region mediating visual awareness of object identity. *Proc Natl Acad Sci U S A*, 96:1790-3.
- BAR, M, KASSAM, KS, GHUMAN, AS, BOSHYAN, J, SCHMIDT, AM, DALE, AM, HAMALAINEN, MS, MARINKOVIC, K, SCHACTER, DL, ROSEN, BR & HALGREN, E. (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*, 103:449-54.
- BAR, M, TOOTELL, RB, SCHACTER, DL, GREVE, DN, FISCHL, B, MENDOLA, JD, ROSEN, BR & DALE, AM. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, 29:529-35.
- BARBAS, H & PANDYA, DN. (1987). Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J Comp Neurol*, 256:211-28.
- BAVELIER, D. (1994). Repetition Blindness between visually different items: the case of pictures and words. *Cognition*, 51:199-236.
- BAYLIS, GC, ROLLS, ET & LEONARD, CM. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res*, 342:91-102.
- BENTIN, S, ALLISON, T, PUCE, A, PEREZ, E & MCCARTHY, G. (1996). Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*, 8:551-565.
- BENTIN, S & GOLLAND, Y. (2002). Meaningful processing of meaningless stimuli: The influence of perceptual experience on early visual processing of faces. *Cognition*, 86:B1.
- BERTELSON, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Q J Exp Psychol*, XIII:90-102.
- BI, GQ & RUBIN, J. (2005). Timing in synaptic plasticity: from detection to integration. *Trends Neurosci*, 28:222-8.
- BICHOT, NP, THOMPSON, KG, CHENCHAL RAO, S & SCHALL, JD. (2001). Reliability of macaque frontal eye field neurons signaling saccade targets during visual search. *J Neurosci*, 21:713-25.
- BIEDERMAN, I, MEZZANOTTE, RJ & RABINOWITZ, JC. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognit Psychol*, 14:143-77.
- BLAKEMORE, C & CAMPBELL, FW. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J Physiol*, 203:237-60.

- BOCH, R & FISCHER, B. (1986). Further observations on the occurrence of express-saccades in the monkey. *Exp Brain Res*, 63:487-494.
- BOCH, R, FISCHER, B & RAMSPERGER, E. (1984). Express-saccades of the monkey: reaction times versus intensity, size, duration, and eccentricity of their targets. *Exp Brain Res*, 55:223-31.
- BOEHM, SG & SOMMER, W. (2005). Neural correlates of intentional and incidental recognition of famous faces. *Brain Res Cogn Brain Res*, 23:153-63.
- BOLZ, J, ROSNER, G & WASSLE, H. (1982). Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. *J Physiol*, 328:171-90.
- BOTZEL, K, SCHULZE, S & STODIECK, SR. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res*, 104:135-43.
- BOUCART, M & HUMPHREYS, GW. (1997). Integration of physical and semantic information in object processing. *Perception*, 26:1197-209.
- BOUSSAOU, D, DI PELLEGRINO, G & WISE, SP. (1995). Frontal lobe mechanisms subserving vision-for-action versus vision-for-perception. *Behav Brain Res*, 72:1-15.
- BRAUN, D, WEBER, H, MERGNER, T & SCHULTE-MONTING, J. (1992). Saccadic reaction times in patients with frontal and parietal lesions. *Brain*, 115 (Pt 5):1359-86.
- BREITMEYER, BG. (1984). *Visual masking: an integrative approach (extracts)*. Oxford: Clarendon Press.
- BREITMEYER, BG & GANZ, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol Rev*, 83:1-36.
- BREITMEYER, BG & OGMEN, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. *Percept Psychophys*, 62:1572-95.
- BRENNER, E & SMEETS, JB. (2004). Colour vision can contribute to fast corrections of arm movements. *Exp Brain Res*, 158:302-7.
- BRIDGEMAN, B. (1978). Distributed sensory coding applied to simulations of iconic storage and metacontrast. *Bull Math Biol*, 40:605-23.
- BRONSTEIN, AM & KENNARD, C. (1987). Predictive eye saccades are different from visually triggered saccades. *Vision Res*, 27:517-20.
- BRUCE, CJ & GOLDBERG, ME. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol*, 53:603-35.
- BUCKLEY, MJ, GAFFAN, D & MURRAY, EA. (1997). Functional double dissociation between two inferior temporal cortical areas: perirhinal cortex versus middle temporal gyrus. *J Neurophysiol*, 77:587-98.
- BUGMANN, G & TAYLOR, JG. (2005). A model of visual backward masking. *Biosystems*, 79:151-8.
- BULLIER, J. (1998). Architecture fonctionnelle du système visuel. In *Vision : aspects perceptif et cognitifs*, ed. M Boucart, AM Henaff & C Belin: Solal, 1998, pp. 11-41.
- BULLIER, J. (2001a). Feedback connections and conscious vision. *Trends Cogn Sci*, 5:369-370.
- BULLIER, J. (2001b). Integrated model of visual processing. *Brain Res Brain Res Rev*, 36:96-107.
- BULLIER, J & NOWAK, LG. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol*, 5:497-503.
- BULLIER, J, SALIN, PA & GIRARD, P. (1992). Le cerveau en temps réel. *La Recherche*, 23:974-981.
- BUSETTINI, C, MASSON, GS & MILES, FA. (1997). Radial optic flow induces vergence eye movements with ultra-short latencies. *Nature*, 390:512-5.
- CALDER, AJ & YOUNG, AW. (2005). Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci*, 6:641-51.
- CAMPBELL, KC & PROCTOR, RW. (1993). Repetition effects with categorizable stimulus and response sets. *J Exp Psychol Learn Mem Cogn*, 19:1345-1362.
- CAMPBELL, KK. (1970). *Body and mind*. New York: Doubleday.
- CARMEL, D & BENTIN, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*, 83:1-29.
- CARPENTER, RH. (1999a). A neural mechanism that randomises behaviour. *J Consciousness Studies*, 6:13-22.
- CARPENTER, RH. (1999b). Visual selection: Neurons that make up their minds. *Curr Biol*, 9:R595-8.
- CARPENTER, RH & WILLIAMS, ML. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, 377:59-62.
- CATTELL, JM. (1885-1886). The inertia of the eye and brain. *Brain*, 8:295-312.
- CAVADA, C & GOLDMAN-RAKIC, PS. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J Comp Neurol*, 287:422-45.
- CAVEGN, D & BISCALDI, M. (1996). Fixation and saccade control in an express-saccade maker. *Exp Brain Res*, 109:101-16.
- CAVEGN, D & D'YDEWALLE, G. (1996). Presaccadic attention allocation and express saccades. *Psychol Res*, 59:157-75.

- CELEBRINI, S, THORPE, S, TROTTER, Y & IMBERT, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci*, 10:811-25.
- CHAVIS, DA & PANDYA, DN. (1976). Further observations on corticofrontal connections in the rhesus monkey. *Brain Res*, 117:369-86.
- CHUN, MM & POTTER, MC. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J Exp Psychol Hum Percept Perform*, 21:109-27.
- CISEK, P & KALASKA, JF. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, 45:801-14.
- CLARK, VP, FAN, S & HILLYARD, SA. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Hum Brain Mapp*, 2:170-187.
- COLTHEART, M. (1980). Iconic memory and visible persistence. *Percept Psychophys*, 27:183-228.
- COLTHEART, V, Editor (1999). *Fleeting memories - Cognition of brief visual stimuli*: MIT Press.
- COSTEN, NP, PARKER, DM & CRAW, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Percept Psychophys*, 58:602-12.
- CRICK, F & KOCH, C. (2003). A framework for consciousness. *Nat Neurosci*, 6:119-26.
- CRUTCHER, MD & ALEXANDER, GE. (1990). Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. *J Neurophysiol*, 64:151-63.
- CURRAN, T, TANAKA, JW & WEISKOPF, DM. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn Affect Behav Neurosci*, 2:1-18.
- DE VALOIS, RL & DE VALOIS, KK. (1980). Spatial vision. *Annu Rev Psychol*, 31:309-41.
- DEBRUILLE, JB, GUILLEM, F & RENAULT, B. (1998). ERPs and chronometry of face recognition: following-up Seeck et al. and George et al. *Neuroreport*, 9:3349-53.
- DECHARMS, RC & ZADOR, A. (2000). Neural representation and the cortical code. *Annu Rev Neurosci*, 23:613-47.
- DEHAENE, S & NACCACHE, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79:1-37.
- DEHAENE, S, NACCACHE, L, COHEN, L, BIHAN, DL, MANGIN, JF, POLINE, JB & RIVIERE, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci*, 4:752-8.
- DEHAENE, S, NACCACHE, L, LE CLEC, HG, KOEHLIN, E, MUELLER, M, DEHAENE-LAMBERTZ, G, VAN DE MOORTELE, PF & LE BIHAN, D. (1998). Imaging unconscious semantic priming. *Nature*, 395:597-600.
- DELORD, S. (1998). Which mask is the most efficient: A pattern or a noise? It depends on the task. *Visual Cognition*, 5:313-338.
- DELORME, A. (2000). *Traitement visuel rapide des scènes naturelles chez le singe, l'homme et la machine*, UPS. Unpublished.
- DELORME, A, RICHARD, G & FABRE THORPE, M. (1999). Rapid processing of complex natural scenes: a role for the magnocellular visual pathways? *Neurocomputing*, 26:663-670.
- DELORME, A, RICHARD, G & FABRE-THORPE, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res*, 40:2187-200.
- DELORME, A, ROUSSELET, GA, MACE, MJ & FABRE-THORPE, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res*, 19:103-13.
- DENNETT, D. (2001). Are we explaining consciousness yet? *Cognition*, 79:221-37.
- DESIMONE, R, ALBRIGHT, TD, GROSS, CG & BRUCE, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci*, 4:2051-62.
- DI LOLLO, V, ENNS, JT & RENSINK, RA. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J Exp Psychol Gen*, 129:481-507.
- DI RUSSO, F, MARTINEZ, A, SERENO, MI, PITZALIS, S & HILLYARD, SA. (2001). Cortical sources of the early components of the visual evoked potential. *Hum Brain Mapp*, 15:95-111.
- DIAMOND, R & CAREY, S. (1986). Why faces are and are not special: an effect of expertise. *J Exp Psychol Gen*, 115:107-17.
- DICARLO, JJ & MAUNSELL, JH. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *J Neurophysiol*, 93:2974-86.
- DINSE, HR & KRUGER, K. (1994). The timing of processing along the visual pathway in the cat. *Neuroreport*, 5:893-7.
- DONCHIN, E, WICKE, JD & LINDSLEY, DB. (1963). Cortical Evoked Potentials and Perception of Paired Flashes. *Science*, 141:1285-6.
- DONDERS, FC. (1969). On the speed of mental processes. *Acta Psychol (Amst)*, 30:412-31.
- DORRIS, MC, PARE, M & MUNOZ, DP. (1997). Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J Neurosci*, 17:8566-79.

- DORRIS, MC, PARE, M & MUNOZ, DP. (2000). Immediate neural plasticity shapes motor performance. *J Neurosci*, 20:RC52.
- DRIVER, J & VUILLEUMIER, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition*, 79:39-88.
- DUNCAN, J & HUMPHREYS, GW. (1989). Visual search and stimulus similarity. *Psychol Rev*, 96:433-458.
- EDELMAN, S. (1999). *Representation and recognition in vision*: MIT Press.
- EIMER, M. (1997). An event-related potential (ERP) study of transient and sustained visual attention to color and form. *Biol Psychol*, 44:143-60.
- EIMER, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *Neuroreport*, 9:2945-8.
- EIMER, M. (2000a). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology*, 17:103-116.
- EIMER, M. (2000b). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Res Cogn Brain Res*, 10:145-58.
- ENNS, JT. (2004). Object substitution and its relation to other forms of visual masking. *Vision Res*, 44:1321-31.
- ENNS, JT & DI LOLLO, V. (2000). What's new in visual masking? *Trends Cogn Sci*, 4:345-352.
- ERDMAN, B & DODGE, R. (1898). In *Visual information processing*, ed. KT Spoehr & SW Lehmkuhle. San Francisco: W.H. Freeman, 1898.
- ERIKSEN, BA & ERIKSEN, CW. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Percept Psychophys*, 16.
- ERIKSEN, CW & SCHULTZ, DW. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Percept Psychophys*, 25:249-63.
- EVANS, KK & TREISMAN, A. (2005). Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform*, 31:1476-92.
- FABRE-THORPE, M, DELORME, A, MARLOT, C & THORPE, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci*, 13:171-80.
- FABRE-THORPE, M, RICHARD, G & THORPE, SJ. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, 9:303-8.
- FARAH, MJ. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behav Brain Res*, 76:181-9.
- FARAH, MJ, WILSON, KD, DRAIN, HM & TANAKA, JR. (1995). The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. *Vision Res*, 35:2089-93.
- FARAH, MJ, WILSON, KD, DRAIN, M & TANAKA, JN. (1998). What is "special" about face perception? *Psychol Rev*, 105:482-98.
- FELLEMAN, DJ & VAN ESSEN, DC. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1:1-47.
- FINDLAY, JM & WALKER, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behav Brain Sci*, 22:661-74; discussion 674-721.
- FISCHER, B & BOCH, R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Res*, 260:21-6.
- FISCHER, B & RAMSPERGER, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Exp Brain Res*, 57:191-5.
- FISCHER, B & WEBER, H. (1992). Characteristics of "anti" saccades in man. *Exp Brain Res*, 89:415-24.
- FISCHER, B & WEBER, H. (1993). Express Saccades and Visual Attention. *Behav Brain Sci*, 16:553-567.
- FISER, J, SUBRAMANIAM, S & BIEDERMAN, I. (2001). Size tuning in the absence of spatial frequency tuning in object recognition. *Vision Res*, 41:1931-50.
- FITTS, PM. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol*, 47:381-91.
- FIZE, D, BOULANOUAR, K, CHATEL, Y, RANJEVA, JP, FABRE-THORPE, M & THORPE, S. (2000). Brain areas involved in rapid categorization of natural images: an event-related fMRI study. *Neuroimage*, 11:634-43.
- FIZE, D, FABRE-THORPE, M, RICHARD, G, DOYON, B & THORPE, SJ. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain Cogn*, accepted.
- FORSTER, KI. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Percept Psychophys*, 8:215-221.
- FOXE, JJ & SIMPSON, GV. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Exp Brain Res*, 142:139-50.

- FRANCIS, G & HERZOG, MH. (2004). Testing quantitative models of backward masking. *Psychon Bull Rev*, 11:104-12.
- FREEDMAN, DJ, RIESENHUBER, M, POGGIO, T & MILLER, EK. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291:312-6.
- GAUTHIER, I, CURRAN, T, CURBY, KM & COLLINS, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat Neurosci*, 6:428-32.
- GAUTHIER, I & LOGOTHETIS, NK. (2000). Is face recognition not so unique after all? *Cognitive Neuropsychology*, 17:125-142.
- GAUTHIER, I, SKUDLARSKI, P, GORE, JC & ANDERSON, AW. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, 3:191-7.
- GAUTHIER, I & TARR, MJ. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res*, 37:1673-82.
- GAUTRAIS, J & THORPE, S. (1998). Rate coding versus temporal order coding: a theoretical approach. *Biosystems*, 48:57-65.
- GAZZANIGA, MS, LEDOUX, JE & WILSON, DH. (1977). Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness. *Neurology*, 27:1144-7.
- GEMBA, H & SASAKI, K. (1989). Potential related to no-go reaction of go/no-go hand movement task with color discrimination in human. *Neurosci Lett*, 101:263-8.
- GEORGE, N. (1997). *Etude des bases neurales de la reconnaissance des visages: apport des potentiels évoqués*, Paris VI. Unpublished.
- GEORGE, N, JEMEL, B, FIORI, N & RENAULT, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport*, 8:1417-23.
- GLIMCHER, PW. (2001). Making choices: the neurophysiology of visual-saccadic decision making. *Trends Neurosci*, 24:654-9.
- GLIMCHER, PW. (2003). The Neurobiology of Visual-Saccadic Decision Making. *Annu Rev Neurosci*, 26:133-179.
- GOFFART, L, QUINET, J, CHAVANE, F & MASSON, GS. (2006). Influence of background illumination on fixation and visually guided saccades in the rhesus monkey. *Vision Res*, 46:149-162.
- GOODALE, MA & MILNER, AD. (1992). Separate visual pathways for perception and action. *Trends Neurosci*, 15:20-25.
- GORE, JL, DORRIS, MC & MUNOZ, DP. (2002). Time course of a repetition effect on saccadic reaction time in non-human primates. *Arch Ital Biol*, 140:203-10.
- GORNO-TEMPINI, ML, CIPOLOTTI, L & PRICE, CJ. (2000). Category differences in brain activation studies: where do they come from? *Proc R Soc Lond B Biol Sci*, 267:1253-8.
- GREEN, DM & SWETS, JA. (1966). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger Publishing Co.
- GREEN, MF, GLAHN, D, ENGEL, SA, NUECHTERLEIN, KH, SABB, F, STROJWAS, M & COHEN, MS. (2005). Regional brain activity associated with visual backward masking. *J Cogn Neurosci*, 17:13-23.
- GRILL-SPECTOR, K & KANWISHER, N. (2005). Visual recognition. *Psychol Sci*, 16:152-60.
- GRILL-SPECTOR, K, KUSHNIR, T, EDELMAN, S, ITZCHAK, Y & MALACH, R. (1998a). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, 21:191-202.
- GRILL-SPECTOR, K, KUSHNIR, T, HENDLER, T, EDELMAN, S, ITZCHAK, Y & MALACH, R. (1998b). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp*, 6:316-28.
- GRILL-SPECTOR, K, KUSHNIR, T, HENDLER, T & MALACH, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci*, 3:837-43.
- GROSS, CG. (1994). How inferior temporal cortex became a visual area. *Cereb Cortex*, 4:455-69.
- GROSS, CG, BENDER, DB & ROCHA-MIRANDA, CE. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, 166:1303-6.
- GROSS, CG, SCHILLER, PH, WELLS, C & GERSTEIN, GL. (1967). Single-unit activity in temporal association cortex of the monkey. *J Neurophysiol*, 30:833-43.
- GUYONNEAU, R. (2006). *Codage par latence et STDP : des stratégies temporelles pour expliquer le traitement visuel rapide*. Unpublished.
- GUYONNEAU, R, KIRCHNER, H & THORPE, SJ. (submitted). *J Vis*.
- GUYONNEAU, R, VANRULLEN, R & THORPE, SJ. (2005). Neurons Tune to the Earliest Spikes Through STDP. *Neural Comput*, 17:859-79.
- HALGREN, E, BAUDENA, P, HEIT, G, CLARKE, JM, MARINKOVIC, K & CLARKE, M. (1994). Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes [corrected]. *J Physiol Paris*, 88:1-50.

- HALGREN, E, RAIJ, T, MARINKOVIC, K, JOUSMAKI, V & HARI, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb Cortex*, 10:69-81.
- HANES, DP & SCHALL, JD. (1996). Neural control of voluntary movement initiation. *Science*, 274:427-30.
- HASBROUCQ, T, BURLE, B, BONNET, M, POSSAMAI, CA & VIDAL, F. (2002). [Treatment dynamics in sensorimotor disorders: the contribution of electrophysiology]. *Can J Exp Psychol*, 56:75-97.
- HE, SQ, DUM, RP & STRICK, PL. (1995). Topographic organization of corticospinal projections from the frontal lobe: motor areas on the medial surface of the hemisphere. *J Neurosci*, 15:3284-3306.
- HELLIGE, JB, WALSH, DA, LAWRENCE, VW & PRASSE, M. (1979). Figural relationship effects and mechanisms of visual masking. *J Exp Psychol Hum Percept Perform*, 5:88-100.
- HELMHOLTZ, HLF. (1853). Über die Methoden, kleinste Zeittheile zu messen, und ihre Anwendung für physiologische Zwecke. *Philosop Mag*, 6:313-325.
- HENDERSON, JM & HOLLINGWORTH, A. (1999). High-level scene perception. *Annu Rev Psychol*, 50:243-71.
- HERSHLER, O & HOCHSTEIN, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Res*, 45:1707-24.
- HICK, WE. (1952). On the rate of gain of information. *Q J Exp Psychol*, 4:11-26.
- HINO, Y & LUPKER, SJ. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *J Exp Psychol Hum Percept Perform*, 26:166-83.
- HUETTEL, SA, SONG, AW & MCCARTHY, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J Neurosci*, 25:3304-11.
- HUMPHREYS, GW, PRICE, CJ & RIDDOCH, MJ. (1999). From objects to names: a cognitive neuroscience approach. *Psychol Res*, 62:118-30.
- HUMPHREYS, GW, RIDDOCH, MJ & PRICE, CJ. (1997). Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. *Philos Trans R Soc Lond B Biol Sci*, 352:1275-82.
- HUPE, JM, JAMES, AC, PAYNE, BR, LOMBER, SG, GIRARD, P & BULLIER, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394:784-7.
- IRVIN, GE, NORTON, TT, SESMA, MA & CASAGRANDE, VA. (1986). W-like response properties of interlaminar zone cells in the lateral geniculate nucleus of a primate (*Galago crassicaudatus*). *Brain Res*, 362:254-70.
- ITIER, R. (2002). *Perception et reconnaissance des visages non familiers chez l'adulte et l'enfant: étude neurophysiologique du traitement de la configuration*, UPS. Unpublished.
- ITIER, RJ, LATINUS, M & TAYLOR, MJ. (2005). Face, eye and object early processing: What is the face specificity? *Neuroimage*.
- IWAI, E & MISHKIN, M. (1969). Further evidence on the locus of the visual area in the temporal lobe of the monkey. *Exp Neurol*, 25:585-94.
- JAMES, TW & GAUTHIER, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Res*, 44:429-39.
- JEFFREYS, DA & MUSSELWHITE, MJ. (1986). A visual evoked potential study of metacontrast masking. *Vision Res*, 26:631-42.
- JOHNSON, JS & OLSHAUSEN, BA. (2003). Timecourse of neural signatures of object recognition. *J Vis*, 3:499-512.
- JOHNSON, JS & OLSHAUSEN, BA. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J Vis*, 5:299-312.
- JOHNSON, KE & MERVIS, CB. (1997). Effects of varying levels of expertise on the basic level of categorization. *J Exp Psychol Gen*, 126:248-277.
- JUTTNER, M & WOLF, W. (1992). Occurrence of human express saccades depends on stimulus uncertainty and stimulus sequence. *Exp Brain Res*, 89:678-81.
- KAHNEMAN, D, TREISMAN, A & GIBBS, BJ. (1992). The reviewing of object files: object-specific integration of information. *Cognit Psychol*, 24:175-219.
- KAITZ, M, MONITZ, J & NESHER, R. (1985). Electrophysiological correlates of visual masking. *Int J Neurosci*, 28:261-8.
- KANWISHER, N. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, 27:117-143.
- KANWISHER, N. (2000). Domain specificity in face perception. *Nat Neurosci*, 3:759-63.
- KANWISHER, N. (2001). Neural events and perceptual awareness. *Cognition*, 79:89-113.
- KANWISHER, N, MCDERMOTT, J & CHUN, MM. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17:4302-11.
- KEYSERS, C. (2000). *The speed of sight : neural correlates of perceptual reports in RSVP*, University of St Andrews. Unpublished.

- KEYSERS, C & PERRETT, DI. (2002). Visual masking and RSVP reveal neural competition. *Trends Cogn Sci*, 6:120-125.
- KEYSERS, C, XIAO, DK, FOLDIAK, P & PERRETT, DI. (2001). The speed of sight. *J Cogn Neurosci*, 13:90-101.
- KIM, JN & SHADLEN, MN. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci*, 2:176-85.
- KIRCHNER, H & COLONIUS, H. (2005). Interstimulus contingency facilitates saccadic responses in a bimodal go/no-go task. *Brain Res Cogn Brain Res*, 25:261-72.
- KIRCHNER, H & THORPE, SJ. (2005). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res*, submitted.
- KOCH, C. (2004). *The Quest for Consciousness: a Neurobiological Approach*. Englewood (CO): Roberts and Company Publishers.
- KOLERS, PA. (1962). Intensity and contour effects in visual masking. *Vision Res*, 2:277-294.
- KOVACS, G, SARY, G, KOTELES, K, CHADAIDE, Z, TOMPA, T, VOGELS, R & BENEDEK, G. (2003). Effects of surface cues on macaque inferior temporal cortical responses. *Cereb Cortex*, 13:178-88.
- KOVACS, G, VOGELS, R & ORBAN, GA. (1995). Cortical correlate of pattern backward masking. *Proc Natl Acad Sci U S A*, 92:5587-91.
- KURATA, K & TANJI, J. (1986). Premotor cortex neurons in macaques: activity before distal and proximal forelimb movements. *J Neurosci*, 6:403-411.
- LACHTER, J & DURGIN, FH. (1999). Metacontrast Masking Functions: A Question of Speed? *J Exp Psychol Hum Percept Perform*, 25:936-947.
- LAMING, DRJ. (1968). *Information theory of choice reaction times*. London: Academic Press.
- LAMME, VA. (2000). Neural Mechanisms of Visual Awareness: A Linking Proposition. *Brain and Mind*, 1:385-406.
- LAMME, VA & ROELFSEMA, PR. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, 23:571-9.
- LAMME, VA, SUPER, H & SPEKREIJSE, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr Opin Neurobiol*, 8:529-35.
- LAVENEX, P, SUZUKI, WA & AMARAL, DG. (2002). Perirhinal and parahippocampal cortices of the macaque monkey: Projections to the neocortex. *J Comp Neurol*, 447:394-420.
- LAVIE, N, RO, T & RUSSELL, C. (2003). The role of perceptual load in processing distractor faces. *Psychol Sci*, 14:510-5.
- LEVICK, WR & ZACKS, JL. (1970). Responses of cat retinal ganglion cells to brief flashes of light. *J Physiol*, 206:677-700.
- LINKENKAER-HANSEN, K, PALVA, JM, SAMS, M, HIETANEN, JK, ARONEN, HJ & ILMONIEMI, RJ. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neurosci Lett*, 253:147-50.
- LIU, J, HARRIS, A & KANWISHER, N. (2002). Stages of processing in face perception: an MEG study. *Nat Neurosci*, 5:910-6.
- LOFTUS, GR & IRWIN, DE. (1998). On the relations among different measures of visible and informational persistence. *Cognit Psychol*, 35:135-99.
- LOGOTHETIS, NK & SHEINBERG, DL. (1996). Visual Object Recognition. *Annu Rev Neurosci*, 19:577-621.
- LU, MT, PRESTON, JB & STRICK, PL. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *J Comp Neurol*, 341:375-392.
- MACÉ, MJ. (2006). *Représentations visuelles précoces dans la catégorisation rapide de scènes naturelles chez l'homme et le singe.*, Toulouse III. Unpublished.
- MACÉ, MJ, BACON-MACE, N, NESPOULOUS, JL & FABRE-THORPE, M. (en préparation). What's seen first: The animal or the bird? *J Cogn Neurosci*.
- MACÉ, MJ, JOUBERT, OR & FABRE THORPE, M. (en préparation). Hierarchy of categories and diagnosticity.
- MACÉ, MJ, ROUSSELET, GA & FABRE THORPE, M. (en préparation). Early ERP differences: what do they reflect?
- MACÉ, MJ, THORPE, SJ & FABRE-THORPE, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *Eur J Neurosci*, 21:2007-18.
- MACKNIK, SL & LIVINGSTONE, MS. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nat Neurosci*, 1:144-9.
- MACMILLAN, NA. (2002). Signal detection theory. In *Steven's Handbook of Experimental Psychology*, ed. H Pashler. New York: John Wiley & Sons, Inc., 2002, pp. 43-90.
- MACMILLAN, NA & CREELMAN, CD. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

- MALACH, R, REPPAS, JB, BENSON, RR, KWONG, KK, JIANG, H, KENNEDY, WA, LEDDEN, PJ, BRADY, TJ, ROSEN, BR & TOOTELL, RB. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A*, 92:8135-9.
- MANGUN, GR & HILLYARD, SA. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J Exp Psychol Hum Percept Perform*, 17:1057-74.
- MARKRAM, H, LUBKE, J, FROTSCHER, M & SAKMANN, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213-5.
- MAROIS, R, YI, DJ & CHUN, MM. (2004). The Neural Fate of Consciously Perceived and Missed Events in the Attentional Blink. *Neuron*, 41:465-472.
- MASQUELIER, T & THORPE, SJ. (en préparation). Unsupervised STDP-based learning of intermediate complexity visual features.
- MASSON, GS, RYBARCZYK, Y, CASTET, E & MESTRE, DR. (2000). Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Vis Neurosci*, 17:753-67.
- MCCARTHY, G, PUCE, A, GORE, JC & ALLISON, T. (1997). Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci*, 9:605-610.
- MCCLELLAND, JL. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychol Rev*, 86:287-330.
- MCCLELLAND, JL & ROGERS, TT. (2003). The parallel distributed processing approach to semantic cognition. *Nat Rev Neurosci*, 4:310-322.
- MCGLINCHY-BERROTH. (1997). Visual information processing in hemispatial neglect. *Trends Cogn Sci*, 1:91-97.
- MCPEEK, RM & SCHILLER, PH. (1994). The effects of visual scene composition on the latency of saccadic eye movements of the rhesus monkey. *Vision Res*, 34:2293-305.
- MERIKLE, PM, SMILEK, D & EASTWOOD, JD. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition*, 79:115-34.
- MEYER, DE, OSMAN, AM, IRWIN, DE & YANTIS, S. (1988). Modern mental chronometry. *Biol Psychol*, 26:3-67.
- MICHAELS, CF & TURVEY, MT. (1979). Central sources of visual masking: indexing structures supporting seeing at a single, brief glance. *Psychol Res*, 41:2-61.
- MILLER, J. (1982). Discrete versus continuous stage models of human information processing: in search of partial output. *J Exp Psychol Hum Percept Perform*, 8:273-96.
- MILLER, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychol (Amst)*, 67:191-257.
- MILNER, AD. (1998). Streams and consciousness: visual awareness and the brain. *Trends Cogn Sci*, 2:25-30.
- MILNER, AD, PERRETT, DI, JOHNSTON, RS, BENSON, PJ, JORDAN, TR, HEELEY, DW, BETTUCCI, D, MORTARA, F, MUTANI, R, TERAZZI, E & ET AL. (1991). Perception and action in 'visual form agnosia'. *Brain*, 114:405-28.
- MITOV, D, VASSILEV, A & MANAHILOV, V. (1981). Transient and sustained masking. *Percept Psychophys*, 30:205-10.
- MIYASHITA, Y. (1993). Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci*, 16:245-63.
- MORRIS, JS, OHMAN, A & DOLAN, RJ. (1998). Conscious and unconscious emotional learning in the human amygdala [see comments]. *Nature*, 393:467-70.
- MORRISON, DJ & SCHYNS, PG. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon Bull Rev*, 8:454-69.
- MOSCOVITCH, M, WINOCUR, G & BEHRMANN, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *J Cogn Neurosci*, 9:555-604.
- MOUCHETANT-ROSTAING, Y. (2001). *Bases neurophysiologiques des processus impliqués dans le traitement visuel des stimuli à forte connotation psychosociale chez l'homme (Visages et Mots)*, Lyon II. Unpublished.
- MOUCHETANT-ROSTAING, Y, GIARD, MH, BENTIN, S, AGUERA, PE & PERNIER, J. (2000a). Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci*, 12:303-10.
- MOUCHETANT-ROSTAING, Y, GIARD, MH, DELPUECH, C, ECHALLIER, JF & PERNIER, J. (2000b). Early signs of visual categorization for biological and non-biological stimuli in humans. *Neuroreport*, 11:2521-5.

- MOUNTCASTLE, VB, LYNCH, JC, GEORGOPOULOS, A, SAKATA, H & ACUNA, C. (1975). Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J Neurophysiol*, 38:871-908.
- MOURET, I & HASBROUCQ, T. (2000). The chronometry of single neuron activity: testing discrete and continuous models of information processing. *J Exp Psychol Hum Percept Perform*, 26:1622-38.
- MULLER, JR, METHA, AB, KRAUSKOPF, J & LENNIE, P. (2001). Information conveyed by onset transients in responses of striate cortical neurons. *J Neurosci*, 21:6978-90.
- MUMMERY, CJ, PATTERSON, K, HODGES, JR & PRICE, CJ. (1998). Functional neuroanatomy of the semantic system: divisible by what? *J Cogn Neurosci*, 10:766-77.
- MUNOZ, DP & WURTZ, RH. (1993a). Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *J Neurophysiol*, 70:559-75.
- MUNOZ, DP & WURTZ, RH. (1993b). Fixation cells in monkey superior colliculus. II. Reversible activation and deactivation. *J Neurophysiol*, 70:576-89.
- MURPHY, GL & SMITH, EE. (1982). Basic level superiority in picture categorization. *J Verbal Learning & Verbal Behavior*, 21:1-20.
- MURRAY, EA & RICHMOND, BJ. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Curr Opin Neurobiol*, 11:188-93.
- NASANEN, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Res*, 39:3824-33.
- NELSON, SB. (1991). Temporal interactions in the cat visual system. II. Suppressive and facilitatory effects in the lateral geniculate nucleus. *J Neurosci*, 11:357-68.
- NEWSOME, WT, BRITTEN, KH & MOVSHON, JA. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341:52-4.
- NOSOFSKY, RM. (1986). Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen*, 115:39-61.
- NOTEBAERT, W & SOETENS, E. (2003). The influence of irrelevant stimulus changes on stimulus and response repetition effects. *Acta Psychol (Amst)*, 112:143-56.
- NOWAK, LG & BULLIER, J. (1997). The timing of information transfer in the visual system. In *Extrastriate visual cortex in primates*, ed. KS Rockland, JH Kaas & A Peters. New York: Plenum Press, 1997, pp. 205-241.
- NOWAK, LG, MUNK, MH, GIRARD, P & BULLIER, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Vis Neurosci*, 12:371-84.
- OGMEN, H. (1993). A neural theory of retino-cortical dynamics. *Neural Netw*, 6.
- OGMEN, H, BREITMEYER, BG & MELVIN, R. (2003). The what and where in visual masking. *Vision Res*, 43:1337-50.
- PALMERI, TJ & GAUTHIER, I. (2004). Visual object understanding. *Nat Rev Neurosci*, 5:291-303.
- PARE, M & MUNOZ, DP. (1996). Saccadic reaction time in the monkey: advanced preparation of oculomotor programs is primarily responsible for express saccade occurrence. *J Neurophysiol*, 76:3666-81.
- PEREA, M, ROSA, E & GOMEZ, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Mem Cognit*, 30:34-45.
- PERRETT, DI, ORAM, MW & ASHBRIDGE, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, 67:111-45.
- PERRETT, DI, ROLLS, ET & CAAN, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res*, 47:329-42.
- PINS, D, MEYER, ME, FOUCHER, J, HUMPHREYS, G & BOUCART, M. (2004). Neural correlates of implicit object identification. *Neuropsychologia*, 42:1247-59.
- POPPEL, E, HELD, R & FROST, D. (1973). Letter: Residual visual function after brain wounds involving the central visual pathways in man. *Nature*, 243:295-6.
- POSNER, MI. (1986). *Chronometric explorations of mind*: Oxford University Press.
- POTTER, MC. (1975). Meaning in visual search. *Science*, 187:965-6.
- POTTER, MC. (1976). Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn]*, 2:509-22.
- POTTER, MC & LEVY, EI. (1969). Recognition memory for a rapid sequence of pictures. *J Exp Psychol*, 81:10-5.
- QUIAN QUIROGA, R, REDDY, L, KREIMAN, G, KOCH, C & FRIED, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435:1102-7.
- RABBITT, PM. (1968). Repetition effects and signal classification strategies in serial choice-response tasks. *Q J Exp Psychol*, 20:232-40.
- RAIGUEL, SE, LAGAE, L, GULYAS, B & ORBAN, GA. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain Res*, 493:155-9.

- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychol Rev*, 85:59-108.
- RATCLIFF, R. (1988). Continuous versus discrete information processing modeling accumulation of partial information. *Psychol Rev*, 95:238-55.
- RATCLIFF, R, CARPENTER, RH & REDDI, BA. (2001). Putting noise into neurophysiological models of simple decision making. *Nat Neurosci*, 4:336-7.
- RATCLIFF, R & ROUDER, JF. (1998). Modeling response times for two-choice decisions. *Psychol Sci*, 9:347-356.
- RATCLIFF, R & SMITH, PL. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev*, 111:333-367.
- RAYMOND, JE, SHAPIRO, KL & ARNELL, KM. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform*, 18:849-60.
- RAYNER, K & POLLATSEK, A. (1987). Eye movements in reading: a tutorial review. In *Attention and performance- The psychology of reading*, ed. M Coltheart. London: Lawrence Erlbaum Associates, 1987, pp. 327-362.
- REDDI, BA. (2001). Decision making: the two stages of neuronal judgement. *Curr Biol*, 11:R603-6.
- REDDY, L, WILKEN, P & KOCH, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *J Vis*, 4:106-17.
- REES, G. (2001). Neuroimaging of visual awareness in patients and normal subjects. *Curr Opin Neurobiol*, 11:150-6.
- REES, G, KREIMAN, G & KOCH, C. (2002). Neural correlates of consciousness in humans. *Nat Rev Neurosci*, 3:261-70.
- RENAULT, B, SIGNORET, JL, DEBRUILLE, B, BRETON, F & BOLGERT, F. (1989). Brain potentials reveal covert facial recognition in prosopagnosia. *Neuropsychologia*, 27:905-12.
- REUTER-LORENZ, PA & FENDRICH, R. (1992). Oculomotor readiness and covert orienting: differences between central and peripheral precues. *Percept Psychophys*, 52:336-44.
- REYNOLDS, JH & DESIMONE, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24:19-29, 111-25.
- REYNOLDS, JH & DESIMONE, R. (2003). Interacting roles of attention and visual salience in v4. *Neuron*, 37:853-63.
- RICHMOND, BJ, OPTICAN, LM, PODELL, M & SPITZER, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol*, 57:132-46.
- RICHMOND, BJ, WURTZ, RH & SATO, T. (1983). Visual responses of inferior temporal neurons in awake rhesus monkey. *J Neurophysiol*, 50:1415-32.
- RIEGER, JW, BRAUN, C, BULTHOFF, HH & GEGENFURTNER, KR. (2005). The dynamics of visual pattern masking in natural scene processing: a magnetoencephalography study. *J Vis*, 5:275-86.
- RIESENHUBER, M & POGGIO, T. (2000). Models of object recognition. *Nat Neurosci*, 3 Suppl:1199-204.
- RO, T, RUSSELL, C & LAVIE, N. (2001). Changing faces: a detection advantage in the flicker paradigm. *Psychol Sci*, 12:94-9.
- ROLLS, ET. (2004). Consciousness absent and present: a neurophysiological exploration. *Prog Brain Res*, 144:95-106.
- ROLLS, ET, JUDGE, SJ & SANGHERA, MK. (1977). Activity of neurones in the inferotemporal cortex of the alert monkey. *Brain Res*, 130:229-38.
- ROLLS, ET & TOVEE, MJ. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B Biol Sci*, 257:9-15.
- ROLLS, ET, TOVEE, MJ & PANZERI, S. (1999). The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci*, 11:300-11.
- ROLLS, ET, TOVEE, MJ, PURCELL, DG, STEWART, AL & AZZOPARDI, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp Brain Res*, 101:473-84.
- ROLLS, ET, TREVES, A & TOVEE, MJ. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp Brain Res*, 114:149-62.
- ROMO, R & SALINAS, E. (2003). COGNITIVE NEUROSCIENCE: Flutter Discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci*, 4:203-18.
- ROSCH, E, MERVIS, CB, GRAY, WD, JOHNSON, DM & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognit Psychol*, 8:382-439.
- ROSS, LE & ROSS, SM. (1980). Saccade latency and warning signals: stimulus onset, offset, and change as warning events. *Percept Psychophys*, 27:251-7.
- ROSS, SM & ROSS, LE. (1981). Saccade latency and warning signals: effects of auditory and visual stimulus onset and offset. *Percept Psychophys*, 29:429-37.

- ROSSETTI, Y. (1998). Implicit short-lived motor representations of space in brain damaged and healthy subjects. *Conscious Cogn*, 7:520-58.
- ROSSION, B, CURRAN, T & GAUTHIER, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition*, 85:189-96.
- ROSSION, B, GAUTHIER, I, GOFFAUX, V, TARR, MJ & CROMMELINCK, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol Sci*, 13:250-7.
- ROSSION, B, GAUTHIER, I, TARR, MJ, DESPLAND, P, BRUYER, R, LINOTTE, S & CROMMELINCK, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11:69-74.
- ROUSSELET, GA, FABRE-THORPE, M & THORPE, SJ. (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci*, 5:629-30.
- ROUSSELET, GA, MACE, MJ & FABRE-THORPE, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *J Vis*, 4:13-21.
- ROUSSELET, GA, MACE, MJM & FABRE THORPE, M. (2004). Comparing animal and face processing in the context of natural scenes using a fast categorization task. *Neurocomputing*, 58-60:783-791.
- ROUSSELET, GA, MACKNIK, SL, THORPE, SJ & FABRE-THORPE, M. (submitted). Temporal course of ERP in fast categorization in natural scenes: a story more complicated than expected? *J Vis*.
- ROUSSELET, GA, THORPE, SJ & FABRE-THORPE, M. (2004a). How parallel is visual processing in the ventral pathway? *Trends Cogn Sci*, 8:363-370.
- ROUSSELET, GA, THORPE, SJ & FABRE-THORPE, M. (2004b). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res*, 44:877-94.
- RUMELHART, DE, HINTON, GE & MCCLELLAND, JL. (1986). *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. Cambridge, MA, USA: MIT Press.
- SAGIV, N & BENTIN, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *J Cogn Neurosci*, 13:937-51.
- SALINAS, E & ROMO, R. (1998). Neuronal representations in a categorization task: sensory to motor transformation. In *Computational Neuroscience: Trends in Research 1998*, ed. J Bower. New York: Plenum, 1998, pp. 599-604.
- SALZMAN, CD, BRITTEN, KH & NEWSOME, WT. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174-7.
- SANTHI, N & REEVES, A. (2004). The roles of distractor noise and target certainty in search: a signal detection model. *Vision Res*, 44:1235-56.
- SASAKI, K, GEMBA, H, NAMBU, A & MATSUZAKI, R. (1993). No-go activity in the frontal association cortex of human subjects. *Neurosci Res*, 18:249-52.
- SCHACTER, DL, BUCKNER, RL & KOUTSTAAL, W. (1998). Memory, consciousness and neuroimaging. *Philos Trans R Soc Lond B Biol Sci*, 353:1861-78.
- SCHALL, JD. (1995). Neural basis of saccade target selection. *Rev Neurosci*, 6:63-85.
- SCHALL, JD. (2001). Neural basis of deciding, choosing and acting. *Nat Rev Neurosci*, 2:33-42.
- SCHALL, JD. (2002). The neural selection and control of saccades by the frontal eye field. *Philos Trans R Soc Lond B Biol Sci*, 357:1073-82.
- SCHALL, JD. (2003). Neural correlates of decision processes: neural and mental chronometry. *Curr Opin Neurobiol*, 13:182-6.
- SCHALL, JD & BICHOT, NP. (1998). Neural correlates of visual and motor decision processes. *Curr Opin Neurobiol*, 8:211-7.
- SCHALL, JD, MOREL, A, KING, DJ & BULLIER, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *J Neurosci*, 15:4464-87.
- SCHENDAN, HE, GANIS, G & KUTAS, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology*, 35:240-51.
- SCHILLER, PH. (1965). Metacontrast interference as determined by a method of comparison. *Percept and Mot Skills*, 20:279-285.
- SCHILLER, PH. (1968). Single unit analysis of backward visual masking and metacontrast in the cat lateral geniculate nucleus. *Vision Res*, 8:855-66.
- SCHILLER, PH & CHOROVER, SL. (1966). Metacontrast: its relation to evoked potentials. *Science*, 153:1398-400.
- SCHILLER, PH, SANDELL, JH & MAUNSELL, JH. (1987). The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *J Neurophysiol*, 57:1033-49.

- SCHILLER, PH, SLOCUM, WM, CARVEY, C & TOLIAS, AS. (2004). Are express saccades generated under natural viewing conditions? *Eur J Neurosci*, 20:2467-73.
- SCHMOLESKY, MT, WANG, Y, HANES, DP, THOMPSON, KG, LEUTGEB, S, SCHALL, JD & LEVENTHAL, AG. (1998). Signal timing across the macaque visual system. *J Neurophysiol*, 79:3272-8.
- SCHROEDER, CE, MEHTA, AD & GIVRE, SJ. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb Cortex*, 8:575-92.
- SCHWARTZ, M & PRITCHARD, WS. (1981). AERs and detection in tasks yielding U-shaped backward masking functions. *Psychophysiology*, 18:678-85.
- SCHYNS, PG. (1997). Categories and percepts: A bi-directional framework for categorization. *Trends Cogn Sci*, 1:183-189.
- SCHYNS, PG. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, 67:147-79.
- SCHYNS, PG, BONNAR, L & GOSSELIN, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol Sci*, 13:402-9.
- SCHYNS, PG & OLIVA, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69:243-65.
- SEECK, M, MICHEL, CM, MAINWARING, N, COSGROVE, R, BLUME, H, IVES, J, LANDIS, T & SCHOMER, DL. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport*, 8:2749-54.
- SEGRAVES, MA. (1992). Activity of monkey frontal eye field neurons projecting to oculomotor regions of the pons. *J Neurophysiol*, 68:1967-85.
- SERGEANT, C, BAILLET, S & DEHAENE, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci*, 8:1391-400.
- SERGEANT, C & DEHAENE, S. (2004). Neural processes underlying conscious perception: Experimental findings and a global neuronal workspace framework. *J Physiol Paris*, 98:374-384.
- SERGEANT, J & PONCET, M. (1990). From covert to overt recognition of faces in a prosopagnosic patient. *Brain*, 113:989-1004.
- SERGEANT, J & SIGNORET, JL. (1992). Functional and anatomical decomposition of face processing: evidence from prosopagnosia and PET study of normal subjects. *Philos Trans R Soc Lond B Biol Sci*, 335:55-61; discussion 61-2.
- SESTOKAS, AK & LEHMKUHLE, S. (1986). Visual response latency of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. *Vision Res*, 26:1041-54.
- SHADLEN, MN, BRITTEN, KH, NEWSOME, WT & MOVSHON, JA. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci*, 16:1486-510.
- SHADLEN, MN & NEWSOME, WT. (1996). Motion perception: seeing and deciding. *Proc Natl Acad Sci U S A*, 93:628-33.
- SHADLEN, MN & NEWSOME, WT. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol*, 86:1916-36.
- SHEINBERG, DL & LOGOTHETIS, NK. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci*, 21:1340-50.
- SIKALUK, PD, BUCHANAN, L & WESTBURY, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Mem Cognit*, 31:100-13.
- SMITH, PL & RATCLIFF, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci*, 27:161-8.
- SOMMER, MA & WURTZ, RH. (2001). Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. *J Neurophysiol*, 85:1673-85.
- SPENCER, TJ & SHUNTICH, R. (1970). Evidence for an interruption theory of backward masking. *J Exp Psychol*, 85:198-203.
- SPELTING, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74:1-29.
- STERNBERG, S. (1969). The discovery of processing stages: extensions of Donders' method. In *Attention and performance II*, ed. WG Koster. Amsterdam: North Holland, 1969, pp. 276-315.
- STRATFORD, KJ, TARCZY-HORNOCH, K, MARTIN, KA, BANNISTER, NJ & JACK, JJ. (1996). Excitatory synaptic inputs to spiny stellate cells in cat visual cortex. *Nature*, 382:258-61.
- SUBRAMANIAM, S, BIEDERMAN, I & MADIGAN, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cogn*, 7:511-535.
- SUGASE, Y, YAMANE, S, UENO, S & KAWANO, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400:869-73.
- SUGRUE, LP, CORRADO, GS & NEWSOME, WT. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat Rev Neurosci*, 6:363-75.

- TANAKA, JW. (2001). The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen*, 130:534-43.
- TANAKA, JW & CURRAN, T. (2001). A neural basis for expert object recognition. *Psychol Sci*, 12:43-7.
- TANAKA, K. (1996). Representation of Visual Features of Objects in the Inferotemporal Cortex. *Neural Netw*, 9:1459-1475.
- TANAKA, K. (2003). Columns for Complex Visual Object Features in the Inferotemporal Cortex: Clustering of Cells with Similar but Slightly Different Stimulus Selectivities. *Cereb Cortex*, 13:90-99.
- TANAKA, K, SAITO, H, FUKADA, Y & MORIYA, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol*, 66:170-89.
- TARR, MJ & CHENG, YD. (2003). Learning to see faces and objects. *Trends Cogn Sci*, 7:23-30.
- TARR, MJ & GAUTHIER, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci*, 3:764-9.
- THOMPSON, KG, BICHOT, NP & SCHALL, JD. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol*, 77:1046-50.
- THOMPSON, KG, HANES, DP, BICHOT, NP & SCHALL, JD. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *J Neurophysiol*, 76:4040-55.
- THORPE, S. (1994). Les mécanismes de prise d'information - Partie 1 : Modèles de traitement de l'information. In *Neuropsychologie humaine*, 1994, pp. 153-178.
- THORPE, S, DELORME, A & VAN RULLEN, R. (2001). Spike-based strategies for rapid processing. *Neural Netw*, 14:715-25.
- THORPE, S, FIZE, D & MARLOT, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520-2.
- THORPE, S & IMBERT, M. (1989). Biological constraints on connectionist modelling. In *Connectionism in perspective*, ed. RPe al.: Elsevier, 1989, pp. 63-93.
- THORPE, SJ & FABRE-THORPE, M. (2001). Neuroscience. Seeking categories in the brain. *Science*, 291:260-3.
- THORPE, SJ, GEGENFURTNER, KR, FABRE-THORPE, M & BULTHOFF, HH. (2001). Detection of animals in natural images using far peripheral vision. *Eur J Neurosci*, 14:869-76.
- TIPPER, SP & CRANSTON, M. (1985). Selective attention and priming: inhibitory and facilitatory effects of ignored primes. *Q J Exp Psychol A*, 37:591-611.
- TIPPER, SP & DRIVER, J. (1988). Negative priming between pictures and words in a selective attention task: evidence for semantic processing of ignored stimuli. *Mem Cognit*, 16:64-70.
- TONG, F, NAKAYAMA, K, VAUGHAN, JT & KANWISHER, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21:753-9.
- TORRALBA, A. (2003). Contextual priming for object detection. *Int J Comput Vision*, 53:169-191.
- TORRALBA, A & OLIVA, A. (2003). Statistics of natural image categories. *Network*, 14:391-412.
- TOVÉE, MJ & ROLLS, ET. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cogn*, 2:35-58.
- TOVÉE, MJ, ROLLS, ET, TREVES, A & BELLIS, RP. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol*, 70:640-54.
- TREISMAN, A. (1991). Search, similarity, and integration of features between and within dimensions. *J Exp Psychol Hum Percept Perform*, 17:652-76.
- TSAO, DY, FREIWALD, WA, TOOTELL, RB & LIVINGSTONE, MS. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311:670-4.
- TURVEY, MT. (1973). On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli. *Psychol Rev*, 80:1-52.
- ULLMAN, S, VIDAL-NAQUET, M & SALI, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci*, 5:682-7.
- UNGERLEIDER, LG. (1985). The corticocortical pathways for object recognition and spatial perception. In *Pattern recognition mechanisms*, ed. C Chagas, R Gattass & C Gross. Vatican City: Pontifical Academy of Sciences, 1985, pp. 21-37.
- UNGERLEIDER, LG, GAFFAN, D & PELAK, VS. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp Brain Res*, 76:473-84.
- UNGERLEIDER, LG & HAXBY, JV. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol*, 4:157-65.
- UTTAL, WR. (1969). The character in the hole experiment: Interaction of forward and backward masking of alphabetic character recognition by dynamic visual noise (DVN). *Percept Psychophys*, 6:177-181.
- VALENTINE, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *Br J Psychol*, 79 (Pt 4):471-91.

- VANNI, S, DOJAT, M, WARNKING, J, DELON-MARTIN, C, SEGEBARTH, C & BULLIER, J. (2004a). Timing of interactions across the visual field in the human cortex. *Neuroimage*, 21:818-28.
- VANNI, S, REVONSUO, A, SAARINEN, J & HARI, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. *Neuroreport*, 8:183-6.
- VANNI, S, WARNKING, J, DOJAT, M, DELON-MARTIN, C, BULLIER, J & SEGEBARTH, C. (2004b). Sequence of pattern onset responses in the human visual areas: an fMRI constrained VEP source analysis. *Neuroimage*, 21:801-17.
- VANRULLEN, R. (2000). *Une première vague de potentiels d'action, une première vague idée de la scène visuelle*, UPS. Unpublished.
- VANRULLEN, R, GAUTRAIS, J, DELORME, A & THORPE, S. (1998). Face processing using one spike per neurone. *Biosystems*, 48:229-39.
- VANRULLEN, R & KOCH, C. (2003a). Competition and selection during visual processing of natural scenes and objects. *J Vis*, 3:75-85.
- VANRULLEN, R & KOCH, C. (2003b). Visual selective behavior can be triggered by a feed-forward process. *J Cogn Neurosci*, 15:209-17.
- VANRULLEN, R & THORPE, SJ. (2001a). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput*, 13:1255-83.
- VANRULLEN, R & THORPE, SJ. (2001b). The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci*, 13:454-61.
- VANRULLEN, R & THORPE, SJ. (2002). Surfing a spike wave down the ventral stream. *Vision Res*, 42:2593-615.
- VAUGHAN, HG, JR. & SILVERSTEIN, L. (1968). Metacontrast and evoked potentials: a reappraisal. *Science*, 160:207-8.
- VOGELS, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur J Neurosci*, 11:1239-55.
- VOGELS, R. (1999b). Effect of image scrambling on inferior temporal cortical responses. *Neuroreport*, 10:1811-6.
- VOGELS, R & ORBAN, GA. (1994). Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J Neurophysiol*, 71:1428-51.
- VON BONIN, G. (1947). The frontal lobe of primates: cytoarchitectural studies, 1947, pp. 67-83 (Ch. 3).
- WATSON, RT, VALENSTEIN, E, DAY, A & HEILMAN, KM. (1994). Posterior neocortical systems subserving awareness and neglect. Neglect associated with superior temporal sulcus but not area 7 lesions. *Arch Neurol*, 51:1014-21.
- WEBER, H, AIPLE, F, FISCHER, B & LATANOV, A. (1992). Dead zone for express saccades. *Exp Brain Res*, 89:214-22.
- WEBER, H & FISCHER, B. (1994). Differential effects of non-target stimuli on the occurrence of express saccades in man. *Vision Res*, 34:1883-91.
- WEBSTER, MJ, BACHEVALIER, J & UNGERLEIDER, LG. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb Cortex*, 4:470-83.
- WEISKRANTZ, L. (1996). Blindsight revisited. *Curr Opin Neurobiol*, 6:215-20.
- WEISSTEIN, N. (1972). Metacontrast. In *Handbook of sensory physiology: Visual Psychophysics*, ed. DJLM Hurvich. Berlin: Springer-Verlag, 1972, pp. 233-272.
- WHALEN, PJ, RAUCH, SL, ETCOFF, NL, MCINERNEY, SC, LEE, MB & JENIKE, MA. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J Neurosci*, 18:411-8.
- WHITE, IM & WISE, SP. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res*, 126:315-35.
- WOODWORTH, RS. (1899). The accuracy of voluntary movement. *Psychol Rev*, 3:1-119.
- YIN, RK. (1969). Looking at upside-down faces. *J Exp Psychol*, 81:141-145.
- YOUNG, MP & 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*, ed. T Ono, L Squire, DI Perrett & M Fukuda. New York: Oxford Univ. Press, 1993, pp. 47-70.

NOTE DU TRADUCTEUR

VERSION ANGLAISE	VERSION FRANÇAISE
Attentional blink	Cécité attentionnelle
Attentional focus	Focalisation attentionnelle
Backward	Rétrograde
Blindsight	Vision aveugle
Change blindness	Cécité au changement
Feedforward	Qui va de l'avant / antérograde
Feedback	En retour / rétrograde
Forward	Antérograde
Gap interval	(Protocole expérimental particulier dans lequel on introduit un) Intervalle entre deux stimuli (la croix de fixation et la cible)
Priming	Amorçage
Oddball	Recherche de l'élément différent
Repetition blindness	Cécité à la répétition
(Response) Missed	Manquée (réponse incorrecte sur un essai cible)
(Response) False Alarm	Fausse détection (réponse incorrecte sur un essai distracteur)
(Response) Hit	Correcte (réponse correcte sur un essai cible)
(Response) Correctly rejected	Correctement rejetée (réponse correcte sur un essai distracteur)
RSVP (Rapid Serial Visual Presentation)	Présentation visuelle rapide d'une série (d'images)
Scrambled pictures	Images brouillées ou Images puzzle
SOA (Stimulus Onset Asynchrony)	Intervalle de temps après le début de l'affichage du stimulus
Speed-accuracy trade-off	Compromis vitesse-précision
STDP (Spike Timing-Dependant Plasticity)	Plasticité à modulation temporelle relative
Split-brain	Cerveau dédoublé

La version en gras est celle qui a été choisie dans le texte

Résumé

Contraintes temporelles des traitements visuels dans une tâche de catégorisation de scènes naturelles

L'ensemble des résultats de cette thèse a été obtenu grâce à une tâche de catégorisation dans laquelle les sujets devaient effectuer une réponse motrice rapide sur la base d'informations visuelles incomplètes. L'utilisation d'un protocole de masquage a permis de visualiser au cours du temps le déroulement des traitements impliqués dans la reconnaissance d'objets. Les résultats obtenus confirment non seulement que les traitements visuels sont extrêmement robustes et rapides (Chapitre 2), mais aussi que les informations accumulées dans le système sensoriel peuvent être transformées en signal décisionnel pour guider la réponse du sujet de manière efficace, en fonction du type d'effecteur utilisé (manuel ou oculaire - Chapitre 3) et du niveau de représentation requis dans la tâche (détection, catégorisation, identification - Chapitre 4). 40 à 50 millisecondes de SOA semblent être suffisantes pour extraire la majorité de l'information pertinente dans la tâche de catégorisation animal/non-animal, un délai qui est encore plus court lorsqu'il s'agit de catégoriser des visages. Cette latence est tout à fait compatible avec l'idée que la reconnaissance d'objet s'appuie d'abord sur un transfert rapide, feed-forward et massivement parallèle des premières informations tout au long du parcours cortical visuel.

Summary

Temporal constraints of visual processing in a categorisation task of natural scenes

The different experiments of this thesis focused on a categorisation task in which subjects have to initiate response on the basis of incomplete visual information. The use of a masking paradigm allowed us to decompose through time the visual processing implicated in object recognition. Results of these experiments not only confirmed that visual processing is extremely robust and fast (Chapter 2), but also demonstrated that the accumulated sensory information could be transformed into a decisional signal to efficiently guide the response, depending on the motor effector (manual or ocular – Chapter 3) or the level of representation required in the task (Detection – Categorisation – Identification – Chapter 4). About 40-50 ms appear to be sufficient to integrate the most pertinent information on the stimulus to perform an animal/non-animal task and this latency can be even shorter while categorising faces of famous people. This delay is compatible with the idea that object recognition is initially based on the rapid transfer of visual information through the visual system, in a feed-forward and massively parallel way.