UNIVERSITE CATHOLIQUE DE LOUVAIN
Faculté de Psychologie et des Sciences de l'éducation
Unité Cognition & Développement

DECOURS TEMPOREL DE LA PERCEPTION VISUELLE DES VISAGES : DE LA CATEGORISATION FACIALE A L'ENCODAGE D'UNE REPRESENTATION INDIVIDUELLE

Corentin Jacques

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Psychologiques

Promoteurs: Professeur Bruno Rossion
Professeur Marc Crommelinck

Décembre 2007

Me voici enfin au moment d'écrire les remerciements, cette partie d'une thèse habituellement parcourue de tous (même si certains s'arrêtent déjà là), et pour laquelle on est généralement pas convaincu d'avoir exprimé avec justesse ce que l'on voulait dire…

Je vais donc faire de mon mieux pour remercier ici toutes les personnes qui ont contribué à l'achèvement de ce travail.

Je souhaiterais tout d'abord adresser mes plus vifs remerciements à mes promoteurs, Bruno Rossion et Marc Crommelinck pour l'opportunité qu'ils m'ont offerte de pouvoir emprunter les chemins aventureux d'une thèse. Bruno, je tiens à te remercier de m'avoir fait confiance dès le départ, lorsqu'il y a cinq ans tu m'as pris comme doctorant alors que tu m'avais à peine rencontré. Ta vivacité intellectuelle, ta créativité et ton enthousiasme pour la recherche ont largement contribué à mon travail de recherche et à ma formation. Je pense que ta capacité à te couper en 10 est représentative des habiletés particulières que tu as su développer pour pouvoir conjuguer ta nature généreuse et ton ambition scientifique.

Je tiens également à remercier les Professeurs Xavier Seron, Jean-Michel Guérit, Rufin VanRullen et David Perrett d'avoir accepté d'être membres de mon jury de thèse. De façon plus particulière, les contributions scientifiques de David Perrett et Rufin VanRullen ont toujours été une source d'inspiration pour mes propres travaux.

Je ne rendrais pas à César ce qui lui appartient si je ne remerciais pas Valérie Goffaux qui m'a donné accès au monde de la recherche scientifique. Je lui suis reconnaissant non seulement pour ses qualités de chercheuse mais également pour ses qualités humaines telles que son humour et sa disponibilité. Je me réjouis de penser que maintenant on aura peut-être un peu plus de temps pour pouvoir collaborer et faire toutes ces manips dont on parle depuis si longtemps.

J'aimerais exprimer ma gratitude aux autres doctorants de notre joyeuse équipe (Adélaïde, Caroline, Thomas, Laurence, Meike), en particulier Adélaïde (avec les accents!) de Heering, certainement la plus Russe des Belges; Caroline Michel (monte dans ta chambre finir ta thèse); et Thomas Busigny. Votre présence a considérablement égayé mes déambulations dans les couloirs de la fac de Psycho!

Merci également à Christine Schiltz pour ses nombreux commentaires intéressants lors de nos réunions de groupe et pour son contact chaleureux. Je voudrais également remercier Anne-Marie Schüller qui a été présente dans notre groupe durant les deux premières années de ma thèse.

Je remercie également les membres des unités CODE et NESC. Je pense notamment à Peggy, Bénédicte, Carolina, Marina, Fabrizio, Samuel et aussi Coralie (de PSOR). Merci aussi à Yolande, notre super-secrétaire.

Durant cette thèse j'ai été aidé par la présence de Marjorie Paque, Christian Namèche, Olivier d'Arripe, Quentin Duvivier et Adriano Boremanse, stagiaires ou mémorants dans notre équipe. Qu'ils en soient remerciés.

Pour une certaine personne avec laquelle je partage toute ma vie excepté celle de la recherche, la fin de cette thèse est probablement un événement qui la soulage autant que moi. Merci Violaine pour tes encouragements, ta patience, ton amour et ta poésie sans lesquels les pages qui suivent seraient bien pâles. Ces derniers mois, tes talents de cuisinière, de plongeuse, de rangeuse et de prof de français ont été mis à contribution de manière égale à mon effervescence derrière le clavier. La suite de ce que je veux te dire se passe de mots.

Enfin, je tiens à remercier mes parents pour leur présence, leur aide et leurs encouragements constants. Je tiens aussi à vous remercier d'avoir joué le rôle de ceux qui gardent la tête froide face à mon affolement au moment de rendre la première version de ce texte. Cette thèse, c'est surtout à vous que je dois de l'avoir réalisée.

Table des matières

Introduction générale **1**

I. Contexte théorique

Chapitre 1. Organisation et propriétés du système visuel **9**

1. Propriétés anatomiques et fonctionnelles du système visuel 9
   1.1. Organisation hiérarchique. 11
   1.2. Deux voies visuelles 13
   1.3. Organisation en colonnes corticales 15
   1.4. Inter-connectivité dense 17
   1.5. Champs récepteurs 19
   1.6. Rétinotopie 21
2. Le cortex inféro-temporal et la reconnaissance visuelle 22
   2.1. Architecture du cortex inféro-temporal 22
   2.2. Le rôle du cortex inféro-temporal dans la reconnaissance visuelle 24
      2.2.1. Propriétés générales des neurones de IT 25
         a. Sensibilité à des formes modérément complexes 25
         b. Invariance et dépendance 28
         c. Organisation modulaire en colonnes 30
         d. Représentation des objets dans IT 32
      2.2.2. Les cellules sensibles aux visages 34
         a. Localisation des cellules sensibles aux visages 35
         b. Propriétés générales des cellules-visage 36
         c. La représentation de l'identité faciale dans IT: une représentation distribuée 43
3. Un réseau cortical pour la perception des visages chez l'humain 46
4. Conclusions 53

Chapitre 2. Dynamique temporelle de la catégorisation visuelle **55**

1. Décours temporel de l'activation des aires visuelles dans la voie ventrale chez le primate 55
2. Vitesse de catégorisation des objets et des scènes visuelles: évidences psychophysiques et électrophysiologiques 62
3. Vitesse d'activation des représentations faciales: neurophysiologie et psychophysique 70
   3.1. Électrophysiologie chez le singe 71
   3.2. Études comportementales chez l'humain 79
      3.2.1. À quelle vitesse un visage présent dans le champ visuel est-il détecté? 79
      3.2.2. Reconnaissance et discrimination faciale 83

Chapter 3. Face processing in humans as revealed by electrophysiology **89**

1. EEG recording and event-related potentials 91
2. Description of the main visual event-related potentials 93
   2.1 ERP components originating from striate and extrastriate cortex. 94
   2.2 The N170 and the VPP 99
3. Electrophysiological studies of face visual processing 102
   3.1. The N170 as an index of face processing specificity 104
      3.1.1. Underlying neural sources of the larger N170 to faces 105
      3.1.2. Is the larger N170 to faces due to differences in the low-level visual parameters between faces and non-faces? 111

3.1.3. The N170 for faces reflects high-level visual processes 114
3.1.4. Modulations of the N170 elicited by non-face categories 116
3.2. The encoding of face information during the N170 time-window 120
3.2.1. The N170 as a structural encoding stage? 121
3.2.2. The respective representation of facial feature and face configuration during the N170 125
a. A short story that repeats itself… 127
b. The representation of facial features during the N170 time-window 127
c. The representation of face configuration during the N170 time-window 130
d. Featural and configural information are integrated during the N170 133
3.2.3. Evidence for the time course of face categorization and encoding of individual faces 134
a. The categorization of a face as a face occurs during the N170 134
b. How early in the visual processing stream are face preferential responses observed: are faces categorized before the N170? 136
c. How early is information about individual faces represented in the brain? 139
3.3. Summary and conclusions 148

II. Contribution empirique

Chapter 4. Presentation of experimental work 153

Chapter 5. Concurrent processing reveals competition between visual representations of faces 159

1. Paper 1 159
2. Complementary results and discussion 170

Chapter 6. The Speed of Individual Face Categorization (Paper 2) 177

Chapter 7. Using the face inversion effect to assess the time course of face categorization and discrimination 193

1. The time course of the inversion effect during individual face discrimination 193
1.1. Paper 3 193
1.2. Complementary results and discussion 209
2. . Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans 213
2.1. Paper 4 213
2.2. Complementary results and discussion 243
3. Dissociating the time course of face categorization and the coding of individual faces (Paper in preparation) 251

Chapter 8. Individual faces are represented holistically during the N170 component (Paper in preparation) 265

General discussion and perspectives 287

Bibliographie générale 301

Liste des abréviations

CGL	Corps Genouillés Latéraux
DTI	Diffusion Tensor Imaging
EEG	Enlectroencephalogram
ERMF	Event Related Magnetic Field
ERP	Event Related Potential
FFA	Fusiform face Area
fMRI	functional Magnetic Resonance Imaging
IRMf	Imagerie par Résonance Magnétique fonctionnelle
IT	Inferior-temporal cortex / Cortex Infero-temporal
ITG	Inferior Temporal Gyrus / Gyrus inféro-temporal
LGN	Lateral Geniculate Neucleus
LOC	Lateral Occipital Complex
MEG	Magnetoencephalography
OFA	Occipital Face Area
OTC	Occipito-Temporal Cortex
PET-scan	Tomographie par Emission de Positon
RSVP	Rapid Serial Visual Presentation
SOA	Stimulus Onset Asynchrony
STP	Superior temporal Polysensory area
STS	Superior temporal sulcus / Sillon Temporal Superieur
V1,V2,..	Aires corticales visuelles primaire, secondaire

Introduction générale

De toute évidence, nous sommes capables d'identifier un visage familier avec une rapidité extrême. De même, aussi déconcertante que puisse paraître d'un point vue scientifique la facilité avec laquelle l'être humain est capable d'extraire un grand nombre d'informations à partir d'un stimulus aussi complexe qu'un visage, la plupart des individus trouvent étonnant que cette capacité puisse faire l'objet de recherches scientifiques. Sergent & Signoret (1992) ont bien traduit la difficulté à justifier l'intérêt scientifique de l'étude de la perception des visages:

> *"People often express surprise and incredulity when they learn that the perception and recognition of faces are the object of intense research. What is there to investigate, they ask, and how such* ***natural activities*** *as identifying a friend, recognizing an emotion, appreciating the beauty of a face,* ***that call for no particular skills****, raise any questions of scientific interest?..."*
> *"...we are also convinced that every one is endowed with the capacity to identify faces, and, when we pass by an acquaintance who shows no sign that he recognizes us, we attribute this failure to his being too absorbed in his thoughts rather than to his being unable to recognize faces."*
> (Sergent & Signoret, 1992, pp.55).

C'est probablement parce que la perception des visages est une habileté si naturelle qu'elle passe inaperçue que l'on trouve généralement dans l'introduction d'un travail de thèse ou de mémoire réalisé dans le domaine de la perception des visages une forme de justification quant au choix de cette catégorie comme objet d'étude scientifique. Dans la plupart des cas, cette justification est axée autour de deux caractéristiques fondamentales de cette catégorie d'objets visuels.

La première est l'indéniable *importance* que le visage possède d'un point de vue *écologique*. En effet, le visage offre non seulement la possibilité de distinguer entre des individus de manière à la fois efficace et rapide, mais également d'obtenir, à propos de ces individus, un grand nombre d'informations complémentaires telles que leur âge, leur sexe ou encore leur origine ethnique. Le visage transmet également quantité d'informations non verbales cruciales dans la communication interpersonnelle comme l'expression faciale, l'intentionnalité dans la direction du regard, ou le détail de mimiques faciales. En outre, nous sommes généralement capables d'inférer, correctement ou non, les traits de personnalité ou

l'état de santé d'une personne. Pour ce faire, nous pouvons notamment nous baser sur la forme que donne au visage l'utilisation particulière de la musculature faciale, sur la couleur et la texture de la peau ou encore sur la forme de certains traits locaux comme les yeux et la bouche. Enfin, nous sommes généralement capables d'émettre une appréciation à propos de l'attractivité d'un visage, appréciation basée non seulement sur la physionomie du visage (forme globale et formes des traits locaux, texture et couleur de la peau, contraste,…), mais également largement déterminée par des facteurs culturels liés à la biographie de chaque individu ou à l'histoire de la société. De manière intéressante, certains des facteurs qui déterminent l'attractivité d'un visage comme la couleur de la peau, dépendent en partie de la concentration sanguine en œstrogènes chez les femmes, fournissant par là un indicateur implicite de la fertilité. Le visage humain est donc manifestement un stimulus visuel qui fournit des informations particulièrement importantes pour l'individu en tant qu'être social.

La deuxième raison qui justifie l'intérêt scientifique pour l'étude de la perception des visages est que cette catégorie se révèle en fait être un *objet d'étude idéal* lorsqu'il s'agit d'explorer les mécanismes perceptifs impliqués dans la catégorisation visuelle d'objets complexes. En effet, les visages humains sont des stimuli multidimensionnels possédant de nombreux indices visuels qui permettent de les distinguer. Parmi ces indices, on trouve les traits internes (yeux, sourcils, nez, bouche,..) et externes (cheveux, oreilles,…) du visage, l'arrangement spatial entre ces traits (position des traits les uns par rapport aux autres et distance métrique entre les traits), la forme globale du visage (contour du visage et volume en trois dimensions) ou encore les propriétés de surface de la peau (couleur, texture, réflectance). Les visages humains offrent donc la possibilité d'étudier la perception d'une catégorie d'objets dont la reconnaissance nécessite l'assemblage de multiples indices en un tout cohérent. Comme nous l'avons vu plus haut, ces indices fournissent non seulement une indication quant à l'identité de la personne mais permettent également de coder, dans un espace très réduit, de nombreuses informations utiles pour l'être humain social. En fait, le visage humain, de par sa complexité et de par les nombreuses apparences qu'il peut prendre en fonction des variations de point de vue et d'illumination, forme probablement l'objet le plus complexe auquel nous soyons régulièrement confrontés. Notons que malgré les nombreux indices précis qui permettent de discriminer entre des visages, les exemplaires de cette catégorie sont néanmoins fortement similaires du point de vue purement physique de la stimulation. La complexité des visages, couplée à la nécessité de les discriminer au niveau individuel (contrairement aux autres catégories d'objets), fait appel à une capacité perceptive de complexité égale. C'est cette capacité qui fait, à juste titre – et c'est là que nous voulions en venir –, l'objet de nombreuses

recherches en Neurosciences. En outre, d'une manière plus générale, l'étude de la perception des visages permet d'accroître les connaissances sur la manière dont on perçoit les objets non faciaux, qui sont généralement appréhendés et reconnus au niveau de base (chaise, chien,…) plutôt qu'au niveau individuel.

Comme nous l'avons indiqué plus haut, la reconnaissance d'un visage fait appel à des mécanismes perceptifs extrêmement *rapides*. Cependant, malgré l'instantanéité avec laquelle cette information nous semble être disponible, les recherches menées sur la perception des visages révèlent que cette instantanéité n'est qu'apparente. En effet, les opérations mises en place lors de l'élaboration d'une représentation faciale dans le système perceptif sont réalisées en un certain laps de temps.

Dans ce travail de thèse, nous avons abordé les aspects temporels de la perception des visages. Si, pour rendre compte de la perception des visages, il est nécessaire de déterminer comment est représentée dans le système perceptif l'énorme quantité d'informations que l'on peut extraire de cette catégorie, il est également crucial de déterminer la *chronologie* précise avec laquelle cette masse d'informations est représentée. En effet, on peut considérer que la capacité à se percevoir des visages (et l'environnement visuel de manière plus générale) est largement contrainte par la finalité de la perception, qui est de permettre une action adaptée à l'environnement. L'intérêt d'étudier la chronologie (ou le décours temporel) de la perception visuelle repose dès lors sur le fait que la vitesse à laquelle une information devient disponible pour réagir à une stimulation est déterminante pour la survie de tout système biologique. Il en découle que toute description de la manière dont sont représentés les visages au niveau cérébral doit pouvoir rendre compte du décours temporel à laquelle ces représentations sont élaborées. Par ailleurs, définir la chronologie temporelle avec laquelle des traitements perceptifs sont réalisés permet également de déterminer certaines propriétés anatomo-fonctionnelles des structures cérébrales impliquées dans ces traitements.

Nous avons mentionné ci-dessus qu'un visage recèle une quantité énorme d'informations qui peut être appréhendée à de multiples niveaux pour permettre une catégorisation plus ou moins fine du visage. En d'autres termes, l'action à réaliser à partir d'un visage va déterminer la précision minimale de l'information qu'il est nécessaire d'extraire de celui-ci. Par exemple, alors qu'une information relativement *grossière* sera suffisante pour détecter la présence d'un visage dans l'environnement, une information beaucoup plus *fine* devra être extraite du visage pour représenter son identité, ses expressions faciales, son intentionnalité…

Dans cette thèse nous avions pour objectif de tenter de déterminer précisément le décours temporel avec lequel sont élaborées, dans le système visuel, les représentations faciales menant à des décisions perceptives (détection d'un visage, reconnaissance,…) impliquant des informations à différents niveaux de résolution. En particulier, nous nous sommes attaché à déterminer la latence à laquelle est représentée dans le système visuel l'information utile à la détection d'un visage (information grossière) et l'information utile à l'encodage de l'identité faciale (information fine). Pour mener à bien cet objectif nous avons principalement eu recours à la technique d'enregistrement des potentiels évoqués au niveau du scalp chez le sujet humain. Cette technique, basée sur la mesure de l'activité électrophysiologique cérébrale à un niveau macroscopique, permet d'appréhender le fonctionnement neuronal impliqué dans la perception visuelle avec une résolution temporelle proche de la milliseconde.

Cette dissertation s'articule autour de deux parties principales; la première exposant le contexte théorique dans lequel s'inscrit notre contribution expérimentale; la deuxième reprenant les travaux empiriques réalisés pour atteindre notre objectif.

La première partie de notre exposé, décrivant le contexte théorique, est elle-même divisée en trois chapitres.

Dans le **premier chapitre** nous examinerons le contexte anatomique et fonctionnel dans lequel prend place la perception visuelle, et plus particulièrement la reconnaissance des objets et des visages. Nous décrirons d'abord les grands principes qui régissent l'organisation anatomo-fonctionelle du système visuel (par ex: organisation hiérarchique, interconnectivité entre aires visuelles,…). Les connaissances acquises depuis les années 1950-1960 sur l'organisation du système visuel, ont largement déterminé la manière dont les modèles psychologiques ou computationnels ont évolué pour rendre compte de la perception visuelle. Nous nous attarderons ensuite à examiner les propriétés du codage visuel des objets et des visages au niveau du cortex inféro-temporal. Cette région corticale constitue en effet un carrefour crucial pour la formation de représentations visuelles permettant l'identification des objets et des visages. Ce cadre nous amènera à considérer une population de cellules particulièrement sensibles à la catégorie des visages. Nous détaillerons les propriétés fonctionnelles de ces cellules ainsi que la manière dont elles permettent l'encodage d'une représentation faciale individuelle. Alors que ce premier chapitre est principalement consacré à la description des travaux réalisés sur des modèles animaux, nous passerons brièvement en

revue, dans une dernière partie, les connaissances à propos de la localisation des régions cérébrales spécifiquement impliquées dans la perception des visages chez l'humain.

Dans le **deuxième chapitre**, nous décrirons les connaissances actuelles à propos des aspects temporels de la perception des objets et des visages. Nous examinerons premièrement la dynamique d'activation des structures cérébrales impliquées dans l'élaboration des représentations visuelles. Cette dynamique temporelle a été principalement déterminée par l'intermédiaire d'enregistrements de neurones tout au long des voies visuelles chez l'animal. Ces travaux suggèrent que la dynamique temporelle de la perception visuelle est loin de respecter la hiérarchie anatomique des aires visuelles corticales et qu'elle procède de manière parallèle et distribuée. Nous verrons ensuite comment le décours temporel de la catégorisation visuelle d'objets et de scènes naturelles a été appréhendé chez l'humain par des méthodologies comportementales et électrophysiologiques. L'ensemble de ces travaux souligne la rapidité et l'efficacité du système cérébral humain à extraire une information d'après une entrée visuelle permettant la catégorisation. La troisième partie sera plus particulièrement consacrée à décrire la dynamique temporelle du traitement visuel des visages. Nous passerons en revue un ensemble de données recueillies par enregistrements neuronaux dans le cortex inféro-temporal du primate, ainsi que des données issues des approches comportementales chez le sujet humain. Cette dernière partie nous permettra ainsi d'avoir un aperçu des contraintes biologiques et psychophysiques déterminant l'extraction d'informations sur les visages à différents niveaux de résolution.

Le lecteur pourrait trouver surprenant que dans ces deux premiers chapitres nous nous attardions à décrire en détail les propriétés fonctionnelles et temporelles des neurones enregistrés dans le système nerveux central chez le primate non-humain, alors que notre méthodologie expérimentale est basée sur le recueil de données électrophysiologiques chez l'humain. Cependant, le lecteur doit garder en mémoire que d'une part, ces travaux chez le primate non-humain forment la base des connaissances actuelles sur la dynamique temporelle du système visuel, et que, d'autre part, les hypothèses et interprétations de notre contribution empirique ont été très largement inspirées de ces travaux. De plus, les connaissances accumulées sur la latence d'activation des aires visuelles chez le primate non-humain permettent d'évaluer la plausibilité biologique des interprétations formulées sur base de données électrophysiologiques récoltées chez l'humain au niveau du scalp.

La description détaillée des données obtenues par l'enregistrement des réponses électrophysiologiques (potentiels évoqués) au niveau du scalp chez l'humain dans des situations impliquant la perception de visages fait l'objet du **troisième chapitre**. Nous

exposerons tout d'abord brièvement les aspects techniques liés à l'enregistrement de l'électroencéphalogramme et décrirons les principaux potentiels évoqués enregistrés lors d'une stimulation visuelle. Nous verrons que dans le cadre de la perception des visages un composant électrophysiologique, la N170, a davantage retenu l'attention des chercheurs en raison de sa plus forte amplitude en réponse à des visages qu'en réponse aux autres catégories d'objets. Ensuite, nous décrirons l'état actuel des connaissances sur une série de thèmes abondamment investigués dans les recherches menées en électrophysiologie sur la perception des visages, impliquant le composant N170. La confusion quant à la détermination du rôle fonctionnel de la N170 nous mènera à déconstruire l'idée que l'on puisse associer ce potentiel évoqué à une étape fonctionnelle déterminée de la perception des visages. Enfin, dans la dernière partie de ce troisième chapitre nous verrons que malgré le nombre très important de travaux menés dans le domaine de la perception des visages, il existe une incertitude flagrante quant au décours temporel avec lequel sont élaborées les représentations faciales permettant la catégorisation et l'identification d'un visage. La contribution expérimentale présentée dans la deuxième partie de cette thèse a donc pour objectif de clarifier précisément ce décours temporel.

I. Contexte théorique

Chapitre 1
Organisation et propriétés du système visuel

Le substrat cérébral dans lequel prend place la perception visuelle est déterminant lorsqu'on s'intéresse à décrire la manière dont sont élaborées les représentations de notre environnement. En effet, tout modèle qui prétend rendre compte de cette élaboration doit pouvoir tenir comptes des contraintes imposées par l'organisation de ce système.

Dans ce premier chapitre nous examinons ces contraintes anatomiques et fonctionnelles qui pèsent sur la perception visuelle par l'intermédiaire des nombreux travaux réalisés chez l'animal, et plus particulièrement sur le primate non-humain. Nous nous attarderons notamment sur la description des connaissances à propos du cortex inféro-temporal, une région cruciale dans l'élaboration des représentations visuelle des objets et des visages. Enfin, en examinant certains des travaux réalisés chez l'humain pour révéler la neuro-anatomie fonctionnelle de la perception des visages, nous verrons que l'extraction d'une représentaiton faciale individuelle met en jeu des intéraction au sein d'un réseau de régions situées dans le cortex occipito-temporal.

1. Propriétés anatomiques et fonctionnelles du système visuel

Les nombreuses études réalisées sur l'organisation anatomique et fonctionnelle du cortex visuel chez le singe (plus particulièrement le singe macaque) ont révélées l'existence de plus d'une trentaine d'aires visuelles se distinguant principalement sur base de critères histologiques, rétinotopiques, de connectivité avec d'autres aires/régions cérébrales et de propriétés fonctionnelles (Felleman & Van Essen, 1991; Zeki, 1993; voir aussi Van Essen, 2005; Figure 1). Par ailleurs, le développement des techniques de neuroimagerie fonctionnelle (principalement la tomographie à émission de position – ou PET-scan et l'imagerie par résonance magnétique fonctionnelle – ou iRMF) non invasive depuis une quinzaine d'années a permis d'accroître fortement nos connaissances à propos de l'organisation fonctionnelle du cortex visuel chez l'humain (Grill-spector & Malach, 2004; Dougherty et al., 2003; Tootell et al., 1996). Il est particulièrement intéressant de tenter de déterminer la correspondance entre les aires visuelles identifiées chez le singe et celles identifiées chez l'humain (voir Figure 1), dans la mesure où les nombreux travaux réalisés chez le singe permettent d'une part d'interpréter des données de neuroimagerie récoltées chez l'humain, et d'autre part de générer

des hypothèses sur les propriétés fonctionnelles des aires visuelles corticales chez l'humain (Denys et al., 2004; Orban et al., 2004; Tootell, 2003 ; VanEssen, 2003). Notons que la division des aires visuelles chez l'humain est principalement basée sur une combinaison de critères rétinotopiques[1] et de critères de propriétés fonctionnelles[2], contrairement à la cartographie des aires visuelles chez le singe qui est opérée également sur base de critères histologiques et de connectivité.

Nous allons ci-dessous passer en revue les principales propriétés qui régissent l'organisation anatomo-fonctionnelle du système visuel.

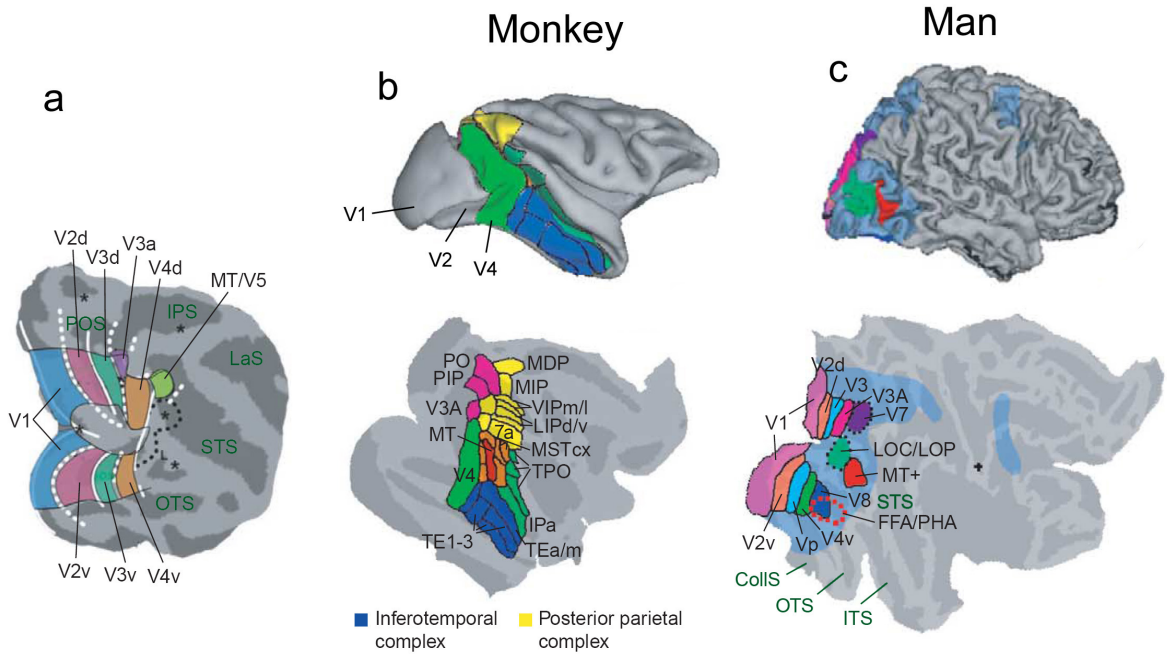

Figure 1: Représentation des aires visuelles corticales chez le singe macaque et l'humain. A. Partie postérieure de l'hémisphère droit d'un cerveau de singe "applati" pour montrer les aires visuelles postérieures (V1, V2, V3, V4, V3A et MT) dorsales et ventrales. B. Aires visuelles de niveau intermédiaire et de haut niveau dans le lobe temporal et la partie postérieure du cortex pariétal. La partie inférieure montre la version applatie du cerveau du dessus. C. Aires visuelles chez l'humain dans le cortex occipital,

[1] La division des aires visuelles dites rétinotopiques (qui possèdent une organisation respectant la topographie du champ visuel, voir plus loin au point 1.4 du chapitre 1) se base sur la cartographie de la représentation du champ visuel au niveau des aires visuelles corticales. Cette cartographie est réalisée en présentant des stimulations diverses dans l'ensemble du champ visuel, permettant ainsi d'établir une correspondance entre la localisation de ces stimulations dans le champ visuel et la localisation anatomique des activations cérébrales provoquées par ces stimulations.

[2] Les critères fonctionnels sont particulièrement utilisés pour la définition des régions visuelles de haut niveau qui ne présentent pas ou peu de propriétés rétinotopiques. Par exemple, les régions qui sont impliquées dans le traitement des visages sont identifiées uniquement sur base du fait qu'elles sont activées de manière plus importante pour les visages que pour d'autres objets. Ces régions sont révélées par soustraction entre l'activation obtenue en réponse à des objets et l'activation obtenue en réponse à des visages. Il en va de même pour des régions répondant davantage pour des objets complexes (visages, objets) que pour du bruit visuel (par ex: le complexe occipital latéral – "lateral occipital complex" – LOC; voir Grill-Spector et al., 2001; Kourtzi et al., 2003) ou des régions répondant spécifiquement à des stimuli en mouvement (MT/V5).

occipito-temporal et pariétal postérieur. Les aires visuelles s'étendant vers le cortex temporal forment la voie "ventrale", alors que les aires visuelles qui s'étendent dans le cortex pariétal forment la voie dorsale. La partie inférieure montre l'hémisphère droit applati et la localisation d'une série d'aires visuelles de bas (par ex: V1, V2) et de haut niveau (par ex: MT+, FFA). Modifié d'après Orban et al., 2004

1.1. Organisation hiérarchique

La trajectoire de l'input visuel dans le système visuel est dite *hiérarchique* car elle traverse séquentiellement une série de structures sous-corticales (corps genouillés latéraux – ou CGL, colliculus supérieur) et corticales (multiples aires visuelles) entre la rétine, qui constitue le niveau inférieur de la hiérarchie, et le sommet de cette hiérarchie. Ce sommet est situé au niveau du cortex inféro-temporal (IT) pour la voie dite « ventrale » et au niveau du cortex pariétal pour la voie dite « dorsale » (voir point suivant ci-dessous) (Figure 1). La notion de hiérarchie réfère plus particulièrement au fait que d'une structure à l'autre, la représentation de l'information sensorielle et les traitements réalisés sur cette information sont de plus en plus complexes. Le résultat des traitements effectués dans une structure donnée convergent sur la structure suivante de plus haut niveau, chaque structure générant une représentation de plus en plus complexe et abstraite de la stimulation par recombinaison des entrées des niveaux précédents (Figure 3). Dès lors, dans une telle structure hiérarchique, les propriétés de réponse et de sélectivité d'un neurone Y à un niveau donné sont principalement déterminées par la combinaison des propriétés de réponse des neurones situés en amont et qui convergent vers ce neurone Y (Kobatake & Tanaka, 1994; Tanaka, 1996). Un exemple, qui a par ailleurs fortement promu l'idée que le système visuel est organisé de manière hiérarchique, est la découverte par Hubel & Wiesel (Hubel & Wiesel, 1968; 1969) de cellules "complexes" dans le cortex visuel primaire (V1) dont les propriétés fonctionnelles (par ex: la structure du champ récepteur) sont partiellement définies par une combinaison des outputs des cellules "simples" qui convergent vers ces cellules "complexes". À des niveaux plus élevés de la hiérarchie, on trouve des neurones répondant de manière spécifique à des traits ou à des combinaisons de traits visuels complexes (Tanaka, 1996) et même à certains stimuli comme des visages ou des mains (Gross, 1972; Bruce et al., 1981; Perrett et al., 1982).

Il en résulte donc que la notion de hiérarchie est directement liée à la spécialisation fonctionnelle des aires visuelles traversées par l'input rétinien. En effet, les informations à propos de la forme, de la couleur, du mouvement ou de la profondeur sont traitées au moins partiellement de manière séparée dans les aires visuelles au bas de la hiérarchie et impliquent donc des représentations neuronales différentes (des "cartes" neuronales) Ces cartes sont par

la suite assemblées en une représentation cohérente et intégrée[3]. Notons que malgré la séparation du traitement de ces propriétés visuelles, elles sont néanmoins appréhendées dans le cortex visuel de manière largement parallèle et interactive. En effet, les représentations générées à chaque niveau de la hiérarchie sont le résultat d'interactions complexes entre les neurones situés à des niveaux différents de la hiérarchie et également entre les neurones situés à même niveau de la hiérarchie (voir 1.4. chapitre 1) (voir Rousselet et al., 2004b; Merigan & Maunsell, 1993; Felleman & Van Essen, 1991; Bullier, 2001; Hochstein & Ahissar, 2002).

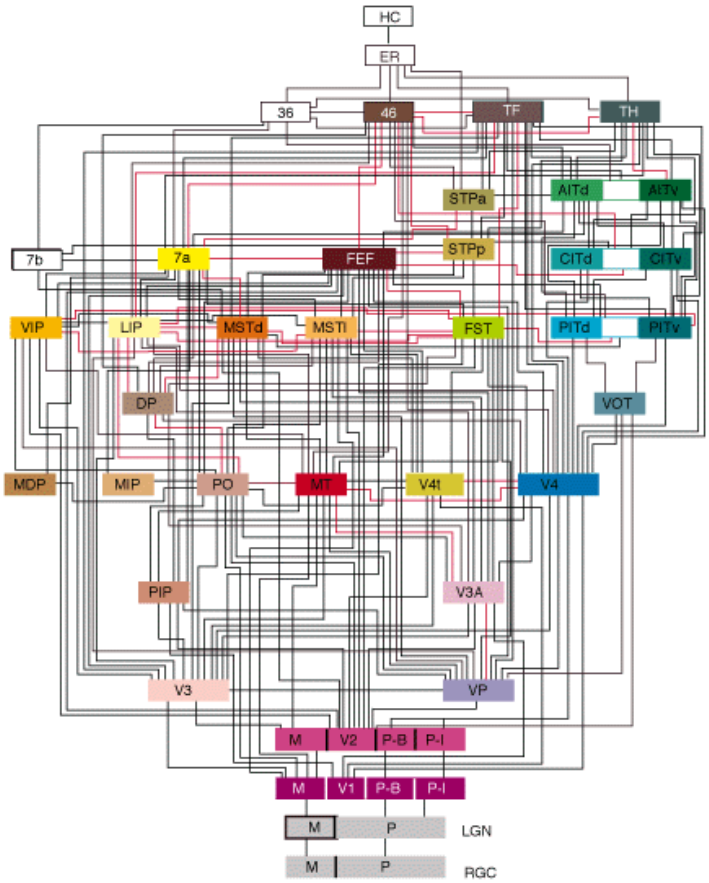

Figure 2: Représentation schématique de la hiérarchie des aires visuelles mises en évidence chez le singe ainsi que les nombreuses connexions entre ces aires. La voie ventrale se situe à droite du graphique. (Felleman & Van Essen, 1991).

[3] La ségrégation entre les représentations de ces différentes propriétés de l'input visuel pose la question du mécanisme neuronal par lequel ces représentations sont combinées pour générer un percept cohérent (le fameux "binding problem"). Le lecteur pourra trouver un compte rendu détaillé des tenants et aboutissants du "binding problem" dans un numéro spécial de la revue *Neuron* paru en 1999 (par ex: Ghose & Maunsell, 1999; Singer, 1999; Reynolds & Desimone, 1999).

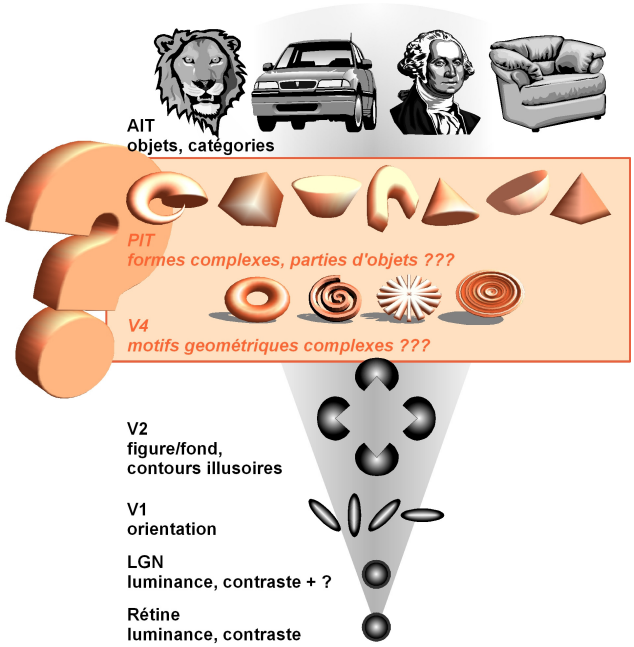

Figure 3: Au fil de la hiérarchie visuelle dans la voie ventrale, les neurones sont sensibles à des stimuli de plus en plus complexes.

La notion de hiérarchie dans le système visuel a eu une influence très importante sur les modèles élaborés pour rendre compte de la reconnaissance visuelle. Le plus connu et le plus influent de ces modèles est sans conteste le modèle développé par David Marr (Marr, 1982), un des pères des Neurosciences computationelles. Ce modèle s'inscrit dans l'approche computationnelle symbolique et postule que la cognition résulte de la manipulation de symboles qui ont une valeur de représentation. Dans le cadre de l'approche hiérarchique telle que définie par Marr, la perception visuelle est appréhendée comme le produit final d'une séquence d'opérations impliquant des représentations successives et de plus en plus complexes. D'autres modèles plus récents (par ex: Kobatake & Tanake, 1994; Riesenhuber & Poggio, 1999) s'inscrivent également largement dans la conception hiérarchique du système visuel pour rendre compte de la catégorisation des objets visuels.

1.2. Deux voies visuelles

Nous venons de voir que la perception visuelle procède notamment par convergence d'informations traitées dans des structures visuelles organisées hiérarchiquement. À la base de la hiérarchie, on trouve la rétine qui tapisse le fond de l'œil et traduit l'information lumineuse

qui y parvient en un code électrophysiologique qui pourra être transmis vers le cerveau[4]. Les axones des cellules ganglionnaires à la sortie de la rétine font un relais au niveau des CGL pour ensuite atteindre le premier niveau cortical majeur: V1. Les neurones de V1 sont sensibles de multiples propriétés de l'input visuel comme l'orientation, la couleur, les fréquences spatiales et temporelles, le mouvement. Dans V1 il existe un regroupement des cellules sensibles à une même caractéristique, de sorte que V1 opère une séparation entre les traitements liés à ces différentes caractéristiques visuelles. L'aire V2, située en aval de la hiérarchie, reçoit la majorité de ses afférences de V1. La ségrégation entre les informations de forme, couleur, mouvement et disparité binoculaire (permettant d'accéder à une information sur la profondeur) est conservée dans V2. Les projections ascendantes de V2 sont séparées en deux grandes voies principales du traitement de l'information visuelle (Figure 4): la voie ventrale et la voie dorsale (Ungerleiger & Mishkin, 1982; Ungerleiger & Haxby, 1994). Ces deux voies sont fortement interconnectées et convergent vers des régions cérébrales communes (par ex: le cortex pré-frontal), mais néanmoins traitent chacune des informations différentes de l'entrée visuelle. Les aires visuelles situées entre V2 et le cortex inféro-temporal (par ex: V4, TEO) forment la voie ventrale, et sont principalement impliquées dans l'identification des objets visuels. La voie ventrale est constituée des projections des neurones des voies parvocellulaire et magnocellulaire (Ferrera et al., 1994). Les aires situées entre V2 et les aires du cortex pariétal postérieur (par ex: V3, MT/V5, MST) forment la voie dorsale. La voie dorsale reçoit la majorité de ces afférences de la voie magnocellulaire et est ainsi particulièrement sensible aux stimuli en mouvement ou aux informations de contraste de luminosité. La voie dorsale est donc impliquée dans la perception du mouvement (artificiel ou biologique), la coordination et la planification visuo-motrice, ou encore dans le contrôle des mouvements attentionnels.

[4] La rétine ne peut cependant en aucun cas être définie comme un simple convertisseur d'un code photonique en un code électrophysiologique. En d'autres termes, l'information qui quitte la rétine ne se résume pas à l'input lumineux que celle-ci subit. En effet, la rétine produit déjà une interprétation de l'environnement visuel qui est contrainte par sa propre organisation. Par exemple, c'est au niveau de la rétine que se trouve l'origine entre la voie **parvocellulaire** (transmettant l'information depuis les cônes de la région centrale de la rétine; la macula) et la voie **magnocellulaire** (transmettant l'information depuis les bâtonnets répartis sur la partie périphérique de la rétine).

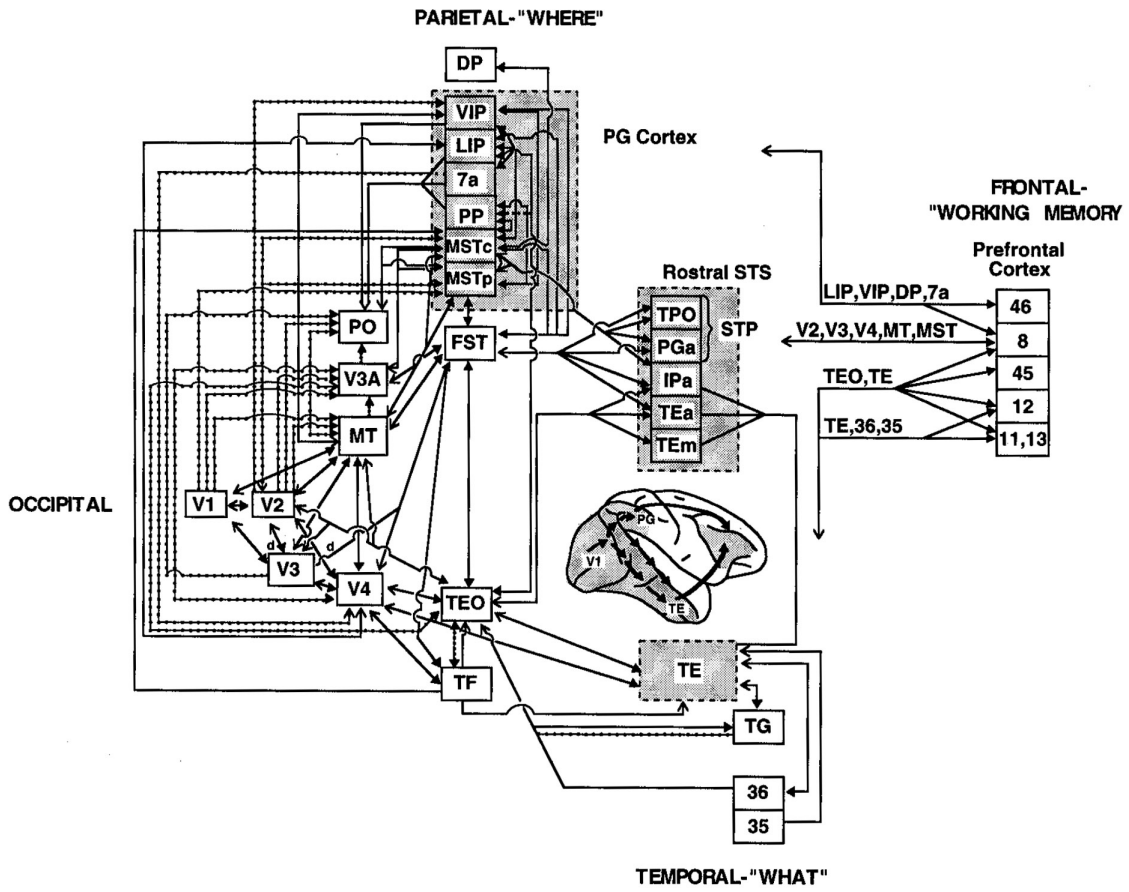

Figure 4: Organisation des voies visuelles en deux voies principales: la voie ventrale dans le cortex temporal et la voie dorsale dans le cortex pariétal. Il existe de nombreuses connexions entre les aires visuelles au sein de chaque voie mais également entre les deux voies. Le sillon temporal supérieur (STS) forme un nœud privilégié pour ces intéractions entre les deux voies. Modifié d'après Ungerleider, 1995 dans Gross, 1998.

1.3. Organisation en colonnes corticales

Les études anatomiques et les enregistrements cellulaires réalisés chez le singe ont mis en évidence que les neurones corticaux qui présentent une sélectivité similaire sont rassemblés en colonnes (perpendiculaires à la surface corticale) traversant généralement l'ensemble des couches corticales (voir Horton & Adams, 2005 pour une revue de la littérature). Les colonnes ont un diamètre moyen d'environ 300 – 600 µm et rassemblent des neurones fortement interconnectés.

La première évaluation de l'organisation en colonnes dans le système visuel provient des travaux de Hubel et Wiesel (1969) qui ont observé que dans V1, les neurones présentant

une même dominance oculaire[5] sont regroupés en bandes étroites. De plus, dans V1 les neurones présentant une même sélectivité à l'orientation sont également rassemblés en colonnes perpendiculaires à la surface du cortex (voir Figure 5). Hubel & Wiesel ont d'ailleurs défini la colonne corticale comme une unité fonctionnelle pouvant traiter un type d'information particulier (par ex: l'orientation ou les fréquences spatiales).

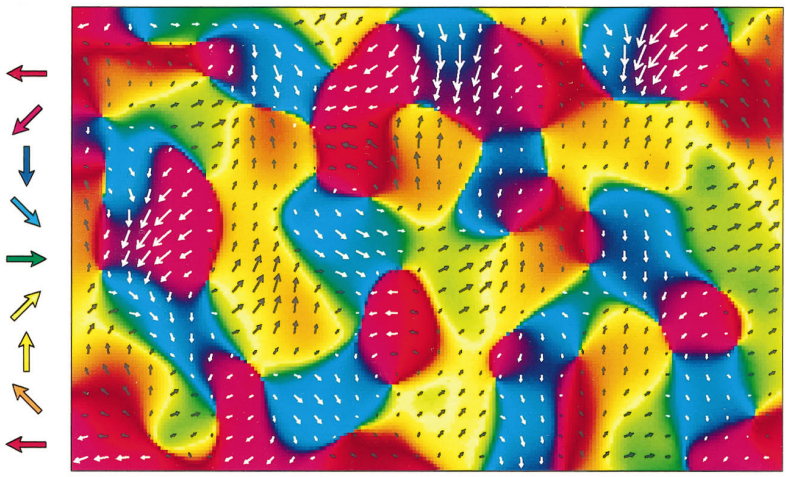

Figure 5: Les neurones présentant une sélectivité pour une même direction de mouvement sont regroupés en colonnes dans V2, chez le chat. La figure représente une cartographie de la sélectivité à une orientation de mouvement sur une surface de cortex visuel chez le chat (V2). L'orientation préférée est codée selon un code de couleur (voir flèches à gauche). L'orientation préférée est également indiquée sous forme de vecteurs sur la figure. On peut observer le regroupement en colonnes des neurones présentant une même sélectivité. Modifié d'après Shmuel & Grinvald, 1996.

Les propriétés de la stimulation visuelle auxquelles les colonnes sont sensibles accroissent en complexité tout au long de la hiérarchie des voies visuelles (Kobatake & Tanaka, 1994). Ainsi, on trouve des colonnes corticales contenant des neurones sensibles à l'orientation d'une barre ou à une gamme spécifique de fréquences spatiales dans V1, alors que dans le cortex inféro-temporal, les colonnes rassemblent des neurones codant pour des formes/patterns plus complexes (Tanaka et al., 1991; Fujita et al., 1992; Wang et al., 1996; Tanaka, 2003).

Notons que malgré le fait que les neurones d'une colonne donnée présentent une sélectivité pour des stimuli similaires, cette sélectivité n'est cependant pas exactement identique. Par exemple, tandis que les neurones de V1 rassemblés dans une même colonne

[5] La plupart des neurones de V1 reçoivent des afférences en provenance des deux yeux (binoculaires) mais leur réponse est généralement dominée davantage par un œil que par l'autre.

sont sélectifs à une gamme de fréquences spatiales ou à une orientation particulière, les propriétés spatio-temporelle[6] du champ récepteur de ces neurones peut fortement varier (DeAngelis et al., 1999). De même, les neurones de IT rassemblés dans une même colonne sont sélectifs à une même forme complexe globale mais la sélectivité précise de chaque neurone varie légèrement d'un neurone à l'autre (Tanaka et al., 1991; Tanaka, 1996). Ces différences de sélectivité pourraient servir notamment à affiner la représentation du stimulus au sein de la colonne.

1.4. Inter-connectivité dense

La connectivité entre les aires visuelles corticales (connexions *extrinsèques*) ou au sein d'une même aire corticale (connexions *intrinsèques*) a principalement été investiguée par des études anatomiques grâce à l'utilisation d'agents de traçage. Ces études ont révélé qu'outre les connexions ascendantes (feedforward) entre les aires visuelles voisines dans la hiérarchie visuelle, ces aires visuelles sont également généralement reliées par des connexions descendantes (feedback). En fait, une large partie des connexions entre aires visuelles est constituée de ces connexions en feedback, qui peuvent posséder des propriétés de vitesse de conduction de l'influx nerveux similaires aux connexions ascendantes (Girard et al., 2001).

De plus, en dehors des connexions entre aires visuelles voisines, il existe un pattern très complexe et dense de connexions (à la fois ascendantes et descendantes) entre des aires visuelles plus distantes (Figure 2 et Figure 4) mais également entre les aires visuelles corticales et les noyaux sous-corticaux thalamiques relayant l'information depuis la rétine vers le cortex visuel (Felleman & Van Essen, 1991). Ces nombreuses connexions existent tant au sein de la voie ventrale et de la voie dorsale qu'entre ces deux voies (Webster et al., 1994; Zhong & Rockland, 2003). Par exemple, l'aire V4 est fortement connectée avec l'aire TEO (située entre V4 et le IT dans la hiérarchie anatomique visuelle - Figure 4), et la partie postérieure de IT dans la voie ventrale mais est également connectée avec les aires du cortex pariétal (Ungerleider et al., 2007). Il existe des connexions directes entre le cortex inféro-temporal dans la voie ventrale et le cortex pariétal dans la voie dorsale (Zhong & Rockland, 2003) ou via un relais important au niveau de la partie antérieure du sillon temporal supérieur (STS) (Harries & Perrett, 1991), qui contient notamment des neurones intégrant partiellement l'information traitée dans les deux voies (par exemple: un codage du mouvement des yeux

[6] Les propriétés spatio-temporelles du champ récepteur correspondent à la manière dont est répartie la sélectivité du neurone dans l'espace de son champ récepteur; sélectivité qui peut également varier en fonction du temps écoulé par rapport à une stimulation visuelle.

d'un visage; Perrett et al., 1985a; 1992; ou du mouvement biologique: Jellema et al., 2004). Il a également été proposé que ces connexions entre voies dorsale et ventrale forment un réseau impliqué dans la perception des objets en trois dimensions (Sereno et al., 2002).

Chez l'humain, le pattern de connectivité des aires visuelles commence à être investigué à l'aide d'une technique de neuroimagerie non-invasive: la DTI ("diffusion tensor imaging"), basée sur la localisation du déplacement (diffusion) anisotrope des molécules d'eau le long des fibres myélinisées (Le Bihan et al., 2001). La technique de DTI a, par exemple, récemment révélé l'existence de connexions directes entre l'aire V3 et une aire située dans le gyrus fusiforme médian activée plus fortement pour les visages que pour les objets (FFA: "fusiform face area") ou encore entre la FFA et le LOC ("Lateral Occipital Complex" – voir Grill-Spector et al., 2001) impliquée dans le traitement de formes visuelles complexes (Kim et al., 2006; Figure 6).

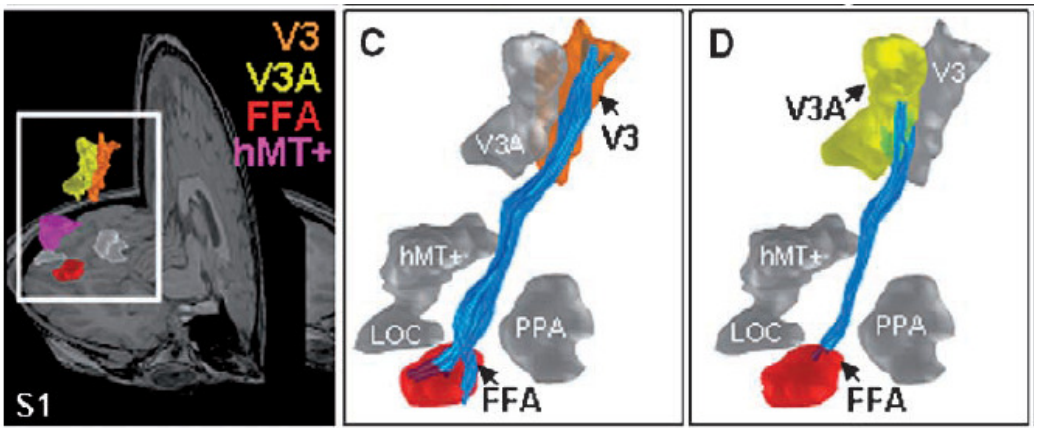

Figure 6: Représentation des connexions directes entre l'aire V3 et la FFA (C) et l'aire V3A (D) et la FFA révélée par la DTI. Les régions sont représentées à gauche sur une image anatomique. Modifié de Kim et al., 2006.

Un autre type de connexions qui représente un pourcentage élevé de la totalité des connexions cortico-corticales sont les connexions horizontales entre les neurones d'une même aire corticale. Ces connexions sont dites *intrinsèques*. Ces connexions peuvent être excitatrices ou inhibitrices et relient fréquemment des colonnes corticales présentant la même sélectivité (voir Tanaka, 1996). Dans IT, il a été proposé que les connexions horizontales inhibitrices jouent notamment un rôle dans la précision de la sélectivité des neurones pour des formes complexes (Wang et al., 2000).

L'observation de l'existence de ces multiples connexions extrinsèques bidirectionelles et intrinsèques, ainsi que le fait qu'elles représentent environs 30% du total des connexions

dans le système visuel (voir Young, 2000) a fortement remodelé la conception de la perception visuelle comme procédant de manière exclusivement ascendante (feedforward) entre la rétine et les niveaux plus élevés de la hiérarchie visuelle (voir Young, 2000 ; Bullier & Nowak, 1995). Des conceptions plus compatibles avec l'organisation anatomique et physiologique du système visuel suggèrent un rôle dominant des connexions bidirectionnelles extrinsèques et des connexions intrinsèques (aussi appelées connexions horizontales) dans toutes les étapes de l'élaboration des représentations visuelles, depuis la ségrégation figure-fond ou la segmentation jusqu'à l'élaboration de représentations visuelles complexes (Bullier, 2001; Bullier & Nowak, 1995; Hochstein & Ahissar, 2002; Lee et al., 1998; Lamme & Roelfsema, 2000; Huppé et al., 2001; Deco & Rolls, 2004). Il a d'ailleurs été proposé que les effets délétères sur la perception d'un stimulus lorsque le stimulus est présenté rapidement et suivi d'un masque visuel pourraient être partiellement liés à une perturbation de ces interactions bidirectionnelles récurrentes (Lamme & Roelfsema, 2000).

1.5. Champs récepteurs

Le champ récepteur d'un neurone du système visuel correspond à la portion du champ visuel à laquelle ce neurone est sensible. Les propriétés du champ récepteur des neurones varient fortement d'un type de neurone à un autre. Ainsi par exemple, le champ récepteur des cellules ganglionnaires de la rétine est concentrique, composé d'une partie centrale ("center") et d'une partie périphérique ("surround"). Ces cellules répondent de manière opposée (antagoniste), à une stimulation du centre et à une stimulation de la périphérie du champ récepteur. La plupart des neurones du cortex visuel ont un point de leur champ récepteur dont la sensibilité est maximale; l'intensité de la réponse diminuant graduellement à mesure que la stimulation s'éloigne de ce point.

Une propriété fondamentale des champs récepteurs est que leur taille augmente d'un facteur d'environ 2.5 à chaque étape de la hiérarchie visuelle. On observe ainsi que les neurones de V1 répondent en moyenne à une portion de 1° du champ visuel alors que les champs récepteurs dans l'aire TE du cortex inféro-temporal peuvent atteindre plus de 50° d'angle visuel (Figure 7). Bien qu'une taille aussi importante des champs récepteurs présente l'avantage de pouvoir représenter les objets avec une certaine invariabilité à la taille ou à la position dans le champ visuel, une telle propriété génère également un problème lorsque plusieurs objets apparaissent simultanément dans le champ récepteur d'un neurone

(Desimone, 1998). En effet, il devient alors difficile pour le neurone de déterminer à quel des objets présentés appartient telle ou telle caractéristique visuelle (Gray, 1999).

Notons par ailleurs que la taille des champs récepteurs peut fortement varier en fonction du type de cellules au sein d'une aire visuelle donnée et en fonction de la position de son champ récepteur dans le champ visuel (voir Gross et al., 1981).

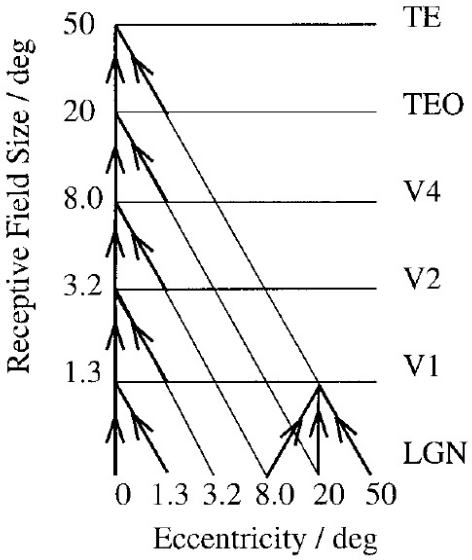

Figure 7: Représentation schématique de la convergence progressive des champs récepteurs au cours des étapes du traitement visuel. La taille moyenne des champs récepteurs augmente à chaque étape. Modifié de Rolls, 2000.

Une autre caractéristique importante du champ récepteur, qui est directement liée à l'existence de connexions horizontales et feedback, est que la sensibilité des neurones ne se réduit pas uniquement à la stimulation présente dans leur champ récepteur. En effet, si les neurones de V1 ont longtemps été considérés comme étant de simples filtres chargés de représenter des variations locales de luminance dans leur champ récepteur, il est maintenant clair que ces neurones intègrent l'information provenant à la fois de leur champ récepteur et des régions entourant ce champ récepteur (Allman, Miezin & McGuiness, 1995; Lamme, Super & Spekerijse, 1998), participant ainsi au phénomène de complétion de contour, de segmentation d'un objet, de ségrégation figure-fond,… (Lee et al., 1998; Lee & Nguyen, 2001; Lamme et al., 1998). Une démonstration bien connue de ce phénomène a été réalisée par Lamme et son équipe (Lamme et al., 1999) à propos du phénomène de la ségrégation figure-fond. Ils ont montré que la réponse des neurones de V1 à un stimulus texturé est influencée par le fait que ce stimulus correspond ou non à une figure se détachant d'un fond (Figure 8).

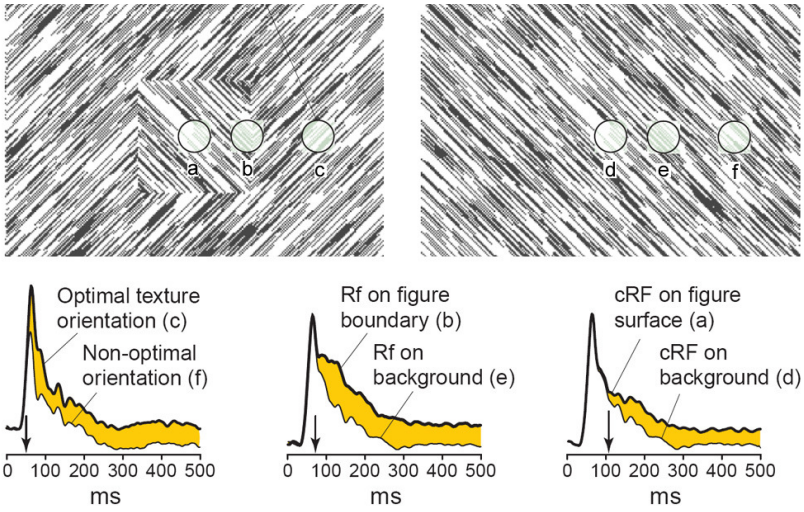

Figure 8: Les cellules de V1 sont sensibles de manière séquentielle à différents aspects du stimulus (a) les réponses neuronales sont comparées selon la position des champs récepteurs (cercles, marqués de a à f) sur le stimulus texturé. La comparaison des réponses c et f montre que la sélectivité à l'orientation de la cellule se manifeste vers 55 ms après l'apparition du stimulus (la flèche indique la différence significative la plus précoce entre les courbes). Ces cellules peuvent signaler la frontière entre la figure et le fond (définis par la différence d'orientation de texture) (b versus e) à 80 ms. Elles augmentent également leur réponse lorsque le champ récepteur couvre la surface de la figure, par rapport à la surface du fond (a versus d). Cet effet émerge à 100 ms, malgré la stricte identité du contenu des champs récepteurs dans les deux situations. Modifié d'après Lamme & Roelfsema, 2000.

D'autres exemples d'influence du pourtour extérieur du champ récepteur sur la réponse des neurones sont illustrés par le fait que la taille du champ récepteur des neurones des IT peut varier selon qu'un objet présent dans le champ récepteur est situé soit sur un fond complexe soit sur une surface uniforme, ou selon que d'autres stimuli sont présents simultanément dans le champ visuel (Rolls et al., 2003).

1.6. Rétinotopie

Les structures du système visuel de la voie ventrale depuis les corps genouillés latéraux jusqu'aux aires V4 et TEO respectent l'organisation rétinotopique selon laquelle il y a une correspondance entre la localisation d'un neurone dans la structure et la position de son champ récepteur dans le champ visuel. Des neurones adjacents ont donc des champs récepteurs qui incluent des portions adjacentes et partiellement superposées du champ visuel. Les structures rétinotopiques forment ainsi une carte -complète ou non- du champ visuel (carte rétinotopique). Les propriétés de rétinotopie des aires visuelles sont notamment utilisées pour cartographier et délimiter ces aires (V1, V2, V3, VP, V3A, V4v,…) chez l'homme par l'intermédiaire de la IRMf. Cependant, la rétinotopie stricte est principalement

présente au niveau des noyaux sous-corticaux (CGL et colliculus supérieur) et du cortex visuel primaire. A partir de V2 la rétinotopie se complexifie et s'estompe peu à peu pour quasiment disparaître dans la partie antérieure du cortex inféro-temporal. Notons également que certaines aires visuelles rétinotopiques ne représentent qu'une partie du champ visuel.

2. Le cortex inféro-temporal et la reconnaissance visuelle

Le cortex inféro-temporal est considéré comme le sommet de la hiérarchie dans la voie ventrale, impliqué principalement dans les traitements visuels menant à la reconnaissance des objets (Tanaka, 1996). Le cortex IT est considéré comme la dernière aire exclusivement visuelle de la voie ventrale.

Nous verrons ci-dessous que c'est au niveau de IT qu'une représentation intégrée des objets apparaît pour la première fois dans le système visuel. Nous commencerons par décrire brièvement l'architecture de IT et son rôle général dans l'élaboration des représentations visuelles des objets. Nous examinerons ensuite plus en détail les propriétés d'une population de neurones de IT répondant spécifiquement à la vue d'un visage. Ces neurones ont en effet un rôle particulièrement important dans la catégorisation et l'identification visuelle des visages. Notons que nous ferons principalement référence aux travaux réalisés par enregistrement de neurones dans le cortex du singe. En effet, même si la correspondance entre les régions identifiées chez le singe et les régions dentifiées chez l'humain (par la IRMf principalement) n'est qu'approximative, la majorité des connaissances actuelles sur les propriétés du cortex visuel des primates provient de ces enregistrements cellulaires réalisés chez le singe.

Même si nous nous intéressont plus particulièrement au décours temporel de la perception des visages, les données présentées dans les sections suivantes sont néanmoins particulièrement importantes pour notre propos. En effet, les propriétés anatomiques et fonctionelles des structures cérébrales accueillant les représentations de ces stimuli complexes sont déterminantes dans la manière dont sont temporellement mises en jeu ces représentations pour générer une décision perceptive rapide.

2.1. Architecture du cortex inféro-temporal

Chez le singe, le cortex IT débute au niveau du sillon occipital inférieur et s'étend antérieurement pour s'arrêter juste avant le pôle temporal qui ne fait pas partie de IT. Sur l'axe

dorso-ventral, il s'étend dans le gyrus inféro-temporal et comprend également la partie ventrale et le fond du sillon temporal supérieur (STS). À l'origine, le cortex IT avait été divisé en une région postérieure (TEO ou pIT: "posterior IT") et une région antérieure (TE, comprenant les régions cIT: "central IT" et aIT: "anterior IT"). Sur base de critères cyto -et myelo - architectoniques et de patterns de connexions afférentes, TE est encore subdivisée en TEav, TEad, Tem, TE1, TE2, TE3 (Figure 9) (Baylis et al., 1987; voir aussi Van Essen, 2003). Des neurones sensibles à des stimuli visuels complexes sont également observés dans la partie dorsale et la lèvre supérieure du STS au niveau des aires TPO, PGa, IPa et IPg. Cette partie du STS a été appelée *aire polysensorielle temporale supérieure* (STP; Bruce et al., 1981) en raison du fait qu'elle contient des neurones répondant à plusieurs modalités sensorielles (visuelle, auditive et somatosensorielle). Dans la suite de ce document, nous utiliserons le terme « IT » pour nous référer aux aires TE et du STS ventral.

Les principales afférences visuelles de IT proviennent de la hiérarchie visuelle sérielle V1-V2-V4-TEO-TE (voir Rolls, 2000; Tanaka, 1996). Il existe cependant des connexions directes depuis les aires V1, V2 et V3 vers TEO, ainsi qu'entre V4 et la partie postérieure de TE. Les aires TE et TEO sont également fortement interconnectées avec le cortex pariétal (Webster et al., 1991; Zhong & Rockland, 2003).

La plupart des projections de IT vers des structures corticales et sous corticales non-visuelles telles que le cortex préfrontal, l'amygdale, l'hippocampe, le cortex perirhinal ou le striatum se font à partir de TE (Webster et al., 1991), plus particulièrement dans sa région antérieure. Il existe également de très nombreuses connexions bidirectionnelles entre le cortex inféro-temporal et les aires du STP dans la partie antérieure du STS, qui sont elles-mêmes connectées avec le cortex pariétal (Harries & Perrett, 1991; Webster et al., 1994).

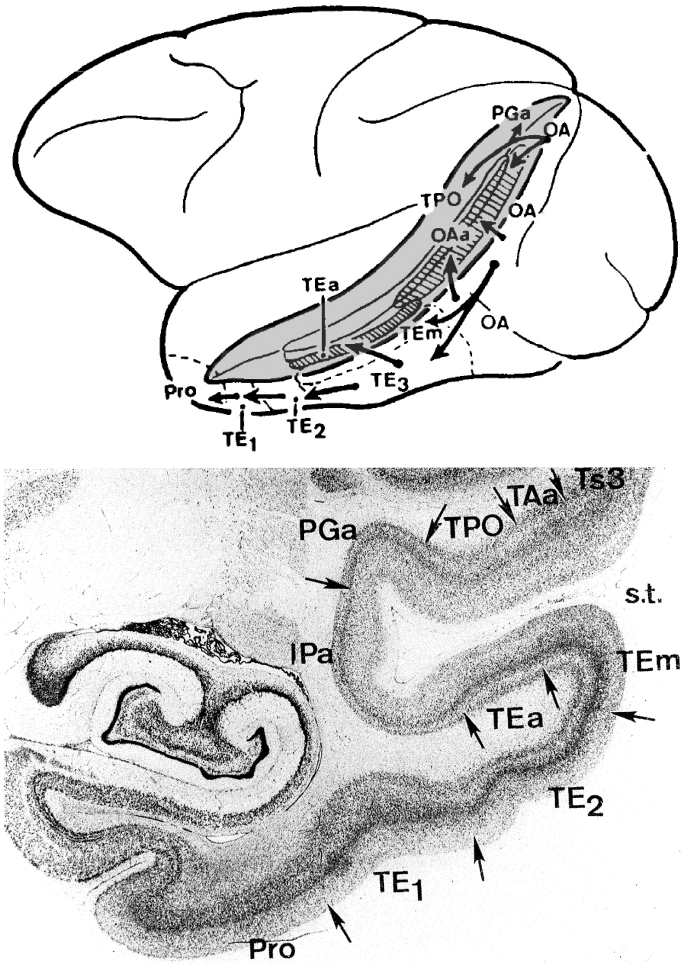

Figure 9: Au dessus: vue latérale du cerveau d'un singe macaque montrant les différentes aires du cortex inféro-temporal et du sillon temporal supérieur (STS). Le STS a été "ouvert" pour montrer les aires normalement enfouies dans le sillon (partie grisée). En dessous : coupe coronale au niveau du cortex inféro-temporal et du STS indiquant les différentes région de IT. Modifié d'après Rolls, 2000.

2.2. Le rôle du cortex inféro-temporal dans la reconnaissance visuelle

Les premières évidences que le cortex inféro-temporal est impliqué dans la reconnaissance visuelle proviennent de l'observation que des lésions de cette région provoquent des agnosies visuelles chez le singe (voir Gross, 1994, 1998 pour une revue de la littérature). L'implication de cette région dans la reconnaissance visuelle a été confirmée au début des années 1970 par l'équipe de Charles Gross (Gross et al., 1969; 1972) qui réalisa les premiers enregistrements cellulaires dans le cortex inféro-temporal chez le singe. Ces premières études en neurophysiologie chez le singe ont été suivies par de nombreux travaux dans les années 1980 et 1990 (dont les plus nombreuses ont été menées par les équipes de

Dave Perrett; Edmund Rolls; Keija Tanaka; Charles Gross; Robert Desimone ; Bruce et al., 1981; Perrett et al., 1982; 1984; 1985a; 1988; 1992; Desimone et al., 1984; Tanaka et al., 1991; Kobatake & Tanaka, 1994; Fujita et al., 1992; Wang et al., 1996; Rolls et al., 1977; 1982, etc.) qui ont permis de révéler un grand nombre de propriétés fonctionnelles et anatomiques des neurones du cortex inféro-temporal. Il existe de nombreuses revues de la littérature qui résument ces propriétés (Tanaka, 1992; 1996; 1997; 2003; Logothetis & Sheinberg, 1996; Desimone, 1991; Rolls, 1992; 2000; 2007; Perrett et al., 1992; Gross, 1992).

L'apparition des techniques d'imagerie cérébrale fonctionnelle non-invasive (PET-scan et IRMf) a largement confirmé la contribution des régions inféro-temporales dans la reconnaissance des stimuli visuels chez l'humain (Sergent et al., 1992; Kanwisher et al., 1997; Grill-spector et al., 1999; Tootel et al., 1996; Haxby et al., 1999; 2000; Rossion et al., 2003b; Grill-spector & Malach, 2004; Grill-spector et al., 2001) et plus récemment chez le singe (Tsao et al., 2003; 2006; Pinsk et al., 2005; Logothetis et al., 1999; Sawamura et al., 2005; voir aussi Orban et al., 2004).

2.2.1. Propriétés générales des neurones de IT

Dans cette section nous allons passer en revue un certain nombre de propriétés générales des neurones du cortex inféro-temporal mises en évidence par les enregistrements cellulaires chez le singe. La sélectivité des neurones de IT pour des formes complexes en font une région particulièrement adaptée à la l'élaboration d'une représentation pouvant mener à la reconnaissance des objets et des visages.

a. Sensibilité à des formes modérément complexes

L'étude des mécanismes de la reconnaissance visuelle dans IT nécessite de pouvoir déterminer le profil de sélectivité des neurones. En d'autres termes, pour connaître ce qu'un neurone encode comme information à propos d'un stimulus, il est impératif de repérer les caractéristiques primaires nécessaires qu'un stimulus doit présenter pour activer ce neurone (Tanaka, 1996; 2003). En effet, de nombreux neurones répondent à des stimuli apparemment très différents mais dans lesquels il est souvent possible de trouver une ou plusieurs caractéristiques communes. De manière à pouvoir investiguer précisément les propriétés de réponse et de sélectivité des cellules de IT, l'équipe de Keiji Tanaka a appliqué une méthode dite "de réduction" (initialement développée par l'équipe de Charles Gross: voir Desimone et al., 1984) permettant de déterminer les caractéristiques critiques qu'un stimulus doit posséder

pour provoquer la réponse d'un neurone donné (Tanaka et al., 1991; Fujita et al., 1992). Cette méthode consiste, pour chaque neurone, à présenter un grand nombre de stimuli dans le champ récepteur du neurone afin de déterminer le stimulus générant la plus forte réponse. Ce stimulus est ensuite graduellement "réduit" en composants géométriques plus simples mais néanmoins efficaces à générer une réponse. Le stimulus critique pour le neurone est celui qui présente la combinaison minimale (la plus simple) de ces composants géométriques générant une réponse maximale[7] (Figure 10). Ce stimulus critique est donc en général celui qu'on ne peut simplifier davantage sans réduire substantiellement la réponse du neurone. Après avoir déterminé le stimulus critique pour un neurone, celui-ci est encore manipulé pour examiner plus en détail le profil de sélectivité du neurone. La détermination du stimulus critique de centaines de neurones dans la partie dorsale de TEO et TE (Tanaka et al., 1991) (principalement dans le gyrus temporal inférieur et la lèvre inférieure du STS) a révélé que la plupart des cellules de TE nécessitent un stimulus modérément complexe (Figure 11) pour être activées de manière maximale mais que ces stimuli étaient trop peu complexes pour représenter des objets entiers. De plus, ces enregistrements ont révélé que les neurones de TE nécessitent un stimulus de plus forte complexité pour être activés que les neurones de TEO ou de V4. La sélectivité des neurones de IT n'est cependant ni absolue ni rigide. En effet, les neurones de IT modifient leur sélectivité suite à des apprentissages perceptifs à discriminer des stimuli similaires (chez le singe: Logothetis et al., 1994; 1995; Sigala & Logothetis, 2002; Freedman et al., 2006; Suzuki et al., 2006; chez l'humain: Schiltz et al., 1999; Kourtzi et al., 2005). La sélectivité des neurones peut même s'adapter de manière dynamique à la distribution des propriétés qui distinguent un set de stimuli, avec une sélectivité plus précise lorsque les stimuli du set sont plus similaires entre eux (De Baene et al., 2007).

[7] D'autres méthodes basées sur une présentation rapide et séquentielle (Rapid serial visual presentation : RSVP) d'un grand nombre de stimuli ont été utilisées plus récemment pour examiner la sélectivité des neurones de IT de manière plus rapide et peut-être plus systématique que la méthode de réduction développée par l'équipe de Tanaka (Keysers et al., 2001; De Baene et al., 2007; Foldiak et al., 2004).

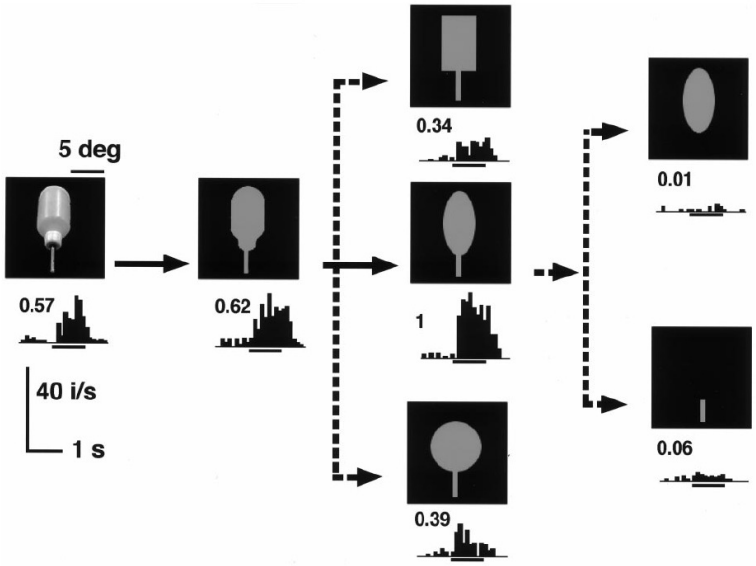

Figure 10: Exemple de procédure de réduction pour déterminer le stimulus critique pour un neurone. Les stimuli utilisés sont présentés avec la réponse qu'ils génèrent dans un neurone de TE (histogramme montrant le taux de décharge au cours du temps). Le chiffre à côté de chaque histogramme représente l'intensité de la réponse par rapport à la réponse maximale que génère le stimulus (ici le stimulus central: 1). (Wang et al., 1998).

L'observation que les cellules de IT répondent à des stimuli complexes mais néanmoins plus simples que des objets entiers a été un argument important pour défendre l'idée qu'il n'existe pas nécessairement dans le cortex inféro-temporal de neurones qui présenteraient une sélectivité très importante leur permettant de répondre spécifiquement à un objet défini ou à un visage unique. L'existence de telles cellules, appelée unités "gnostiques" (Kornoski, 1967), encore appelées cellules "grand-mère" (voir Gross, 2002 pour une historique du terme "grandmother cell"), a été postulée dans le cadre du modèle hiérarchique de la vision. Dans un version "extrême" du modèle hiérarchique de la vision, les neurones à la base de la hiérarchie codent des propriétés simples de la stimulation et convergent vers un très petit nombre de neurones voire un neurone unique capable de reconnaître un objet unique (Barlow, 1972). Nous verrons plus loin que le codage des objets dans IT semble plutôt être distribué sur un nombre important de neurones.

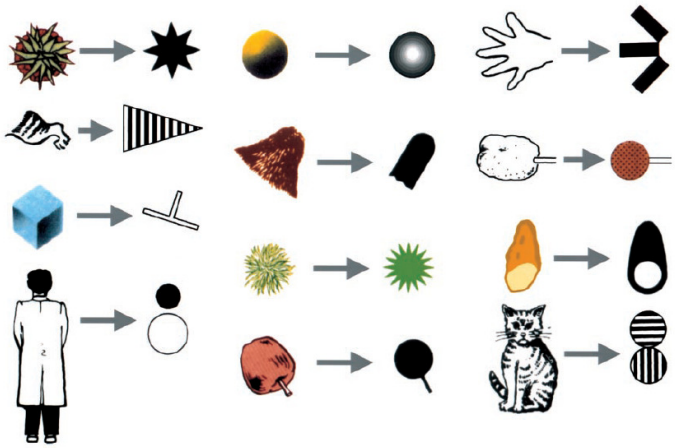

Figure 11: Exemples de stimuli critiques "modérément complexes" obtenus après réduction pour 12 cellules de TE. L'image de gauche représente l'image originale de l'objet générant la plus forte réponse pour chaque cellule et le stimulus de droite représente le stimulus critique déterminé par la méthode de réduction. (Tanaka, 2003).

b. Invariance et dépendance

Une caractéristique importante de la reconnaissance visuelle est la capacité à catégoriser et reconnaître les objets malgré des changements très importants dans les conditions de stimulation rétinienne (changement de taille, de position dans le champ visuel, d'illumination,…). Les neurones de IT semblent posséder les propriétés fonctionnelles nécessaires à l'accomplissement de la reconnaissance invariante à ces changements visuels.

Les neurones du cortex inféro-temporal ont généralement des champs récepteurs très larges (en moyenne de 20° à 70°) qui incluent quasi toujours la fovéa. La taille importante du champ récepteur permet à ces neurones de répondre sélectivement à certains stimuli quelle que soit la *position* du stimulus dans le champ visuel. Cependant, l'amplitude de la réponse est maximale au centre du champ récepteur (généralement la fovéa[8]) (Ito et al., 1995; Op de Beek & Vogels, 2000; Tovee et al., 1994) et l'amplitude de la réponse décroît à mesure que l'on s'éloigne du centre du champ récepteur (Ito et al., 1995; Op de Beek & Vogels, 2000; Rolls et al., 2003). Ce profil de réponse pourrait permettre de coder la position du stimulus, en particulier pour les stimuli apparaissant dans la région centrale du champ visuel (Op de Beek & Vogels, 2000). Notons également que la taille des champs récepteurs peut varier énormément lorsque plusieurs stimuli sont présents simultanément dans le champ récepteur ou lorsqu'un stimulus est présenté dans un contexte de scène naturelle (Rolls et al., 2003). Il a

[8] Ce phénomène est généralement attribué à la plus forte représentation de la fovéa dans le système visuel par rapport au champ visuel périphérique. Cette "magnification corticale" de l'information fovéale permet notamment d'assurer que les stimuli apparaissant à la fovéa soient traités préférentiellement par rapport aux stimuli apparaissant à la périphérie (Rolls & Tovee, 1995a; Sheinberg et Logothetis, 2001; Rolls et al., 2003)

été proposé que cette modification de la taille des champs récepteurs pourrait avoir un rôle important dans la sélection des objets à traiter dans IT (Deco & Rolls, 2004; Rousselet et al., 2004b). Malgré cette variation dans l'amplitude absolue de la réponse neuronale en fonction de la position dans le champ récepteur, la sélectivité des neurones est largement préservée sur l'ensemble du champ récepteur (Ito et al., 1995; Wang, 1998).

Une large proportion des cellules de IT présente une sélectivité relativement invariante à des changements importants de la *taille* du stimulus (par ex: plus de 4 octaves, Ito et al., 1995) (Tanaka et al., 1991; Ito et al., 1995; Rolls & Baylis, 1986; Lueschow et al., 1994; Sary et al., 1993). D'autres neurones tolèrent des changements de taille beaucoup plus restreints (moins de 2 octaves, Ito et al., 1995), et certains ne répondent de manière maximale qu'à des stimuli ayant une taille définie[9] (Tanaka et al., 1991; Lueschow et al., 1994 ; Ashbridge et al., 2000) (Figure 12). Ainsi, les réponses des neurones de IT permettent d'encoder à la fois les caractéristiques du stimulus présenté (indépendamment de leur taille) et certaines modifications apportées au stimulus (position, taille,…).

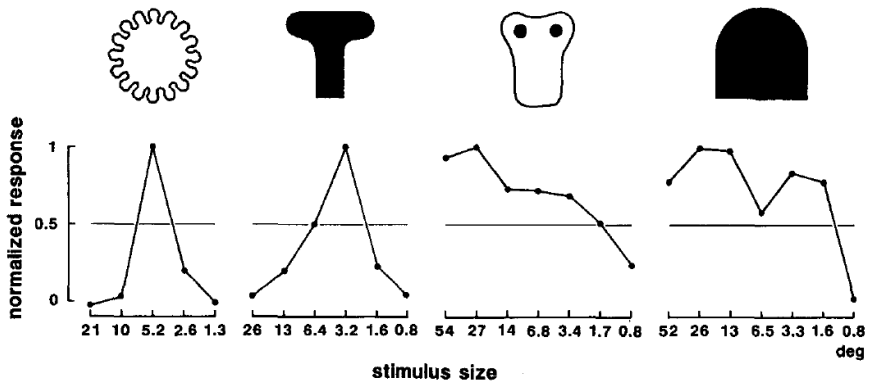

Figure 12: Sensibilité de 4 cellules de TE à la taille du stimulus. Certaines cellules de TE sont fortement sensibles à la taille (à gauche) et d'autres moins (à droite). (Tanaka, 1996).

La caractéristique première des neurones de IT est qu'ils codent la forme des stimuli. Ainsi, Ito et al. (1994) ont observé qu'un large pourcentage (40%) des cellules enregistrées répondent encore à plus de 50% de leur réponse maximale après l'inversion du contraste du

[9] Il est intéressant de constater qu'un grand nombre de neurones répondant préférentiellement au corps humain dans le STS sont activés de manière maximale lorsque qu'un corps humain de taille humaine est présenté (Ashbridge et al., 2000). Ces cellules déchargent de manière moindre lorsque le stimulus est réduit en taille. Cependant, certaines de ces cellules présentent également une relative invariance à la taille du stimulus. Il serait dès lors intéressant de voir s'il y a un lien entre la nature du stimulus et la sélectivité à la taille des neurones qui code pour ce stimulus. Il est en effet possible que la taille d'un stimulus soit un élément diagnostique de sa catégorie d'appartenance ou que cette propriété soit particulièrement informative par rapport à la distance à laquelle se situe ce stimulus par rapport à l'observateur.

stimulus critique pour ces cellules (voir aussi Rolls & Baylis, 1985). De plus, certaines cellules de IT préservent leur sélectivité pour un stimulus, que la forme soit définie par la luminance, le pattern du mouvement de textures ou encore la disparité binoculaire (Sary et al., 1993).

En contraste avec la relative invariance de la réponse des neurones de IT au changement de position dans le champ visuel, de taille ou de contraste, un changement d'orientation dans le plan vertical ou horizontal a généralement un effet important sur la réponse de ces neurones (Tanaka et al., 1991; voir aussi Gross, 1992).

c. Organisation modulaire en colonnes

Nous avons vu que le regroupement sur un axe perpendiculaire au cortex de neurones présentant une sélectivité similaire (colonnes corticales) est une propriété que l'on retrouve dans la majorité des aires corticales visuelles. Alors que les colonnes corticales de V1 contiennent des neurones ayant une sélectivité pour une orientation similaire (mais étant sensibles à des paramètres différents de fréquence spatiale, de phase,…), les neurones de IT sont rassemblés en colonnes sur base de leur sélectivité à un stimulus possédant des caractéristiques similaires. Cette organisation de IT a été mise en évidence par les travaux de l'équipe de Tanaka (Fujita et al., 1992), qui ont montré que lorsqu'une électrode est introduite de manière verticale dans le cortex (pénétration perpendiculaire au cortex), les neurones enregistrés le long de cette pénétration répondent à un stimulus critique (après réduction) semblable (Figure 13). En revanche, lors d'une pénétration oblique d'une électrode, le profil de sélectivité des neurones pour un stimulus critique change environ tout les 400 μm, révélant ainsi le diamètre approximatif des colonnes.

L'organisation en colonnes a été confirmée par la même équipe grâce à des enregistrements de l'activité des neurones de IT par la technique d'imagerie optique[10] (Wang et al., 1996; 1998). Dans l'étude originale (Wang et al., 1996), le stimulus critique pour une série neurones était d'abord déterminé et ensuite l'activité de la région où se trouvait chacun des neurones était enregistrée par imagerie optique. En général, la présentation du stimulus critique provoquait l'activation de plusieurs petites régions (entre 2 et 8 régions) d'une taille d'environs 500 μm et comprenant le site d'enregistrement du neurone pour lequel le stimulus

[10] L'imagerie optique consiste à illuminer la surface corticale d'une lumière infra-rouge et à capter la réflexion de cette lumière par une caméra CCD. Les régions corticales activées consommant plus d'oxygène reflètent moins bien la lumière et apparaissent ainsi comme des zones sombres sur l'image captée par la caméra. L'avantage de l'imagerie optique est qu'elle permet d'appréhender de manière simultanée le fonctionnement d'une zone corticale entière (Wang et al., 1998; Ts'o et al., 1990).

critique avait été déterminé (Figure 14). Le fait que seule l'activation d'un grand nombre de neurones localisés à un même endroit peut produire une réponse observable par imagerie optique confirme que les neurones avec une sélectivité similaire pour un stimulus critique sont rassemblés en colonnes.

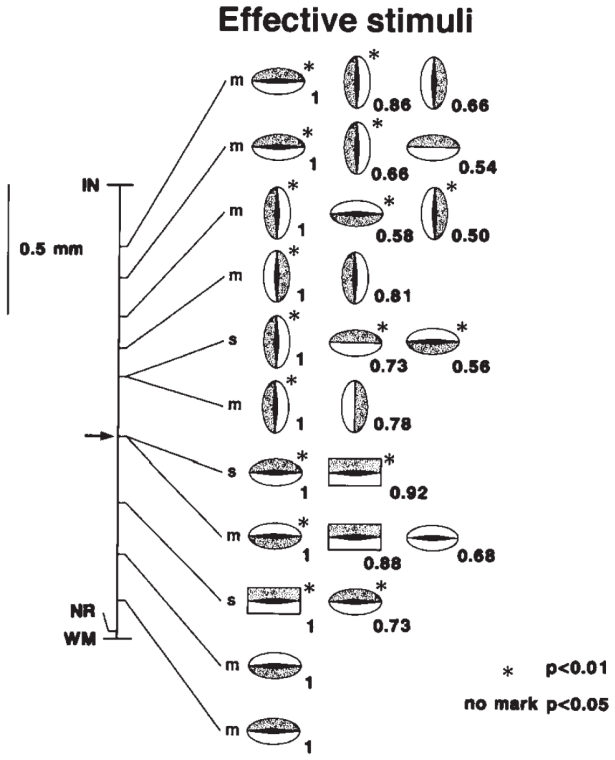

Figure 13: Organisation en colonnes des cellules de TE présentant une sélectivité pour des stimuli critiques similaires. La figure montre les variations dans le stimulus critique pour les différentes cellules localisées le long d'une même pénétration verticale. Le stimulus générant la réponse la plus forte pour une série de cellules situées le long d'une même pénétration verticale est présenté dans la colonne de droite. Le chiffre à côté de chaque stimulus représente l'intensité de la réponse par rapport à la réponse maximale que génère le stimulus. Modifié d'après Tanaka, 1996.

Une autre observation intéressante réalisée au moyen de l'imagerie optique par l'équipe de Tanaka est que des stimuli critiques différents mais possédant des caractéristiques communes (par ex: changement de luminosité entre les parties du stimulus), activent des régions corticales -chacune d'un diamètre d'environ 500 µm- qui se recouvrent partiellement (Figure 14). Dans cette situation, chacune des régions activées par les différents stimuli comprennent le site où une électrode a été placée préalablement (pour déterminer le stimulus critique des neurones), mais ces régions s'étendent dans des directions différentes, suggérant qu'il existe une forme d'espace de représentation continu pour certaines caractéristiques visuelles complexes dans IT.

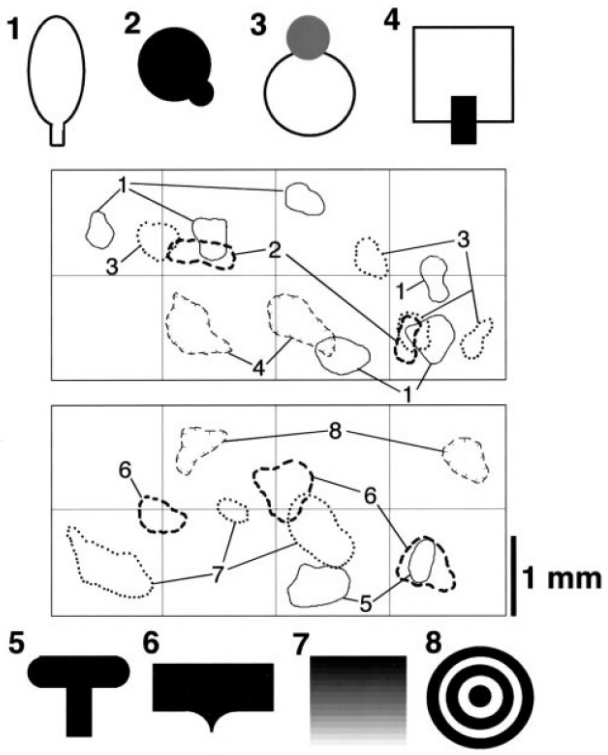

Figure 14: Chaque stimulus représenté (1 à 8) active plusieurs petites régions (colonnes) dans le cortex IT. Les cartes sont obtenues par imagerie optique lors de la présentation des stimuli. On voit que les régions activées par différents stimuli se recouvrent partiellement. Modifié d'après Tanaka, 2003.

d. Représentation des objets dans IT

Nous avons vu que les neurones de IT encodent des stimuli modérément complexes qui ne sont ni des objets entiers, ni des parties d'objets. De plus, les observations de l'équipe de Tanaka indiquent d'une part qu'un stimulus modérément complexe active plusieurs colonnes corticales et d'autre part que des stimuli différents activent des régions qui se superposent en partie (Figure 14). Cette propriété particulière du codage des objets visuels dans IT a constitué un des arguments pour soutenir l'hypothèse que les objets sont représentés de manière distribuée dans le cortex. Cette hypothèse a été également soutenue par l'observation que la réduction systématique d'un objet ou des parties d'un objet en traits plus simples provoque généralement une diminution du nombre de colonnes activées par rapport à l'objet entier (Tsunoda et al., 2001). En outre, la présentation des traits simplifiés d'un objet active fréquemment des colonnes qui ne sont initialement pas activées par l'objet entier. Ces résultats suggèrent que la représentation d'un objet entier dans IT aurait comme base non

seulement l'activation conjointe de plusieurs colonnes corticales chacune codant pour une caractéristique plus simple de l'objet, mais également l'inhibition de neurones présents dans d'autres colonnes répondant seulement lorsque un trait simplifié de l'objet est présenté. La représentation d'un objet dans IT nécessite donc à la fois des connexions horizontales excitatrices entre colonnes distantes activées par un même objet (>1 mm) ainsi que des connexions inhibitrices qui n'existent que sur de courtes distances de moins de 1 mm. L'implication des connexions inhibitrices (via interneurones GABAergiques) dans la sélectivité des neurones de IT à été mise en évidence chez le singe par l'injection dans IT d'une substance (la bicuculline) bloquant cette inhibition (Wang et al., 2000). L'effet principal de l'injection de la bicuculline a été de diminuer la sélectivité de 75% des neurones (i.e. chaque neurone répondait à un nombre plus élevé de stimuli). Le résultat le plus intéressant était qu'après injection de bicuculline, certains neurones répondaient à des stimuli auxquels ils n'étaient normalement pas sensibles mais qui étaient apparentés sur certaines dimensions (par exemple : couleur ou forme globale) aux caractéristiques du stimulus préférentiel pour ces neurones dans des conditions normales. Pour prendre un exemple, si un neurone répondait normalement de manière maximale à un rond blanc sur fond noir ; sous bicuculline, ce neurone devenait également fortement sensible à d'autres stimuli possédant une forme ronde ou ovale et qui générait peu à pas de réponse dans des conditions normales. Cette observation était particulièrement vraie pour des neurones situés le long d'une même pénétration ou à une distance de moins de 0.7 mm l'un de l'autre, indiquant que les connexions inhibitrices entre neurones au sein d'une même colonne ou entre neurones de colonnes voisines ont un rôle crucial dans la détermination de la sélectivité neuronale pour des stimuli visuels dans TE (Wang et al., 2000; Tanaka, 2003; Tamura et al., 2004).

Outre ces connexions inhibitrices, il existe également des connexions excitatrices entre neurones situés dans une même colonne. Ces connexions pourraient permettre de propager l'activation d'un ou plusieurs neurones à une population plus importante de neurones proches codant pour des stimuli critiques similaires mais légèrement différents sur certaines dimensions (différence d'orientation, de couleur, de polarité du contraste, etc), générant ainsi une représentation visuelle invariante à des changements physiques de la stimulation comme l'orientation ou l'illumination (Tanaka, 1996; 2003). Il est par ailleurs probable que des colonnes codant pour des caractéristiques différentes mais néanmoins conjointement présentes dans de nombreux objets entretiennent des liens excitateurs, ce qui permettrait de générer une représentation complète d'un objet même si certaines parties de l'objet ne sont pas visibles.

Le pattern d'excitation/inhibition entre des neurones proches présentant une sélectivité légèrement différente permettrait donc à la fois de générer des représentations très précises nécessaires à la discrimination fine entre objets (Tanaka, 1996; 2003; Wang et al., 2000) ainsi que des représentations permettant de catégoriser comme équivalentes des stimulations rétiniennes différentes ou des stimulations visuellement similaires mais contenant des caractéristiques diagnostiques de catégories différentes (voir Sigala & Logothetis, 2002). De plus, ce type de représentation satisferait aux exigences de flexibilité du système visuel et permettrait donc de rendre compte des phénomènes d'apprentissage perceptif.

Les connexions bidirectionnelles que IT entretient avec les cortex préfrontal, entorhinal et perirhinal, qui contiennent des neurones sélectifs pour des catégories d'objets ou des visages (Freedman et al., 2001; voir Kreiman et al., 2000 et Quiroga, 2005 pour des enregistrements réalisés chez des sujets humains épileptiques) pourraientt constituer un réseau permettant de générer une représentation plus abstraite et invariante que la représentation des objets dans IT (voir Tanaka, 1996).

Pour résumer cette section, il semble que les objets soient représentés dans IT par l'activation conjointe de plusieurs colonnes corticales codant pour des stimuli de complexité moindre que les objets. Le codage des objets dans IT semble reposer sur l'activité d'ensembles restreins de cellules (Tanaka, 1991; Gross, 1992), et non sur l'activité de neurones qui encoderaient l'objet dans son entièreté (cellules "grand-mère").

Dans IT et le STS il existe une certaine proportion de neurones qui répondent de manière préférentielle à la présentation d'un visage par rapport à d'autres stimuli visuels complexes. Dans la section suivante, nous examinons plus en détail les propriétés de ces neurones et le rôle qu'ils jouent dans la représentation de l'identité faciale au niveau de IT.

2.2.2. Les cellules sensibles aux visages

C'est à l'équipe de Charles Gross, effectuant des enregistrements de cellule unique dans le cortex inféro-temporal, qu'on doit la découverte d'un petit nombre de cellules répondant de manière plus importante à la présentation de visages ou de stimuli similaires à une main[11] (Gross et al., 1972; voir Figure 15).

[11] Gross et al. (1972) racontent que la première cellule répondant à une main pointée vers le haut a été découverte de manière accidentelle lorsque, lassés de ne pas parvenir à trouver de stimulus pouvant activer la

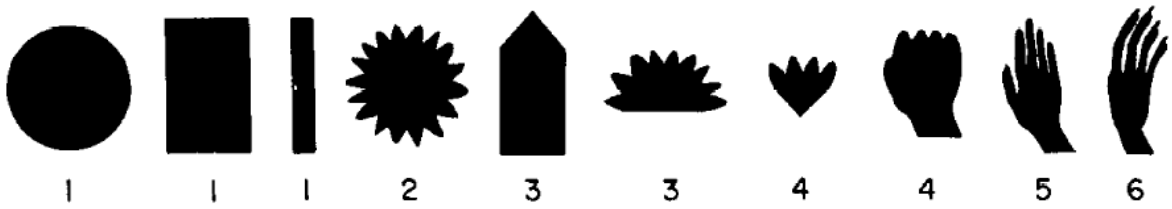

Figure 15: Exemple de stimuli utilisés pour tester la sélectivité d'un neurone répondant préférentiellement à une main de singe. Les stimuli sont ordonnés par ordre croissant (1 à 6) suivant l'intensité de la réponse qu'ils génèrent. (Gross et al., 1972).

Ces découvertes ayant généré un scepticisme généralisé dans la communauté scientifique, il faudra attendre le tout début des années 1980 pour voir la première description détaillée de propriétés des cellules répondant aux visages (Bruce et al., 1981; Perrett et al., 1982; Desimone et al., 1984).

a. Localisation des cellules sensibles aux visages

Suite aux nombreux enregistrements cellulaires réalisés chez le singe, il semble que les cellules sensibles aux visages soient réparties sur une région relativement vaste au niveau du cortex temporal antérieur. Il semble que la plus forte proportion de cellules-visage se trouve au niveau du STS (entre 5 et 20% des cellules selon la région corticale; Perrett et al., 1982; 1984; Baylis et al., 1987 ; Rolls, 2000), dans sa partie dorsale (aires TPO et PGa), ventrale (aires IPa, TEa et la partie supérieure de TEm) et également dans le fond du sillon (séparant les aires IPa et PGa). On trouve également des cellules-visage dans des proportions moins importantes (5 à 10%) dans les autres aires du cortex inféro-temporal (TEm, TE1, TE2, TE3), situées au niveau du gyrus inféro-temporal (ITG : inferior-temporal gyrus) et du sillon temporal antéro-médian (AMTS). Cependant, en raison du fait que le STS est plus facile d'accès que le gyrus inféro-temporal[12] (ITG), la plupart des études réalisées sur les cellules-visage concentrent les enregistrements au niveau du STS (Figure 16). L'estimation de la proportion de cellules-visage dans le ITG est donc basée sur un nombre beaucoup plus faible d'enregistrements. De plus, les quelques études ayant réalisé des enregistrement dans les deux régions au cours de la même expérience rapportent des proportions semblables entre le STS et

cellule, ils ont fait un signe de la main dans le champ récepteur de la cellule, provoquant une forte réponse de la cellule.

[12] En effet, le STS se trouve au-dessus (dorsalement) du gyrus inféro-temporal (ITG). Dès lors, la plupart des électrodes insérées dans IT traverse d'abord le STS avant d'atteindre le ITG. Il est donc probable que dans de nombreux cas la pénétration de l'électrode soit stoppée au niveau du STS lorsque les premières cellules-visage sont enregistrées, et ne soit donc pas poursuivie jusque dans le ITG.

le ITG (Eifuku et al., 2004: 34% dans STS et 29% dans ITG; voir aussi Young et Yamane, 1992; Hasselmo et al., 1989).

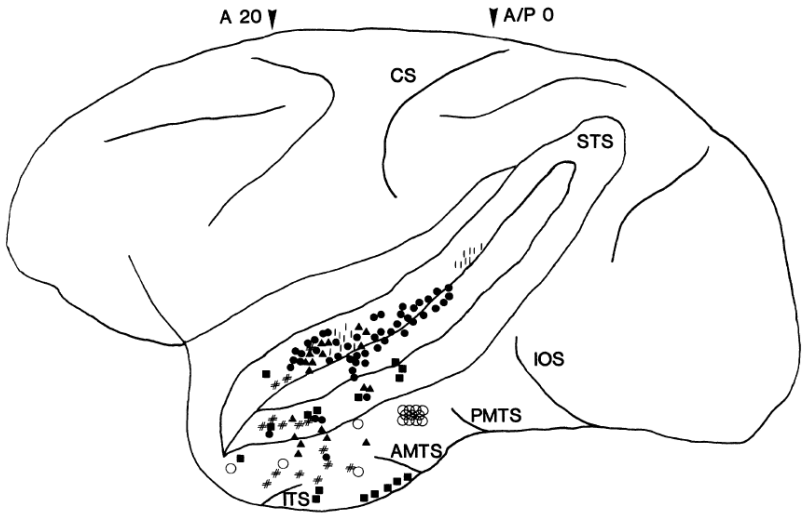

Figure 16: Vue latérale d'un cerveau de singe macaque. Localisation des cellules-visage dans le STS et IT enregistrées dans une série d'études des années 1980 et début 1990. Modifié d'après Perrett et al., 1992.

Plus récemment, des études en IRMf chez le singe ont également révélé qu'il existe plusieurs petites régions (2 ou 3) dans le STS d'une taille supérieure à une colonne corticale (Logothetis et al., 1999; Tsao et al., 2003; 2006; Pinsk et al., 2005) et contenant une forte concentration de cellules spécifiques aux visages (Tsao et al., 2006). Il est cependant encore trop tôt pour établir une relation directe entre les régions sensibles aux visages identifiées chez le singe et celles mises en évidence chez l'humain par la même technique (dans le gyrus fusiforme médian: FFA, le gyrus occipital inférieur: OFA, et une région dans le STS postérieur; voir Haxby et al., 2000). Une description approfondie des propriétés de réponse et de sélectivité de chacune des régions décrites chez le singe permettra éventuellement de clarifier la relation entre les régions identifiées chez le singe et chez l'humain.

Enfin, des cellules sensibles aux visages sont également présentes dans d'autres régions en dehors du cortex IT, notamment au niveau du cortex préfrontal et dans le lobe temporal médian (Rolls, 1984).

b. Propriétés générales des cellules-visage

1. Une cellule-visage est définie comme telle si sa réponse est significativement supérieure (entre 2 et 10 fois supérieure: Baylis et al., 1985; 1987; Rolls, 1984; Perrett et al.,

1982) lors de la présentation d'un visage par rapport à d'autres objets et formes géométriques simples ou complexes (Gross, 1992) (Figure 17). De plus cette sélectivité doit persister même lors de changements de taille, de couleur[13] ou de contraste[14]. Il faut noter que la plupart des cellules-visage répondent plus fortement lors de la présentation de visages réels ou de modèles en trois dimensions (de singe ou d'homme) par rapport à des photographies (Perrett et al., 1984; 1985a). De plus, un grand nombre de cellules-visage enregistrées dans le STS ne répondent que faiblement ou pas du tout à des dessins schématiques de visages (Perrett et al., 1984).

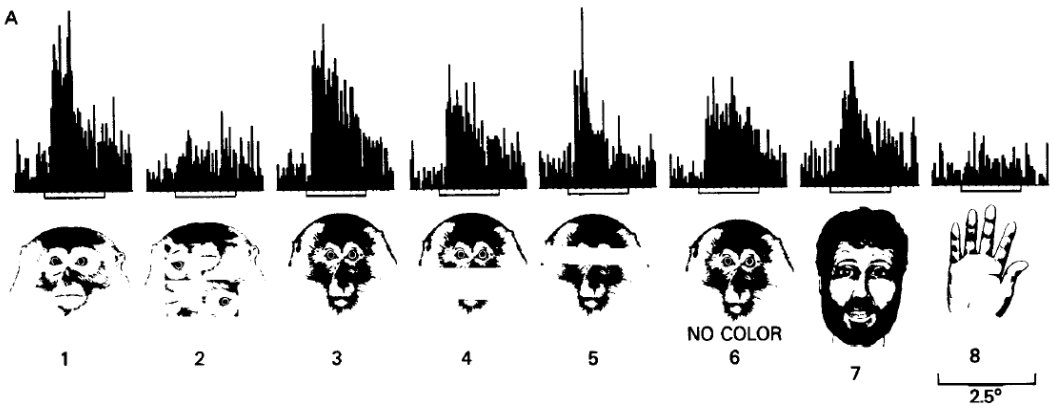

Figure 17: Exemple d'une cellule qui répond davantage à un visage qu'à n'importe quel autre stimulus. Les histogrammes représentent le taux de décharge en fonction du temps pour un visage de singe en couleur naturelle (1), le même visage avec les traits réarrangés (2), un autre visage de singe en couleur (3), le même visage sans le museau (4) ou sans les yeux (6), en noir et blanc (6), un visage humain (7) et une main (8). (Adapté de Desimone et al., 1984).

2. La plupart des cellules-visage ont une réponse de type excitatrice et soutenue (Perrett et al., 1982). Ainsi, les cellules continuent à décharger parfois même pendant plusieurs secondes si le visage reste présent dans le champ récepteur. Comme nous le verrons également plus loin, la latence de réponse de la plupart de ces cellules est comprise dans un intervalle entre 70 et 130 ms après l'apparition du visage (Oram & Perrett, 1992; Kiani et al., 2005; Perrett et al., 1982; Sugase et al., 1999; Baylis et al., 1985; Edwards et al., 2003), avec une moyenne d'environs 100 ms.

[13] Une étude récente (Edwards et al., 2003) a cependant montré que pour une large proportion de cellules-visage, les visages en couleur génèrent une réponse plus importante que les visages en niveaux de gris, ces derniers provoquant une réponse plus forte que des visages aux couleurs anormales.

[14] Même si les cellules-visage conservent leur sélectivité lors de l'inversion ou la réduction du contraste, leur réponse est néanmoins réduite par rapport à un contraste normal (Perrett et al., 1984; Rolls & Baylis, 1986).

3. Un pourcentage important de cellules-visage répond lorsqu'un ou plusieurs traits faciaux (yeux, bouche, nez, cheveux) sont présentés isolément (Perrett et al., 1982; 1984; Desimone et al., 1984). De même, la plupart des cellules répondent également si un trait facial est masqué (Perrett et al., 1982; 1985a; Desimone et al., 1984). Cependant, pour la quasi-totalité des cellules, la réponse à un visage présenté en entier est supérieure à la réponse aux traits isolés ou aux visages auxquels un trait à été masqué[15] (Figure 18).

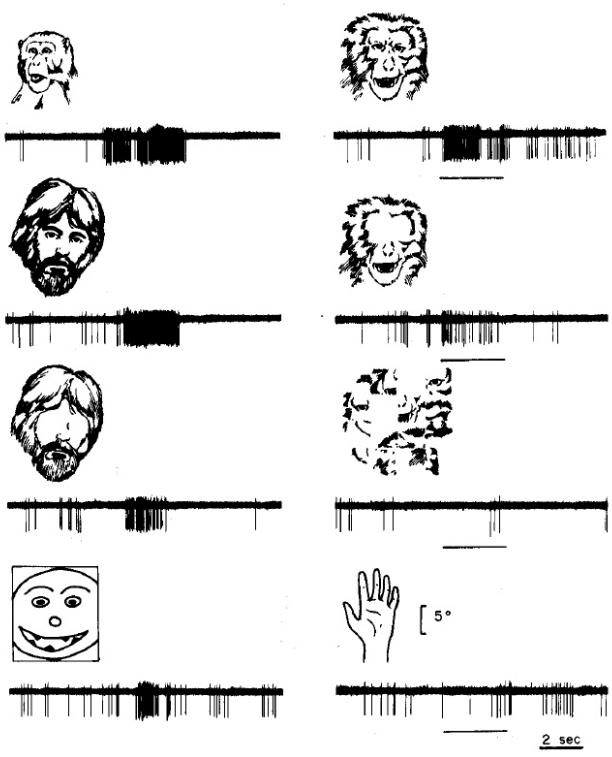

Figure 18: Réponse d'une cellule sensible aux visages dans le STP. La cellule répond de manière maximale en réponse à un visage en entier; répond moins fortement lorsque les visages sont présentés sans les yeux; et ne répond pas lorsque l'image du visage est découpée en morceaux. De plus, cette cellule répond moins fortement à la présentation d'un visage schématique. Modifié d'après Gross et al. 1981

De plus, une forte proportion de ces cellules ont leur réponse fortement réduite lorsque les traits internes du visage sont réarrangés dans le visage (Perrett et al., 1982; Desimone et al., 1984; Rolls et al., 1994), et aucune cellule ne semble répondre plus fortement pour les visages aux traits réarrangés par rapport aux visages avec les traits normalement configurés (Rolls et al., 1994). Enfin, pour la plupart de ces cellules, le visage ne peut être "réduit" (Tanaka et al., 1991) en un stimulus plus simple sans réduire la réponse de la cellule (Tanaka,

[15] De nombreuses cellules sensibles à la direction du regard dans l'aire STP répondent cependant de la même manière à des yeux isolés qu'à un visage entier (Perrett et al., 1985a).

1997; Wang, 1996). Ces observations révèlent donc que les cellules-visage préfèrent en général un visage entier et dont les traits sont configurés normalement, suggérant qu'elles ne sont pas simplement des détecteurs de traits faciaux. Cette propriété de réponse des cellules-visage est particulièrement importante si l'on considère que les visages sont traités dans leur globalité (et non par simple assemblage de traits), comme suggéré par de nombreuses études comportementales chez l'humain (Young et al., 1987; Tanaka & Farah, 1993; Hole, 1994; Goffaux & Rossion, 2006). Nous aurons l'occasion de revenir sur cette propriété du codage de l'information faciale dans le troisième chapitre, ainsi qu'à l'occasion d'une de nos études empiriques présentée dans le Chapitre 8.

4. La grande majorité des cellules-visage répondent de manière maximale à un point de vue particulier du visage (Perrett et al., 1984; 1985a; 1991; Harries & Perrett, 1991; Oram & Perrett, 1992; Wang et al., 1996; Desimone et al., 1984). Les vues les plus représentées sont face, profil, arrière de la tête, tête levée et tête baissée (Perrett et al., 1985a; 1991; 1992), avec une proportion plus forte pour les vues de face et de profil (Figure 19). Certaines cellules sont également sensibles aux visages présentés à 45° de profil (Eifuku et al., 2004; Perrett et al., 1991), même si, à l'origine, aucune cellule répondant de manière maximale à cette vue n'avait été enregistrée (Perrett et al., 1985a). La réponse de la plupart de ces cellules diminue à mesure que la vue du visage s'éloigne de la vue préférée, pour être réduite de moitié pour une vue située à 60° de la vue optimale dans le plan horizontal (Perrett et al., 1991; Foldiak et al., 2004). Il existe également une faible proportion de neurones qui répondent de manière équivalente aux différentes vues du visage et qui ont une latence de réponse légèrement plus tardive (~10 ms), suggérant que ces neurones pourraient collecter l'activité de plusieurs neurones sensibles à des point de vue différents de manière à créer une représentation du visage invariante au point de vue (Perrett et al., 1992). Il semble que les cellules sensibles au point de vue soient plus prédominantes dans la partie supérieure et le fond du STS que dans le IT, bien que peu d'études aient testé la sensibilité au point de vue dans IT (voir Eifuku et al., 2004). Enfin, les cellules sensibles à la même vue semblent groupées en colonnes corticales (Harries & Perrett, 1991).

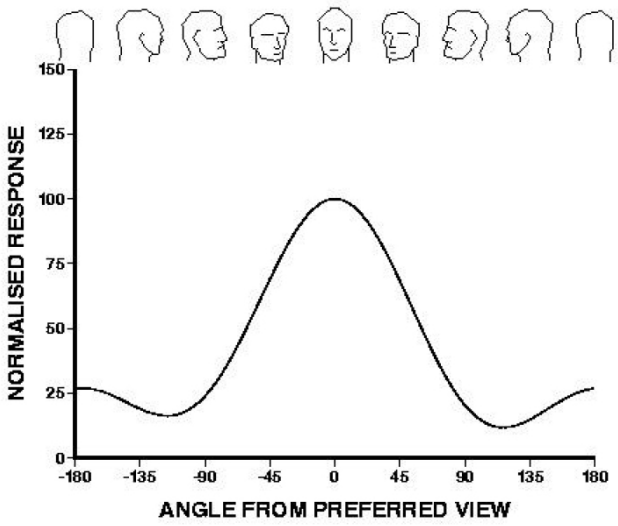

Figure 19: Exemple de courbe de sélectivité d'un neurone-visage à un point de vue de visage. La réponse est maximale pour un visage de face et décroît progressivement jusqu'à environs 60° de part et d'autre du point de vue de face. Adapté d'après Perrett et al., 1991.

5. Les cellules-visage présentent les mêmes propriétés d'invariance à la taille et à la position dans le champ visuel que les autres cellules visuelles de IT (Rolls & Baylis, 1986; Perrett et al., 1982; Tanaka et al., 1991; Tovee et al., 1994). Cependant, la latence de réponse des neurones à un visage présenté en périphérie est plus élevée que la latence enregistrée pour un visage présenté à la fixation (Tovee et al., 1994). Ces cellules conservent également leur sélectivité après diminution ou inversion du contraste même si ces manipulations réduisent globalement l'amplitude de la réponse (Baylis & Rolls, 1985). De plus, de la même manière que pour les autres neurones de IT, les cellules-visage sont également organisées en colonnes d'une taille moyenne de 0.5 à 1 mm (Perrett et al., 1984; Wang et al., 1996), voire parfois davantage (3 à 6 mm: Harries & Perrett, 1991). Cette organisation qui a d'abord été révélée par des enregistrements cellulaires dans le STS (Perrett et al., 1984; Harries & Perrett, 1991), a été confirmée par imagerie optique par l'équipe de Tanaka (Wang et al., 1996; 1998).

6. Selon certaines études, la plupart des cellules-visage sont insensibles à la rotation du visage dans le plan vertical (Desimone et al., 1984; Perrett, 1982, 1984, 1985a). D'autres études indiquent que la majorité des cellules sensibles aux visages sont sensibles à l'orientation avec un nombre beaucoup plus important de cellules répondant spécifiquement aux visages présentés à l'endroit (Tanaka et al., 1991; Ashbridge et al., 2000; voir aussi Perret et al., 1998) (Figure 20). Cependant, parmi cette deuxième série d'études, Ashbridge et al. (2000) ont investigué la sensibilité à l'orientation pour une majorité de cellules répondant à la

fois aux visages et au corps entier, rendant difficile la comparaison avec les études précédentes. Dans une étude récente (Tsao et al., 2006 ; Figure 20), les enregistrements réalisés dans une région présentant une forte concentration de cellules-visage dans le STS ont révélé que la plupart des cellules testées (N=69) répondaient moins fort et légèrement plus tardivement pour les visages présentés à l'envers par rapport aux visages présentés à l'endroit (voir aussi Perrett et al., 1982; 1985a). D'une manière générale, ces différents résultats semblent indiquer que les cellules enregistrées dans le STS en dehors de IT (partie dorsale du STS – dans les aires STP) sont relativement insensibles à l'orientation, alors que les cellules situées dans IT sont davantage sensibles à l'orientation. Comme nous aurons l'occasion de le voir dans le chapitre 3, le composant électrophysiologique enregistré chez humain au niveau du scalp en réponse à des visages (la N170) est, de manière paradoxale, plus ample pour les visages inversés que pour les visages présentés à l'endroit (Rossion et al., 1999a; Eimer, 2000a; Itier & Taylor, 2002). Il n'existe pas encore à l'heure actuelle de conscenscus sur la raison pour laquelle ce composant est plus ample pour les visages inversés alors que les cellules-visage chez le singe répondent généralement moins fortement pour des visages inversés. Dans une série d'études que nous avons réalisées dans le cadre de cette thèse, nous nous sommes attaché à caractériser le phénomène de l'effet de l'inversion des visages sur la réponse électrophysiologique enregistrée au niveau du scalp. Etant donné que l'inversion d'un visage affecte profondément la capacité à le reconnaître (Yin, 1969), nous avons notamment tenté de déterminer la dynamique temporelle avec laquelle l'inversion d'un visage affecte l'encodage de l'identité faciale, telle que révélée par l'analyse de la réponse électrophysiologique.

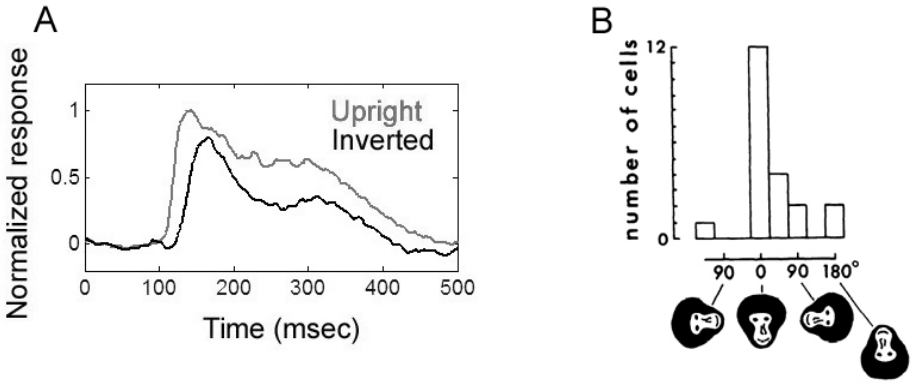

Figure 20: A. Réponse moyennée sur un groupe de 69 neurones dans le STS en réponse à des visages présentés à l'endroit et à des visages présentés à l'envers (modifié d'après Tsao et al., 2006). B. Nombre de cellules (parmi 21) de TE répondant à différentes orientations verticales de visages (modifié d'après Tanaka et al., 1991).

7. Il existe des différences dans les propriétés des cellules-visage en fonction de la région du cortex dans laquelle elles sont situées. Une première différence est liée aux traitements de l'expression et de l'identité faciale. Il apparaît que la majorité des cellules sensibles à l'expression faciale sont localisées dans le STS, tant dans sa partie ventrale et dans le fond du STS (qui font partie intégrante de IT) que dans sa partie dorsale (qui fait partie du STP) (Hasselmo et al., 1989; Sugase et al., 1999), alors que l'information sur l'identité (i.e. les propriétés "physiques" des visages) semble davantage codée dans IT (partie ventrale du STS et ITG) (Hasselmo et al., 1989; Young & Yamane, 1992; Eifuku et al., 2004; Tsao et al., 2006; Yamane et al., 1988). Notons cependant que certaines cellules sont sensibles aux deux types d'information faciale. Cette ségrégation des régions cérébrales codant l'expression et l'identité faciale est en accord avec des données empiriques obtenues chez l'humain en neuroimagerie (Winston et al., 2004 ; Andrews & Ewbank, 2004) et découle directement de la proposition faite par Bruce & Young dans leur modèle cognitif du traitement de l'information faciale (Bruce & Young, 1986) qui proposait une analyse séparée de ces informations. Cependant, l'existence ou non de deux "voies" (ou deux régions cérébrales) distinctes pour traiter l'information de l'expression et de l'identité faciale fait actuellement toujours l'objet de débats dans la littérature (voir Calder & Young, 2005 pour une revue critique). Une seconde différence importante est le fait que la région du STP semble impliquée dans le traitement de l'information relevante pour les interactions sociales. En effet, on y trouve une forte proportion de cellules-visage sensibles à l'orientation de la tête dans le plan horizontal ainsi qu'à la direction du regard (Perret et al., 1985; Perrett et al., 1992). De nombreuses autres cellules dans le STP sont également fortement sensibles au mouvement biologique (Perrett et al., 1982; 1985a;b; Jellema et al. 2004), à la position du corps (Jellema & Perrett, 2003) et à la taille de présentation du corps (Ashbridge et al., 2000).

8. La plupart des cellules-visage répondent à plusieurs visages, mais pas de manière équivalente (Perrett et al., 1984; Young & Yamane, 1992; Baylis et al., 1985). Il semble donc que ces cellules soient également impliquées dans la représentation de l'identité faciale. Nous examinons le codage de l'identité faciale plus en détail ci-dessous.

c. La représentation de l'identité faciale dans IT: une représentation distribuée

Même si les cellules-visage sont fortement spécifiques à la catégorie des visages, chacune des cellules répond à un grand nombre de visages différents (Figure 21). Si, au départ, la découverte de ces neurones codant pour une catégorie définie d'objets semblait soutenir l'hypothèse de cellules "grand-mère"[16] (Gross, 2002), il a rapidement été observé que ces cellules répondent à plusieurs visages différents (Perrett et al., 1982; 1984; Baylis et al., 1985; Rolls & Tovee, 1995b), suggérant un codage distribué de l'identité faciale, à l'instar du codage distribué probablement utilisé pour représenter les objets non-faciaux dans IT. En effet, les cellules-visage répondent différemment à différents visages (Figure 21; Baylis et al., 1985; Perrett et al., 1982), suggérant que l'identité faciale est représentée par une population de neurones, chaque neurone étant considéré comme un élément de la représentation (Young & Yamane, 1992; Foldiak, 2002).

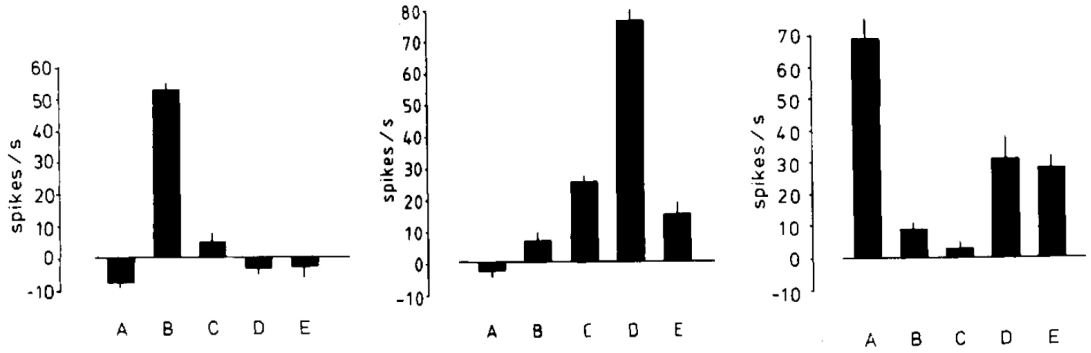

Figure 21: Réponses (taux de décharge) de 3 neurones-visage (gauche, milieu et droite) dans le STS à 5 visages différents (notés de A à E). Adapté d'après Baylis & Rolls, 1985.

Il a ainsi été montré que, bien qu'un neurone unique encode une information en moyenne de 0.3 bit[17] à propos de quel visage, parmi 20, a été présenté (Tovee & Rolls, 1995; Rolls et al., 1997; Rolls et al., 1999; Rolls et al., 2006), l'information (en bit) disponible augmente plus ou moins linéairement (avec une pente d'environs 0.3) avec le nombre de neurones utilisés pour coder cette information. Étant donné que le nombre de possibilités d'encodage croit de manière exponentielle avec le nombre de bit codés (# possibilités = $2^{\#\ bits}$), le nombre de visages différents qu'il est possible d'encoder croît de manière exponentielle

[16] Un stimulus visage ne peut en général pas être décomposé ou réduit sans fortement atténuer la réponse des neurones-visage (Tanaka et al., 1991; Tanaka, 2003).

[17] Le bit correspond à une unité d'information pouvant encoder 2 valeurs (code binaire). Le nombre de bits nécessaires à encoder N valeurs est défini par le logarithme en base 2 de N: $\log_2 N$.

avec le nombre de neurones dans la population (Rolls et al., 1997; Rolls, 2007). Le fait que chaque neurone encode 0.3 bit d'information indique qu'il existe une redondance assez importante dans le codage de l'information par ces neurones[18]. En effet, dans des conditions idéales où la réponse de chaque neurone est indépendante de la réponse des autres neurones, chaque neurone pourrait encoder 1 bit d'information (si l'on considère qu'il n'existe que 2 possibilités de réponse pour chaque neurone : soit il répond, soit il ne répond pas du tout, ce qui n'est évidemment pas le cas). Cette redondance dans le codage de l'information à propos de l'identité faciale a également été estimée en mesurant le degré avec lequel la représentation d'un visage particulier est distribuée sur une population de neurones; les extrêmes étant d'un côté un codage où chaque visage est représenté par un neurone (i.e. un codage par cellule grand-mère) et de l'autre un codage distribué sur l'ensemble de la population des cellules enregistrées. À l'aide du calcul d'un coefficient de "sparseness[19]", Rolls & Tovee (1995b) ont ainsi estimé que chaque cellule-visage enregistrée était activée par environ 33% des visages de leur set de stimuli[20]. Plus récemment, des enregistrements cellulaires chez l'humain dans le lobe temporal médian (Figure 22) lors de la présentation de stimuli visuels (pas uniquement des visages) ont estimé que dans cette région, le coefficient de "sparseness" est proche de 1% (chaque cellule n'étant activée que par 1% des stimuli présentés) (Waydo et al., 2006), ce qui est bien en dessous de l'estimation réalisée par Rolls et ses collègues dans le cortex inféro-temporal, et cela avec des cellules-visage[21] uniquement. L'observation qu'un nombre relativement faible de neurones puisse encoder un nombre important de visages a généré l'idée que l'identité des visages puisse être représentée de manière distribuée mais en utilisant une population restreinte de neurones (i.e. "sparse population coding") (Baylis et al., 1985;

[18] En effet, étant donnée la redondance qui existe dans le codage neuronal (les réponses des neurones individuels ne sont que partiellement décorrélées) le nombre de bits encodés par une population de neurones n'est pas défini par $\log_2 N$. Si c'était le cas il suffirait de 16 neurones pour coder 2^{16} visages dans des conditions idéales d'absence de bruit et de réponse neuronale binaire.

[19] La mesure avec laquelle une représentation est distribuée ou non est quantifiée par un coefficient de "sparseness", qui s'étend de 0 (représentation purement locale) à 1 (représentation distribuée sur l'ensemble des neurone).

[20] Rolls & Tovee (1995b) et Baylis et al. (1985) rapportent également des valeurs de coefficient de "sparseness" bien plus élevées (entre 60 et 70%), mais obtenues sans correction pour le taux de décharge spontanée des neurones (Rolls, 2000; 2007).

[21] Les cellules-visage ne représentant qu'environ 10% des cellules dans IT, il est probable que si Rolls et ses collègues avaient considéré l'ensemble des neurones enregistrés dans IT (dont 90% de neurones non sélectifs aux visages), ils auraient trouvé un coefficient plus proche de celui obtenu par Waydo et al. (2006). Notons également que les neurones enregistrés dans le lobe temporal médian sont en général beaucoup plus sélectifs en terme d'identité faciale que les neurones de IT. De plus, certains de ces neurones répondent au visage d'une personne familière, quel que soit le format de présentation du visage. Ces neurones répondent même au nom de la personne présenté par écrit (Quiroga et al., 2005). Quiroga et al. (2005), ainsi que Waydo et al. (2006), suggèrent cependant que ces observations ne sont pas forcément une évidence pour l'existence de cellules grand-mère dans le lobe temporal médian.

Young & Yamane, 1992; Rolls & Tovee, 1995b; Rolls et al., 1997; voir aussi Waydo et al., 2006). Ce type de codage implique que l'identité de chaque visage est représentée par un profil unique d'activité dans un ensemble restreint de neurones; des visages différents étant codés par des ensembles de neurones se recouvrant partiellement.

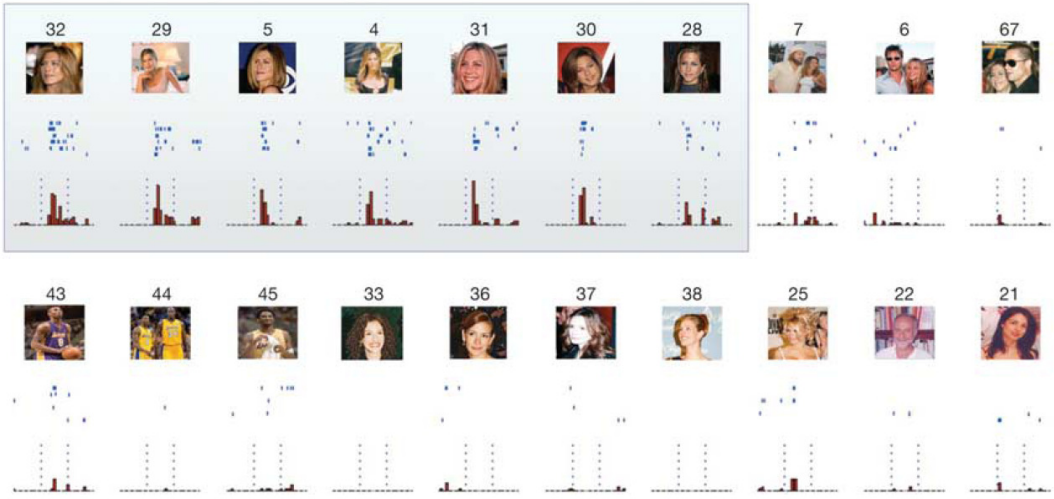

Figure 22: Exemple de réponse d'une cellule dans le lobe temporal médian qui est activée fortement par des images contentant le visage de l'actrice Jennifer Aniston, et pas ou très peu par le visage d'autres personnes. Modifié d'après Quiroga et al., 2005.

Les avantages du "sparse coding" par rapport aux deux extrêmes sont nombreux (Rolls, 2000; 2007; Foldiak, 2002). Premièrement, la capacité de représentation est très élevée même avec un nombre restreint de neurones. Deuxièmement, la représentation générée par la population est extrêmement précise et flexible, permettant de représenter les visages de manière grossière ou très fine selon l'action à réaliser à partir de la stimulation. Troisièmement, la représentation étant distribuée et redondante, l'ajout de bruit dans la réponse neuronale ou la perte d'une partie des neurones ou des synapses provoquera des déficits graduels de reconnaissance (et non des déficits *tout ou rien* comme prédirait un codage local). Quatrièmement, une représentation distribuée permet l'apprentissage de nouveaux exemplaires et la généralisation de la représentation à des exemplaires d'une même catégorie d'objets. Cinquièmement, une représentation distribuée peut-être activée dans son entièreté via l'activation de n'importe quelle partie de la représentation (via des connexions horizontales entre les neurones), permettant ainsi la complétion des objets fragmentés ou partiellement masqués. Enfin, quel que soit le code neuronal utilisé pour transmettre l'information d'un niveau de la hiérarchie à un autre (taux de décharge, codage temporel des décharges,… voir plus loin, note 33), un codage distribué de l'information permet de propager

l'information très rapidement. En effet, étant donné que l'information est distribuée sur un grand nombre de neurones, un nombre très faible de décharges (entre 1 et 3) fournies par chaque neurone contient énormément d'informations. De plus, en raison de la forte redondance qui existe dans la représentation, une information précise peut-être transmise en intégrant seulement l'information d'une partie des neurones constituant la population (par ex: les neurones déchargeant le plus précocement).

3. Un réseau cortical pour la perception des visages chez l'humain

Les travaux présentés jusqu'ici sont issus des enregistrements cellulaires chez l'animal, principalement chez le singe macaque. D'une manière générale, on considère que le traitement visuel chez le singe et chez l'humain est relativement similaire. D'ailleurs, de nombreuses aires visuelles ayant des propriétés similaires sont décrites chez les deux espèces (voir Figure 1 par exemple). La nature invasive des enregistrements neuronaux rendent impossible son application chez l'homme, à part lorsque des électrodes sont implantées dans le tissu cérébral de manière à localiser des foyers épileptogènes (par exemple: Alisson et al., 1999; McCarthy et al., 1999; voir Engel et al., 2005 pour une revue de la littérature). La plupart des données sur la neuro-anatomie fonctionnelle chez l'humain proviennent des enregistrements indirects de l'activité cérébrale réalisés par l'intermédiaire de la neuroimagerie fonctionnelle (PET-scan, IRMf). Dans cette section nous décrivons brièvement les régions visuelles qui sont impliquées dans la perception des visages et mises en évidence notamment par la neuroimagerie fonctionnelle.

La plupart des connaissances actuelles à propos de la neuro-anatomie fonctionnelle de la perception des visages provient de trois sources d'informations différentes: l'étude de personnes souffrant de troubles spécifiques de reconnaissance de visages à la suite de lésions du tissu cérébral (prosopagnosie), l'enregistrement de cellules sélectives aux visages dans le cortex inféro-temporal du singe et les études menées sur le sujet humain à l'aide de techniques de neuroimagerie fonctionnelle.

Les premiers cas de troubles de reconnaissance faciale on été décrit au 19ème siècle (voir Mayer & Rossion, 2007), mais c'est à Bodamer (Bodamer, 1947) que l'on doit le terme *prosopagnosie* pour référer à ce trouble de reconnaissance faciale sans déficit visuel de bas niveau. La plupart des cas de prosopagnosie ont été décrits durant les années 1970-1980,

avant l'apparition des techniques de neuroimagerie fonctionnelle. La forte hétérogénéité de la localisation des lésions cérébrales provoquant un déficit prosopagnosique a été un des premiers indices révélant le caractère distribué du traitement perceptif des visages. Les régions impliquées dans la prosopagnosie sont en général localisées au niveau du cortex occipito-temporal (par ex: gyrus fusiforme, gyrus lingual, gyrus occipital inférieur) avec une forte dominance de l'hémisphère droit (voir Figure 23). Cette dominance de l'hémisphère droit et plus particulièrement le fait qu'une lésion unilatérale droite est suffisante pour provoquer un déficit prosopagnosique a également été l'une des premières évidences du rôle prédominant de cet hémisphère dans la perception des visages.

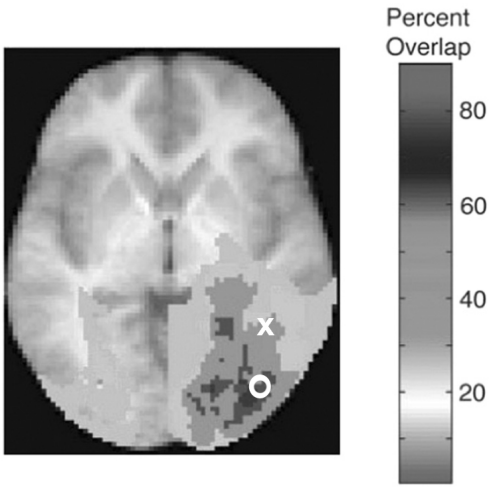

Figure 23: Pourcentage de recouvrement des lésions cérébrales chez huit patients prosopagnosiques. Les lésions sont principalement localisées dans le cortex occipito-temporal inférieur avec le recouvrement le plus important (>60%) au niveau du gyrus occipital inférieur. Les deux marqueurs indiquent la position moyenne de la FFA (croix) et de la OFA (cercle) (Bouvier & Engel, 2005; modifié d'après Mayer & Rossion, 2007).

La découverte de cellules sélectives aux visages dans le cortex inféro-temporal du singe au début des années 1970 (Gross et al., 1972) et les nombreuses études qui ont suivi durant les années 1980 et 1990 constituent également l'une des premières et principales sources d'information à propos des corrélats neuronaux de la perception des visages. Nous avons décrit certaines propriétés de ces cellules-visage dans la section précédente. Dans l'ensemble, ces études indiquent la présence de cellules-visage largement distribuées sur la surface du cortex inféro-temporal et au niveau de la partie supérieure du STS (aires du STP). Par ailleurs, la sélectivité de ces cellules semble varier en fonction de leur localisation; les cellules du gyrus inféro-temporal et partie ventrale du STS étant plus particulièrement impliquées dans l'encodage de l'identité faciale (Hasselmo et al., 1989; Eifuku et al., 2004),

alors que les cellules du STS dorsal et STP sont davantage sensibles à l'information visuelle importante pour déterminer les interactions sociales (expression faciale, mouvement biologique, direction du regard – Hasselmo et al., 1989; Perrett et al., 1985a; b; 1992; Jellema et al., 2004). Ces données soulignent encore une fois le caractère distribué du traitement des visages au niveau cérébral. Nous avons également vu que des études récentes en IRMf chez le singe indiquent l'existence dans le STS de plusieurs régions d'une taille supérieure à une colonne corticale, et qui sont particulièrement sensibles aux visages (Logothetis et al., 1999; Pinsk et al., 2003; Tsao et al;, 2003; 2006).

Le développement des techniques de neuroimagerie offrant la possibilité d'appréhender le fonctionnement cérébral du sujet humain "normal" a permis de confirmer et de développer l'idée que le traitement perceptif des visages implique un réseau de multiples aires cérébrales. Dans les années 1990, les premières études en PET-scan (Sergent et al., 1992 ; Sergent & Signoret, 1992) et puis en IRMf (par ex: Kanwisher et al., 1997 ; Gauthier et al., 2000 ; voir Haxby et al., 2000), ont permis de révéler l'existence de plusieurs régions dans le cortex visuel, activées plus fortement en réponse à la présentation de visages par rapport à d'autres objets complexes[22] (voir Haxby et al., 2000 pour une revue de littérature). Les trois principales régions sont situées bilatéralement[23] dans le gyrus fusiforme médian (Sergent et al., 1992; dénommé plus tard comme '*FFA*' par Kanwisher et al., 1997), le gyrus occipital inférieur (*OFA*: Gauthier et al., 1998) et dans la partie postérieure du *STS* (Puce et al., 1998) (Figure 24). Ces trois régions sont décrites comme principalement impliquées dans l'élaboration d'une représentation visuelle d'un visage et forment le système "central" du modèle hiérarchique du traitement des visages développé dans Haxby et collègues (Haxby et al., 2000; Figure 25). D'après ce modèle, les représentations élaborées par ces régions servent d'input à d'autres régions cérébrales impliquées dans le traitement de l'information émotionnelle, sémantique ou encore liées à la compréhension du langage parlé (système "étendu").

[22] Ces régions sont donc définies de manière purement fonctionnelle en soustrayant l'activation obtenue en réponse à des objets de l'activation obtenue en réponse à des visages. Cette approche permet donc de définir des régions d'intérêt dans lesquelles seront testées des hypothèses plus spécifiques, ce généralement pour diminuer le risque de faux-positifs dans les analyses statistiques. En effet, ce risque est réduit si un plus petit nombre d'échantillons sont comparés entre des conditions par rapport à la situation où l'ensemble du cerveau est inclus dans l'analyse statistique (voir Friston et al., 2006 et Saxe et al., 2006 pour un débat sur la validité des méthodes reposant sur la définition de régions d'intérêts).

[23] Cependant, on trouve généralement une dominance de l'hémisphère droit.

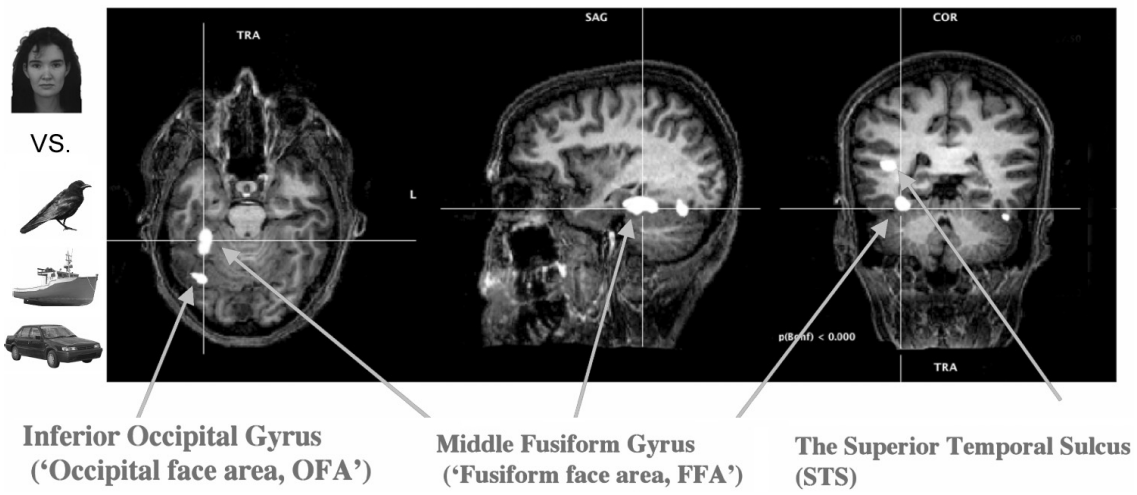

Figure 24: Les trois régions (en blanc) dans le cortex visuel de l'humain qui présentent une activation plus forte en réponse à des visages qu'à d'autres catégories d'objets sont représentées au niveau de l'hémisphère droit (situé à gauche sur la figure) sur des coupes axiale (gauche), sagittale (milieu) et coronale (droite). Ces trois régions sont situées dans le gyrus occipital inférieur (OFA: occipital face area), dans le gyrus fusiforme médian (FFA: fusiform face area) et dans la partie postérieure du STS.

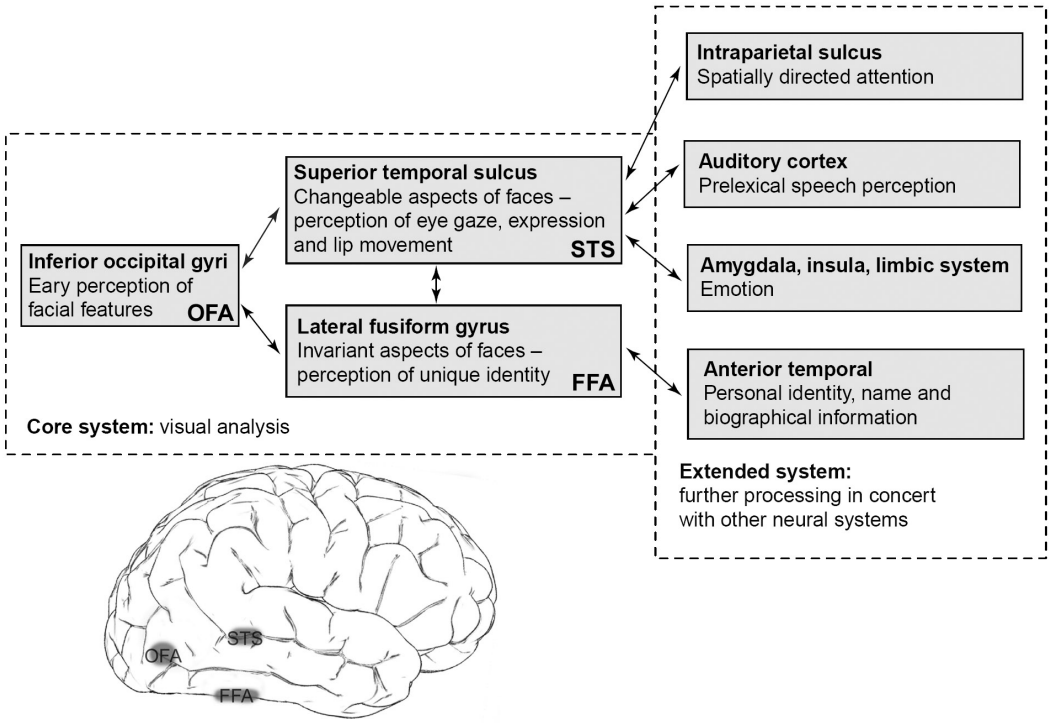

Figure 25: Modèle hiérarchique distribué du traitement neuronal des visages chez l'humain (Haxby et al., 2000). Le modèle comprend un système central (core system) composé de la OFA, FFA et STS et un système étendu (extended system) en dehors du cortex visuel et impliqué dans des traitement post-perceptif des visages (récupération d'informations sémantiques, perception des émotions du visage,…). Modifié d'après Haxby et al., 2000.

Parmi les trois régions impliquées dans la perception des visages dans le cortex visuel, c'est la FFA qui a été la plus étudiée (voir Kanwisher & Yovel, 2006 pour une revue de la littérature), en particulier parce que cette région a été au départ définie comme un module dédié uniquement au traitement des visages (Kanwisher et al., 1997; voir Kleinschmidt & Cohen, 2006). Outre le codage de la catégorie "visage" dans ces régions (Kanwisher et al., 1997; Tong et al., 1998), de nombreuses données indiquent que la FFA et la OFA sont également impliquées dans le traitement de l'identité faciale (Gauthier et al., 2000; Winston et al., 2004; Schiltz et al., 2006; Andrews & Ewbank, 2004; Mazard et al., 2006; Grill-Spector et al., 2004; Schiltz & Rossion, 2006). La plupart des études montrant un codage de l'identité faciale dans ces régions ont recours au paradigme d'adaptation, dans lequel la répétition d'un même stimulus (par exemple: une même identité) provoque une diminution de l'activation dans les régions codant pour ce stimulus (Grill-Spector & Malach, 2001; Grill-Spector et al., 2006; voir aussi Ringo, 1996). Les régions dont le niveau d'activité est plus élevé lorsque deux stimuli différents par une ou plusieurs dimensions (par exemple l'identité faciale) sont présentés successivement, sont considérées comme sensibles aux dimensions manipulées. Un effet d'adaptation à l'identité faciale (comparaison entre le signal obtenu lorsque le même visage est répété par rapport à la présentation successive de deux visages différents) a été trouvé principalement dans la FFA et la OFA (Gauthier et al., 2000; Winston et al., 2004; Schiltz et al., 2006; Andrews & Ewbank, 2004; Mazard et al., 2006). Il a également été montré que la FFA et la OFA sont plus fortement activées lors des essais où les sujets identifient correctement un visage que lors des essais où le visage n'est pas identifié correctement (Grill-Spector et al., 2004), confirmant la sensibilité de ces deux régions à l'identité faciale. De plus, la localisation de lésions entraînant un déficit prosopagnosique recouvre largement le territoire de ces deux régions (Figure 23), soulignant leur rôle dans l'encodage de l'identité faciale.

Le STS postérieur a fait l'objet d'un nombre plus restreint d'études en neuroimagerie, particulièrement en comparaison de la FFA. D'une manière générale, il semble que cette région soit plus particulièrement impliquée dans le traitement de l'information dynamique d'un visage comme le mouvement ou la direction des yeux et le mouvement de la bouche (Hoffman & Haxby, 2000; Puce et al., 1998; Andrews & Ewbank, 2004). D'autres études suggèrent également que le STS dans sa partie postérieure (Andrews & Ewbank, 2004) ou dans une partie plus antérieure (Winston et al., 2004) pourrait être impliqué dans le traitement des expression faciales. Tant les études en neuroimagerie (Yovel & Kanwisher, 2005; Mazard et al., 2006; Andrews & Ewbank, 2004; Grill-Spector et al., 2004) que l'observation de la

localisation des lésions provoquant un déficit de reconnaissance de l'identité faciale (voir Mayer & Rossion, 2007; Bouvier & Engel, 2006) indiquent que le STS n'est pas une région déterminante dans le codage de l'identité faciale chez l'humain.

D'autres études en neuroimagerie ont mesuré l'effet de manipulation de stimuli générant des effets comportementaux (temps de réponse et performances) robustes sur le niveau d'activation de ces régions. Par exemple, un des effets les plus connu dans la littérature sur le traitement des visages est l'effet d'inversion par lequel la reconnaissance ou la discrimination de visages à un niveau individuel sont réduites si ces visages sont présentés à l'envers (Yin, 1969). Il a ainsi été observé que le codage de l'identité faciale dans la FFA et la OFA est sensible à l'orientation verticale des visages. Plus précisément, alors qu'un effet d'adaptation à l'identité faciale est observé dans ces régions lorsque les visages sont présentés à l'endroit (Figure 26), cet effet d'adaptation est réduit ou disparaît entièrement lorsque les visages sont présentés à l'envers (Mazard et al., 2006; Yovel & Kanwisher, 2005; voir aussi Schiltz & Rossion, 2006). L'effet d'inversion sur la reconnaissance des visages est supposé être lié à l'impossibilité de percevoir les visages inversés de manière holistique (Young et al., 1987). En accord avec cette proposition, il a récemment été montré que les visages sont représentés de manière holistique dans la OFA et plus particulièrement dans la FFA (Schiltz & Rossion, 2006). Il est également intéressant de souligner que l'interaction entre l'orientation des visages et l'effet d'adaptation à l'identité faciale est également légèrement plus robuste dans la FFA que dans la OFA (Mazard et al., 2006; Yovel & Kanwisher, 2005).

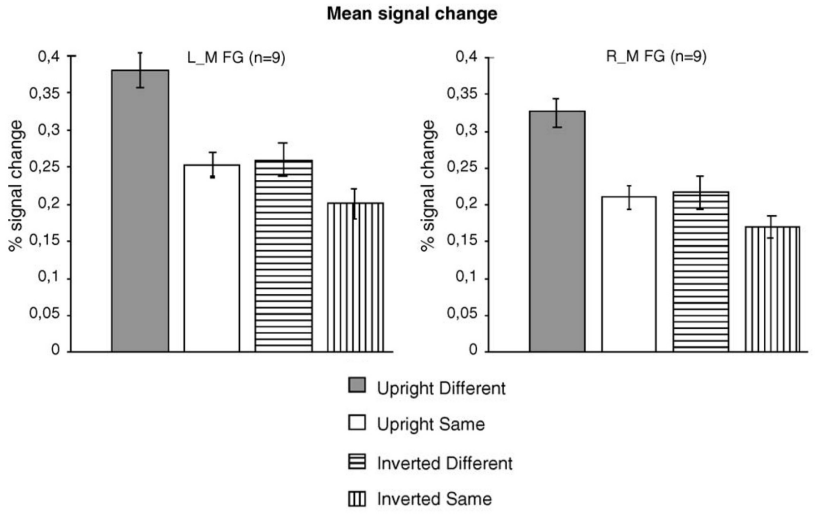

Figure 26: Niveau d'activation dans la FFA de l'hemisphère gauche (gauche) et de l'hémisphère droit (à droite) en réponse à des séquences de visages identiques (same) en comparaison à des séquences de visages différents (different) présentés à l'endroit ou à l'envers. Notons la plus grande différence d'activation entre la présentation de mêmes vs. différentes identités (effet d'adaptation) à l'endroit qu'a l'envers. Modifié d'après Mazard et al., 2006.

Un série d'études récentes mettent également en avant l'intérêt de combiner la neuroimagerie fonctionnelle à l'étude de patient souffrant de prosopagnosie acquise (suite à une lésion cérébrale) pour déterminer la fonction des régions "visages" identifiées chez le sujet sain dans le traitement visuel des visages (Rossion et al., 2003b; Schiltz et al., 2006; Dricot et al., submitted; Steeves et al., 2005). Ainsi par exemple, la patiente prosopagnosique PS présente des lésions englobant la OFA dans l'hémisphère droit et la FFA dans l'hémisphère gauche (Rossion et al., 2003b; Sorger et al., 2007). Malgré la lésion de la OFA droite - située en amont de la FFA dans le modèle hiérarchique de Haxby et al. (2000) (Figure 25) –, le profil d'activation de la FFA en réponse à des visages ou à des objets est comparable à celui observé chez des sujets normaux (Rossion et al., 2003b; Schiltz et al., 2006). Par contre, malgré le fait que la FFA chez PS discrimine normalement les visages des autres catégories d'objets, cette région ne discrimine plus l'identité des visages, contrairement aux sujets normaux (Schiltz et al., 2006; Dricot et al., submitted). En effet, chez cette patiente, la FFA n'est pas plus fortement activée en réponse à la présentation des visages d'identités différentes qu'en réponse à la présentation de visages identiques (effet d'adaptation). Ces résultats (c.–à-d. catégorisation normale de visages et absence de codage de l'identité dans la FFA droite) combinés avec la lésion au niveau de la OFA droite, suggèrent d'une part que la OFA n'est pas nécessairement un point d'entrée obligé pour activer la FFA (Figure 27), et d'autre part que la OFA est nécessaire à un codage normal de l'identité faciale (Rossion et al., 2003b; Schiltz et al., 2006; voir aussi Rossion, in press). En fait, alors qu'une lésion de la OFA droite semble suffisante pour générer un trouble prosopagnosique (voir Figure 23), il y a encore peu d'évidences permettant de déterminer si une lésion unique dans le gyrus fusiforme droit (FFA droite) peut également entraîner un tel déficit (Mayer & Rossion, 2007).

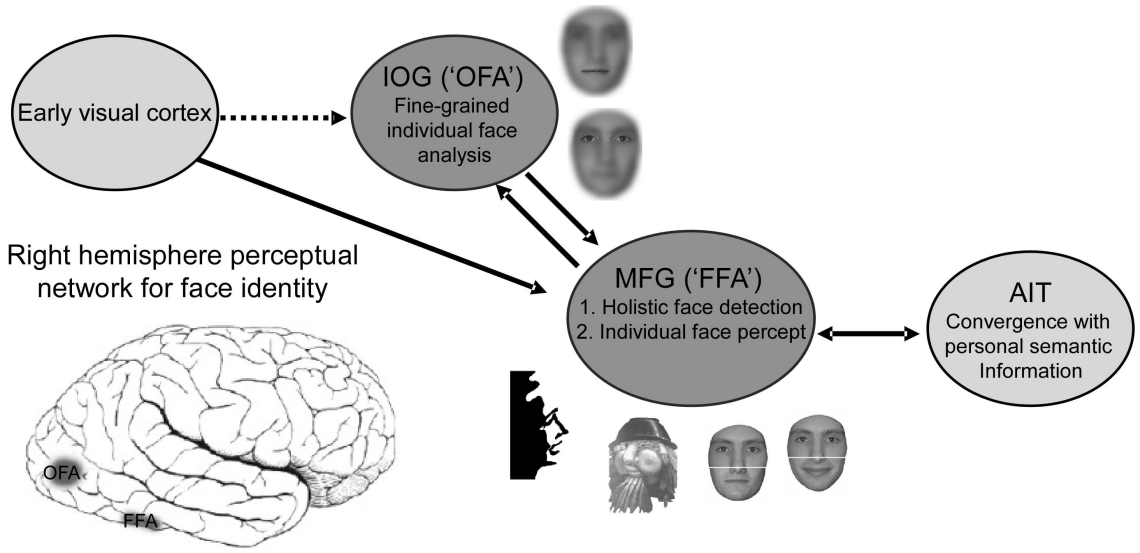

Figure 27: Modèle revisité de l'organisation neuro-fonctionnelle impliquée lors de l'encodage de l'identité faciale dans le cerveau humain (Rossion, in press). Les visages sont catégorisés de manière holistique (voir visages de type Mooney et visages "légumes" d'Arcimboldo) dans la FFA (Schiltz & Rossion, 2006) après les traitements visuels dans le cortex visuel de bas niveau. L'activation de la FFA est indépendante de la OFA située postérieurement. La OFA reçoit ou non des inputs directs du cortex visuel de bas niveau (ligne pointillée). Ce modèle peut donc rendre compte de l'activation normale de la FFA en l'absence de OFA telle qu'observé chez des patients prosopagnosiques (Rossion et al., 2003b; Steeves et al., 2005), contrairement à des modèles purement hiérarchiques où la OFA est un point d'entrée nécessaire pour activer la FFA (Haxby et al., 2000; Fairhall & Ishai, 2007). La représentation globale et grossière du visage dans la FFA est ensuite raffinée par des interactions avec les neurones de la OFA qui possèdent des champs récepteurs de plus petite taille, permettant ainsi de représenter l'information faciale avec plus de précision. L'interaction entre ces deux régions permet de générer une représentation faciale individuelle. Modifié d'après Rossion, in press.

En conclusion, les travaux décrits dans cette section indiquent que le traitement spécifique de l'identité faciale met en jeu une série de régions dans le cortex visuel (principalement la OFA et la FFA). De plus, les études en neuroimagerie fonctionnelle de patients souffrant d'un déficit prosopagnosique suite à des lésions cérébrales, suggèrent un rôle critique de la OFA dans le codage de l'identité faciale. Ces études indiquent également que la discrimination entre des visages différents dans la FFA dépend de l'intégrité de la OFA, et suggèrent ainsi que le codage de l'identité résulte d'une interaction entre ces deux régions.

4. Conclusions

Ce premier chapitre nous a permis d'appréhender dans leur ensemble, les propriétés anatomiques et fonctionnelles du système visuel, tant à un niveau macroscopique qu'à un niveau microscopique. Retenons en particulier que dans le cortex inféro-temporal du singe, il existe une large proportion de neurones adaptés à l'encodage de nombreuses informations faciales, et notamment des informations sur l'identité. Ces neurones possèdent des propriétés

d'invariance qui leur permettent de signaler la présence d'un visage dans l'environnement visuel indépendamment des conditions de visibilité. Ces cellules peuvent encoder des traits faciaux isolés mais sont néanmoins en général préférentiellement activées par des visages présentés dans leur entièreté. Par ailleurs, nous avons également mentionné que les représentations individuelles de visages reposent sur un codage distribué au sein de ces cellules-visage. Un tel codage qui permet à ces représentations d'être activée dans leur entièreté via l'activation d'une partie de la représentation; probablement la partie qui reçoit les afférences les plus précoces en provenance des aires de plus bas niveau. Cette propriétés de codage distribué permettrait donc de répondre à une caractéristique essentielle de la peception des visages, et que nous allons abordé plus en détail de le chapitre suivant: les visages sont perçus et reconnus à un niveau individuel avec une étonnante rapidité. Enfin, dans la dernière partie nous avons vu que chez l'humain (et probablement chez le singe), l'élaboration des représentations faciales individuelles repose également sur l'intéraction entre plusieurs régions du cortex occipito-temporal.

Chapitre 2
Dynamique temporelle de la catégorisation visuelle

Dans le chapitre précédent nous avons décris certaines propriétés structurelles du système visuel, ainsi que la conception actuelle sur la manière dont cette structure réalise le codage des objets et des visages dans le cortex inféro-temporal.

Dans ce deuxième chapitre nous allons aborder les aspects temporels de la catégorisation visuelle. Nous examinerons les données issues des enregistremments réalisés chez le singe ainsi que et les données issues des mesures comportementales dans des paradigmes psychophysiques chez l'humain. Ensuite, nous nous attarderons plus en détail sur la description de la dynamique temporelle du traitement visuel des visages, par l'intermédiaire à la fois des enregistrements de cellules-visages chez le singe, et des données issues des approches comportementales chez le sujet humain.

Nous verrons que l'ensemble de ces données met en évidence la rapidité de la perception visuelle en général, et du traitement des visages en particulier. De plus, ces résultats permettent de définir une série de contraintes biologiques et psychophysiques dont doit pouvoir rendre compte toute description du décours temporel de la perception des visages.

1. Décours temporel de l'activation des aires visuelles dans la voie ventrale chez le primate.

Comme nous l'avons vu dans la section précédente, l'input lumineux parvenant sur la rétine lorsqu'un objet apparaît dans le champ visuel est d'abord transformé en signal électrophysiologique pour ensuite parcourir un certain nombre de structures cérébrales sous-corticales et corticales organisées de manière hiérarchique. Ce trajet au travers des régions visuelles de la voie ventrale se termine au niveau du cortex inféro-temporal contenant des neurones codant des caractéristiques complexes de l'input visuel. Une conséquence directe de cette organisation hiérarchique est que les neurones des régions situées en aval de la hiérarchie seront activés plus tardivement que les neurones des régions de plus bas niveaux. Ce délai dans la latence des réponses neuronales est lié au temps nécessaire pour transmettre l'information visuelle d'une étape de la hiérarchie à une autre; chaque niveau impliquant un

ou plusieurs relais synaptique. Ce délai entre l'apparition d'un objet dans le champ visuel et l'activation de sa représentation dans les régions visuelles de haut niveau est déterminant lorsqu'on s'intéresse à la vitesse à laquelle les objets ou les visages sont catégorisés dans le cerveau.

Il est difficile d'établir les temps de latence exacts d'activation des différentes aires visuelles car la plupart des études effectuées sur le singe ne réalisent pas d'enregistrements simultanés dans plusieurs aires visuelles. En outre, les études rapportant des données sur la latence d'activation d'aires visuelles ont souvent été menées par des laboratoires différents et des conditions de stimulation[24] et d'enregistrement (e.g. singes éveillés vs. anesthésiés) différentes (voir la revue de la littérature de Bullier, 2001 et la méta-analyse réalisée par Lamme & Roelfsema, 2000).

Dans cette section nous nous attacherons à décrire la latence de la première activation des aires visuelles liée à une stimulation rétinienne. Il est néanmoins à noter que les neurones d'une aire visuelle donnée ne répondent pas tous au même moment et les latences de réponse s'échelonnent parfois sur une centaine de millisecondes (Figure 28). On observe qu'en moyenne, la variance dans les latences de réponse d'une aire visuelle augmente au fil de la hiérarchie. En raison de cette variance, les distributions des latences de réponse de différentes structures visuelles se recouvrent fortement.

[24] La latence de réponse des neurones dans la plupart des aires visuelles est fortement dépendante des propriétés visuelles de bas niveau de la stimulation (contraste, luminance, spectre de fréquences spatiales,…) (voir Oram et al., 2002). Dans les régions de plus haut niveau, la latence de réponse neuronale peut également être dépendante de la catégorie de stimulus présenté (Kiani et al., 2005).

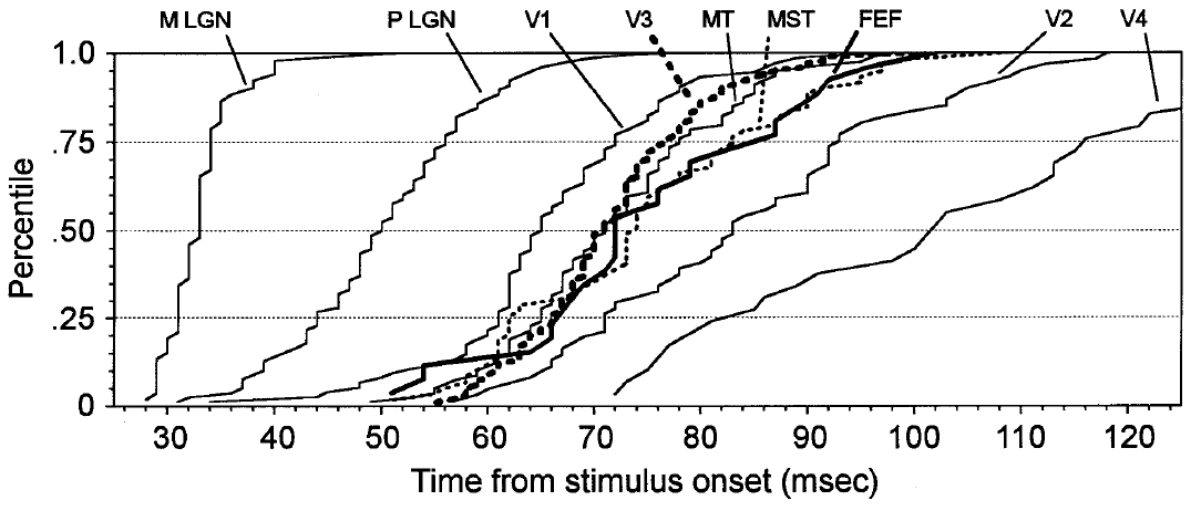

Figure 28: Distributions cumulées des latences de réponses évoquées par un stimulus visuel dans le cortex visuel primaire (V1), les corps genouillés latéraux (LGN sur la figure : « LGN M » se réfère à la voie magnocellulaire et « LGN P » à la voie parvocellulaire), et différentes aires visuelles corticales dans la voie ventrale (V2, V4) et dorsale (V3, MT, MST). FEF se réfère à « Frontal Eye Fields ». Tant la latence de réponse que la variance dans les latences de réponse augmentent au fil de la hiérarchie (par ex: V1 – V2 – V4). Adapté d'après Schmolesky et al., 1998.

Lorsqu'une stimulation visuelle apparaît sur la rétine, la latence des réponses les plus précoces des neurones dans le corps genouillé latéral est d'environ 20 à 30 ms pour les neurones des voies magno- et parvocellulaire. Cependant, la différence de vitesse d'activation entre les voies magno- et parvocellulaire se marque dès le CGL dans la mesure où la quasi-totalité des neurones de la voie magnocellulaire ont répondu dans les premières 40 ms alors qu'il faudra 10 à 30 ms de plus pour une activation similaire de la voie parvocellulaire du CGL (Schmolesky et al., 1998 ; Maunsell et al., 1999; Figure 28). Cette différence de latence d'activation des aires visuelles entre voies magno- et parvocellulaire se marque également au niveau de V1. En effet, alors que les neurones de la couche 4Cα recevant des afférences magnocellulaires répondent dès 35 ms avec une latence moyenne de 45 ms (Schmolesky et al., 1998 ; Maunsell & Gibson, 1992 ; Nowak et al., 1995), les neurones de la couche 4Cβ recevant des inputs parvocellulaires ne répondent en moyenne qu'environ 65 ms après l'apparition du stimulus. Les premières réponses de cellules de V2 se situent vers 55 ms avec une latence moyenne d'environs 80 ms (voir Figure 28, Figure 29 et Figure 30). À nouveau, les neurones de la voie magnocellulaire (bandes épaisses) répondent plus rapidement (~60 ms) que ceux de la voie parvocellulaire (bandes étroites et bandes claires : ~80 ms). Les

premières réponses des neurones de V4[25] se situent aux alentours de 60 à 70 ms (moyenne de réponse: ~100 ms) et celles de IT varient fortement entre les différentes sous-régions (Baylis et al., 1987); les régions antérieures de IT étant activées plus tardivement (premières réponses : de 60 à 90 ms ; latence moyenne des réponses: de 110 à 140 ms). Des latences de réponses similaires sont observées dans les différentes sous-régions visuelles du STS. Il semble donc qu'un des facteurs déterminants pour la latence d'activation des régions de IT soit la proximité des connexions avec les aires extrastriées (Baylis et al., 1987).

Les enregistrements de cellules répondant aux visages dans le cortex inféro-temporal (principalement dans la partie antérieure du ITG et du STS) indiquent que la latence moyenne de réponse de ces neurones (~100 ms ; Kiani et al., 2005; Oram et Perrett, 1992 ; Baylis et al., 1985) est relativement similaire à la latence de réponse mesurée pour des neurones présentant d'autres types de sélectivité. Les réponses les plus précoces des cellules-visage apparaissent à environs 70-80 ms après la présentation d'un visage, alors qu'on observe peu de ces cellules répondant avec une latence supérieure à 130 ms. Il a cependant été récemment montré que chez le singe, ces neurones déchargent en moyenne 15 à 20 ms plus rapidement en réponse à des visages de primates (humain et singe) qu'à des visages d'animaux non-primates (Kiani et al., 2005). De même, la latence de réponse des cellules-visage semble légèrement inférieure à celle des neurones répondant à d'autres objets (Edwards et al., 2003). Notons également que les latences de réponse varient selon la région dans laquelle les neurones-visage sont enregistrés. En effet, dans une étude récente, Eifuku et al. (2004) rapportent des latences de réponse entre 90 et 120 ms pour le STS antérieur (site d'enregistrement de nombreuses études sur les neurones-visage) et des latences entre 120 et 200 ms pour la partie ventrale de l'aire TEa (TEav), située à proximité du pôle temporal (voir aussi Tamura & Tanaka, 2001). De plus, ces auteurs décrivent également l'existence de certains neurones-visage situés dans TEav dont la latence de réponse varie entre 200 et 300 ms; une latence bien supérieure à la latence habituellement décrite pour les neurones visages.

[25] Au-delà de V2 dans la voie ventrale, vu qu'il n'y a plus réellement de ségrégation entre la voie magnocellulaire et la voie parvocellulaire, il devient difficile d'évaluer dans quelle mesure les neurones recevant leur input de la voie magnocelulaire répondent effectivement plus rapidement que ceux de la voie parvocellulaire. Par ailleurs, la différence apparente de latence d'apparition des réponses parvo- et magnocellulaire dans le corps genouillé latéral ou dans V1 pourrait ne pas refléter la latence avec laquelle l'information visuelle est effectivement transmise dans le cortex visuel (Maunsell et al., 1999). En effet, au niveau de la vision centrale il y a environ 10 fois plus de neurones parvocellulaires que de neurones magnocellulaires. Etant donné que l'émission d'un potentiel d'action par un neurone dépend de la sommation spatiale et temporelle des potentiels postsynaptiques, un neurone qui reçoit ses inputs d'un nombre très important de neurones parvocellulaires pourrait répondre plus rapidement qu'un autre neurone qui reçoit ses afférences d'un plus petit nombre de neurones magnocellulaires. Ce phénomène pourrait dès lors limiter fortement l'avantage temporel de la voie magnocellulaire par rapport à la voie parvocellulaire (Maunsell et al., 1999).

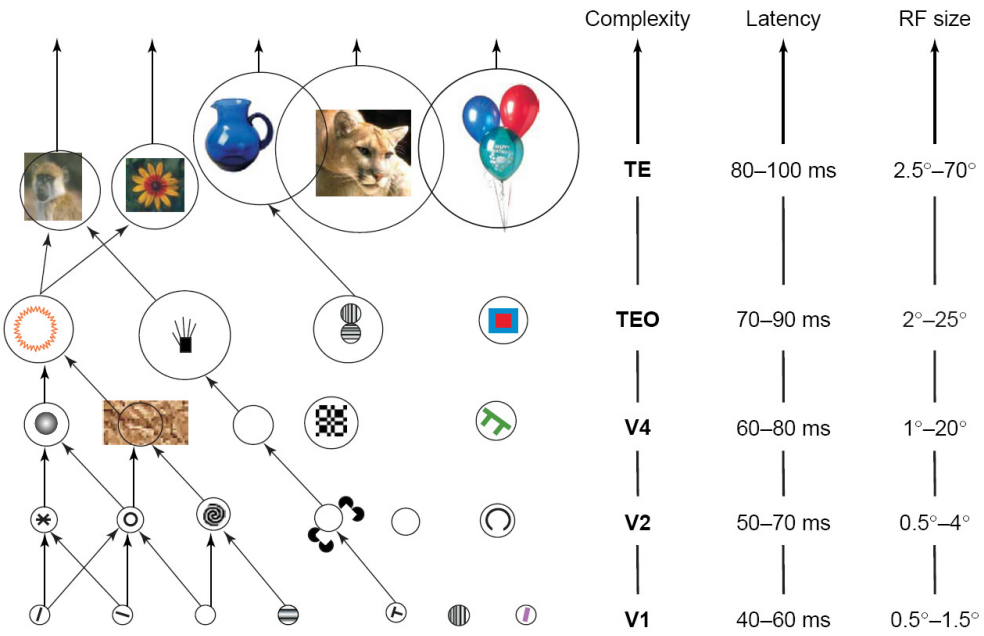

Figure 29: Représentation schématique de la hiérarchie dans la voie ventrale. Les représentations de haut niveau dans TE sont progressivement construites en intégrant les inputs convergents des niveaux inférieurs. La figure montre le lien entre l'augmentation de la latence de réponse, de la taille des champs récepteurs et de la complexité de la sélectivité des neurones. Modifié d'après Rousselet et al. (2004b).

Bien qu'en première approximation les latences d'apparition des premières réponses dans les aires visuelles semblent respecter le principe de la hiérarchie dans les aires visuelles, il est important de préciser que ces latences sont en fait loin de respecter précisément la hiérarchie anatomique (voir Figure 29 et Figure 30). Quels sont les arguments qui soutiennent cette proposition?

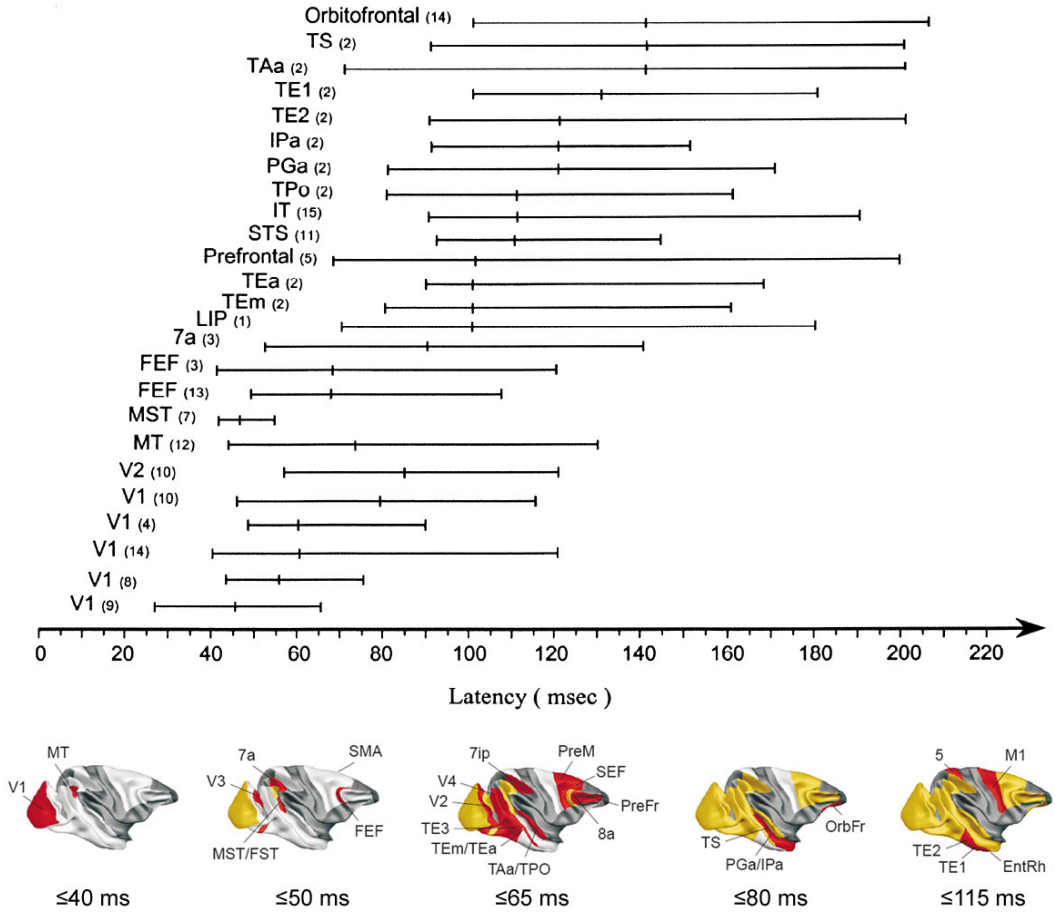

Figure 30: Latences d'activation des aires corticales. En haut, pour chaque aire représentée, le trait central représente la latence médiane d'activation et les traits aux extrémités, les percentiles 10 et 90%. Les latences des aires représentées plusieurs fois correspondent aux résultats obtenus dans des études différentes (modifié d'après Bullier, 2001). Partie inférieure: latences d'activation des aires corticales représentées sur un cerveau de macaque entre 40 ms et 115 ms. Les zones foncées sont les zones activées endéans la latence indiquée sous chaque cerveau. Les zones colorées en blanc sont les zones qui n'ont pas encore été activées (modifié d'après Lamme & Roelfsema, 2000). Cette figure montre que les latences d'activation ne suivent pas nécessairement l'ordre de hiérarchie visuelle anatomique.

Premièrement, le traitement de l'information dans les voies visuelles opère de manière largement parallèle et distribuée (Bullier & Nowak, 1995), des informations différentes étant traitées dans des "canaux" différents (par ex: les voies magno- et parvocellulaires; ségrégation des informations sur la forme, couleur, mouvement et profondeur dans V1,…) et distribuées sur plusieurs aires corticales. Un exemple frappant est la séparation entre la voie dorsale et la voie ventrale. De nombreuses données attestent que les aires visuelles de la voie dorsale (MT, MST, 7a) sont activées beaucoup plus rapidement que les aires de la voie ventrale (Schmolesky et al., 1998; voir aussi Lamme & Roelfsema, 2000). En fait, les premières réponses neuronales dans la majorité des aires visuelles pariétales et même dans des aires non-visuelles, comme les "frontal-eye-field" ou le cortex préfrontal, sont observées avant ou simultanément aux premières réponses de V2 dans la voie ventrale (Figure 30). Sur base de

cette observation, il a été par exemple proposé que l'activation rapide de la voie dorsale et éventuellement du cortex préfrontal pourrait influencer par feedback la sélectivité des réponses des neurones de la voie ventrale avant même que ceux-ci ne soient activés (Bullier, 2001; voir aussi Bar, 2003; 2004). La proposition que le traitement visuel procède de manière parallèle et distribuée est largement compatible avec l'observation qu'il existe une forte variabilité dans la latence de réponse des neurones au sein d'une aire donnée, de sorte que les distributions de latences de réponse d'aires visuelles différentes se recouvrent fortement (voir Figure 28).

Deuxièmement, nous avons vu que les aires visuelles sont densément interconnectées, certaines de ces connexions permettant à l'information de transiter d'une aire à une autre en contournant certains niveaux de la hiérarchie (par ex: de V2 à TEO, de V4 à IT, …). Ceci confirme également le caractère parallèle de l'activation des aires visuelles (Bullier & Nowak, 1995). De plus, pour un même "parcours" le long de la hiérarchie visuelle, certains neurones sont activés beaucoup plus rapidement que d'autres et transmettent ainsi l'information plus rapidement.

Troisièmement, outre la latence d'activation des neurones individuels, un facteur important à prendre en compte dans la vitesse d'activation des voies visuelles est la densité des projections d'une aire à l'autre (Maunsell et al., 1999). En effet, étant donné que les neurones doivent intégrer spatialement et temporellement les inputs des neurones dont ils reçoivent des afférences avant d'émettre eux-mêmes un potentiel d'action (voir la note n°25), deux aires visuelles qui reçoivent des inputs à une latence égale ne répondront pas effectivement à la même latence si le nombre d'inputs qu'elles reçoivent chacune diffère fortement.

En conclusion de cette section, les données sur les latences d'activation des aires visuelles ainsi que les remarques formulées ci-dessus indiquent que l'élaboration des représentations visuelles à un niveau global dans le système visuel procède d'une interaction complexe entre des contraintes anatomiques, des contraintes liées à la fonction de chaque aire visuelle et des contraintes liées à la dynamique temporelle de l'activation de ces aires. De plus, l'interconnectivité bidirectionnelle des aires visuelles entre elles ainsi qu'avec un nombre élevé de structures cérébrales non visuelles, mais également la densité des connexions horizontales entre les neurones d'une même aire visuelle, suggèrent que la perception visuelle

ne peut se résumer à l'input sensoriel issu de la rétine[26] (Young, 2000). En effet, le contexte (attentes, tâches ou actions à réaliser,..) dans lequel prend place ce traitement sensoriel est déterminant probablement à quasi tous les niveaux de la hiérarchie visuelle pour orienter et organiser la perception.

Dans les sections suivantes, nous nous contentons d'examiner le résultat de cette interaction: la construction de représentations menant à la catégorisation visuelle.

2. Vitesse de catégorisation des objets et des scènes visuelles: évidences psychophysiques et électrophysiologiques.

Les premières évidences psychophysiques collectées à propos la vitesse de catégorisation d'objets présents dans des scènes visuelles proviennent des études réalisée par Potter (Potter, 1975 ; 1976; Potter & Levy, 1969) avec un paradigme de présentation visuelle sérielle rapide (" Rapid Serial Visual Presentation " : RSVP). Ce paradigme dans lequel des stimuli sont présentés l'un après l'autre suivant un rythme variable (par ex : de 100 à 300 ms par stimulus) réduit le temps imparti au traitement visuel de chaque image en raison de la succession temporelle rapide des stimuli. Ces études ont montré notamment que la détection d'une image cible parmi une série d'images présentées très rapidement (~100 ms par image) est réalisée au-dessus du niveau du hasard. Par ailleurs, la fréquence maximale à laquelle une lettre ou un visage peut être identifié dans une séquence RSVP n'est pas déterminée par le rapport signal/bruit, manipulé en variant des paramètres de bas niveau comme la taille, le contraste ou la luminance des stimuli (Näsänen et al., 2006). Ces dernières observations suggèrent que la vitesse de traitement d'objets présentés dans une séquence RSVP est limitée par la résolution temporelle des processus visuels de hauts niveau, plutôt que par la résolution temporelle des processus traitant l'information visuelle de bas niveau, tels que le contraste ou la luminance[27] (Näsänen et al., 2006). Cette proposition a été récemment étayée par une étude en neuroimagerie (fMRI) qui montre que la fréquence de présentation de stimuli à laquelle

[26] Cette proposition est notamment illustrée par le fait que dans la couche corticale de V1 (chez le chat) recevant les afférences des CGL, seulement 5% des synapses excitatrices proviennent réellement des CGL; les 95% restant provenant de neurones situés dans d'autres partie V1, dans d'autres aires corticales ou d'autres noyaux thalamiques (voir Young, 2000).

[27] Notons cependant que dans des séquences RSPV, des paramètres de bas niveau augmentant la saillance d'une image par rapport à une autre peuvent attirer l'attention et dès lors moduler la capacité à détecter une cible dans ces images. En d'autres termes, si une image qui précède une image contenant une cible est particulièrement saillante (par exemple si le contraste est élevé), la cible sera moins fréquemment détectée (Einhauser et al., 2007).

l'activation maximale est observée diminue progressivement d'une étape de la hiérarchie visuelle à une autre: V1 répondant de manière maximale à une fréquence de stimulation plus élevée (~20-25 stimuli/sec), que V4v (~10 stimuli/sec) ou que la FFA (~5 stimuli/sec) (Mckeef et al., 2007). Ces observations à propos de la résolution temporelle des aires visuelles de haut niveau (FFA) sont également concordantes avec des résultats obtenus chez le singe qui montrent que dans des séquences RSVP, la sélectivité des neurones-visage dans IT est progressivement réduite lorsque la fréquence de présentation est progressivement augmentée de 5 à 71 images par seconde, et ce parallèlement à la baisse de performance en détection de cibles dans ces séquences (Keysers et al., 2001; voir aussi De Baene et al., 2007). Notons cependant que malgré cette diminution de sélectivité (notamment liée à une réduction de la durée de la réponse), 65% des neurones de IT restaient sélectifs même à très de hautes fréquences de présentation (71 stimuli/sec).

Un autre paradigme largement utilisé pour investiguer le temps nécessaire à la catégorisation ou la reconnaissance d'objets ou de scènes visuelles est le paradigme de présentation rapide masquée (la présentation du stimulus est suivie d'un autre stimulus appelé "masque" : " backward-masking "). La procédure de backward-masking repose sur le fait que la durée allouée au traitement visuel d'un stimulus présenté brièvement est réduite lorsqu'un masque apparaît rapidement après ce stimulus (Breitmeyer & Ogmen, 2000). Cette procédure a donc été appliquée principalement pour investiguer les propriétés temporelles de la perception visuelle.

Dans le domaine de la catégorisation d'objet inclus ou non dans des scènes visuelles, les études de backward-masking révèlent que les performances des sujets dans des tâches de reconnaissance ou de catégorisation sont proportionnelles à la durée de présentation du stimulus ou de l'intervalle entre l'apparition du stimulus et l'apparition du masque ("Stimulus Onset Asynchrony" : SOA). Ces études indiquent que d'une manière générale, un temps de présentation ou un SOA entre 40 et 50 ms permet d'obtenir des performances bien au-dessus du niveau du hasard (~80% dans des tâches à deux possibilités de réponse), et que les performances commencent à plafonner au-delà de 80-100 ms de temps de présentation (Grill-spector & Kanwisher, 2005; Bacon-Macé et al., 2005; voir Figure 31C).

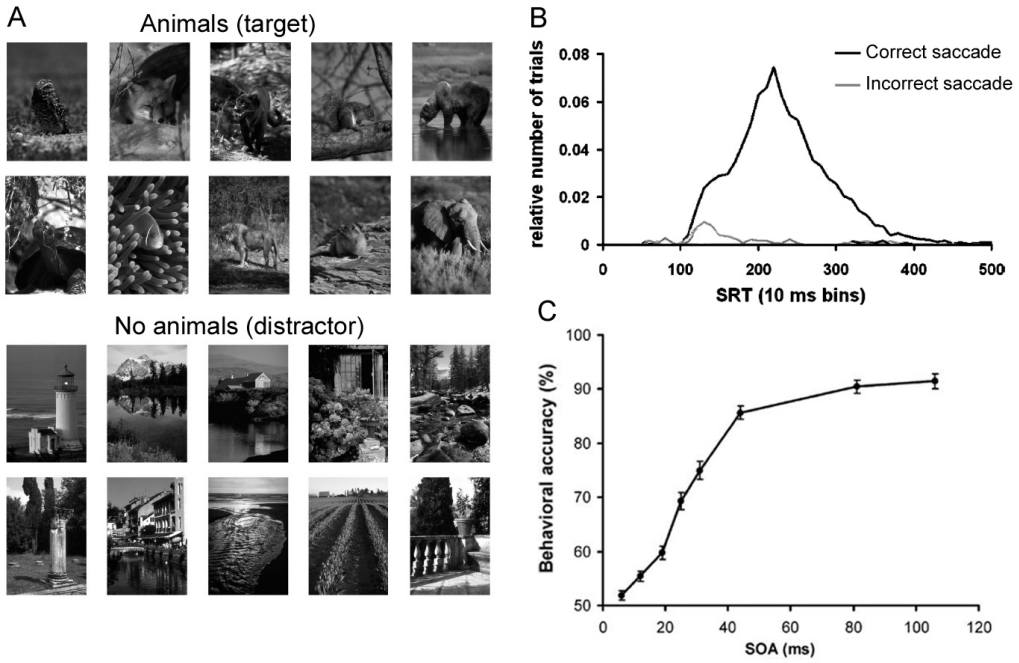

Figure 31: A. Exemples de scènes naturelles contenant un animal (images cibles) ou ne contenant pas d'animal (images distracteurs) utilisées dans Kirchner & Thorpe (2006). B. Distribution sur l'ensemble des sujets des temps de réponses saccadiques (SRT) à une tâche de détection d'animal dans des scènes naturelles (voir A) pour les essais corrects et incorrects (Kirchner & Thorpe, 2006). La proportion d'essais corrects dépasse la proportion d'essais incorrects pour les saccades réalisées à partir d'environs 120 ms après l'apparition de l'image (modifié d'après Kirchner & Thorpe, 2006). C. Performance à une tâche de détection d'animal dans des scènes naturelles en fonction de temps entre l'apparition du stimulus et l'apparition d'un masque visuel (SOA: 6, 12, 18, 25, 31, 43, 81 et 106 ms). Modifié d'après Bacon-Macé et al., 2005.

Àun niveau neurophysiologique, le masque réduit ou interrompt la réponse neuronale liée au traitement du stimulus précédant le masque (Kovacs et al., 1995 ; Keysers et al., 2001 ; Rolls et al., 1994 ; Rolls et al., 1999; voir Keysers & Perrett, 2002 pour une discussion sur l'origine neurophysiologique du masking). De manière intéressante, dans ces paradigmes de présentation masquée, le pattern de modulation de la sélectivité des neurones de IT en fonction de la durée de présentation du stimulus est comparable au pattern de performances comportementales de singes ou de sujets humains (Kovacs et al., 1995 ; Rolls et al., 1994, voir aussi Keysers & Perrett, 2001). Des résultats similaires ont été obtenus dans une étude en IRMf chez l'humain (Grill-spector et al., 2000), où le pattern de performance en dénomination d'objet en fonction du temps de présentation était similaire au pattern de modulation de l'activation du LOC ("Lateral Occipital Complex") impliqué dans la représentation des objets (Grill-spector et al., 2001). De plus, pour un temps de présentation donné (40 ms), les performances des sujets étaient positivement corrélées au niveau d'activation dans cette région (Grill-spector et al., 2000). Plus récemment, une étude menée en potentiels évoqués où les sujets devaient détecter la présence d'un animal dans une scène

présentée pour un durée variable (6 à 106 ms) montre que la différence entre la réponse électrophysiologique évoquée par une cible et celle évoquée par un distracteur diminue progressivement avec la diminution du SOA (Bacon-Macé et al., 2005). De plus, le pattern de modulation des potentiels évoqués en fonction du temps de présentation des images était fortement corrélé avec la performance des sujets en détection d'animal dans ces images.

Une troisième série d'évidences pour une catégorisation visuelle rapide provient des études d'enregistrements neuronaux chez le singe dont nous avons déjà abondamment parlé. Ces études montrent que les neurones de IT présentent une sélectivité pour des stimuli complexes (Gross, 1972; Bruce et al., 1981; Perrett et al., 1982; Desimone et al., 1984; Logothetis & Sheinberg, 1996; Tanaka et al., 1991; Tanaka, 1996; Tamura & Tanaka, 2001; Freedman et al., 2003) et que cette sélectivité est exprimée dans la partie initiale de la réponse des neurones, dès 80-100 ms après l'apparition du stimulus (Sugase et al., 1999; Tamura & Tanaka, 2001; Vogels, 1999; Matsumoto et al., 2005; Tsao et al., 2006; Edwards et al., 2003; Baylis et al., 1987; Hung et al., 2005; Oram & Perrett, 1992; Keysers et al., 2001; Rolls et al., 2006).

Bien que les résultats des études de masking, de RSVP, ainsi que les enregistrements réalisés dans IT chez le singe fournissent une indication sur le temps nécessaire au système visuel pour encoder/représenter un objet, ces études ne donnent pas d'indication sur la vitesse à laquelle ces stimuli sont effectivement *catégorisés* (i.e. incluant le traitement perceptif et décisionnel lié à la catégorisation) par le sujet[28]. La catégorisation rapide d'objets dans des scènes visuelles a été largement investiguée par l'équipe de Simon Thorpe à l'aide d'un paradigme go/no-go dans lequel le sujet ne répond que lorsqu'une cible apparaît[29]. Dans une étude pionnière (Thorpe et al., 1996), les sujets recevaient comme instruction de détecter le plus rapidement possible la présence d'un animal dans un scène visuelle naturelle présentée pendant 20 ms (non masquée). Dans ces conditions, les sujets étaient à la fois performants

[28] En fait, il a été suggéré que les neurones de IT seraient davantage impliqués dans la représentation précise et sélective des objets plutôt que dans leur catégorisation en tant que telle. Des neurones du cortex préfrontal recevant des afférences des neurones de IT seraient quant à eux davantage impliqués dans l'encodage de l'information à propos des catégories définies par leur caractère fonctionnel pour le comportement (Freedman et al., 2003).

[29] Le recours à des designs go/no-go permet de réduire le temps nécessaire à effectuer la commande motrice par rapport à une tâche où les choix de réponse sont assignés à des clefs de réponse motrice différentes. De plus, le fait que l'instruction donnée aux sujets (fournir une réponse motrice *uniquement* dans une des deux conditions) maximise en quelque sorte la différence entre ces conditions, permet également probablement d'accélérer la décision perceptive.

(entre 90 et 98% de réponses correctes) et rapides (temps de réaction médian de 445 ms). De plus, les potentiels évoqués en réponse aux essais "go" (scènes contenant un animal) commençaient à diverger des potentiels évoqués pour les essais "no-go" (scènes ne contenant pas d'animal) aux alentours de 150 ms, indiquant un traitement ultra rapide des scènes visuelles (voir Figure 32). Ce résultat suppose en effet qu'il faut au maximum 150 ms au système visuel pour établir une distinction entre une image contenant et une image ne contenant pas d'animal. De plus, cette activité différentielle sur les potentiels évoqués apparaissant vers 150 ms n'est pas liée à des différences de bas niveau (e.g. luminance, contraste, spectre d'amplitude des fréquences spatiales) entre les images contenant des catégories différentes d'objets (animaux, véhicules), contrairement à des différences plus précoces (~75-150 ms) observées sur les potentiels évoqués (VanRullen & Thorpe, 2001a; Johnson & Olhausen, 2003).

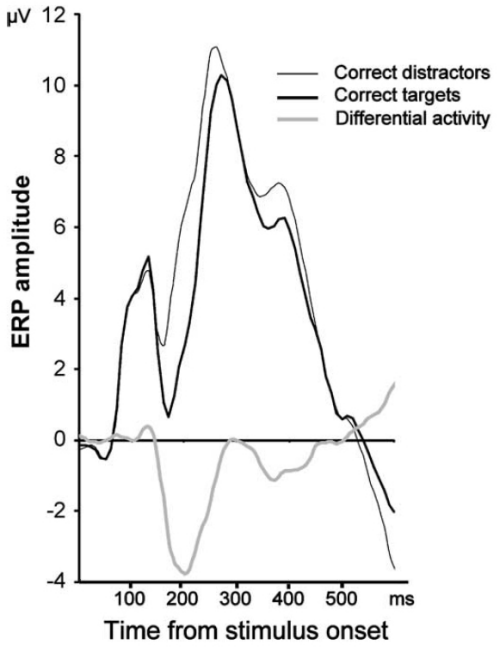

Figure 32: Réponse électrophysiologique (potentiels évoqués) dans une tâche de détection d'animal dans des scènes naturelles. Le sujet ne doit répondre que lorsqu'un animal est présent dans la scène. Dans ces conditions, la comparaison des essais corrects où un animal était présent (target) aux essais où auncun animal n'était présent (distractor), génère une activité différentielle (en gris) qui débute aux alentour de 150 ms. Modifié d'après Bacon-Macé et al., 2005.

Les résultats de l'étude de Thorpe et al. (1996) ont été répliqués à de nombreuses reprises dans des conditions différentes. Par exemple, en utilisant un dispositif de réponse plus sensible et une tâche de catégorisation (animal vs. moyen de transport), VanRullen & Thorpe (2001b) ont obtenu des temps de réaction médians d'environ 350 ms seulement. En

analysant la distribution des temps de réaction chez l'ensemble des sujets, ils ont pu montrer que le temps minimal nécessaire à des humains pour réaliser cette tâche au dessus du niveau du hasard est inférieur à 250 ms (voir aussi Delorme et al., 2000; Macé et al., 2005a). Dans des conditions similaires, des singes entraînés à la même tâche sont encore plus rapides, avec des temps de réaction minimaux entre 180 et 200 ms (Fabre-Thorpe et al., 1998; Delorme et al., 2000; Macé et al., 2005a). Un certain nombre de propriétés de la catégorisation rapide de scènes visuelles ont été mises en évidence par les travaux de cette équipe. Il a ainsi été observé que les performances à ce type de tâches diminuent linéairement avec l'excentricité, mais restent néanmoins très nettement supérieures au niveau du hasard même à forte excentricité (93% vs. 71% de réponses correctes respectivement à 0° et 60° d'excentricité, Thorpe et al., 2001). De même, les performances, temps de réactions et activités électrophysiologiques différentielles (potentiels évoqués) varient de manière non linéaire avec le contraste des stimuli présentés (Macé et al., 2005b). Étonnamment, les performances et temps de réaction sont relativement insensibles à l'orientation des stimuli présentés (Guyonneau et al., 2005; Rousselet et al., 2003).

Plus récemment, les travaux du même groupe se sont également concentrés sur la mesure du temps nécessaire pour effectuer une saccade vers une scène visuelle contenant un animal (Kirchner & Thorpe, 2006; Guyonneau et al., 2006) ou un visage humain (Thorpe et al., 2007; Crouzet et al., 2007). Dans ces études, il a été montré que le temps minimal nécessaire pour effectuer correctement une saccade vers une scène (parmi deux présentées) contenant un animal est de l'ordre de 120 ms (Kirchner & Thorpe, 2006; Guyonneau et al., 2006; voir Figure 31 plus haut), soit la moitié du temps nécessaire à fournir une réponse manuelle dans une tâche similaire. Dans le même type de paradigme expérimental, les sujets sont capables d'effectuer correctement une saccade vers une scène contenant un visage humain en seulement 100 ms (Thorpe et al., 2007; Crouzet et al., 2007).

En résumé, les résultats décris ci-dessus, en particulier ceux qui concernent la catégorisation de stimuli isolés ou contenus dans une scène visuelle, imposent de fortes contraintes temporelles qu' il est nécessaire de prendre en compte lors de la description de processus impliqués dans la catégorisation visuelle.

Ces travaux mettent en effet en évidence la rapidité des traitements impliqués dans la catégorisation visuelle. Sur base des données concernant la vitesse d'activation des régions visuelles chez le singe il a été avancé que ce type de catégorisation se base probablement en grande partie sur l'information visuelle transmise durant la première volée afférente feed-

forward depuis la rétine jusqu'au cortex inféro-temporal, pour activer ensuite les aires impliquées dans le contrôle moteur afin de fournir la réponse. Etant donnés le nombre important de relais à franchir, il a été estimé que chaque aire visuelle traversée ne peut intégrer l'activité neuronale que pendant 10 à 20 ms. Vue la fréquence maximale de décharge d'un neurone (~100 Hz), Thorpe & Imbert (1989) ont proposé que chaque niveau de la hiérarchie n'a le temps d'intégrer qu'un seul potentiel d'action provenant du niveau précédent avant de devoir relayer l'information au niveau suivant[30] (voir aussi Thorpe et al., 1996). La conséquence de ce type de traitement visuel est d'une part que la catégorisation qui en résulte doit nécessairement être basée sur une information limitée et probablement relativement pauvre en détails (voir Figure 33; Fei-Fei et al., 2007), et d'autre part que les neurones doivent présenter une forte sélectivité dès la première décharge, ce qui semble être -au moins partiellement- le cas.

Les travaux de l'équipe de Simon Thorpe, en particulier les résultats obtenus en potentiels évoqués (Thorpe et al., 1996; VanRullen & Thorpe, 2001a; Macé et al., 2005b; Bacon-Macé et al., 2005) et avec des réponses saccadiques (Kirchner & Thorpe, 2006; Guyonneau et al., 2006; Thorpe et al., 2007; Crouzet et al., 2007), et qui montrent une catégorisation de scènes complexes en moins de 100-150 ms incluant la décision (et la commande occulo-motrice pour les réponse saccadiques), posent directement la question du type d'informations utilisées dans les images pour réaliser la tâche. En effet, Torralba & Oliva (2003) ont montré que certains paramètres visuels de bas niveau des images (e.g. variations régionales de la distribution de l'orientation des fréquences spatiales dans les images) sont corrélés avec le type de scènes visuelles (e.g. scènes de forêts, de villes, de plages,…) et qu'un algorithme basé sur une analyse en composantes principales de ces paramètres est capable de catégoriser les scènes avec une performance supérieure à 80%. Même si cet algorithme n'est capable que de catégoriser des types de scènes visuelles plutôt que des objets dans ces scènes, il est possible que l'analyse de ces paramètres par les humains et les singes participent en partie à l'efficacité et à la rapidité de catégorisation observée. Notons que l'analyse d'un grand nombre de paramètres statistiques (luminance moyenne, contraste, kurtosis, énergie,…) des images utilisées par Kirchner & Thorpe (2006) n'a pas permis de révéler de différences entre

[30] Cette proposition néglige cependant le fait que la catégorisation rapide telle que mise en évidence dans les paradigmes go/no-go pourrait en partie bénéficier du traitement de l'information transmise par des connexions qui contourneraient certains niveaux de la hiérarchie visuelle. Notons également que d'autres auteurs sont en désaccord avec l'idée que les représentations visuelles activées de manière feed-forward soient basées sur l'intégration d'une seule une décharge par neurone (par ex: Roll et al., 2006). Selon Rolls et al., 2006, une transmission rapide de l'information pourrait être basée sur un temps d'intégration d'environs 20 ms à chaque niveau de la hiérarchie, ce qui permettrait d'accumuler l'information transmise par les 1, 2 ou 3 premières décharges.

celles qui contenaient un animal (cibles) et celles qui n'en contenaient pas (distracteurs) qui permettraient de rendre compte de leurs résultats. Malgré ces analyses, il est cependant encore difficile à ce stade d'appréhender la mesure avec laquelle des paramètres de bas niveau des images peuvent être utilisés pour réaliser la tâche. De plus, la détection d'animaux dans des scènes visuelles pourrait être basée en partie sur la détection de traits de complexité intermédiaire (Rousselet et al., 2003; 2007b; voir Ullman et al., 2002) qui ne seraient pas décrits dans les statistiques réalisées sur leurs images par Kirchner & Thorpe (2006).

PT 27 ms	There was a range of dark sploches in the middle of the picture, running from most of the way on the left side, to all the way on the right side. This was surrounded primarily by a white or light gray color. (Subject: KM)	Couldn't see much; it was mostly dark w/ some square things, maybe furniture. (Subject: AM)	Looked like something black in the center with four straight lines coming out of it against a white background. (Subject: AM)
PT 40 ms	I saw a very brght object, shaped in a pyramidal shape. There was somethin black in the front, but I couldn't tell what it was. (Subject: JB)	This looked like an indoor shot. Saw what looked like a large framed object (a painting?) on a white background (i.e., the wall). (Subject: RW)	The first thing I could recognize was a dark sploch in the middle. It may have been rectangular-shaped, with a curved top. . .but, that's just a guess. (Subject: KM)
PT 67 ms	Possibly outdoors. maybe a few ducks, or geese. Water in the background. (Subject: JL)	I saw the interior of a room in a house. There was a picture to the right, that was black, and possibly a table in the center. It seemed like a formal dining room. (Subject: JB)	A person, I think, sitting down or crouching. Facing the left side of the picture We see their profile mostly. They were at a table or were some object was in front of them (to their left side in the picture). (Subject: EC)
PT 500 ms	It was definately on a coast byt hte ocean with a large [r]ock in the forground and atleast three bird sitting on the rock. (Subject: CC)	Some fancy 1800s living room with ornate single seaters and some portraits on the wall. (Subject: WC)	This looks like a father or somebod helping a little boy. The man had something in his hands, like a LCD screen or laptop. they looked like they were standing in a cubicle. (Subject: WC)

Figure 33: Exemples de descriptions données par des sujets à propos des trois images présentées au-dessus en fonction du temps de présentation des images (27, 40, 67 ou 500 ms, présentation en backward masking) (Fei-Fei et al., 2007). Ces descriptions illustrent le type de représentations qu'il est possible d'extraire d'une scène visuelle présentée très rapidement.

Enfin, quelle que soit l'information utilisée pour catégoriser des objets ou des scènes, il est probable que cette catégorisation résulte d'une accumulation progressive d'informations à propos de la stimulation (Bacon-Macé et al., 2005). Il est dès lors envisageable que les images donnant lieu aux réponses les plus rapides possèdent des caractéristiques qui nécessitent d'accumuler peu d'information et que les réponses sont dans ce cas basées sur une analyse de

ces caractéristiques de bas niveau ou de niveau intermédiaire[31]. En revanche, les réponses produites avec des temps de réaction plus élevés pourraient être basées sur une représentation de plus haut niveau de la scène (résultant d'une plus forte accumulation d'information) élaborée dans le cortex inféro-temporal. Sur base de cette proposition, on pourrait inférer que dans ces tâches de catégorisation rapide, les temps de réaction moyens et médians (~350-400 ms pour les réponses manuelles et ~230-250 ms pour les réponses saccadiques) fournissent une indication plus conservatrice (et peut-être plus représentative) du temps nécessaire pour catégoriser un objet ou une scène visuelle. En outre, même si le système visuel est capable d'extraire une information suffisante pour réaliser une tâche de catégorisation, il est plus que probable que la représentation continue à s'affiner dans les aires visuelles après que la réponse ait été fournie. En d'autres termes, même si une représentation précise est élaborée lorsqu'une scène visuelle est présentée, il est probable que suffisamment d'informations soient disponibles pour prendre une décision correcte avant que ne soit "terminée" l'élaboration de cette représentation, et ce probablement avant que la perception de cette information ne soit consciente (Dehaene & Naccache, 2001).

En conclusion, rappelons que l'ensemble des travaux décrits dans cette section souligne la rapidité avec laquelle l'information est accumulée dans le système visuel. Ces travaux suggèrent également que la catégorisation rapide se base probablement sur une information limitée à partir de laquelle des inférences (par ex: tel que formulé dans le cadre des probabilités Bayesiennes; voir Kersten et al., 2004) sont produites par rapport au stimulus rencontré (Young, 2000).

3. Vitesse d'activation des représentations faciales: neurophysiologie et psychophysique

Nous venons de voir que le système visuel peut extraire extrêmement rapidement une information suffisante pour opérer une catégorisation grossière d'objets ou des scènes visuelles. Dans cette section, nous allons examiner si le même type de catégorisation très

[31] On pourrait alors penser que dans le cas des réponses très rapides, la décision ne soit pas basée sur une représentation de la catégorie à proprement dite, mais plutôt sur une représentation partielle de moindre complexité (voir Ullman et al., 2002). Il est à noter que l'observation du temps de réponse saccadique de l'ordre de 100-120 ms est relativement incompatible avec l'idée que la décision serait basée sur une représentation du stimulus dans IT (dont la latence moyenne d'activation est de 100 ms chez le singe). Pour tenir compte de cette incompatibilité, Kirchner & Thorpe (2006) proposent que les réponses les plus rapides se basent probablement sur l'information représentée au niveau de V4 et dans les aires visuelles antérieures dans la hiérarchie visuelle.

rapide est également possible lorsque que le stimulus visuel est un visage. Les données présentées dans les sections suivantes nous permettrons d'avoir un aperçu plus précis du temps nécessaire à la catégorisation d'un visage ou à l'élaboration d'une représentation faciale individuelle nécessaire à la reconnaissance. Nous décrirons tout d'abord les données issues des travaux réalisés chez le singe par enregistrement cellulaire de neurones-visage. Nous aborderons ensuite les données comportementales obtenues chez l'humain dans des tâches de catégorisation faciale et de reconnaissance individuelle.

3.1. Électrophysiologie chez le singe

À quelle vitesse la catégorie "visage" est-elle représentée par les neurones du cortex inféro-temporal et à quelle vitesse la réponse de ces mêmes neurones contient-elle une information sur l'identité faciale? Ce sont ces deux questions que nous allons aborder dans cette section en parcourant les évidences issues des travaux d'enregistrement de neurones uniques chez le singe. Comme déjà indiqué précédemment, ces questions sont particulièrement importantes lorsqu'il s'agit de déterminer le décours temporel de le perception des visages chez l'humain. En effet, une description de ce décours temporel devra tenir compte des contraintes imposées par les propriétés temporelles de la structure biologique qui représente ces informations faciales. Les données les plus précises quant à cette dynamique temporelle sont celles obtenues par enregistremments cellulaires chez le singe macaque.

Nous avons vu que les premières réponses des neurones répondant spécifiquement aux visages dans le cortex inféro-temporal du singe sont en général observées aux alentours de 70-80 ms[32] avec une moyenne sur l'ensemble des neurones située aux alentours de 100 ms (Oram & Perret, 1992; Kiani et al., 2005; Rolls et al, 2006; Edwards et al., 2003; Sugase et al., 1999; Matsumoto et al., 2005; Afraz et al., 2006). Il semble qu'en moyenne, lors de la présentation d'un stimulus visuel complexe, les neurones-visage présentent une sélectivité à la catégorie des visages qui débute vers 80-90 ms après l'apparition du stimulus (Figure 34) (Sugase et al., 1999; Matsumoto et al., 2005; Afraz et al., 2006; Hung et al., 2005; Tsao et al., 2006; Baylis et al., 1987). Une étude récente montre qu'une micro-stimulation électrique

[32] Certaines études rapportent également des neurones-visage qui commencent à décharger encore plus précocement, aux alentours de 50 à 60 ms (Edwards et al., 2003; Sugase et al., 1999; Baylis et al., 1987). Cependant, ces neurones représentent probablement une proportion minime (~5%) de l'ensemble des neurones enregistrés dans IT (voir Baylis et al., 1987).

appliquée à des petits groupes de neurones dans IT affecte de manière maximale la catégorisation de visages présentés dans du bruit visuel lorsque cette stimulation est appliquée entre 50 et 100 ms après l'apparition du stimulus (Afraz et al., 2006). De plus, les auteurs de cette étude ont également observé une forte corrélation entre la sélectivité aux visages (par rapport à d'autres catégories) des neurones stimulés et les performances à la tâche de catégorisation faciale. En d'autres termes, plus les neurones stimulés étaient sélectifs aux visages, plus les singes avaient tendance à répondre qu'ils avaient perçu un visage dans la stimulation visuelle. Ces résultats confirment que l'activité spécifique aux visages observée avant 100 ms dans ces neurones-visage est impliquée de manière déterminante dans la catégorisation faciale.

La dynamique temporelle du codage de l'identité faciale par les neurones-visage de IT chez le singe a été principalement investiguée par des approches basées sur la théorie de l'information (Matsumoto et al., 2005; Tovee & Rolls, 1995; Rolls et al., 1999; Rolls et al., 1997; Rolls et al., 2006; Sugase et al., 1999; Tsao et al., 2006). Dans ces études, la quantité et la qualité de l'information transmise par un neurone ou une population de neurones sont mesurées en fonction du temps écoulé depuis l'apparition du stimulus ou depuis le début de la réponse des neurones. L'information transmise représente la mesure avec laquelle la réponse des neurones[33] est sensible aux paramètres manipulés dans le set de stimuli présentés (catégorie du stimulus, identité, expression faciale,…) et permet de discriminer entre ces stimuli.

[33] La réponse neuronale est généralement exprimée en termes de nombre de décharges/potentiels d'action par unité de temps, mais il est également possible de déterminer l'information transmise si seulement une seule décharge par neurone est considérée, ou encore si uniquement un codage de l'ordre de la première décharge de chaque neurone à travers une population de neurone est considéré (voir VanRullen et al., 2005 pour une revue récente des données en faveur de cette hypothèse). Un dernier type de code neuronal peut être exprimé dans le profil temporel de la réponse neuronale (Richmond et al., 1990).

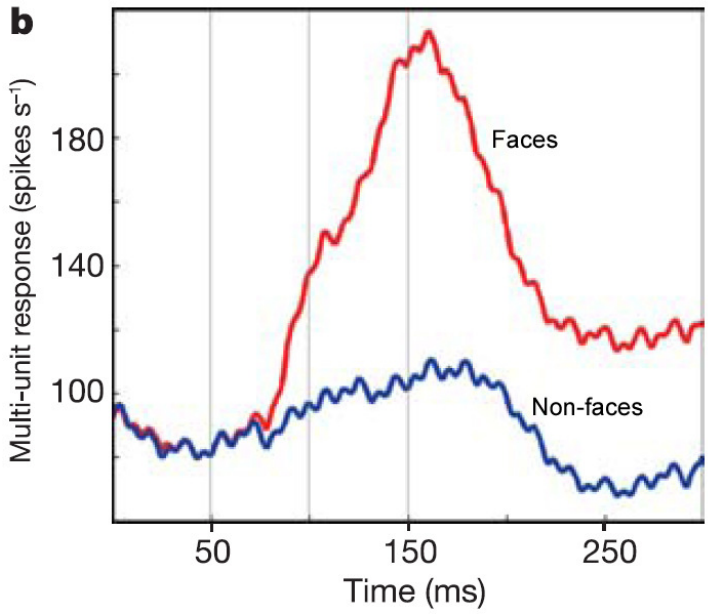

Figure 34: Histogramme de l'activité d'un groupe de neurones-visage en réponse à des visages (courbe "faces") et à des objets non-faciaux en fonction du temps après l'apparition du stimulus. Les courbes se séparent vers 80 ms. (Adapté d'après Afraz et al., 2006).

En mesurant l'information transmise à propos de l'identité faciale dans des fenêtres temporelles successives ou des fenêtres de durées variables, les travaux menés notamment par Edmund Rolls ont montré que la partie initiale de la réponse des neurones de IT contient un pourcentage important de l'information contenue dans des fenêtres temporelles plus longues. Par exemple, Tovee & Rolls (1995) observent que les 20 premières millisecondes de la réponse de neurones individuels de IT (100-120 ms) contiennent en moyenne environ 35% (55% pour une fenêtre de 50 ms) de l'information contenue dans une fenêtre s'étendant de 100 à 500 ms après la stimulation (voir aussi Tovee et al., 1993; Rolls et al., 1997; 1999; 2006). Dans une de ces études (Rolls et al., 1997), l'information présente dans le pattern de réponse d'une population de 14 neurones entre 100 et 150 ms après l'apparition du stimulus permettait de prédire avec une performance de 40% quel visage parmi 20 visages différents avait été présenté au singe (niveau du hasard = 5%). En comparaison, une fenêtre de 100 à 600 ms permettait d'atteindre 70% de réponses correctes.

Une autre équipe de chercheurs a abordé le décours temporel de la transmission de l'information dans les neurones de IT en tentant de déterminer le type et la résolution de l'information représentée en fonction du temps (Sugase et al., 1999; Matsumoto et al., 2005). Dans leur étude initiale, Sugase et al. (1999) ont utilisé un set de stimuli comprenant trois catégories (visages de singes, visages d'humains, formes géométriques). Les stimuli de chaque catégorie se différenciaient sur deux dimensions (identité et expression pour les visages;

forme et couleur pour les formes géométriques). Ces auteurs se sont intéressés aux neurones qui présentaient à la fois une sélectivité dans leur réponse pour la catégorie "globale" (visage de singe vs. visage humain vs. forme) et une sélectivité pour un ou plusieurs exemplaires des différentes catégories (information "fine": par exemple sélectivité pour une des trois identités de visage humain). De manière intéressante, l'information globale était représentée plus précocement dans la réponse des neurones individuels (latence moyenne pour l'information globale: 90 ms, voir aussi Tamura & Tanaka, 2001), alors que l'information fine à propos de l'identité était représentée plus tardivement (latence moyenne: 154 ms pour les visages humains; voir Figure 35). Ces résultats ont été répliqués par Matsumoto et al. (2005) avec le même set de stimuli, en analysant cette fois la réponse des neurones au niveau de la population (et non plus au niveau des neurones individuels comme dans Sugase et al., 1999) avec une approche par analyse en composantes principales. Dans cette étude, l'information globale et fine était représentée de manière maximale dans des fenêtres temporelles respectivement de 90 à 140 ms et de 140 à 190 ms. En outre, l'information globale était toujours représentée dans la fenêtre où l'information fine était maximale (140-190 ms) indiquant que l'information globale et fine est accumulée (au moins partiellement) en parallèle.

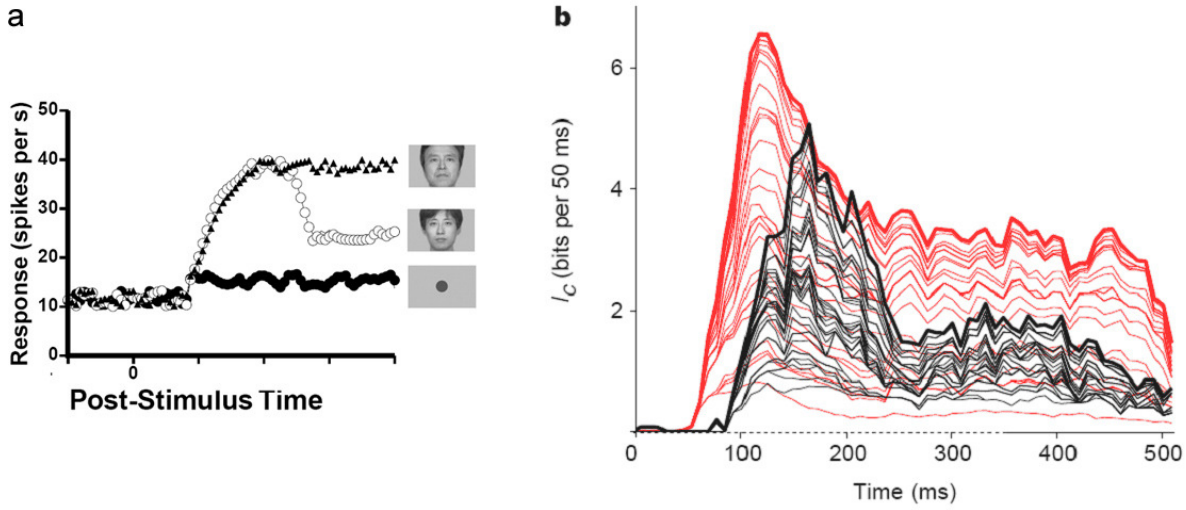

Figure 35: A. Représentation schématique de la manière dont un neurone du cortex inféro-temporal pourrais successivement discriminer entre des stimuli de catégories différentes (visage vs. forme) et ensuite entre des visages d'identités différentes. Modifié d'après Oram & Richmond, 1999. B. Histogrammes représentant la sommation (cumulée) de l'information transmise à propos de la catégorie de stimulus (en gris: information grossière: visage de singe vs. visages d'humain vs. forme carrée) et de l'information fine à propos des stimuli (en noir: identité faciale et expression faciale) pour un groupe de 32 neurones. On voit le décalage temporel entre l'accumulation d'information fine par rapport à l'information grossière. Modifié d'après Sugase et al., 1999.

A première vue, la proposition selon laquelle l'information sur l'identité faciale n'apparaîtrait dans la réponse neuronale qu'après un certain délai par rapport au codage de l'information globale peut sembler contradictoire avec les observations de l'équipe de Rolls qui montrent qu'une information significative sur l'identité faciale est présente dès le début de la réponse de ces neurones. Cependant, une série de facteurs distinguant ces deux groupes d'études pourrait rendre compte des différences de résultats. Premièrement, dans les études de Rolls référencées ci-dessus, l'analyse de l'information transmise est réalisée avec un set de 20 stimuli composés de visages humains et de visages de singes (Tovee & Rolls, 1995; Rolls et al., 1997) ou un set composé d'objets de catégories différentes (visages, objets, scènes naturelles, formes géométriques – Rolls et al., 2006), confondant alors partiellement l'information transmise à propos de la catégorie du stimulus (information "globale") et l'information transmise à propos de l'identité des stimuli[34]. Notons que des latences similaires de codage d'un stimulus particulier on été récemment obtenues (~90-100 ms), mais encore une fois en utilisant à la fois des stimuli de catégories différentes et des exemplaires de stimuli d'une même catégorie (Hung et al., 2005). Deuxièmement, il est possible que le faible nombre d'identités de visages humains (N=3) utilisés par Sugase et al. (1999) et Matsumoto et al. (2005) (par rapport à 20 visages dans Tovee & Rolls, 1995; Rolls et al., 1997) ait provoqué une sous-estimation de l'information transmise par les neurones enregistrés. Troisièmement, dans Sugase et al (1999), les latences auxquelles une information significative apparaît dans la réponse des neurones sont des latences moyennes établies sur l'ensemble des neurones enregistrés. Au vu des larges écarts-types mentionnés pour ces latences moyennes (par ex: information sur l'identité des visages humain: 150 +/- 35 ms), il est probable que certains neurones encodent l'identité faciale et la catégorie globale de manière quasi-simultanée. Pour soutenir cette proposition, notons que dans cette étude la différence moyenne entre l'apparition de l'information globale et l'information fine est de 51 ms avec un écart-type de 39 ms. Cette remarque ne remet cependant pas en cause le fait qu'*en moyenne*, l'information globale est significativement présente dans la réponse des neurones avant l'information plus fine. En revanche, dans un modèle de reconnaissance visuelle où l'information est d'abord transmise de manière feed-forward à travers la hiérarchie visuelle, il est possible que l'information effectivement transmise par IT et utilisée lors d'une décision rapide à propos de l'identité faciale soit déterminée par la réponse des neurones dont la sélectivité pour l'identité est la plus précoce. Enfin, notons que Rolls indique que l'information sur l'identité du visage,

[34] Sugase et al. (1999) ont également montré que les neurones individuels codaient la catégorie "visages de singes" vs. "visages d'humains" plus rapidement que l'identité des visages dans ces catégories respectives.

même si elle est présente dès les premières décharges, continue à s'accumuler de manière importante entre 100 et 200 ms après l'apparition du visage. De la même manière, lorsque la durée de réponse des neurones est réduite par l'apparition d'un masque (Kovacs et al., 1995; Rolls et al., 1999) ou l'utilisation de séquences RSVP très rapides (Keysers et al., 2001), les neurones ne conservent que partiellement leur sélectivité par rapport à la situation où aucun masque n'apparaît après le stimulus. Dans des conditions semblables, des sujets obtiennent des performances en identification de visages au-dessus du niveau du hasard (Rolls et al., 1994; Keyser et al., 2001); mais néanmoins bien inférieures aux performances obtenues lorsque les stimuli sont présentés plus longtemps.

Les données revues ci-dessus s'inscrivent parfaitement dans le cadre de modèles de la prise de décision dans des systèmes neuronaux comme résultant d'une accumulation d'évidence à propos de la stimulation rencontrée (Gold & Shadlen, 2001; Shadlen & Newsome; 1996; Bogacz, 2007; Reddi & Carpenter, 2000 ; Perrett et al., 1998 ; Tjan, 2001) (voir aussi les modèles de diffusion en psychologie cognitive: Ratcliff, 1988, 2006). Selon les modèles d'accumulation, la prise de décision (au niveau neuronal) qui détermine la perception d'un stimulus ou le déclenchement d'un comportement se base sur l'accumulation progressive d'informations (définies en termes d'intensité d'activité neuronale) dans les systèmes de neurones impliqués dans le traitement de ces informations. La décision survient lorsque suffisamment d'informations ont été accumulées pour franchir un seuil (de décision[35]). La vitesse d'une décision donnée dépend essentiellement de la vitesse d'accumulation des informations et du seuil de décision. Le taux d'accumulation dépend de facteurs jouant sur l'intensité du signal présent dans la stimulation[36], de facteurs attentionnels qui peuvent augmenter le rapport signal/bruit dans la stimulation, et également de la relation entre la stimulation et les représentations internes de l'observateur[37]. Le seuil de décision est fonction

[35] Le terme *décision* doit être appréhendé ici dans un sens très large. Il peut correspondre aussi au fait qu'une information suffisante est accumulée dans le système visuel dans sa globalité pour générer une action, mais pourrait également simplement correspondre au fait que suffisamment d'activité a été accumulée dans un neurone pour générer un potentiel d'action.

[36] L'intensité du signal est lui-même fonction des propriétés de bas niveau de la stimulation telles que la contraste ou la luminance, du nombre de stimulation présentes simultanément dans le champ visuel, mais également de la visibilité ou disponibilité des éléments de la stimulation diagnostiques à la décision (par exemple: les yeux pour un visage).

[37] La vitesse d'accumulation sera fonction du nombre de neurones impliqués dans cette accumulation, et ce nombre dépend directement de l'expérience visuelle avec un stimulus particulier. Par exemple, un nombre plus important de neurones est sélectif pour des stimuli présentés dans leur orientation ou leur vue naturelle (par ex: visages à l'endroit en vue de face ou de ¾) (Perrett et al., 1985; Oram & Perrett, 1992; Ashbridge et al., 2000). Dés lors, au niveau d'une population de neurones, la réponse est plus forte pour des visages à l'endroit (correspondant à l'orientation préférentielle des neurones) que pour des visages à l'envers, ce qui provoque une

de la nécessité de privilégier soit la rapidité de réponse soit la performance (rapport vitesse/performance), de facteurs de type top-down comme le contexte, la prédictibilité de la réponse ou la tâche à réaliser, et de la similarité entre les stimulations présentées. Ainsi, par exemple, un rapport signal/bruit plus élevé (i.e. une représentation à plus haute résolution) sera davantage nécessaire pour une discrimination fine (comme l'identification d'un visage) que pour une discrimination plus "globale" comme la simple détection d'un visage. De même, dans une tâche de reconnaissance ou de discrimination de visages, les stimulations présentées sont généralement plus similaires (deux visages d'identités différentes) que les stimulations présentées dans une tâche de catégorisation (par exemple: un visage vs. une voiture). Par conséquent, l'identification nécessitera l'accumulation d'informations sur un temps plus long que la détection. De plus, la manipulation des facteurs jouant sur la décision peut empêcher certaines décisions d'être prises de manière performante, si par exemple il y a trop de bruit dans la stimulation ou si trop peu de temps est accordé pour accumuler des informations nécessaire à la réalisation de la tâche.

Ce cadre permet donc d'envisager que l'accumulation d'informations à propos de l'identité faciale est moins rapide et atteint le seuil de décision de manière retardée par rapport à l'accumulation d'information à propos de la catégorie de stimulus présenté (par exemple: visage vs. voiture). De plus, le fait qu'une information sur l'identité faciale semble présente dans la réponse des neurones de IT dès les premières décharges (Tovee & Rolls, 1995; Rolls et al., 1999; Tsao et al., 2006), tout comme l'est l'information sur la catégorie du stimulus (Sugase et al., 1999; Afraz et al., 2006; Rolls et al., 2006), indique que l'information à propos de ces deux niveaux de catégorisation commence probablement à s'accumuler de manière simultanée (voir Perrett et al., 1998). Dans cette situation, étant donné qu'une représentation plus fine est nécessaire pour effectuer une décision sur l'identité par rapport à une décision sur la catégorie, les informations accumulées à propos de l'identité ne deviennent effectivement utilisables par le système visuel que plus tardivement par rapport aux informations sur la catégorie.

On verra également dans la section suivante ainsi que dans le chapitre suivant que dans certaines situations, ce cadre permet d'envisager que des informations suffisantes soient accumulées dans le système visuel pour signaler la présence d'un visage avant même que ne soient activées les représentations neuronales spécifiques au visage (c.-à-d. les représentations

---

accumulation d'information beaucoup plus rapide pour les visages à l'endroit que pour les visages présentés à l'envers (Perrett et al., 1998; Ashbridge et al., 2000). Ces différences de sélectivité neuronale sont le résultat d'un apprentissage et correspondent aux propriétés statistiques visuelles de l'environnement naturel (voir Logothetis et al., 1995; Perrett et al., 1998).

liées à l'activité des neurones-visage dans IT ou le cortex occipito-temporal chez l'humain). Cette décision pourrait être en effet basée sur l'analyse des propriétés visuelles de bas niveau des visages ou de niveau "intermédiaire" qui sont statistiquement plus fréquemment associées avec la catégorie des visages qu'avec d'autres catégories d'objets. Dans une telle situation, la décision serait donc basée sur l'accumulation d'informations dans des aires visuelles de plus bas niveau (V1, V2, V4,…) non directement sensible à la catégorie du stimulus présenté.

Pour résumer cette section, nous avons vu que le traitement visuel procède par une accumulation et une intégration rapide d'informations et que nombreux facteurs peuvent moduler la vitesse à laquelle ces informations sont accumulées et la vitesse à laquelle la décision est prise. De plus, la quantité d'informations qui doivent être accumulées ainsi que le temps nécessaire à l'accumulation de ces informations est fonction du type de décision à réaliser. Nous avons également vu que les données de la littérature chez le singe à propos du traitement de l'information faciale sont compatibles avec ce modèle d'accumulation. En effet, les neurones accumulent les informations sur la catégorie de stimulus présenté et sur leur identité de manière partiellement parallèle, représentant *significativement* les deux types d'informations dans des fenêtres temporelles distinctes mais se chevauchant partiellement. Il semble donc que la représentation d'un visage devienne de plus en plus précise à mesure que le temps s'écoule et que l'information est accumulée. Enfin, ces données indiquent que les informations sur l'identité faciale apparaissent rapidement dans la réponse des populations de neurones dans IT, suggérant que des informations suffisantes à l'identification faciale sont présentes dans la réponse de ces neurones entre 100 et 200 ms après l'apparition du visage. Notons cependant que ces informations continuent à être accumulées tout au long de la réponse neuronale, du moins jusqu'à au moins 400 ou 500 ms après l'apparition du visage (Rolls et al., 1999; 1997; 2006).

Enfin, les résultats obtenus chez le singe, et qui peuvent probablement être au moins partiellement transposés à l'humain, mettent en évidence une contrainte importante qui pèse sur la dynamique temporelle de la perception des visages. Premièrement, la catégorisation faciale est liée à l'activation des neurones-visage qui ne répondent pas avant 70-90 ms chez le singe. Deuxièmement, les informations utiles à la catégorisation et à l'identification sont accumulées de manière simultanée, et il n'existe qu'un faible délai (moins de 40-50 ms) entre la catégorisation globale du visage et l'activation d'une représentation individuelle. A ce stade, le lecteur doit garder à l'esprit que cette contrainte temporelle identifiée chez le singe est

essentielle et qu'elle a largement inspiré les travaux empiriques réalisés dans le cadre de cette thèse; ce tant dans les hypothèses que dans les interpétations.

3.2. Études comportementales chez l'humain

Nous avons vu dans la section précédente qu'au niveau des neurones de IT, les informations sur la catégorie du stimulus présenté (e.g. visage vs. objet) sont accumulées plus rapidement que les informations à propos de l'identité faciale. Nous verrons ci-dessous que cette différence de taux d'accumulation d'informations est reflétée au niveau du temps nécessaire pour réaliser une tâche impliquant la détection ou la catégorisation faciale par rapport à une tâche impliquant des jugements sur l'identité faciale. Étant donné que les mesures comportementales ne fournissent pas une mesure en continu des processus visuels, décisionnels et moteurs prenant place entre l'apparition d'une image et la réponse motrice, les différences de temps de réaction entre catégorisation faciale et reconnaissance ou discrimination de l'identité ne devraient pas être interprétées comme des évidences pour un traitement sériel et discret de l'information faciale.

3.2.1. A quelle vitesse un visage est-il détecté dans le champ visuel ?

Dans cette section, nous allons passer en revue les travaux réalisés sur la vitesse à la laquelle un visage est catégorisé ou détecté[38] dans l'environnement visuel. Nous aborderons également brièvement la question du type d'informations qui sont susceptibles d'être utilisées dans ce processus de catégorisation.

Dans les études menées récemment sur le temps nécessaire à la détection d'un visage lors d'une stimulation visuelle, les temps de réaction moyens oscillent entre 400 et 500 ms (Lewis & Edmonds, 2003; 2005; Rousselet et al., 2003; Halit et al., 2006). En utilisant un

[38] Nous utiliserons ici les termes *détection faciale* et *catégorisation faciale* comme des synonymes. En effet, si l'on peut généralement faire une distinction entre la détection d'une forme "segmentable" dans le champ visuel (y a t'il un objet dans la scène présentée?) et la catégorisation de cette forme (quel est cet objet ?) (Nakayama et al. 1995; Mack et al., in press), cette distinction n'est pas spécifiquement liée au domaine des visages. Nous utiliserons le terme *détection faciale* pour faire référence au fait qu'un stimulus présent dans le champ visuel a été détecté comme faisant partie de la catégorie des visages. Par ailleurs, même si de manière générale la segmentation des objets/formes présents dans une scène visuelle débute avant la catégorisation et nécessite d'accumuler moins d'informations (Mack et al., 2007; voir aussi Fei-Fei et al., 2007), l'élaboration d'une représentation complexe dépend probablement d'un processus interactif où les formes segmentées sur base d'une information visuelle de bas niveau activent partiellement des représentations internes de plus haut niveau (liées aux catégories) qui elles mêmes vont agir de manière rétro-active (feedback) pour orienter la segmentation (Peterson & Gibson, 1994). Dans ce cadre, il est dès lors difficile d'établir une distinction franche entre détection et catégorisation.

paradigme go/no-go dans une tâche de détection de visages et une analyse de la distribution des temps de réaction, Rousselet et al. (2003) ont montré que les temps de réaction minimums présentant un biais vers une proportion plus forte de réponses correctes sont de seulement 260 ms (Figure 36). Ces temps de réaction sont très similaires à ceux mesurés dans des conditions semblables lors de tâches de détection ou de catégorisation d'objets non-faciaux dans des scènes visuelles dont nous avons parlé précédemment (VanRullen & Thorpe, 2001b).

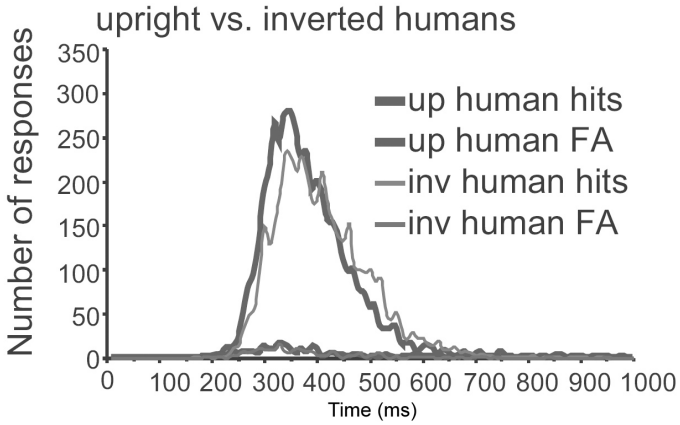

Figure 36: Distribution des temps de réactions (nombre de réponses données par séquence 10 ms) dans une tâche de détection de visages (go/no-go) dans des scènes naturelles présentées à l'endroit ou à l'envers. Les distributions des temps de réaction pour les essais corrects (hits – distributions les plus larges) et les essais incorrects sont présentées. Modifié d'après Rousselet et al., 2003.

Notons que de nombreux facteurs affectent la vitesse à laquelle un visage est détecté dans une scène visuelle (dans le sens que ces facteurs influencent la vitesse d'accumulation d'informations dans le système). Lewis & Edmonds (2003) ont par exemple montré que le caractère signifiant de la scène accélère la détection d'un visage, suggérant que la présence du corps sous le visage accélère sa détection[39]. De même, la manipulation de propriétés de bas niveau (contraste, distribution des couleurs, spectre de fréquence spatiale, présence des yeux[40]) qui sont statistiquement fréquemment associés à la catégorie des visages mène à une modulation des performances en détection.

Une question transversale dans la littérature sur la détection faciale est de savoir si la détection des visages fait appel à des processus spécifiques qui permettraient par exemple

[39] Des nombreux travaux mettent en avant l'importance du contexte général d'une scène visuelle (contenu sémantique ou position relative des objets) dans l'orientation de l'attention et la détection d'objet dans ces scènes visuelles (Oliva & Torralba, 2006; Chun & Jiang, 1998; voir Henderson, 2003 pour une revue de littérature).

[40] Un effet lié à la présence des yeux dans l'image peut-être interprété comme un effet de bas niveau en tant qu'ils constituent la région la plus contrastée du visage, ou comme un effet de haut niveau, lié à la configuration faciale normale. Le fait que l'inversion d'un visage affecte les performances en détection suggère qu'il existe au moins en partie un facteur de haut niveau dans l'utilisation de la région des yeux pour détecter un visage.

d'accélérer la détection faciale (Farah et al., 1998, Tovee, 1998). Cette question a notamment été abordée avec des paradigmes de recherche visuelle permettant de mettre en évidence l'efficacité avec laquelle un stimulus cible est détecté parmi des distracteurs[41]. Cette littérature s'est principalement attachée à déterminer la mesure avec laquelle un éventuel effet "pop-out" pour les visages (Nothdurft, 1993) ou pour les expression faciales négatives (peur ou colère, Foxe et al., 2000; Purcell et al., 1996; voir également Vuillemier, 2005) est lié à l'utilisation de représentations de haut niveau spécifiques aux visages (par ex: le traitement holistique[42]) ou bien s'il résulte de paramètres confondus tels que les caractéristiques visuelles de bas niveau des stimuli, généralement mal contrôlés dans ces études (Purcell et al., 1996; Hershler & Hochstein, 2005; VanRullen, 2006). Par exemple, deux études récentes ont trouvé un effet "pop-out" pour les visages dans une tâche de recherche visuelle (Hershler & Hochstein, 2005; VanRullen, 2006), mais ces deux études diffèrent complètement dans leur interprétation de cet effet. D'une part, pour Hershler & Hochstein, l'effet pop-out pour les visages repose sur le traitement holistique spécifique à cette catégorie. D'autre part, VanRullen (2006) réplique l'effet de pop-out pour les visages mais montre que cet effet est lié à des différences de bas niveau entre le visage (la cible) et les distracteurs (spectre d'amplitude, similarité visuelle plus grande entre les cibles et les distracteurs quand les cibles sont des voitures que quand les cibles sont des visages[43]). Il montre également que l'inversion des stimuli (une manipulation qui perturbe le traitement holistique des visages: Tanaka & Farah, 1993; Young et al., 1987; Sergent, 1984; McKone, 2004) affecte l'efficacité de la recherche visuelle, mais de manière

[41]Si la vitesse de détection de la cible est peu influencée par le nombre de distracteurs (<10 ms/item), la recherche est dite parallèle (effet "pop-out"). Par contre, lorsque chaque item distracteur ajouté dans le display amène à une augmentation plus forte (>20 ms/item) du temps de détection de la cible, la recherche est dite sérielle (Treisman & Gelade, 1980).

[42] Le traitement holistique correspond au fait que les informations faciales (traits locaux, distance entre les traits, forme volumétrique (3D) du visage,…) ne sont pas perçues indépendamment les unes des autres mais sont plutôt intégrées dans une représentation *holistique*, une *gestalt*. De ce fait, si un changement est opéré sur une de ces informations, la perception des autres traits sera également modifiée (voir Young et al., 1987; Tanaka & Farah, 1993; Farah, 1998).

[43] Dans leur étude, Heschler & Hochstein (2005) observent que la recherche d'un visage parmi des distracteurs (voitures ou maison) est plus rapide que la recherche d'une voiture parmi des visages ou des maisons, concluant à un avantage pour la détection de visage. Cependant, VanRullen (2006) note que cet effet d'asymétrie dans la recherche visuelle est probablement lié au fait que dans la tâche de recherche d'une voiture, la cible (voiture) est fortement similaire visuellement aux distracteurs (maisons et visages), alors que dans la tâche de recherche de visages, la cible (visage) diffère fortement des distracteurs (voitures et maisons). VanRullen montre que cette asymétrie disparaît lorsque la similarité entre cible et distracteurs est équivalente pour les visages et les voitures. Cependant, il trouve également que la recherche d'un visage parmi des distracteurs de tous types (voitures, autres moyens de transport, animaux et scènes naturelles) est beaucoup plus rapide que la recherche d'une voiture parmi des distracteurs de tous types (visages, autres moyens de transport, animaux et scènes naturelles). Ce résultat est en fait concordant avec l'idée que les visages possèdent des caractéristiques visuelles particulières qui ne sont majoritairement pas partagées par d'autres catégories d'objets. Il est donc probable que le système visuel ait appris à associer ces caractéristiques visuelles avec la catégorie des visages et que cette association soit utilisée précocement par le système visuel pour détecter un visage dans le champ visuel.

non significative (3 ms/item vs. 8 ms/item pour stimuli à l'endroit et stimuli à l'envers respectivement).

Plus généralement, ces résultats posent la question du type d'information utilisée pour détecter un visage dans l'environnement visuel. Nous avons vu que certaines caractéristiques de bas niveau comme le contraste, la couleur ou le spectre d'amplitude peuvent influencer la recherche visuelle de visages. Ces types de paramètres sont par ailleurs utilisés par les algorithmes de détection automatique des visages dans des photos ou des vidéos[44] (Hsu et al., 2002). L'observation récente que dans une tâche de détection de visages dans des scènes visuelles, les sujets sont capables d'initier une réponse saccadique vers le visage en un minimum de 100 ms (140 ms en moyenne) (Thorpe et al. 2007; Crouzet et al., 2007), suggère que la détection rapide de visages repose en partie sur la détection de paramètres de bas niveau fréquemment associés avec les visages (par exemple: une tache ovale rose/brune fortement contrastée et avec plus d'énergie dans les basses fréquences spatiales – voir aussi note de bas de page n°43). En effet, nous avons vu que les premières réponses sélectives aux visages dans IT chez le singe sont aux alentours de 80-90 ms. Ces latences étant probablement plus élevées chez l'homme (de l'ordre de 5/3 plus élevées, Shroeder et al., 1998; 2004; Foxe & Schoeder, 2005), il est improbable que les réponses saccadiques les plus rapides (100-140 ms) soient basées sur l'activité des neurones-visage dans IT.

Dans des conditions naturelles, où la visibilité des traits diagnostiques d'un visage peut être réduite (faible illumination, position ou vue particulière du visage, visage partiellement occulté, etc.) et la résolution des traits internes peut être faible (Loftus & Harley, 2005), la seule analyse de ces paramètres de bas niveau est probablement rarement suffisante à une détection faciale rapide et performante. Dans de telles situations, des représentations ou des processus de haut niveau (traitement holistique) peuvent probablement guider la détection d'un pattern facial. Par exemple, la perception des visages dont le contraste a été "binarisé" (visages "mooney"), n'est possible que par l'application d'une représentation holistique pour détecter le visage (George et al., 2005; Latinus & Taylor, 2005; Andrews & Schluppeck, 2004; Mckone, 2004). L'inversion de ces images mooney -empêchant ainsi le traitement holistique- provoque généralement la disparition du percept de visage. Une autre indication

[44]Les paramètres de couleur et contraste sont particulièrement utilisés pour une première approximation de la localisation du/des visage(s). En effet, les régions des yeux et de la bouche sont des régions de fort contraste. De même, la distribution des couleurs des visages occupant un espace réduit dans le spectre des couleurs, celle-ci est particulièrement diagnostique (Edwards et al., 2003; voir également Goffaux et al., 2005a pour des résultats similaires sur la catégorisation de scènes naturelles). En général, des étapes ultérieures d'appariement à des templates ou de décomposition en composantes principales sont réalisées pour affiner la détection.

que le traitement holistique est impliqué dans la détection faciale est qu'on observe généralement que des sujets mettent en moyenne entre 20 et 50 ms (selon les conditions de stimulation) de plus pour détecter un visage inversé dans une scène (inversée également - Lewis & Edmonds, 2003; Rousselet et al., 2003). De manière tout à fait intéressante, les temps de détection minimaux dans l'étude de Rousselet et al. (2003) sont peu affectés par l'inversion alors que les temps de réaction moyens et médians montrent un effet d'inversion. Une possibilité pour rendre compte de ce résultat serait que les temps de réaction les plus rapides sont liés à des stimuli où le visage est détecté sur base de propriétés de bas niveau et peuvent donc être détectés aussi rapidement à l'envers qu'à l'endroit. En revanche, lorsque le visage est moins saillant en termes de caractéristiques visuelles de bas niveau (donnant lieu à des temps de réaction plus élevés), le traitement holistique donne un avantage aux visages à l'endroit (en accélérant l'accumulation d'information). De plus, Rousselet et al. (2003) montrent également que l'effet d'inversion est réduit lorsque les sujets doivent détecter des animaux ou des visages d'animaux sur lesquels le traitement holistique n'est pas applicable. Une évidence supplémentaire pour l'importance du traitement holistique des visages à l'endroit provient de l'observation du fait que lorsque le temps de traitement visuel est fortement réduit par un masque visuel, les visages à l'endroit sont détectés plus fréquemment que les visages inversés (Purcell & Stewart, 1988). De la même manière, le temps de présentation nécessaire pour obtenir une performance de 75% en détection est inférieur pour un visage à l'endroit (~35 ms) par rapport à un visage inversé (~48 ms) ou avec les traits internes anormalement réarrangés (~54 ms) (Purcell & Stewart, 1988).

3.2.2. Reconnaissance et discrimination faciale

Il existe relativement peu d'études comportementales ayant étudié en détail le temps nécessaire pour reconnaître un visage. La plupart des études dans le domaine de la reconnaissance faciale s'attachent en effet à investiguer le type d'informations utilisées pour reconnaître un visage ainsi que les processus perceptifs/cognitifs mis en jeu. À cette fin, de nombreuses études mesurent l'effet de diverses manipulations apportées aux visages sur les performances et les temps de réaction en reconnaissance (mesurée dans des tâches d'appariement immédiat ou délayé, des tâches de choix forcés, des tâches de dénomination, etc). Ces manipulations impliquent notamment des changements d'orientation (Yin, 1969; Valentine & Bruce, 1988; Bruyer et al., 1993;), des modifications de traits faciaux ou de

distance entre ces traits (Leder & Bruce, 2002; Freire et al., 2000; Goffaux et al., 2005b), la présentation sélective de certaines bandes de fréquences spatiales (Näsanen, 1999; Goffaux et al., 2005b; Goffaux & Rossion, 2006; Liu et al., 2000), des modifications sélectives de la forme ou des propriétés de surfaces (texture, couleur, réflectance,…) des visages (Hill et al., 1997; Jeffery et al., 2006; Jiang et al., 2006; Lee & Perret, 2000; Russell et al., 2007).

D'une manière générale, les temps de réaction moyens obtenus chez des sujets humains dans des tâches impliquant un traitement de l'identité faciale oscillent entre 500 et 1000 ms selon la tâche (appariement, choix forcés, reconnaissance "old/new") et les images de visages (photographies vs. dessins, présence ou absence de traits externes comme les cheveux) utilisés (Valentin & Bruce, 1988; Rossion et al., 1999a; Itier & Taylor, 2002; 2004c; Michel et al., 2006; Schiltz et al., 2006). Notons que ces temps de réponse sont obtenus lorsque des visages *non familiers* sont présentés sous le *même point de vue* (généralement de face). Dans ces études, les réponses les plus rapides sont obtenues dans des tâches de discrimination d'identité (tâches de matching ou tâches de détection de répétition de visages: "1-back"). Lorsqu'un changement de point de vue est introduit entre deux visages à apparier, les performances se dégradent et les temps de réaction augmentent (chez l'humain: Hill et al., 1997; Pilz et al. 2006; chez le singe: Eifuku et al., 2004), et ce particulièrement quand les visages ne sont pas familiers (Jiang et al., 2006).

Les temps de réaction obtenus dans des tâches de *décision de familiarité* (qui impliquent nécessairement la reconnaissance de l'identité faciale) des visages célèbres, appris ou personnellement familiers, varient selon les études. En effet, alors que Anaki et al. (2007) rapportent des temps de réaction moyens d'environ 400 ms[45], d'autres études rapportent des valeurs proches de 500 ms (Caharel et al., 2005; 2006) ou encore supérieures à 700-800 ms (Young et al., 1985; Schweinberger et al., 2002a; b). En outre, les visages personnellement familiers sont reconnus plus rapidement que les visages célèbres (Caharel et al., 2005). Notons qu'une tâche de décision de familiarité est généralement plus complexe et demande plus de temps pour être réalisée qu'une tâche de discrimination car la décision de familiarité nécessite de récupérer une information en mémoire, ce qui n'est pas le cas pour une tâche de discrimination.

À notre connaissance il n'existe pas d'étude ayant investigué de manière précise à l'aide d'un recueil des temps de réponses, le temps nécessaire à discriminer ou à reconnaître

[45] Cependant, ces auteurs n'incluent pas les temps de réponse déviant de plus de 2.5 écarts-types par rapport à la moyenne dans leurs analyses. Cela a généralement pour conséquence de diminuer légèrement les valeurs obtenues pour les temps réponse moyens.

des visages sur base de leur identité. Les résultats décrits ci-dessus sont en effet obtenus sans imposer de contrainte temporelle forte aux sujets et aucune analyse précise de la distribution des temps de réaction n'a été réalisée. Comme nous l'avons vu dans la section précédente, l'analyse de la distribution des temps de réaction est particulièrement informative pour déterminer le temps minimal nécessaire à la réalisation d'une tâche donnée. Dans le cas du traitement de l'identité faciale, les données reprises ci-dessus nous indiquent qu'il faut en moyenne environ 500-600 ms (ou plus dans certaines études) pour reconnaître un visage ou discriminer deux visages mais sans indication précise du temps minimal requis. Dès lors, sur la simple base de ces temps de réaction moyens il nous est impossible de déterminer si les contraintes temporelles énoncées dans la section précédente d'après les travaux effectués chez le singe sont réalistes pour le sujet humain. Dans la partie empirique de cette thèse, en plus de l'analyse des potentiels évoqués, nous aurons l'occasion d'examiner les distributions des temps de réaction dans des tâches de discrimination d'identité, et ce de manière à apporter une information complémentaire aux travaux comportementaux réalisés jusqu'à présent dans la littérature.

Outre le receuil des temps de réaction moyens, le temps minimal de traitement perceptif nécessaire à la reconnaissance faciale a également été investigué au moyen des procédures de présentation masquée ou de présentation dans des séquences RSVP.

Les études limitant le temps de traitement visuel en présentant un masque constitué de bruit visuel montrent que les performances des sujets en reconnaissance ou discrimination sont fortement réduites lorsque l'intervalle de temps entre l'apparition du visage et l'apparition du masque (SOA) est extrêmement court (20 ms dans Rolls et al., 1994; 17, 33 et 50 ms dans Tanskanen et al., 2007; 53 ms dans Lehky, 2000). Ces études montrent également que les performances des sujets ne s'améliorent plus (ou très peu) lorsque le SOA entre le visage et le masque est supérieure à 80-100 ms.

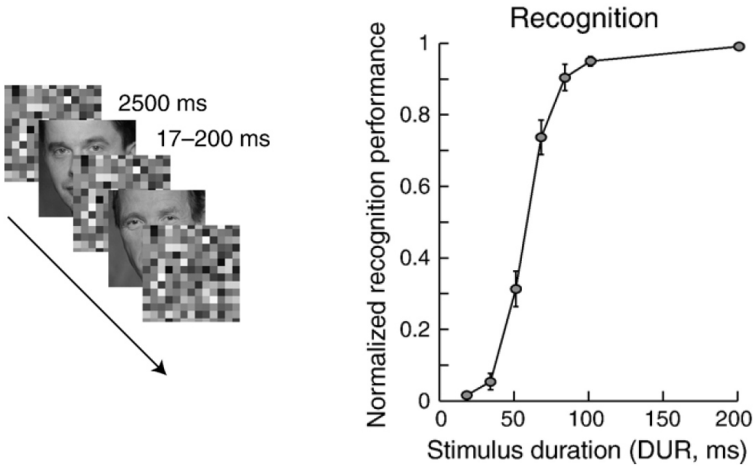

Figure 37: Performance (normalisée) à une tâche de reconnaissance à 6 choix de réponses possibles en fonction de la durée de présentation du visage (17, 33, 50, 67, 83, 100, 200 ms). Les visages sont présentés entre deux masques de bruit visuel (voir à droite). Modifié d'après Tanskanen et al., 2007.

Les études en RSVP indiquent un temps nécessaire au traitement de l'information faciale pour la reconnaissance relativement similaire aux résultats obtenus dans les études de masking. Par exemple, Näsänen et al. (2006) ont observé que pour que les sujets détectent un visage cible dans une séquence RSVP avec 80% de réponse correcte, la fréquence de présentation des visages ne devait pas excéder 10 par secondes (100 ms par visage). Il est cependant probable que le temps de présentation nécessaire pour obtenir des performances égales soit légèrement supérieur en RSVP qu'en masking avec du bruit visuel. En effet, alors que 100 ms de présentation sont nécessaires pour obtenir 80% de réponse correcte en RSVP (Näsänen et al., 2006; voir aussi Keysers et al., 2001; Mckeef et al., 2007a), moins de 40 ms de présentation sont nécessaires pour obtenir les mêmes performances dans l'étude de Rolls et al. (1994)[46] et environs 70 ms dans Tanskanen et al. (2007). De même, Lehky (2000) observe que lorsque le SOA n'est que de 53 ms, les sujets atteignent 80% de réponses correctes dans une tâche d'appariement (même vs. différent) lorsque les visages à apparier sont séparés de 17% sur un continuum de visages morphés (la différence maximale entre deux visages étant de 100%), alors que seulement 7% d'écart est nécessaire avec 100 ms de présentation. Le fait que la durée de présentation nécessaire à une performance égale en reconnaissance faciale soit plus importante en RSVP qu'en présentation masquée est probablement lié d'une part au design RSVP lui-même qui implique de traiter davantage d'information que dans un

[46]Il est cependant probable que les résultats de Rolls sous-estiment le temps de présentation nécessaire à la reconnaissance faciale. En effet, une partie des visages utilisés dans cette tâche (5 visages) pouvaient probablement être discriminée sur la seule base des cheveux. C'est d'autant plus probable que les performances des sujets en reconnaissance restent au-dessus du niveau du hasard dans la condition de SOA de 20 ms même lorsque les traits internes des visages sont réarrangés (les cheveux restant inchangés).

paradigme de masking, et d'autre part au fait que le RSVP est une forme de présentation masquée où les autres visages présents dans la séquence constituent en eux-mêmes des masques. Il a en effet été montré que l'efficacité d'un masque appliqué sur un visage dépend du type de masque utilisé[47] et plus particulièrement de l'information faciale présente dans le masque (Loffler et al., 2005; Bachmann et al. 2004). En utilisant plusieurs types de masques, Loffler et al. (2005) observent par exemple dans leur expérience qu'un masque de bruit visuel a très peu d'efficacité pour réduire la performance, alors qu'un visage à l'endroit (présenté à 20° de profil pour minimiser les effets de masking de bas niveau par rapport au visage cible présenté de face) est le masque le plus efficace; nettement plus efficace que le même visage présenté à l'envers. Dans cette expérience, l'influence du masque constitué d'un visage à l'endroit est maximale à des SOA de 27 et 80 ms et disparaît si le SOA dépasse 133 ms. Des résultats similaires ont été obtenus dans une étude de Lee & Perrett (2000) où les visages cibles étaient présentés pendant une durée variable entre deux autres visages qui servaient de masques (forward et backward masking), les sujets n'atteignaient environ 80% de réponses correctes que lorsque les visages cibles étaient présentés pendant 100 voire 133 ms.

Les données présentées ici indiquent que l'information sur l'identité d'un visage apparaissant à la fovéa[48] et sous un point de vue familier (-45° à 45°) est accumulée rapidement dans le système visuel, autorisant une reconnaissance des visages en moins d'une seconde. En particulier, les études de masking et de RSVP permettent d'imposer une contrainte temporelle forte sur le traitement visuel des visages. Ces études suggèrent ainsi qu'une accumulation d'information faciale pendant un maximum de 100 ms serait suffisante pour obtenir des bonnes performances en reconnaissance/discrimination de visage. Cette estimation est par ailleurs entièrement compatible avec les données obtenues en enregistrement de neurones dans IT chez le singe (Tovee et Rolls, 1994; Rolls et al., 1999;

[47] La question de l'efficacité d'un masque est directement liée à la question de savoir à quel niveau de la hiérarchie visuelle le traitement visuel du masque est supposé interférer avec le traitement du stimulus masqué, ainsi qu'à la question du mécanisme neurophysiologique par lequel le masque produit son effet (interruption, intégration, compétition – voir Keysers & Perrett, 2002; voir aussi Lamme & Roelfsema, 2000). Si l'on considère que l'un des mécanismes par lequel le masque peut interférer avec le traitement du stimulus est un mécanisme de compétition entre les représentations du stimulus et du masque (Keysers et Perrett, 2002), le masque entrera en compétition avec le stimulus de manière maximale si ces deux stimuli sont codés par des populations de neurones similaires (voir Desimone, 1998; Reynolds et al., 1999). Si le masking d'un visage est plus efficace quand le masque est un autre visage que quand c'est un pattern de bruit, on devra également postuler que le masque produit son effet au niveau des populations de neurones qui codent les visages (voir Loffler et al., 2005; Näsänen et al., 2006).

[48] La présentation de visages en périphérie visuelle révèle que les sujets sont plus rapides pour reconnaître un visage lorsque celui-ci est présenté dans l'hemichamp visuel gauche, indiquant une dominance de l'hémisphère droit dans la reconnaissance des visages (Young et al., 1985).

Rolls et al., 2006; Sugase et al., 1999; Matsumoto et al., 2005). Notons également, en regard des données exposées dans la section précédente, que les tâches nécessitant un accès à l'identité du visage sont réalisées plus lentement (temps de réaction moyen en détection: ~380-400 ms – Rousselet et al., 2003; en discrimination ou reconnaissance: >500 ms) et nécessitent des temps de présentation plus élevés (35-40 ms pour 75% de réponses correctes en détection – Purcell & Stewart, 1988 vs. ~70 ms en reconnaissance).

Il est cependant à noter que ces résultats sont obtenus dans les conditions idéales de laboratoire et qu'il est probable que la vitesse de reconnaissance soit fortement variable dans des conditions de vision naturelle. En effet, de nombreux facteurs comme la visibilité des traits diagnostiques du visage, les éléments de contexte, la présence d'éléments diagnostiques extérieurs au visage (corps, vêtements), vont moduler la vitesse à laquelle la reconnaissance faciale aura lieu[49].

Dans le chapitre suivant nous abordons l'abondante littérature sur la perception des visages ayant eu recours à l'enregistremment des réponses electrophysiologiques au niveau du scalp chez l'humain.

[49] Même si des éléments externes permettent d'inférer l'identité d'une personne, la reconnaissance faciale sera généralement la seule possibilité pour identifier sans faute cette personne.

Chapter 3

Face processing in humans as revealed by electrophysiology

In the empirical work of the present dissertation, we have attempted to precisely determine the time course of early face processing relying on an electrophysiological methodology. In this third chapter, we will therefore examin current evidence from the human electrophysiology literature about functional and temporal properties of face processing. Electrophysiological methods in human provide a main way to investigate the precise time course of brain processes involved in various perceptual or cognitive functions.

We will start by describing some of the main event-related potential (ERP) components elicited on the scalp after the presentation of a visual simulus. One of these component, which forms the so-called VPP/N170 complex, has generated a particular interest in the field of face processing, owing to its larger amplitude in response to faces than to other non-face categories. Because the ERP research on face processing has mostly focused on the N170 component, we will mainly concentrate our review of the ERP literature on the description of this ERP component. However, we will also advocate the view that focusing on specific ERP components to describe their functional role in visual processing, largely misses the fact that these components reflect the activation of multiple visual areas spread out on a large cortical surface, and probably involved in various functions.

Toward the end of the chapter, we will examin current evidence about the time course with which the human brain processes and represents information about the generic face category and about individual faces. We will tentatively indicate that these current evidence are mostly at odd with the biological constraints imposed on visual face processing as identified in the second Chapter.

The first electroencephalogram (EEG) was recorded by Hans Berger in 1929. By placing an electrode on the scalp of a human subject, he was able to observe modulations of electrical potentials as a function of time. Even though these observations were not taken seriously by the neurophysiologist community at that time, they were replicated some years later in independent laboratories (see Luck, 2005). The EEG is generated by the simultaneous activation of tens of millions of neurons located in various places on the cortical surface.

Analysis of the raw EEG is widely used to perform clinical diagnosis related mostly to sleep or various neurological disorders. However, this type of global raw EEG analysis is not sensitive enough to allow investigating precise neural responses such as those involved in visual processing. This analysis of event-related-potentials (ERP) from the EEG, following the averaging method developed by Dawson (1954), provides a way to isolate the neural responses specific to a particular event (sensory, motor, cognitive) precisely located in the time domain.

The reader can find a large number of information both on EEG recording and ERP analysis in the following books: Regan (1989), Guérit (1998) and Luck (2005).

The major strength of EEG recording is the ability to sample neural events with a very high temporal resolution (on the order of one millisecond). This high temporal resolution contrasts with the very limited precision with which the neural sources of scalp ERPs can be located in the brain. This limitation in spatial resolution is related to the distance between the recording electrodes and the neural sources generating the ERP, and to the resistance of the various materials through which the electrical signal is diffused (brain tissue, skull,…). Another limitation in the spatial resolution of ERPs come from the fact that ERP source localisation is based on inferences as to the number of neural sources, their location and the orientation of the electrical field (i.e. dipole) generated by these various sources. One possibility to overcome this limitation in spatial resolution while preserving a high temporal resolution is to record the magnetic field that accompanies any electrical field (Magnetoencephalography: MEG). Whereas the electrical fields are distorted by the resistance of the various materials they are passing through, associated magnetic fields are insensitive to these material properties. Thus, the recording of event-related magnetic fields (ERMF) allows avoiding the spatial blurring observed for electrical fields, and therefore achieving a much more precise brain localization of the events recorded at scalp level. Another way to improve the precision of ERP (or ERMF) source localization is to perform the same experiment in the same subjects with both EEG (or MEG) recording and functional neuroimaging (fMRI). The location of brain activation measured with fMRI may constrain the ERP source localization and thus may allow for a spatially more precise investigation of the dynamics of brain functions (Martinez et al., 1999; 2001; Di Russo et al., 2003).

Owing to its high temporal resolution, EEG is a complementary tool to functional neuroimaging techniques such as PET-scan or fMRI; methods that have a much higher spatial resolution (i.e. <1 mm; see Grill-Spector et al., 2006) but have a rather poor temporal resolution (generally of 1 to several seconds, although it can be lowered to a few hundreds of

ms in some circumstances, see e.g. Formisano & Goebel, 2003). Moreover, these neuroimaging techniques being based on the measure of hemodynamic properties related to neural activity, they provide only an indirect measure of brain activity.

Although the recording of single-cell responses in the brain of non-human animals allows to precisely investigate the functional properties of these neurons at a microscopic level, the invasive nature of these recordings makes it impossible to perform such recordings in humans, except in clinical circumstances (e.g. Allison et al., 1999;). Moreover, whereas the recording of single neurons or small groups of neurons provides information about brain functioning at a more local level (i.e. when recording from a cell, one dos not have a clear picture of what is happening somewhere else in the system), EEG recording allows apprehending brain functioning at a global scale.

More generally, ERP analysis may help to achieve a central goal in Cognitive Neuroscience which is to clarify the time course with which the processing of a stimulus evolves from the onset of this stimulus to the motor output in reaction to the stimulus. A major interest of EEG recording is to allow performing a direct and continuous measure of electrical brain activity at a high temporal resolution. Therefore, whereas behavioural measures do not provide direct information about the perceptual or cognitive processes taking place between the onset of the stimulus and the motor response, EEG recording allows a continuous monitoring of these processes. This continuous measurement of brain activity offers the opportunity to apprehend the precise time course of neuro-cognitive events inter-locked in time between stimulus and response onset. Moreover, ERP analysis provides a way to assess the time course of perceptual processes before decisional components or motor processes can take place, or even without the need to record a motor response. This method thus allows measuring perceptual processes without being influenced by decisional response bias that may strongly affect behavioural outputs.

1. EEG recording and event-related potentials

EEG is recorded by means of electrodes located on the scalp following a standard montage based on the 10-20 system[50] (Figure 38) or on its extension to electrode montage with a larger number of electrodes (10-10 or 10-5 systems). The EEG mostly reflects the

[50] The 10-20 system is a standard electrode montage according to which the midline prefrontal (FPZ) and the occipital (OZ) electrodes are positioned at a distance from the nazion/inion equal to 10% of the distance between the two landmarks (nazion – inion). The distance between the other electrodes is equal to 20% of the nazion-inion distance.

variations in post-synaptic potentials generated along the apical dendrites of cortical pyramidal cells oriented perpendicularly to the scalp surface. These post-synaptic potentials are accompanied by modulations of intra- or extracellular local field potentials. The local field potentials generated by the activation of multiple neurons sum up if the dendrites of these neurons are oriented in the same way. Thus, if post-synaptic potentials are generated simultaneously in a large group of such neurons, the resulting electrical field becomes strong enough to be picked up at scalp level. The ERPs recorded at scalp level can thus be considered as perturbations of the EEG triggered by and external or an internal event. It is also possible to record these electrical field with intra-cranial electrodes generally implanted in the brain of epileptic patients awaiting surgery.

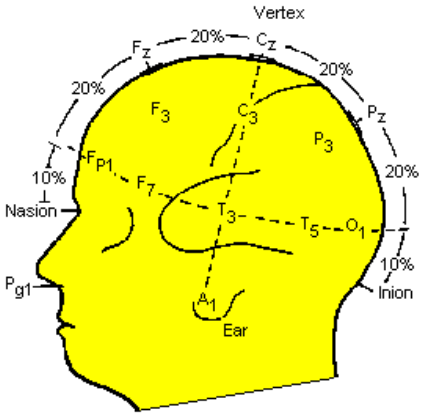

Figure 38: Schematic representation of the 10-20 EEG electrode system.

Given the weak amplitude of the ERP responses (generally around 5 to 20 µV) relative to EEG amplitude variations (100 to 200 µV), certain techniques are needed to isolate and reveal these ERP responses. The most classical method to isolate ERP responses relies on the averaging in the time domain of EEG segments time-locked to the studied event (e.g. a visual stimulation). This method relies on the idea that the ERP is a transient event which is both time- and phase-locked to the stimulation and which is superimposed on background EEG unrelated to the event. If a sufficiently large number of EEG segments are averaged, the random modulations of the background EEG will approach zero, therefore only revealing the EEG event related to the stimulation (Figure 39). The signal-to-noise ratio thus increases as a function of the number of EEG segments included in the averaging.

Following time-domain averaging, ERP waveforms are generally displayed as modulations of electrical potential (in µV) as a function of time. The ERP parameters the most frequently used as dependant variables are the amplitude and the latency of specific ERP

components about which hypothesis have been formulated. Other parameters that can be measured and characterized are the slope of an ERP component, the EEG variations in the frequency domain as a function of time, the precision with which the latency or the phase of an ERP component is synchronized to the stimulation from trial to trial, etc. The interpretation of these various parameters of sometimes very complex, given that an ERP component almost always reflects the sum of the activity of multiple neural sources which vary or not independently from each other.

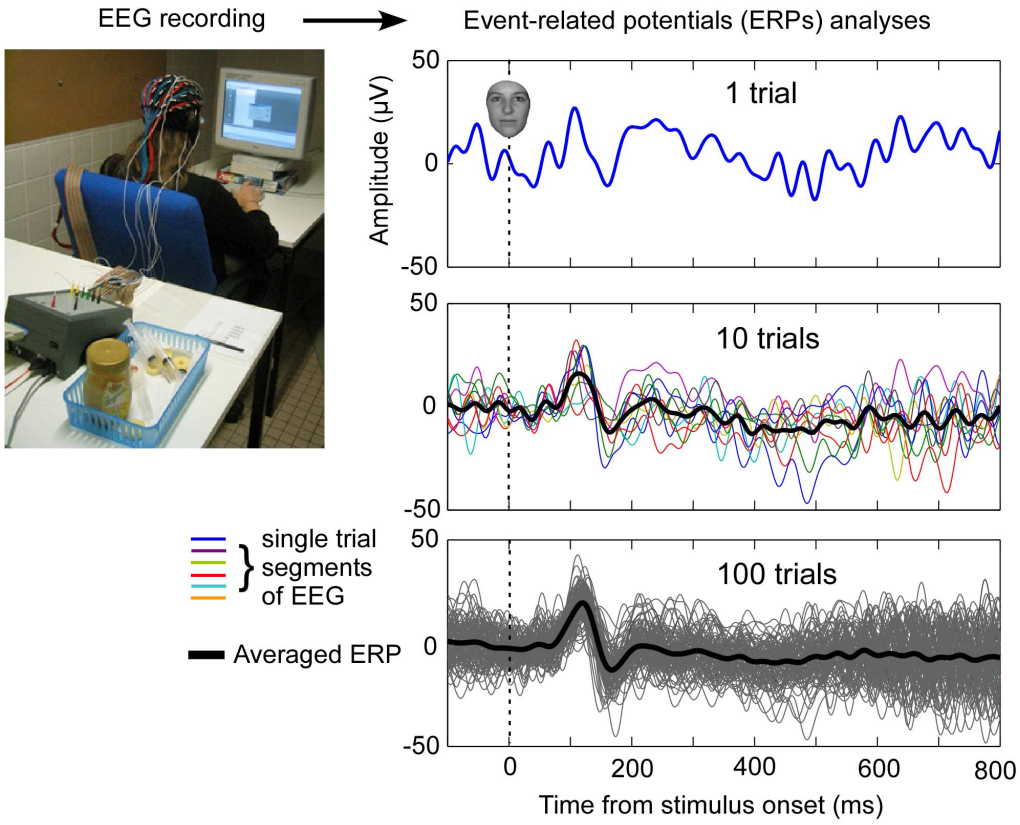

Figure 39: Illustration of the time-domain EEG averaging. After EEG recording, EEG segments time-locked to a stimulation are averaged in the time-domain to reveal the ERPs. This figure shows that increasing the number of averaged EEG segments (thin traces) will yield more stable and noise-free ERP components.

2. Description of the main visual event-related potentials

The onset of a visual stimulation elicits at scalp level a series of sequential waveforms (the ERP components) related to the activation of the visual cortex. The main visual ERPs are

named C1 (or N70), P1 and N1[51] (Figure 40). Each of these visual components is characterized by a different onset and peak latency and a different spatial distribution on the scalp. It is generally acknowledge that these components reflect the activation of visual areas from the primary visual cortex to the higher-level cortical visual areas in the ventral and dorsal stream. Note however, that the electrical activity of some visual brain regions cannot be easily recorded due to their location in the brain relative to the electrodes on the scalp (e.g. regions located on the medial ventral surface of the brain).

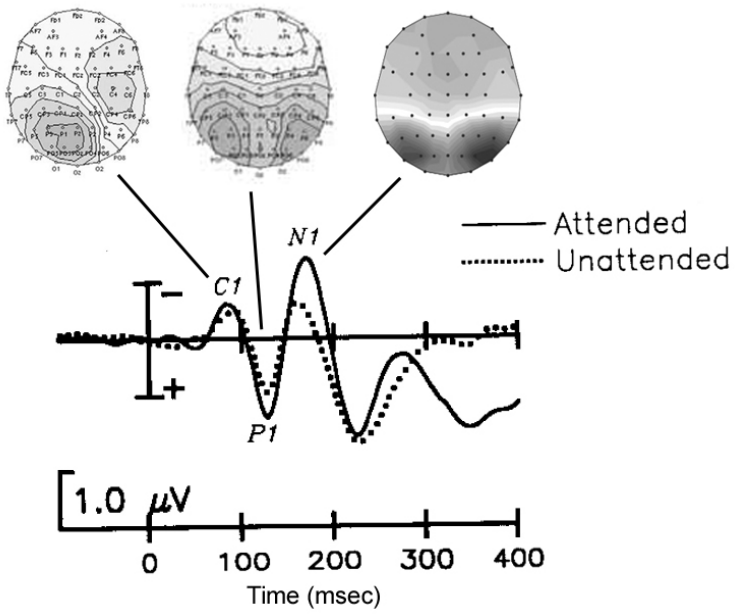

Figure 40: ERP components recorded as a function of whether the stimulus appears or not at an attended location. The main visual ERP components (C1, P1, N1) are depicted. An illustration of the scalp distribution of each component is provided. Note that the topography of the C1 changes as a function of the location of the stimulus in the visual field (see main text). ERP waveforms modified from Gomez-Gonzalez et al., 1994.

2.1 ERP components originating from striate and extrastriate cortex.

The C1 and P1 components are the first two ERP components measured following the onset of a visual stimulation. The properties of these components as function of the stimulation have been widely investigated since the 1970's (see Regan, 1989). Indeed, given their early latency, these components are considered as mainly exogenous ERPs, the study of

[51] ERPs are generally characterized by their positive (P) or negative (N) polarity, and either by their mean peak latency (e.g. the P100 component is a positivity maximal at around 100 ms after stimulus onset) or by their position order relative to other components of the same polarity (the P100 is also called P1 because it is the first positivity observed after the onset of a visual stimulation).

which provides information about functional and anatomical properties of the retinotopic striate (V1) and extrastiate visual cortices.

These components are generally not considered to reflect neural activity specific or particularly tuned to faces. Nevertheless, in the next paragraph we provide some information about these components that needs to be considered when studying the time course of visual face processing, as in the present thesis.

The C1 component (or N70) starts between 40 and 70 ms, depending on the stimulation conditions, and is maximal at around 80 ms (Jeffreys & Axford, 1972; Clark et al., 1995; Foxe & Simpson, 2002; Martinez et al., 1999; Di Russo et al., 2003). Its distribution on the scalp (topography) is maximal over medial parieto-occipital regions. Jeffreys & Axford (1972) first described this component as reflecting the activity of the primary visual cortex. Indeed, the polarity (positive or negative) of the component is reversed depending on the location of the stimulation in the upper or lower visual field. Jeffreys & Axford concluded that only the particular retinotopic organisation of V1 on either side of the calcarine sulcus could account for this polarity reversal of the C1. Hence, a stimulation in the upper visual field excites V1 neurons in the lower bank of the calcarine sulcus (the *negative* pole of the electrical dipole being oriented upward), while a stimulation in the lower visual field activates V1 neurons in the lower bank of the calcarine sulcus (the *positive* pole of the electrical dipole being oriented upward). Attempts to localize the sources of the C1 by dipole localization (Clark et al., 1995; Foxe & Simpson, 2002; Martinez et al., 1999; Di Russo et al., 2003) or combined analyses of EEG and fMRI data obtained in similar conditions (Martinez et al., 1999; 2001) have confirmed the origin of the C1 in the primary visual cortex. A recent study has however suggested that only the initial part of the C1 (from 50 to 60 ms) would reflect the activity of V1 alone; the late part (60 to 80 ms) reflecting the activation of both V1 and surrounding extrastriate areas (Foxe & Simpson, 2002; see also Vanni et al., 2004).

When the stimulation is at the centre of the visual field, the C1 is generally of very weak amplitude or has a positive polarity, being therefore superimposed with the subsequent positive P1 component. Therefore, the C1 is generally absent or is not analyzed in most ERP experiments investigating the processing of complex stimuli (objects or faces) when presented at the fovea.

The P1 is a component with a positive polarity that starts at around 80 ms after the onset of a visual stimulus and peaks between 90 and 120 ms. Its distribution on the scalp is

bilateral, more lateral than the C1, and with a maximum at the level of occipital electrodes in the left and right hemisphere. Numerous studies have attempted to localize the sources of the P1 in the visual cortex, either by localization of current dipole or by constraining dipolar localization by fMRI activation maps (Clark et al., 1995; Gomez-Gonzalez et al., 1994; Di Russo et al., 2002; Martinez et al., 1999; 2001). Whereas current dipole localization suggests that the P1 originates from extrastriate regions in the lateral occipital cortex (Clark et al., 1995; Clark & Hylliard, 1996; Gomez-Gonzalez et al., 1994), co-localization of EEG and fMRI suggest that the early part of the P1 originates from the lateral occipital cortex and the late part originates from ventral visual cortex in the posterior fusiform gyrus (Di Russo et al., 2002; Martinez et al., 1999; 2001). Note however that the sources of the P1 are much more complex to identify and localize than the sources of the C1 because most visual areas are activated within the first 100 ms following the onset of a visual stimulus. It is therefore likely that the activity from multiple ventral and dorsal stream areas, including V1, participate in generating the P1 and that the use of dipole localization methods are to constraining to reveal these multiple sources. Other localization methods based on distributed sources (e.g. LAURA: Michel et al., 2001, or LORETA, Pasqual-Marqui et al., 1994; see Michel et al., 2004 for a review on EEG source localization techniques) have also localized the P1 at the level of the lateral occipital cortex (e.g. Pourtois et al., 2004).

The P1 component has been largely investigated in the framework of studies on the time course with which spatial attention modulates early visual processing (e.g. Luck et al., 1994; Gomes-Gonzalez et al., 1994; Hillyard et al., 1998; Hopfinger & Mangun, 1998; Di Russo et al., 2003; Schuller & Rossion, 2005; Martinez et al., 1999; see Mangun, 1995; Luck et al., 2000 for reviews). The main finding of these studies is that the P1 elicited by lateralized simple stimuli is enhanced in amplitude when subjects are instructed to covertly attend to the stimulated location, or if central or peripheral cues predict the forthcoming stimulus location. While the earliest spatial attention effects on the P1 are observed starting at around 70-80 ms, the C1 component is not modulated by spatial attention, suggesting that the earliest inputs to V1 are immune to spatial attention effects (Gomes-Gonzalez et al., 1994; Martinez et al., 1999; 2001; Di Russo et al., 2003; Schuller & Rossion, 2005). It has been shown however, that the activity in V1 is modulated by spatial attention, but only after the initial activation and would be caused by feedback connections from extrastriate areas at around 150 ms (Martinez et al., 1999; 2001; Di Russo et al., 2003). These ERP spatial attention effects are generally associated with behavioural effect facilitating or suppressing the processing of a

visual stimulus appearing respectively and an attended vs. unattended location (Posner, 1980; Cheal & Lyon, 1991; Lu & Dosher, 1998; Lee et al., 1999; Pestilli & Carrasco, 2005).

The activity of V1 and other early extrastiate areas (e.g. V2, V3) has been widely investigated by means of measuring the C1 and P1 components elicited by simple visual patterns presented one-by-one, such as chess-like stimuli or gratings of various spatial frequencies, sizes, contrasts, colour or visual field position or even simple light flashes. Another stimulation method used mainly because it allows a very high EEG signal/noise ratio is the repeated presentation of these simple stimuli. In this case, the stimulus is either flashed several times per second (e.g. at 10 Hz), or the polarity of the stimulus contrast is reversed at the same frequency. David Regan is the first to have used such a rapid visual presentation technique (Regan, 1966) and coined the term "steady-state potentials" to designate the ERP generated in this type of stimulus presentation (see Regan, 1989). In general, the evoked potentials generated by such presentation strongly overlap, and specific components (like the C1 or P1) cannot be identified anymore. When the presentation frequency is high enough (e.g. > 6-7 Hz), the electrophysiological response takes the form of a sinusoidal waveform oscillating at the same frequency as the stimulation (Figure 41). The amplitude of the steady-state response is generally defined by decomposing the signal frequency spectrum with a Fourier analysis and measuring the signal amplitude of the second harmonic of the stimulation frequency (e.g. Di Russo & Spinelli, 1999; Di Russo et al., 2007; Regan, 1989; Hoffmann et al., 2003). Among one of the current use of the steady-state potential technique is the simultaneous presentation of tens of stimuli at various visual field locations. To be able to isolate the signal corresponding to each stimulus in the EEG, each stimulus is presented at a determined frequency not correlated with the frequency at which other stimuli are presented. This type of stimulation is called "multifocal visual evoked potentials" (see Hoffman et al;, 2003) and has mostly clinical applications such as determining the presence of visual field scotoma or detecting blindness in individual that cannot communicate (e.g. babies). Because it has an excellent signal-to-noise ratio, this technique is also used to maximize the reliability of very weak electrophysiological responses, such as when estimating visual acuity or visual detection threshold.

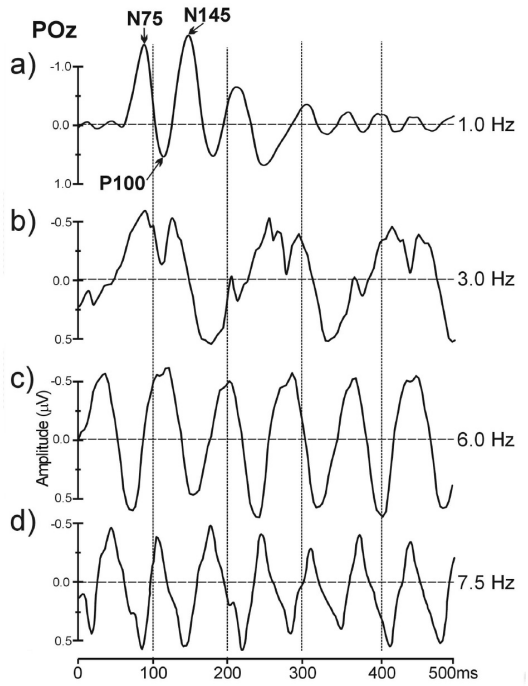

Figure 41: Visual evoked potential elicited by pattern reversal of a grating located in the upper left visual field. The figure shows the transition from transient VEP (upper row) to steady-state response when the stimulation frequency increases (from 1 to 7.5 Hz). Modified from Di Russo et al., 2007.

The amplitude and latency of the early visual ERPs C1 and P1 are strongly sensitive to manipulations of low level visual properties of the stimulation such as spatial frequency composition, contrast, luminance, size and visual field position. The sensitivity of early striate and extrastriate responses to these visual parameters has been mostly studied either by the steady-sate potential methodology or by measuring the ERP response to a contrast- or pattern-reversal of the stimulus (see Regan, 1989). These studies have for instance allowed determining some functional properties of the geniculo-cortical magno- and parvocellular pathways in humans or the sensitivity of striate and extrastriate regions to colour information.

Other studies in which single transient stimulation is used also indicate the sensitivity of the P1 (or the M1 recorded in MEG) to these low-level visual properties (Mihaylova et al. 1999; Musselwhite & Jeffreys, 1985; Tanskanen et al., 2005; Busch et al., 2004; Schneider et al., 2007; Halgren, 2000). These observations are in line with single-cell recording data in visual areas and subcortical structures (LGN, Superior colliculus). However, sensitivity to these parameters (contrast, luminance, size …) decreases as one progresses along the hierarchy of visual areas (Avidan & Behrmann, 2002). The sensitivity to these low-level parameters is one the major reason why these components are generally thought as reflecting low-level visual processes.

A number of studies have observed a larger P1 in response to faces than other complex objects (Eimer, 1998; 2000c; Goffaux et al., 2003; Itier & Taylor, 2004a; Liu et al., 2002; Herrmann et al., 2005a; b), suggesting that there are face-specific processes taking place at that latency. However, all these studies show a stronger face-sensitivity of the subsequent ERP component, the N170, discussed in the next section. We will come back later on to the issue of whether P1 responses also reflect or not face-specific processes, and we will argue that these P1 effects are largely due to low-level parameters differing between faces and objects rather than the faceness of the stimulus.

2.2 The N170 and the VPP

Even though a large number of evidence for a specificity of face visual processing had been accumulated for more than a century from the study of prosopagnosic patients (Bodamer, 1947; Quaglino & Borelli, 1867; Damasio et al., 1990), it is only at the end of the 1980's and beginning of 1990's that a series of studies have systematically described an ERP component that was particularly sensitive to images of faces. Jeffreys (Jeffreys, 1989; 1993; 1996; Jeffreys et al., 1992; Jeffreys & Tukmachi, 1992) and Grusser & Botzel (Bötzel & Grusser, 1989; Seeck & Grüsser, 1992; Bötzel et al., 1995) first observed a positive component maximal at the vertex (termed the VPP for "Vertex positive potential" by Jeffreys, 1989) with a peak latency varying between 140 and 200 ms, and whose amplitude is systematically larger in response to face stimuli (photographs or drawings) than to other non-face object categories (trees, chairs, shoes, vase, flowers,…) (Figure 42).

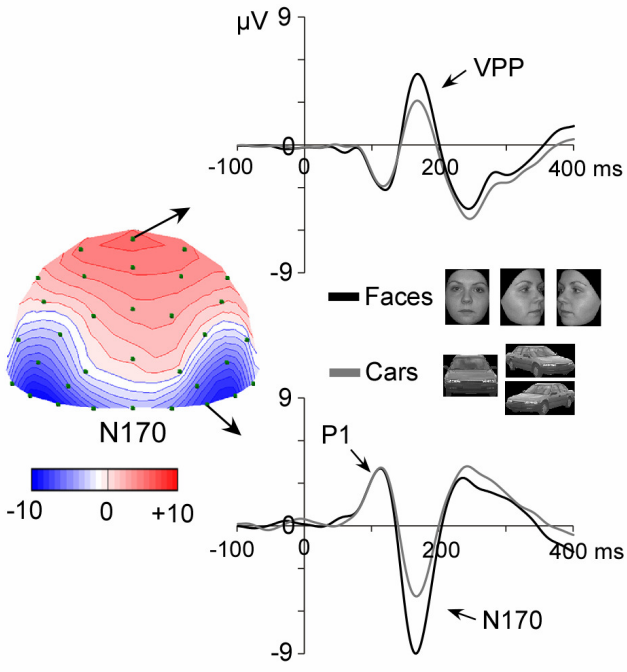

Figure 42: Examples of ERP responses elicited by faces and cars. The waveforms on the lower plot show the P1 and N170 components elicited by these stimuli. The N170 is maximal over occipito-temporal electrodes between 160 and 170 ms on average (see scalp topography on the left – backview of the head). The N170 is temporally coincident with a positive component observed at the vertex (VPP) which has similar response properties as the N170 (see Joyce & Rossion, 2005). Figure adapted from Rossion & Jacques (in press).

Jeffreys noted in his very first study on the N170 (Jeffreys, 1989) that when using a reference electrode placed on the nose rather than on the mastoids to analyse the ERP data, a component of negative polarity and with a similar latency as the VPP could be observed in some subjects over temporal electrodes. This author (Jeffreys, 1989; 1996) therefore suggested that the neural sources of the VPP were probably located on the ventral surface of the temporal cortex, or at the level of the STS. The origin of the VPP in the temporal regions was confirmed afterward by numerous studies (Botzel et al., 1995; Halgren et al., 2000; Itier and Taylor, 2002; Rossion et al., 1999b, 2003a; Schweinberger et al., 2002a; Watanabe et al., 2003). However, the use of a reference electrode located on the mastoids or on the earlobes as well as the small number of electrodes located over posterior temporal regions in the initial studies on the VPP has strongly limited the observation and the description of this posterior temporal negative component[52] (Botzel et al., 1995; see also Joyce & Rossion, 2005). Using a

[52] Given that the EEG is computed as the electrical potential difference between the electrical activity measured at each scalp electrode and the activity measured at a reference electrode, the ERPs measured at an electrode located close to the reference electrode are of weak amplitude. In the case of the posterior temporal negativity, the use of a mastoid or an ear lobe reference electrode strongly reduces this component.

larger number of electrodes widely distributed on the scalp, Botzel et al. (1995) noted that the amplitude difference between faces and objects at the VPP was largely accounted for by a negative component recorded over temporal regions which showed a larger face vs. object amplitude difference than the VPP.

Later, the use of an average or nose reference[53] to analyse the ERPs, and the availability or EEG recording systems with a larger number of electrodes favoured the investigation of this negative component maximal around 150-190 ms over bilateral occipito-temporal regions, and termed the N170 by Bentin et al. (1996). A large number of evidence point toward the possibility that the VPP and the N170 share part – if not all – of their neural sources in the occipito-temporal cortex. As such, these components may represent the two "end points" of the same dipole (Joyce & Rossion, 2005), these components being of opposite polarity and having a very similar latency. Moreover, in the studies describing both components in the same experiments, they show a similar sensitivity to the stimulus category (Rossion et al., 2003a; Itier & Taylor, 2004a) or to various stimulus manipulation such as vertical inversion (Rossion et al., 1999a; 2003; Itier & Taylor, 2004a; Eimer, 2000a), contrast inversion (Itier & Taylor, 2002), modifications in the position of inner facial features (Caharel et al., 2006; Macchi-Cassia et al., 2006) or the presentation of isolated facial features (Taylor et al., 2001; Jemel et al., 1999). Finally, the amplitude modulations of the two components as a function of the reference electrode (e.g. nose, mastoid, ear lobes, average reference) are strongly correlated (Joyce & Rossion, 2005). Note however that the view that the VPP and the N170 are functionally equivalent is not shared by all ERP researchers on face processing (Bentin et al., 1996; Bötzel et al., 1995; Eimer, 2000c; George et al., 1996; George et al., 2005). In addition, the poor spatial resolution of ERPs would not allow distinguishing between different neural sources for these components if the sources are located close to each other. It is therefore currently still difficult to determine the degree of commonality between the neural sources of the VPP and the N170. As a final note, one should be aware that the N170 and the VPP probably represent the summed activation of a large number of visual areas whose dipolar electrical activity is oriented in various directions and strength. It is therefore most likely that the summed dipolar activity in occipito-temporal regions (the N170) is slightly different than that picked up at the vertex (the VPP). Actually, even when considering only the occipito-temporal region, one can often observe that the experimental sensitivity and the latency of the N170 is slightly different from electrode to the other. In the

[53] When using an average reference the electrophysiological signal is computed by subtracting the signal averaged over all scalp electrodes to the signal measured at each electrode.

near future, the use of high density EEG recording systems (128 channels or more) which captures the N170 on more than 10 electrodes in each hemisphere, will probably call for a revision of "the N170" as a unitary functional phenomenon. Indeed, this view of ERP components as discreet functional processing stage is missing the fact that these components (such as the N170) reflects the activation of multiple visual areas which are spread out on a large cortical surface and are activated at various latencies.

Given the uncertainty of the correspondence between the N170 and the VPP, a fair amount of empirical work on the VPP performed before the N170 labelling by Bentin et al. (1996) has been largely ignored. This is rather unfortunate given that Jeffreys (Jeffreys, 1996) remarkably summarized and integrated the observations on the VPP with the known properties of face-cells in the monkey, the data from intra-cranial recordings in epileptic patients[54] and behavioural data on face perception. Even though one must remain careful when comparing these two components, it seems that the empirical work on the VPP could readily be used to interpret N170 data in relation to face perception or to generate new hypothesis. Finally, we note that a large number of studies, including Bentin et al. (1996) and subsequent studies, are by and large replications of previous observations gathered on the VPP.

3. Electrophysiological studies of face visual processing

Since these early work on the VPP and N170[55], more than a hundred (maybe several hundreds) studies have used ERPs to investigate different functional properties of face processing[56]. When attempting to organize this literature, one can derive a number of general topics of research in face processing according to which most of them can be classified. What are these topics?

[54] In the 1990's a very large number of intracranial recordings in epileptic patients have been performed over almost the entire surface of the human dorsal and ventral temporal lobe. These recordings have revealed that a large number of cortical "patches" throughout the temporal cortex (especially over superior temporal sulcus, medial temoral gyrus and fusiform gyrus) generate a local field potential (similar to ERPs bu at a smaller scale) of negative polarity around 200 ms (N200), that is much larger for face stimuli than for all kind of different objects or shapes (Alisson et al., 1994; 1999; 2002; Puce et al., 1999).

[55] Before the late 1980's, several studies had been conducted on face processing (e.g. Barrett et al., 1988; Barrett & Rugg, 1989), but these studies either only recorded ERP to faces, or recorded ERPs to both faces and non-face objects but did not describe any difference on early ERPs (Neville et al., 1982; Small, 1983).

[56] The paper from Bentin et al., (1996), cited by most ERP studies on face processing, has currently more than 400 citations.

1. Is the larger N170 to faces reflecting domain specific mechanisms (i.e. a face module, dedicated to process faces) or domain general mechanisms (i.e. that can be recruited for any object class the needs to be categorized at the individual level)? These studies include experiments comparing the ERP response to faces vs. non-face objects, attempting to determine what causes the largest N170 to faces. Other experiments measure the effect of visual expertise with a non-face object category (e.g. cars, birds, dogs) on the ERP response to this category compared to non-expert categories.

2. What is the nature of the face representation that subtends face perception during the time-window of the N170? Most of the studies in this "theme" have focused around determining the relative contribution of information about individual facial features (mostly the eyes) vs. facial spatial configuration in triggering face specific processes. One of the ensuing question is whether features or global face configuration is accessed and processed first.

3. What is the time course of different facial categorization processes (e.g. identity, expression, gender processing …)?

4. Can early face processing (again as indexed by the N170) be modulated by top-down influences, such as visual attention, semantic or contextual priming, visual experience, subject's expectations, or task demands?

5. What can we learn from ERPs about the development of face processing from birth to adulthood? How does face processing develops from birth to adulthood, and how is this reflected in the ERP response?

6. What "functions" in face processing are impaired in populations with mental or developmental disorders[57] (e.g. children with autism or Asperger syndrome, people with schizophrenia, social phobia…) or in people with specific developmental or acquired impairments in face processing (i.e. prosopagnosia)?

In the next parts of the text we will review and discuss the findings in some of these research topics that we believe are important to comprehend the experimental part of this thesis. More specifically, we will concentrate on the first three topics, focusing mostly on the studies involving characterization of face processing at a perceptual level, thus mainly reflected on the N170 component.

[57] ERPs are particularly interesting here because face processing can be investigated without the need to record a behavioural response from the subject.

3.1. The N170 as an index of face processing specificity

We have seen that early studies have found that the VPP component is larger in amplitude in response to faces than to other non-face objects such as trees, chairs, shoes, vase, flowers, airplane,… (Jeffreys, 1989; Jeffreys, 1992; 1993; 1996; Jeffreys & Tukmachi, 1992; Botzel & Grusser, 1989; Seek & Grusser, 1992; Botzel et al., 1995). Numerous studies conducted since then and focussing on the N170 have confirmed the larger amplitude of the response to pictures of faces than to pictures of a wide array of other complex non-face objects (Bentin et al., 1996; Carmel and Bentin, 2002; Eimer, 1998; 2000a; b; c; d; Goffaux et al., 2003; Itier et al., 2006a; Itier & Taylor, 2004a; Rossion et al., 2000; Rossion et al., 2004a; Rousselet et al., 2005; 2007a; Zion-Golumbic & Bentin, 2007; Boutsen et al., 2006; Rebai et al., 2001). Similar observations have been described on a magnetoencephalographic component termed the "M170" (Halgren et al., 2000; Swithenby, 1998; Liu et al., 2000), although its relation to the N170 is still debated[58] (Taylor et al., 2001; Itier et al., 2006b; 2007; Deffke et al., 2007). The N170 difference between faces and object is generally larger in the right hemisphere (Bentin et al., 1996; Itier & Taylor, 2004a; Rossion et al., 2003a).

As noted above, some studies found that faces elicit a larger P1 component (i.e. before the N170). However this effect is not always observed and the N170 appears to be the first ERP response that systematically distinguishes faces from non-face object categories. This amplitude difference between faces and objects usually starts at about 130 ms after stimulus onset, although the precise onset time has rarely been measured (see Rousselet et al., 2005 for the N170; Seek & Grusser, 1992 for the VPP) or even discussed. We will discuss the issue of the time course of face preferential responses more thoroughly later in this thesis.

Why is the N170 larger for faces and what is the functional meaning of this larger amplitude? We will try to briefly address these questions in the next section.

[58] Since MEG records magnetic fields which are oriented *orthogonally* compared to electrical fields to which EEG recording is sensitive, in some cases the activity from some brain areas can be apprehended with MEG recording and only poorly with EEG recording and vice versa. In the case of the N170/M170, one proposal is that the M170 (from MEG) primarily arise from the fusiform gyrus (Itier et al., 2006b; Liu et al., 2002), but that the position of this region makes it difficult to record its activity with EEG. This is however a still ongoing debate (Deffke et al., 2007; Watanabe et al., 20003).

3.1.1. Underlying neural sources of the larger N170 to faces

Interpreting a differential amplitude of a scalp ERP component between two conditions is not straightforward (Luck, 2005) because it can be related to many underlying neural events. This is especially true for ERP components that reflect the resultant activity of many neural sources such as the N170.

There are mostly two ways to account for the larger N170 to faces than objects as revealed by time averaging of EEG segments time-locked to stimulus onset (Figure 43).

According to the **first**, traditional view, ERP components originates from a massive synchronized increase of post-synaptic neural activity time-locked and phase-locked to stimulation onset, superimposed onto background electrophysiological activity unrelated to the stimulation (the signal + noise model; Figure 43; left part). In this framework, the larger N170 to faces could simply reflect a stronger increase of neural activity as compared to objects, in visual areas from which this component originates. This would be reflected as a stronger increase of amplitude at a fixed latency and polarity at the scalp level for face stimuli (Figure 44). The possible sources of this stronger increase are discussed below. Moreover, even though this phasic increase of neural activity generating an ERP is largely time-locked to stimulus onset, there is some degree of variation in the onset and peak latency of a given ERP component from trial to trial ("time jitter") (see Figure 45). Given the nature of time averaging, the amplitude of a component will be negatively correlated to the amount of jitter from trial to trial (i.e. the larger the jitter, the more "smeared" and small the ERP will appear, see Figure 45). Therefore, part of the larger N170 amplitude for faces after ERP averaging could also be accounted for by a *lower* variance in the peak latency (i.e. a smaller jitter) of the N170 from trial to trial for faces compared to non-face objects.

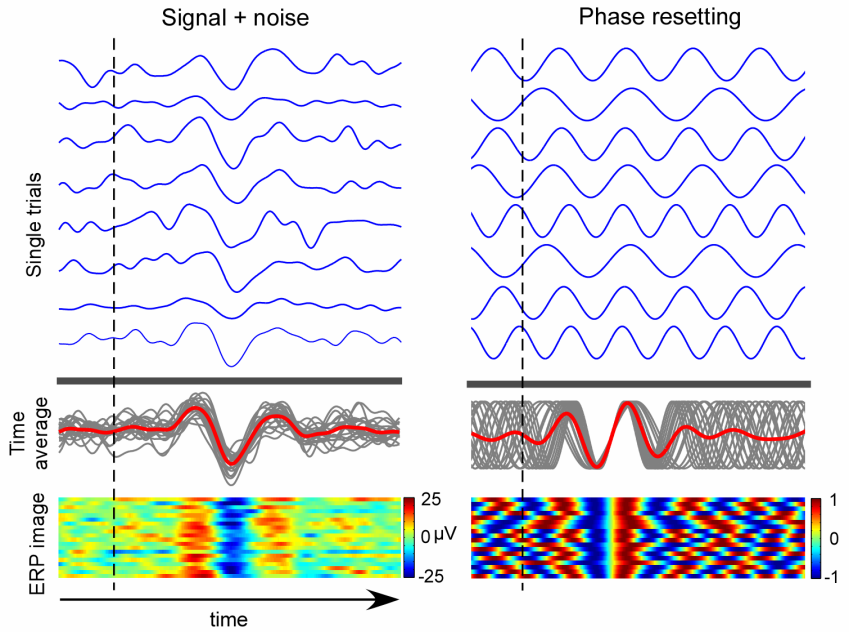

Figure 43: Two models for the generation of ERP responses in time averages. Left: the traditional model proposes that ERPs correspond to a phasic increase of neural activity at a fixed latency and polarity that is superimposed on background EEG noise. The averaging of single trials (upper row) allows revealing of the ERP by increasing signal-to-noise ratio (middle-row). Middle row: Thin traces are individual trials and thick trace is an average of 20 trials. The ERP is then considered as an increase of amplitude at a fixed latency (see bottom row). Bottom row: ERP image of 20 single trials. The ERP image is a 3D representation of EEG single trials with *x* = time, *y* = trials and *z* = color-coded EEG amplitude. Right: simplified illustration of the phase-resetting model. This model posits that the phase across single trials in the EEG is randomly distributed before stimulus onset. Then, the onset of the stimulus (vertical dashed line) biases the phase of the EGG so that it becomes similar across all trials. In the phase resetting model, time averaging reveals an ERP response which is not associated with an increase of amplitude in the EEG (see middle and bottom rows).

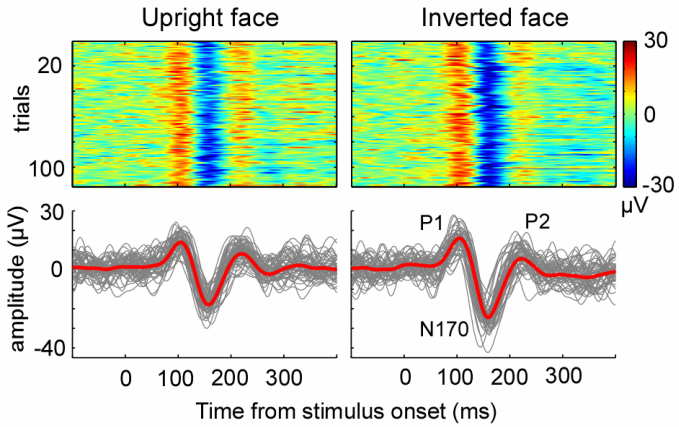

Figure 44: This figure shows that the N170 response is best described as a strong increase of power time-locked and phase-locked to stimulus onset. <u>Upper row</u> depicts ERP images for 115 EEG single trials time-locked to the onset of a stimulus. Note the strong increase of amplitude during P1, N170 and P2 components relative to the 100 ms before stimulus onset. Bottom row shows an average of the trials depicted in each ERP image. The figure also illustrates that difference between type of stimuli (here upright and inverted faces; left and right column respectively) at the N170 is also accounted for by a change in amplitude in each individual trial. This is also true for the N170 face effect (Rousselet et al., 2007a). <u>Bottom row</u> shows individual trials (thin grey traces) and an average of all trials (thick red trace). This also shows the increase of amplitude during the time windows of the P1, N170 and P2 components.

According to a **second** view, ERPs would be generated by a phase resetting of the ongoing EEG oscillations (i.e. that precede the onset of the stimulus; see Figure 43 – right part) due to the onset of a stimulus. That is, stimulus onset would trigger a realignment ("resetting") of the phase of the EEG which otherwise would be out of phase from trial to trial. This realignment of the EEG ongoing oscillations due to the stimulation would occur without any substantial increase of amplitude, leading to a visible (time-locked) ERP component only after averaging in the time domain (Sayers and Beagley, 1974). The larger N170 to faces could then be due to faces yielding a more precise temporal alignment of the EEG phase from trial to trial as compared to non-face objects. The precision of the phase-alignement across time points from trial to trial can be quantified by computing the "phase-locking factor"[59] (Tallon-Baudry et al., 1996; Makeig et al., 2002). It ranges from 1 (EEG at a given time point perfectly in phase from trial to trial) to 0 (No relationship between the phase of the EEG from trial to trial). Interestingly, it has been proposed that the visual N1 component to simple stimuli can indeed be largely generated by such a phase resetting of EEG ongoing oscillations in the alpha range (Makeig et al., 2002).

The issue of which of the two model for ERP generation best explains face vs. objects N170 amplitude difference has functional implications because the traditional model (signal + noise) directly implies an *increase* in the summed neuronal activation (see below), whereas according to the phase-resetting model (Klimesch et al., 2004; Makeig et al., 2002), the larger N170 to faces is simply due to single trials for faces being better aligned temporally at 130-200 ms[60].

Tackling this issue, it was recently found that the pure phase-resetting account for a higher N170 to faces over objects appears to be incorrect: this effect can be accounted for completely by a modulation of ERP amplitude centred in the 5 to 15 Hz EEG frequency band (Rousselet et al., 2007a), in line with the traditional view of the ERP generation. Hence, the larger N170 to faces than objects appears to be due to a time- and phase-locked increase of power, compatible with a fixed increased of neural activity in localized brain areas.

[59] The phase-locking factor can also be used to quantify the amount of jitter in the ERP latency as defined by the signal + noise model.

[60] The reason why face would yield a better alignment compared to object is unknown, but one can speculate that it could be due for instance to the fact that the face category has a higher degree of visual similarity between its exemplars, to the fact that face would activate a more homogeneous neural population, or to the fact that faces are highly salient and behaviourally relevant stimuli that may be processed in a more automatic and precisely timed fashion.

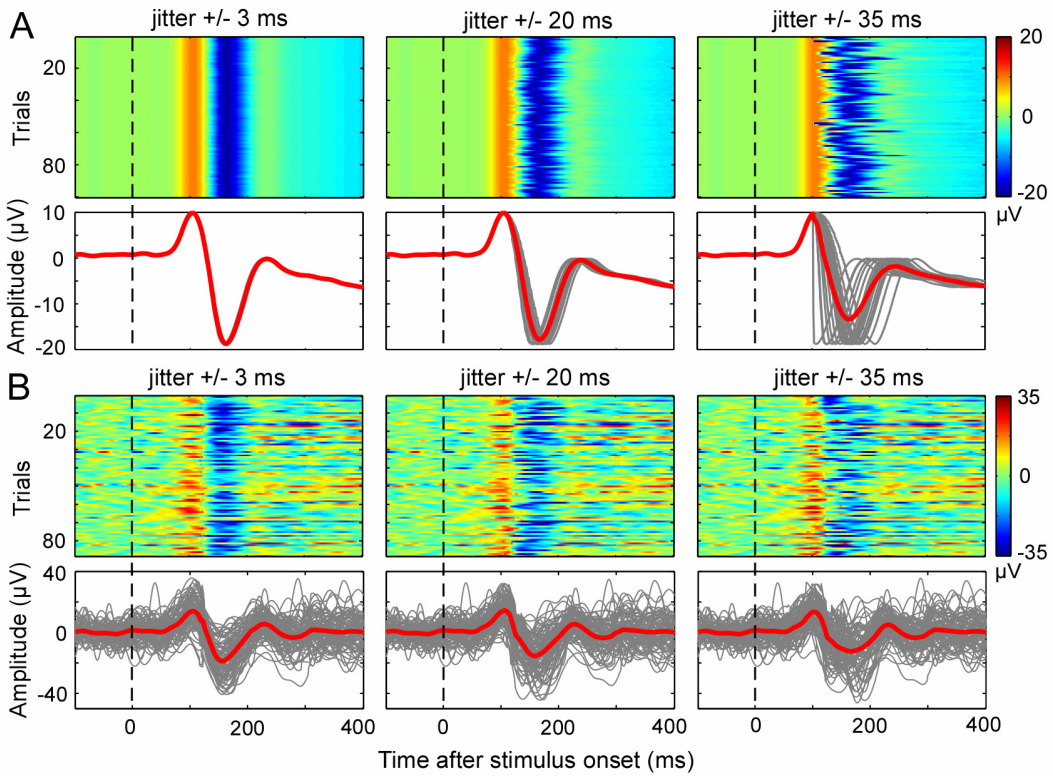

Figure 45: Effect of inter-trial time jitter in the peak latency of the N170 on its amplitude. A. Simulation of jitter in the slope between the P1 peak and the N170 peak, based on a grand averaged ERP response to face stimuli. The jitter was generated by randomly stretching/compressing the ERP from 100 to 160 ms after stimulus onset in the range of +/- 3 ms (left), +/- 20 ms (middle) and +/- 35 ms (right). Upper row shows ERP images (number of trials generated = 100) of the jitter simulation. Lower row shows a subset of individual trials (thin gray traces) as well as an average of 100 simulated trials (thick red trace). B. Simulation of inter-trial jitter using real EEG data. Note in both simulations the reduction of amplitude and the smearing/widening of the N170 as time-jitter increases. Modified from Rossion & Jacques, in press.

If the larger N170 for faces is accounted by an increase of power at a fixed latency in the 130-200 ms time window, it remains to be determined what are the underlying neural sources causing this increased amplitude. What may these sources be?

First, we need to determine whether this larger amplitude is due to the recruitment of additional brain sources for faces or whether the same sources are recruited for faces and objects but with a higher intensity for faces. Different attempts to localize the neural generators of the N170 to faces (either by dipole localization or distributed sources models (e.g. LORETA – Pasqual-Marqui et al;, 1994 or LAURA- Michel et al., 2001) have suggested that multiple sources are simultaneously active during the 130-200 ms time window, including the fusiform gyrus and lateral occipito-temporal cortex (Rossion et al., 2003a; Schweinberger et al., 2002a; Shibata et al., 2002; Herrmann et al., 2005a; Deffke et al., 2007),

as well as the STS region (Watanabe et al., 2003; Itier & Taylor, 2004b). The contribution of a source in the fusiform gyrus has also been found for the M170 in MEG (Halgren et al., 2000; Deffke et al., 2007; Itier et al., 2006b). The locations of small face sensitive cortical sites where N200s are observed in intracranial recoding are largely compatible with these sources (Allison et al., 1999; see Figure 46).

The comparison of faces and non-face objects generally suggest that at least some of the sources generating the N170 are different between these categories. For instance, the dipole solutions that best explained the N170 for faces, cars and words (Rossion et al., 2003a) were found at similar locations in the posterior fusiform gyrus, but differed in strength, orientation and hemispheric lateralization, suggesting the involvement of slightly different neural sources for each category. Moreover, using the LAURA algorithm to reveal the sources of the N170 to faces and several other categories (e.g. flowers, houses, tools, lions, …), it was found that while a region in the lateral occipital cortex is similarly active for all categories, upright and inverted faces recruited an additional region in posterior STS[61]. Additional evidence for slightly different sources (i.e. additional) for faces than non-faces, come from spatio-temporal cluster analyses of the scalp topographies (i.e. "segmentation", Pasqual-Marqui et al., 1995), showing that the N170 for faces is associated with an extra activation (i.e. a different scalp topography) compared to objects (Itier et al., 2004a; Rousselet et al., 2004c).

[61] Note however, that in this experiment, non-face categories elicited an N170 of particularly small amplitude compared to faces, probably due to differences in the stimulus properties between faces (upright and inverted) and other categories. Indeed, while all face stimuli were presented in a front view on a black background, other stimulus categories were of different size and views and most were embedded in a background visual scene. As noted by the authors, the source intensity for non-face stimuli were close to noise-level, rendering the localisation less reliable.

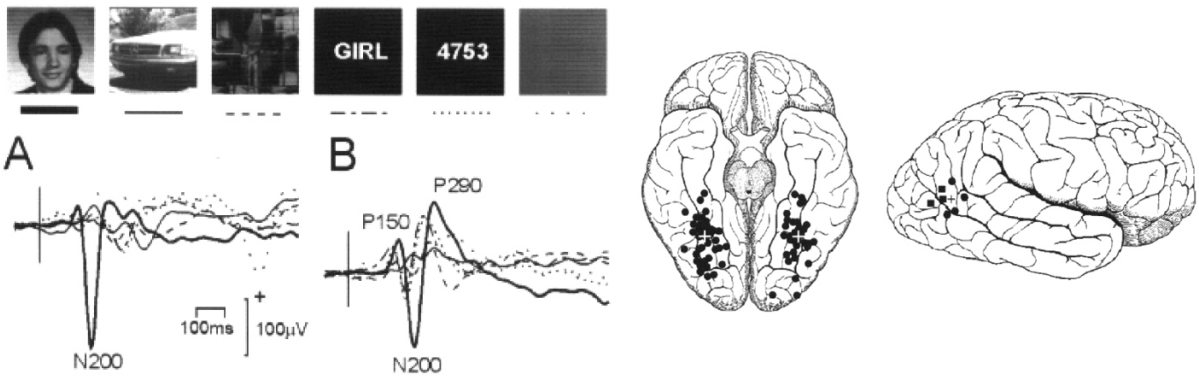

Figure 46: Right (A and B). Examples of intra-cranial N200 potentials recorded on the cortical surface at ventral sites (see ventral view of the brain). The ventral view in the middle shows the location where face-specific N200 components have been found. On the right, location of face specific N200 on the lateral surface of the brain arounf the medial temporal gyrus. The relationship between N200 and the N170 recorded on the scalp is currently unknown. Modified from Allison et al., 1999.

Next, turning to neuroimaging data may bring further understanding to the question at hand. As introduced in section 3 of Chapter 1, several areas in the occipito-temporal cortex have been identified that respond preferentially to faces, in the posterior part of the superior temporal sulcus (pSTS), in the middle fusiform gyrus ('fusiform face area', FFA) and in the inferior occipital cortex ('occipital face area', OFA) (Kanwisher et al., 1997; Gauthier e al., 2000; Puce et al., 1998; see Haxby et al., 2000 for a review). As for objects, to our knowledge, the only region that responds in more to pictures of objects than faces is located in the parahippocampal gyrus. Other visual areas in high level cortex, such as LOC (See Grill-spector et al., 2001 for a review), respond similarly to faces and objects. Compatible with these neuroimaging data, intracranial recording found that whereas most site where N200 responses to objects can be observed also respond to faces, the opposite is not true. That is, almost no response to objects is observed at face specific sites (Allison et al., 1999).

Altogether, evidence from source localisation of the N170 to faces and objects, neuroimaging and intracranial recordings, suggest that multiple cortical sources whose timing of activation overlap, interact and add up to contribute to the N170 recorded over the occipito-temporal region of the scalp. Moreover, the larger N170 to faces compared to non-face is probably, at least party, due to the recruitment of additional brain regions responding preferentially to faces.

Other, additional, possibilities for a larger N170 to faces should be considered. First, it is possible that the sources of the N170 to faces are located in the head in such a way as to be more easily picked-up by EEG scalp electrodes. As mentioned above, one of the regions potentially recruited more for objects than faces is located very deep and medially in the

parahyppocampal gyrus, rendering its activity difficult to measure at scalp level with EEG. On the contrary, the STS region that most probably contribute a great extend to the face N170 is located much upper and closer to the scalp than the parahippocampal region. Second, we have seen that faces are represented in the cortex by single-neurons that respond to the object "face" clustered in columns (Perrett et al.; 1984; Harries & Perrett, 1991; Tanaka et al., 1991; Wang et al., 1996), while most non-face objects are represented by combinations of cortical columns that code for simpler features than a whole object (Tanaka et al., 1991; 1996; Wang et al., 1996; Tsunoda et al., 2001). It may thus well be that responses of face-cells in the inferior-temporal cortex are more synchronous than responses of cells coding for simpler features. Given that the amplitude of ERPs is dependent on the temporal summation of post-synaptic potentials, the putatively more synchronous neural responses to faces may yield a larger ERP component at scalp level, independently of the respective number of face vs. "non-face" neurons.

3.1.2. Is the larger N170 to faces due to differences in the low-level visual parameters between faces and non-faces?

An important question to address when comparing the ERPs generated by different stimulus categories is whether these differences really reflects a difference in higher level object categorization processes, or if they can rather be accounted for by differences in the low-level visual properties (e.g. luminance, local and global contrast, spatial frequency amplitude spectrum,…) between these categories. This distinction is crucial to address because if one postulates that some neural activity reflects a *categorization* process, just as the larger N170 is supposedly related to face *categorization*, this activity should be mostly independent of modifications made to these lower-level properties, rather reflecting a processing of the category, and not just its physical properties.

In most ERP studies comparing faces to other categories, low-level visual properties of each category, known to influence the amplitude of early visual potentials (e.g. C1, P1 and N1; see Regan, 1989), are usually not tightly controlled or matched. Given that the low level properties between exemplars of a given category are generally highly correlated (e.g. most faces are highly contrasted and share a common spatial frequency amplitude spectrum), the fact that only a finite number of categories can be tested in a single experiment makes it

difficult to be certain that the N170 is indeed sensitive to the face category and not to some other more low-level feature that just happen to be present in faces and not in other objects.

In some studies however, in which the low-level properties have been controlled as much as possible between faces and the control object category compared (e.g. houses in Rousselet et al., 2005; 2007a), the N170 was still much larger in response to faces, strongly suggesting that low-level parameters do not account for the N170 face effect. Moreover, the larger N170 for faces over occipito-temporal regions is largely independent of the format of face presentation, supporting the view that the larger N170 for faces represents neural processes specifically involved in face processing and is not due to differences in low-level properties between faces and non-faces. Specifically, the larger amplitude is found for schematic or sketch faces, photographs or paintings or faces (Sagiv & Bentin, 2001; Latinus et al., 2006) (see Figure 47), Mooney figures perceived as faces (George et al. 2005, Latinus & Taylor, 2005; see Figure 49) or even animal or great ape faces[62] (Carmel & Bentin, 2002; de Haan et al., 2002; Itier et al., 2006a; Rousselet et al., 2004a).

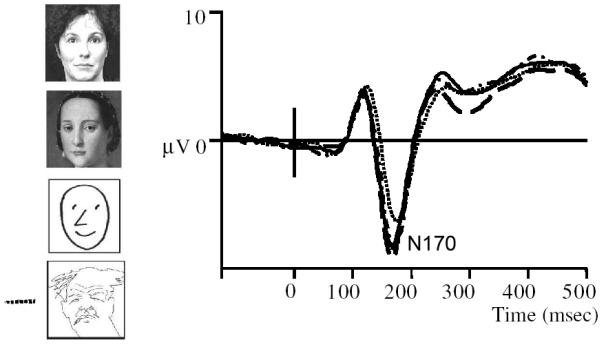

Figure 47: The N170 is of equal amplitude for face photographs or paintings and schematic faces. Scketches of faces elicit a slightly smaller N170 respons. Note that the original figure does not contain a legend to identify which waveform corresponds to which stimulus. Modified from Sagiv & Bentin, 2001.

A recent ERP study has challenged the view that the N170 is indeed larger for faces than for non-face categories, claiming that its larger amplitude to faces is merely due to an uncontrolled methodological artifact in previous studies (Thierry et al., 2007). More precisely, according to the authors of this study, previous ERP studies observed a larger N170 in response to faces simply because the pictures of faces used are much more similar to each

[62] This indicates that the overall N170 sensitivity to face is not species-specific. However, human face sensitivity of the N170 is revealed through other stimulus manipulation such as picture-plane vertical inversion which strongly increases the N170 amplitude to inverted human face while either only slightly increasing (Rousselet et al., 2004a), strongly decreasing (Itier et al., 2006a), or not modulating at all (dehaan et al., 2002) the N170 to non-human faces.

other as compared to object pictures (e.g. all faces in the sets of pictures would be presented in the same view-point and size while pictures of objects would vary in view-point, size and eccentricity – Figure 48). Following the author's reasoning, this lack of control would have caused a larger inter-trial jitter in the peak latency of the N170 for the non-face object conditions (see above), and thus a reduction of the averaged N170, as compared to the face condition (Figure 45). Thierry et al. argue that when physical variance is controlled, the larger N170 in response to faces is abolished. The view of Thierry et al. (2007) that previous studies used uncontrolled sets of stimuli was rapidly dismissed in a short (Bentin et al., 2007b) and a long (Rossion & Jacques, in press) paper. In the latter paper, among other things, we demonstrate that (1) the physical variance between pictures of faces and of non-faces were controlled in many previous studies that found a larger N170 to faces, and (2) stimulus sets with different inter-stimulus variance may elicit N170s of similar amplitude.

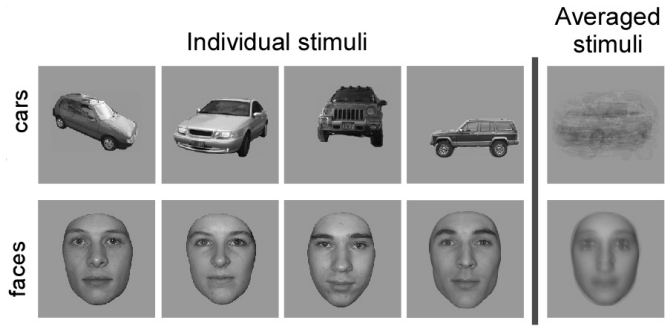

Figure 48: Thierry et al. (2007) argued that in traditional studies of the N170, while pictures of faces are presented always at the same position, viewpoint and size, there is much more physical variability in the non-face stimuli that compared to faces (e.g. pictures of cars used in many studies). Hence, the average of 16 full- front pictures of faces (left) can still be recognized as a face stimulus, while the average of 16 highly variable pictures of cars leads to an unrecognizable stimulus. Modified from Rossion & Jacques (in press).

Finally, although a careful control for low-level parameters is important to establish the validity of the early differences between object categories, in particular the larger response to face stimuli, such control is not always desirable in each and every ERP experiment. Indeed, when controlling for all low-level properties, the stimuli may become highly simplified or degraded, and the experiment may lose part of its ecological validity. Once it is established that an effect such as the larger electrophysiological responses to faces than objects is not simply due to low-level visual properties left uncontrolled in a single study, it may be very interesting to test the response properties of the component with stimuli that present as much as the characteristics of natural face pictures. For instance, face stimuli may

have relatively more power or may convey diagnostic information in different spatial frequency bands than object stimuli or visual scenes in general (Bosworth, Bartlett & Dobkins, 2006).

3.1.3. The N170 for faces reflects high-level visual processes.

Rather than reflecting difference in low-level visual properties between categories, plenty of observations in the literature indicate that the N170 amplitude is driven more by high-level processes (i.e. the *perception* of a face). One striking example is the finding that a stimulus perceived as a face when it is in the upright orientation, such as a two-tone "Mooney" face stimulus, elicits a decreased N170 amplitude when the exact same stimulus is presented upside-down and is no longer perceived as a face (George et al., 2005; Latinus and Taylor, 2005; Latinus et al., 2006; Figure 49A). Similarly, the famous "Vegetable garden" painting by the Italian painter Giuseppe Arcimboldo, where a face stimulus is perceived due to the combination of fruits and vegetables for instance, elicits a large N170 when the painting is presented upright. However, when the exact same stimulus is presented upside-down, the face is usually not perceived, leading to a reduced N170 (Caharel et al., in preparation; see Figure 49B). These effects occur despite the fact that picture-plane inversion of a stimulus keeps constant all low-level properties but for the phase orientation of the stimulus.

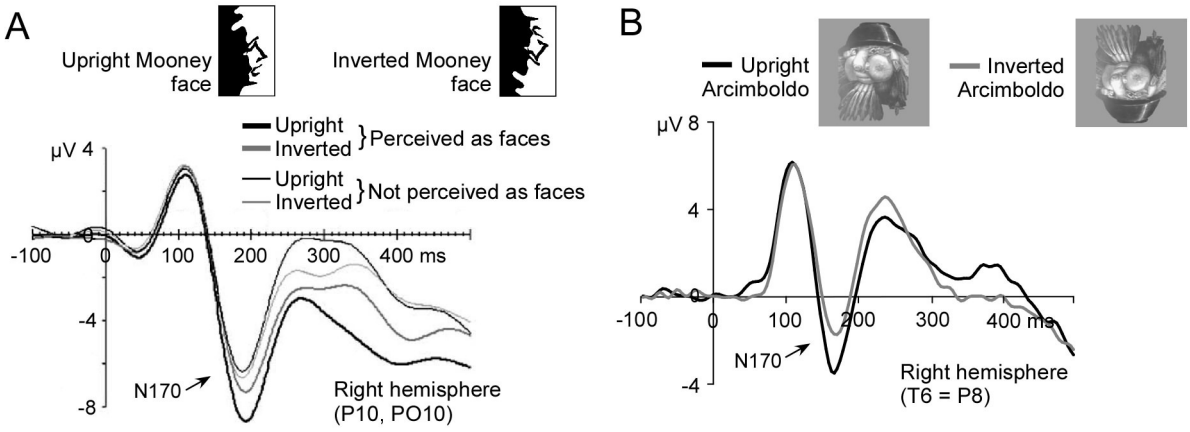

Figure 49: A. When two-tone ("Mooney") images are presented upright, they usually lead to the perception of a face stimulus, yielding a larger N170 than when the exact same pictures are presented inverted and do not lead to the perception of a face (Figure adapted from George et al., 2005). B. The same effect is observed for pictures of the paintings of G. Arcimboldo, where the face stimulus is perceived as emerging from the organization of nonface features such as fruits and vegetables (Caharel et al., in preparation). Modified from Rossion & Jacques (in press).

Similarly, the large modulations of N170 amplitude to non-face objects induced by visual expertise (Tanaka and Curran, 2001; Scott et al., 2006), the reduction of the N170 amplitude to a face presented peripherally while subjects are fixating an objects of expertise (Rossion et al., 2007; Rossion et al., 2004; see also below), or the absence of N170 face effect in some individuals suffering from congenital or developmental prosopagnosia in contrast to normal controls (Bentin et al., 2007a; Harris et al., 2005; Minnebusch et al., 2007) cannot be accounted for at all by low-level properties, again because the exact same stimuli are compared. The same reasoning applies for studies in which the N170 to otherwise meaningless stimuli (e.g. two dots located side-by-side on an horizontal plane) is increased in amplitude after conceptually priming the participants' awareness to the physiognomic value of these stimuli (e.g. the two dots can be perceived as two eyes), even though the same stimuli are used before and after priming (Bentin & Golland, 2002; Bentin et al., 2002).

Other sources of evidence suggest that the N170 reflect high-level visual processes involved in the perception of the face category. That is, many stimulus manipulations (e.g. inverting the face in the vertical plane, inverting the contrast of the face image, presenting facial features in isolation, "jumbling" the facial features into an abnormal configuration, removing facial features thought to strongly affect face recognition) often lead to an N170 of larger or equal amplitude as compared to a "normal" face (Rossion et al., 1999a; 2000; Eimer, 1998; 2000b; Likenkeart-hansen et al., 1998; Itier & Taylor, 2002; 2004a; c; Bentin et al., 1996; Jemel et al., 1999; Taylor et al., 2001; Shibata et al., 2002; George et al., 1996; Zion-Golumbic & Bentin, 2007; Itier et al., 2007), strongly suggesting that the neural activity underlying the larger N170 is involved in processing aspects of faces such as the facial features and their configuration. The effect of the face manipulation on the N170 in relation to face perception will be discussed in a later section.

Altogether, these results support the view that the N170 is more correlated with the perception of a face than with low-level properties of face images. Similar properties are usually found for high-level areas that respond more to faces in the visual cortex (FFA) (e.g. Winston et al., 2004; Tong et al., 1998; Andrews & Schluppeck, 2004).

Finally, other evidence that the N170, or the VPP recorded at the vertex, reflect high level processing comes from the finding that for a fixated face, a large change of size does not greatly affect the amplitude of the VPP (Jeffreys et al., 1992) or the N170 (Rossion et al., unpublished data). This is true unless the face is presented away from the fixation point, leading to a reduction of the component's amplitude (Jeffreys et al., 1992; for the N170, see Rousselet et al., 2005). However, this eccentricity effect appears to be linked to the reduced

cortical representation of peripheral visual field locations, as compensating for cortical magnification[63] differences between fovea and periphery by enlarging the peripheral stimuli leads to an N170 of roughly equal amplitude for a face presented at fixation or at 10° of eccentricity, and to a recovery of the N170 amplitude difference between faces and non-face objects (Rousselet et al., 2005). The invariance of the N170 to changes in size and position (as long as the size is matched for the cortical magnification factor) is perfectly compatible with the view that the N170 reflects the activation of visual areas associated with higher-level face visual processes. Indeed, face neurones in the high-level IT cortex of the monkey brain exhibit similar size and position invariant properties[64].

3.1.4. Modulations of the N170 elicited by non-face categories.

Apart from being larger in response to face pictures, the N170 can also distinguish between different object categories (e.g. houses vs. shoes, Rossion et al., 2000; Figure 50), suggesting that general perceptual categorization or discrimination processes takes place at around that latency (see Vogels & Luck, 2001 for similar conclusions about the N1 component). For instance, pictures of cars often yield a larger N170 than other categories (e.g. cars *vs.* houses and chairs in Itier et al., 2006a; *vs*. birds and furnitures in Carmel and Bentin, 2002; or *vs.* chairs, shoes, greebles, and houses in Rossion et al., 2000), probably due to their high familiarity, complexity and internal visual similarity comparable to faces. Sometimes, pictures of cars elicit N170 as large as for pictures of faces (Schweinberger et al., 2004), although a larger N170 to faces than cars is usually observed (Bentin et al., 1996; Carmel and Bentin, 2002; Goffaux et al., 2003; Itier et al., 2006a; Philiastides and Sajda, 2006; Rebai et al., 2001; Rossion et al., 2000; 2003a).

Such amplitude differences between non-face categories, or strong sensitivity to non-face stimulus repetition (i.e. "adaptation"), can also be observed in face-sensitive areas

[63]The cortical magnification factor refers to the fact that in retinotopic visual areas, the cortical surface devoted to process information appearing at the fovea far exceeds the surface devoted to the periphery (see Daniel & Whitteridge, 1961; Levi et al., 1985).

[64] The observation that the larger N170 to faces is reduced when stimuli are presented at increasing eccentricity may appear inconsistent with the finding that the *face-selectivity* of face-cells in IT cortex of the monkey is relatively translation-invariant, especially for stimuli presented at less than 10° of eccentricity (Tovee et al., 1994). This discrepancy between ERP and single-cell studies can partly be accounted for if one considers the large differences in the size of the stimuli used in ERP (usually around 3-5°) and single cell (8 to 17° in Tovee et al., 1994). This is supported by the finding of Rousselet et al. (2005) that compensating for cortical magnification differences between fovea and periphery lead to a recovery of the N170 amplitude difference between faces and non-face objects. Moreover, even though the selectivity of face neurons in terms of response-amplitude is invariant to stimulus position, the latency of the response in these cells is generally increased by 20-30 ms (Tovee et al., 1994), comparable to what is observed in ERPs (Rousselet et al., 2005).

revealed by fMRI, such as in the FFA (Grill-spector et al., 2006; Downing et al., 2006; Avidan et al., 2002) in line with the view that object representation in the visual cortex is widely distributed (Tanaka et al., 1996; Ishai et al., 1999; Haxby et al., 2000).

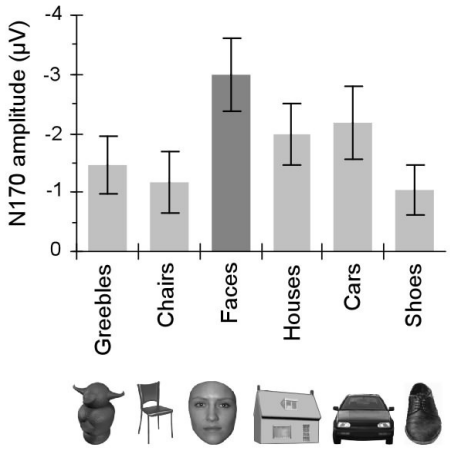

Figure 50: N170 amplitude to six different categories, showing both higher amplitude for faces and significant differences between different non-face categories (Rossion et al., 2000). This figure also illustrates the generally higher amplitude for car stimuli than for other categories. The N170 amplitude is averaged over 2 pairs of electrodes located over occipito-temporal cortex in left and right hemispheres. Modified from Rossion & Jacques, in press.

At this state of knowledge, it is difficult to interpret in functional terms these amplitude differences between non-face objects both in the N170 and in fMRI data. Many factors that differ between object categories such as low-level visual properties, general shape (e.g. number and shape of external parts, concavity and convexity of the edges/angles, "roundedness" vs. "sharpeness" of the angles, non-accidental vs. metric properties[65] – see Kayaerts et al., 2003; 2005), surface properties (e.g. texture, reflectance, colour), behavioural relevance, or familiarity/experience could account for these amplitude differences.

Several evidences suggest that differential familiarity, expertise or perceptual experience with different categories could partly account for these differences as well.

First, several object categories (body parts, whole body, printed words, biological movement) that are highly familiar and behaviourally relevant, trigger specific responses in the human or the monkey brain as revealed by fMRI (Downing et al., 2006), single-cell recording (Gross et al., 1972; Desimone et al., 1984; Perrett et al., 1985b; Jellema et al.,

[65] Non-accidental properties are properties of a stimulus do not vary when the object in rotated in depth (e.g. whether a part of the contour of an object is curved or straight). Therefore, these properties are considered important for invariant object recognition (Biederman, 1987). Metric properties are stimulus properties that are modified when the object is rotated in depth (e.g. if the contour is curved, the degree of curvature will change with rotation).

2004), or intracranial recordings of N200s (Allison et al., 1999; 2002; McCarthy et al., 1999). These categories also usually elicit larger responses than object in the time window of the N170 (hands: Eimer, 2000b; Mouchetant-rostaing et al., 2000; Words: Rossion et al., 2003a; Bodies: Thierry et al., 2006; eye movements: Puce et al., 2000; see Figure 51), although the latency and the topography of the component for each category is usually slightly different than that for the face-N170. Aside from these highly familiar and behaviourally relevant categories, a recent study (Caharel et al., in preparation), also found a larger N170 in response to common objects (house, car, lamp,…) relative to "non-common" objects (catapult, telescope, harp,…).

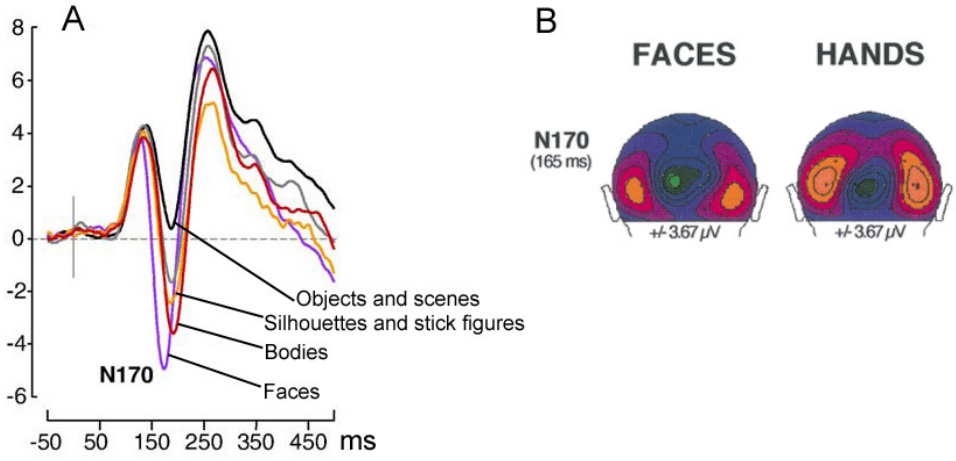

Figure 51: A. N170 response to faces, bodies, silhouettes and stick figures, objects and scenes. Modified from Thierry et al., 2006. B. Topography (back-view of the head) of the N170 to faces and hands. Note the similar amplitude but slightly different topographies of the N170 elicited by the two categories. Modified from Mouchetant-Rostaing et al., 2000.

Second, a number of studies have show that the N170 can be strongly modulated by the level of expertise with non-face categories. For instance, testing bird and dog experts, Tanaka & Curran (2001) found that the N170 is larger in response to bird pictures than dog pictures in bird expert while the opposite is found for dog expert (Figure 52). In another study, people expert at discriminating fingerprint exhibited a delayed N170 for inverted compared to upright fingerprint (similar to what is found for faces), an effect that wasn't observed in people having no experience with fingerprints. Similar results were found when people are trained over several sessions with bird pictures (Scott et al., 2006) and artificial objects ("greebles", Rossion et al., 2002). More direct evidence that faces and object of expertise recruit similar functional processes comes from studies in which the N170 substantially suppressed when subjects are fixating novel exemplars of an non-face object

category on which they had been trained ("greebles", Rossion et al., 2004) or for which they have acquired expertise over many years ("Cars", Rossion et al., 2007; see Figure 53). These findings are consistent with numerous observations of change in stimulus selectivity in IT neurons due to perceptual learning and stimulus features diagnosticity (e.g. Logothetis et al., 1995; Sigala & Logothetis, 2002; Freedman et al., 2003; 2006).

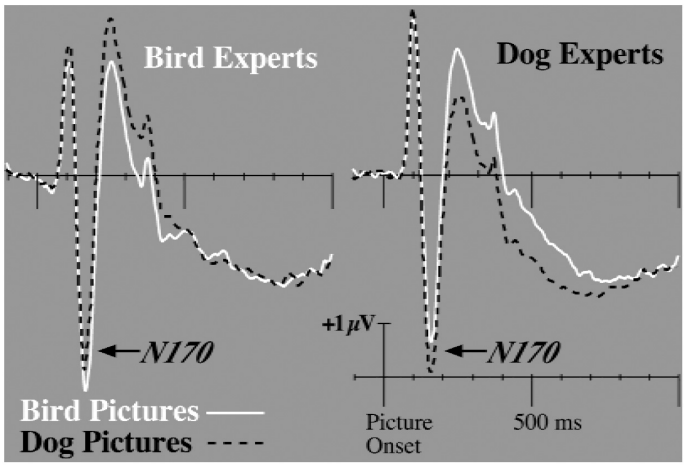

Figure 52: The N170 elicited by bird pictures is larger than that elicited by dog pictures in subjects who are bird experts. The opposite result is found in dog experts. Modified from Tanaka & Curran, 2001.

Third and finally, while attempts to disclose selective attention effects on the N170 response to face stimuli have been inconclusive (Eimer, 2000c; Severaq-Cauquil et al., 2000; Carmel & Bentin, 2002), the N170 response to a non-face object category (e.g. cars) may be increased when subjects selectively attend to this category (Carmel & Bentin, 2002), thereby reducing the N170 amplitude difference with faces.

To summarize, in this section we have seen that the N170 is the first component that systematically distinguishes faces from other non-face object categories. In fact, this larger N170 to faces is probably the most well-known property of the N170. This larger N170 to faces is probably due to the recruitment of a larger number of areas in the visual cortex compared to other objects. Moreover, the N170 is more correlated with the perception of a face (i.e. reflects high-level processes) than to the low-level visual properties of the stimulation. Finally, although the N170 is larger to faces, this component can be modulated by experience with non-face categories.

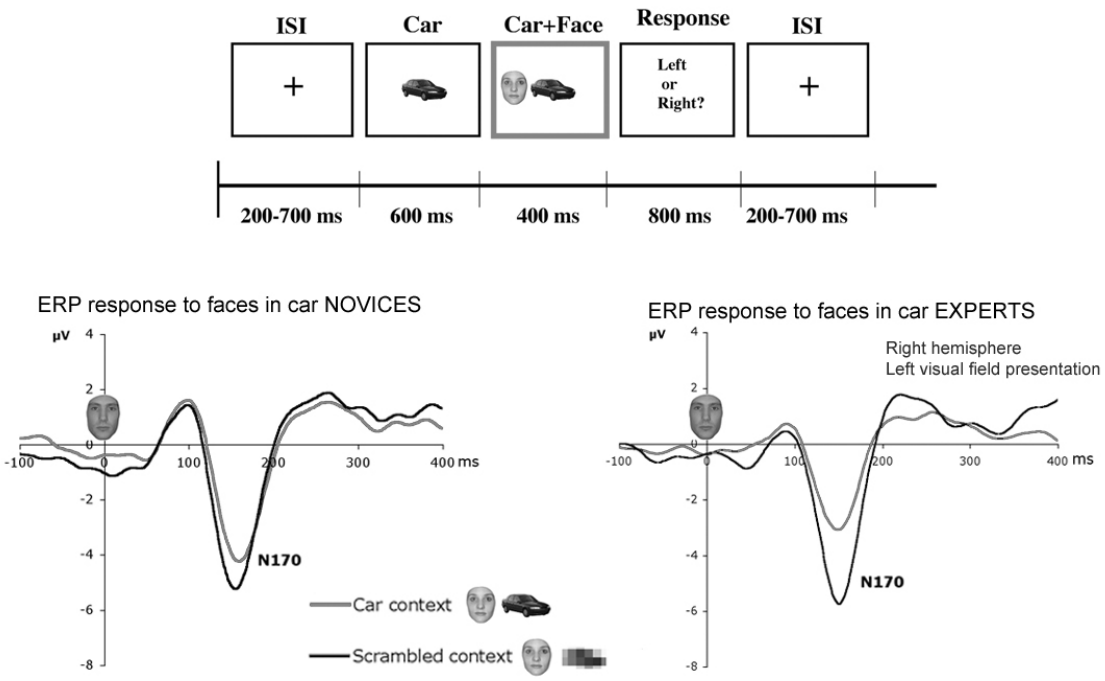

Figure 53: The N170 elicited by a face presented in the periphery (~3° from fixation) is reduced when subjects are fixating a car compared to a scrambled car. The N170 reduction is much stronger when subjects are experts with the car category. Moreover, the amount of reduction is correlated with a behavioural measure of car expertise. This suggests that faces and object of expertise compete for early neural representation during the N170 time window. Upper row displays the stimulus trial sequence: a car first appears at fixation for 600 ms, followed by a face for 400 ms. Modified from Rossion et al., 2007.

Given the particular sensitivity of the N170 to face, this component seems to be an ideal time-window (i.e. that can be used as a "tool") to investigate high-level human face processing. In the next sections we review findings in the ERP literature about the functional processes involved in face perception taking place during the time-window of the N170.

3.2. The encoding of face information during the N170 time-window.

We have seen that the presentation of face stimuli elicit a large N170 over occipito-temporal regions, suggesting that specific visual encoding of faces takes place during this time-window[66]. The question we to turn to here is: *what are the face processes taking place during the time-window of the N170?*

[66] Although faces elicit a large N170, we have seen that non-face categories also generate N170 of varying amplitude, suggesting that this time window is particularly important in visual categorization in general.

Although it is tempting to link an ERP component – and the N170 in particular – to a functional "stage" in visual processing, we will defend the idea that there is in fact no particular need to refer to distinct processing "stages" to investigate the time course of face processing. As for the N170, most studies refer to this component as a "structural encoding" stage, which we believe is mostly too restrictive and generally counter-productive when aiming at clarifying how faces are represented during this time-window. Our arguments are presented in the next section.

3.2.1. The N170 as a structural face encoding stage?

In the original papers on the VPP component recorded for face stimuli, Jeffreys (Jeffreys, 1989; 1996) associated this component, with the *structural encoding* stage of Bruce & Young's model of face processing (Bruce & Young, 1986). In Bruce & Young's model, the description of the processing steps leading to face recognition is derived from more general object recognition models in which the visual input of an object is initially described as a 3D object-based[67] structure, before it is compared to representations stored in memory for recognition and further higher-level processing (see e.g. Marr, 1982; Biederman, 1987). In their model, Bruce & Young (1986) suggested that the structural encoding stage builds a representation (a "structural code") "*which captures those aspects of the structure of a face essential to distinguish it from other faces*". Once built, the structural representation can be compared to stored face representations in order to recognize the face[68].

Since the pioneer work of Jeffreys, it has been generally claimed in the ERP literature on face processing that the VPP and N170 represents somehow the structural encoding stage of Bruce & Young's model (e.g. Jeffreys, 1996; Bentin et al., 1996; Eimer, 2000b; d). However, whereas Bruce & Young (1986) did not refer to the structural encoding stage as a face detection stage (see previous paragraph), this is in fact how it is generally understood in the ERP literature: a stage at which a generic facial structure is activated, common for unfamiliar and familiar faces, before the latter can activate internal (stored) representations. For instance, according to Bentin & Deouell (2000, P.51) "…*the brain mechanism with which*

[67] An object-based representation (or observer independent) is a 3D representation of an object (or a face) that is independent of the particular point of view or viewing condition of the observer. It is opposed to image-based or observer-dependent representations which are contingent on the particular viewing conditions.

[68] However, given that face perception is particularly sensitive to changes in viewpoint, Bruce & Young (1986) did not claim that the structural encoding stage corresponded to the extraction of an object-based representation. Rather, they acknowledged that multiple viewpoint-dependent representations may co-exist at this level.

the N170 is associated is triggered by the presence of a human face in the visual field but is not directly involved in face recognition." (see also e.g. Eimer, 2000d; Bentin & Deouell, 2000; Sagiv & Bentin, 2001; Carmel & Bentin, 2002; Schweinberger et al., 2002a; b). This view is based on the following observations: (1) the N170 is evoked by faces displayed in many different format (photographic, schematic, drawings,…) suggesting a relative independency of viewing conditions; (2) Prosopagnosic patients for which the structural encoding (in the sense of face detection) is preserved display a "normal" larger N170 to faces than to non-face objects (Eimer & McCarthy, 1999; Harris et al., 2005); (3) the N170 is mostly insensitive to face familiarity indicating that it represents a stage of processing before the activation of stored individual representations of familiar faces; (4) it is strongly influenced by manipulations of local facial features and/or face configuration (see next section) suggesting it is a stage where facial features and configuration are visually encoded.

However, the interpretation of the functional significance of the neural processes occurring during the N170 as a *structural encoding* stage is complicated by several factors.

First, there is no real agreement, both among researchers and even across studies by the same researchers, on what this structural face encoding stage means. That is, while all researchers generally agree that this structural encoding is a stage at which faces are first represented specifically (i.e. they are categorized as faces), there is some confusion in the literature about the type of process(es) performed during this stage and about the nature of the information it processes. For instance, some authors mention that the structural encoding would entail a multi-stage process: "*Its early stages involve in "the detection of heads in the visual field and the processing of individual face components, whereas later stages are responsible for the configurational analysis of faces.*" (Eimer, 2000b, p. 2319). Others suggest that the structural encoding stage contains two distinct components: one that encodes facial features and one that process faces in a holistic manner acting on configural information (Sagiv & Bentin, 2001). In a more recent report, the same authors make a distinction between a "*configural structural encoding*" and the detection/encoding of the facial features, with only the latter being represented at the N170[69] (Zion-Golumbic & Bentin, 2007).

[69] A good illustration of this confusion about which information is represented during early perceptual processing of faces can be found in the Introduction of George et al. (2005, p. 664): "*Bentin et al. (1996) proposed that the N170 may be related to the detection of eye features or to the mere detection of faces based on first-order relations and holistic processing (Bentin et al., 1996; Sagiv & Bentin, 2001). However, other authors provided some evidence that the N170 is related to late stage of face structural encoding (Eimer, 2000), and it has been proposed that earlier activities – around 100 ms – would be associated to facial categorisation or perceiving general facial configuration (Itier & Taylor, 2002; Liu et al., 2001; Likenkeart-; 1998; Halit et al., 2000).*"

Moreover, in the ERP literature on face processing, while most authors refer to the classical hierarchical distinction between a stage in which a *structural representation* (i.e. perceptual coding of faces) is built from visual inputs and a subsequent independent *recognition* stage (i.e. that involves the comparison/matching between the structural representation and the representations stored in memory – the "*face recognition units*" in Bruce & Young's model), the fact that information about an individual face may be represented during this structural encoding stage is not addressed at all. More precisely, as reflected in Bentin & Deouell's citation above, the authors usually state that there is a structural encoding stage (VPP/N170) and then a subsequent recognition stage[70] (e.g. Eimer, 2000d; Bentin & Deouell, 2000; Sagiv & Bentin, 2001; Carmel & Bentin, 2002), largely missing the fact that a (structural) representation that is supposed to be later matched to another representation should in itself contain the information that is to be matched (in this case the individual aspects of the face).

A **Second** reason to remain cautious when interpreting the N170 as a structural encoding stage is that it is commonly assumed in ERP investigations that an ERP component reflects a functional "stage" or a component of a cognitive architecture model such as Bruce & Young's model. Moreover, in many ERP studies cognitive models are used as a frame of reference to derive both hypothesis and data interpretations. Here we would like to advocate the view that this mapping of cognitive models onto ERP responses is missing the important fact that most ERP components reflect the summed activity in highly parallel and distributed neural networks. Cognitive models do not usually take fully into account such a dynamic, parallel, distributed and non-linear nature of the brain systems (see Henson, 2006 for a discussion). Part of the temptation to associate an ERP component with a fixed functional stage probably stems from the apparent similarity between the sequentially activated components/modules identified in cognitive psychology, and the sequence of apparently discreet and sequential deflections making up the scalp ERP response.

As for the N170, even though we fully agree with the view that this component represents the neural activity involved in the visual/perceptual encoding of faces, we believe

[70] For example: from Eimer, 2000 (p. 695): "*In their influential model of face recognition, Bruce and Young (1986) distinguish between a perceptual structural encoding stage, where individual face features and their spatial configuration are analysed, and a subsequent recognition stage, where structural representations are compared with stored face representations (face recognition units).* ", and from Sagiv & Bentin, 2001 (p. 937): "*several models of face recognition incorporate an initial 'structural encoding' mechanism with physiognomic features as input and an integrated, but not fully identified, representation of a face as output*".

that a definition of this component as a "structural encoding stage" is vague and too restrictive at the same time, for the reasons mentioned above, and thus generally counter-productive Consider the following arguments: (1) as already mentioned the N170 represents the summation of electrical activity from many sources in the visual cortex, and (2) the time window of the N170 extends from about 120-130 ms to 190-200 ms suggesting a considerable amount of computation is performed during such a "long" time window. It is therefore highly improbable that the neural processes underlying the N170 only reflects a single function of face processing. In addition, referring to the N170 as a structural encoding stage distinct from a "recognition stage" does not capture the fact that visual information processing in the visual system is probably more continuous than discreet (i.e. it proceeds in a continuous accumulation of information).

Unfortunately, in the ERP literature on face processing, very few attempts have actually been made to question the link between a structural encoding and the N170. This may lead to a certain amount of circularity. Indeed, because many researchers use the definition of the N170 as a structural encoding stage as a building block for experimental designs and interpretation schemes of ERP studies, they can only derive functional properties of face processing compatible with this framework.

To summarize, we would like to reinstate that there is actually no need to refer to a cognitive model[71] as a framework to derive functional interpretations of ERP data in basic research on face or object visual processing. In the case of the N170, the view of this component as a structural encoding stage is problematic, firstly because it has generated a loath of confusion given the lack of agreement among researchers as to what this stage represents, and secondly because it strongly constrains and biases the description and empirical investigation of face visual processing in ERP research.

Finally, if the N170 cannot be described as reflecting a given functional stage, as we advocate here, there is probably no reason to design experiments targeted to determine what neural or cognitive process this particular component may reflect. We believe a more fruitful approach in ERP research on face processing is rather to use the N170 response as a "tool", or as a critical time window, to investigate the functional and temporal properties of face perception. This approach should prove sufficient to address issues such as the following:

[71]Our goal, however, is certainly not to question the cognitive approach in psychology or neuropsychology, but rather to emphasize that this approach cannot be applied to electrophysiology without strongly constraining investigations in this domain.

What is the time course and evolution of face information processing from stimulus onset to response? Are there multiple representations or a unique representation that evolves/refines over time? What is the nature of this/these representations (what format)? How is information about faces encoded by populations of neurons and how individual face representations relate to each other (i.e. how is the "representational" face space organized)? What facial cues are represented to categorize a face or encode its identity?

This is the type of approach that we have attempted to use and develop in our empirical work for this dissertation.

3.2.2. The respective representation of facial feature and face configuration during the N170.

Human face recognition or discrimination is based on the processing of multiple internal and external cues such as facial features (e.g. eyes, mouth, nose, hair,…) and on the spatial integration of these features. In the literature, the processing of the local facial features (also called "featural processing", "piecemeal processing" or "analytic processing") is thought to be somewhat segregated from the processing of the spatial relationship between these features (also called "configuration" or "configural processing" or “holistic processing”) (see Maurer et al., 2002). Usually, local facial features are defined as the obvious circumscribed components of a face such as the eyes, the nose, the mouth, the eyebrows, the hair… Others author consider properties of the skin surface (e.g. texture, color, reflectance) or the general shape of the contour as a facial feature as well. The terms "configuration" or "configural information/processes" in the face processing literature has generated a host of confusion, mostly because the very same terms are used to refer to any process or face information that involves the spatial relationship between facial features. More precisely, three different concepts have been classified under the "face configuration" label (see Maurer et al., 2002; Rossion & Gauthier, 2002 for reviews): (1) the general spatial layout of the face (i.e. a face is composed of two eyes above the nose and the mouth), also called "first-order relations" and shared by all upright human faces; (2) the relative metric distances between facial features (e.g. inter-ocular or nose-mouth distances), also termed "second-order relations"; (3) the fact that the facial cues (local features, skin surface information, metric distances, global shape related to shape-from-shading[72]) are not perceived/processed independently from one another

[72] Shape from shading is the information about an object's shape or 3D volume that can be retrieved through the particular shading due to the interaction between light-source and the shape's 3D volume. Given the stability of light orientation in natural environment (i.e. from above), the processing of shape-from-shading in faces is crucial to derive information about the face's shape. Manipulation that modifies the "normal" configuration of

but are rather integrated into a gestalt, or a so-called *holistic* representation. Therefore, in contrast to local features or first and second order relationship, holistic processing is not an information in faces (there are no such thing as holistic information), it is a process that is used to integrate facial information into a gestalt – a whole – representation (Tanaka & Farah, 1993). As such, holistic processing is the operation of combining both facial features and the spatial configuration. These three concepts have been distinguished based on different kind of (debated) empirical evidence and are thought to represent, at least partly, different functional processes involved in face perception which we won't detail here.

While the first-order facial configuration is used to detect a face in the visual environment (Lewis & Edmonds; 2003; Purcell & Stewart, 1988), it cannot be used to recognize the face given that this configuration is common to all human faces. On the contrary, the processing of metric distances between features is not critical for face detection (i.e. the categorization of a face as a face), unless the distances between features are grossly exaggerated and is rather involved in representing an individual face. The third concept - holistic processing – is involved in both face detection (e.g. holistic processing is necessary to integrate and bind the facial features in normal or "degraded" faces like Mooney face stimuli or a face in the paintings of Arcimboldo for instance) and in representing individual faces (i.e. the identification of individual facial features is strongly influenced by the other face parts or the whole face context (Young et al., 1987; Hole, 1992; Tanaka & Farah, 1993; Goffaux & Rossion, 2006)).

In this section we will mainly be interested in the type of face cues that are represented in the perceptual encoding taking place during the N170 time window. To our knowledge, no ERP studies have been conducted to investigate the time course with which these different facial cues (i.e. local or global features, metric-distances, 3D shape) are encoded in an *individual* face representation. Rather, most of the ERP studies we will review here have measured the rather coarse effect of manipulating featural or configural (only first-order relations, but *not* metric-distances) information on the N170 without any manipulation of face identity. Therefore, these studies can only address the question of the type of facial cues that are involved in the categorization of a face as a face (i.e. face detection). The question of how holistic processing may be involved in the coding of individual faces is investigated in one of the experiments of this thesis (see chapter 8).

---

shape-from-shading in faces, such as contrast inversion, disrupts face holistic processing (Lewis & Johnston, 1997).

a. A short story that repeats itself…

As we already mentioned, Jeffreys (Jeffreys, 1989; 1996) originally associated the VPP component with the *structural encoding* stage of Bruce & Young's model of face processing (Bruce & Young, 1986) during which facial feature and configuration are integrated.

Later, based on the finding that an isolated pair of eyes (i.e. without any other facial features) elicit a larger N170 than a whole face (see Figure 54), Bentin et al. (1996)'s suggested that the N170 rather reflects a stage during which physiognomic features (most particularly the eyes) are detected. The conception of the N170 as an eye detector was rapidly dismissed by Eimer's finding (Eimer, 1998) that faces presented with or without the eyes elicit N170s of equal amplitudes (although slightly delayed for a face without eyes) (see similar results in Itier et al., 2007). These results were also confirmed by the observation that a face with eyes open vs. closed elicit an N170 of equal amplitude (although again slightly delayed for eyes-closed) (Taylor et al., 2001). The interpretation of the N170 as reflecting the *structural encoding* stage was then favoured in the ERP literature of face processing, including Bentin and colleagues (Bentin & Deouell, 2000; Sagiv & Bentin, 2001). However, more recently the history started to repeat itself, and it was again claimed that the N170 amplitude is mostly determined by the eye region (Schyns et al., 2003; Smith et al., 2004; 2007), or local features in general (Zion-Golumbic & Bentin, 2007; Harris & Nakayama, 2007), and that the neural processes underlying the N170 do not represent – or only minimally – the processing of face configuration.

Let's look at the evidence about whether facial features and/or face configuration are represented during the N170 time-window.

b. The representation of facial features during the N170 time-window

Probably the strongest evidence that facial features are represented during the N170 comes from the finding that the presentation of a pair of eyes in isolation leads to a face-like N170 (i.e. larger than non-face objects), which is in fact usually larger than the N170 recorded to a whole face (Bentin et al., 1996; Taylor et al., 2001; Shibata et al., 2002; Itier et al., 2007; see Figure 54). There is much fewer evidence that other features such as lips and nose do elicit a N170 larger than the N170 to non-face objects (see Bentin et al., 1996; Taylor et al.,

2001). Other "features" like global facial shape/contour (including hair) also elicit a large N170 (Eimer, 2000b; Shibata et al., 2002; Zion-Gollumbic & Bentin, 2007; see also Jeffreys, 1996 for the VPP). The sensitivity of the N170 to the eyes has also been demonstrated recently using the "Bubbles"[73] technique (Gosselin & Schyns, 2001). Using this technique it has been shown that when the parts of a face revealed through bubbles apertures include the eye region, a large N170 is observed. Conversely, when the revealed areas do not include the eye region, a smaller N170 is observed (Schyns et al., 2003; Smith et al., 2004; 2007).

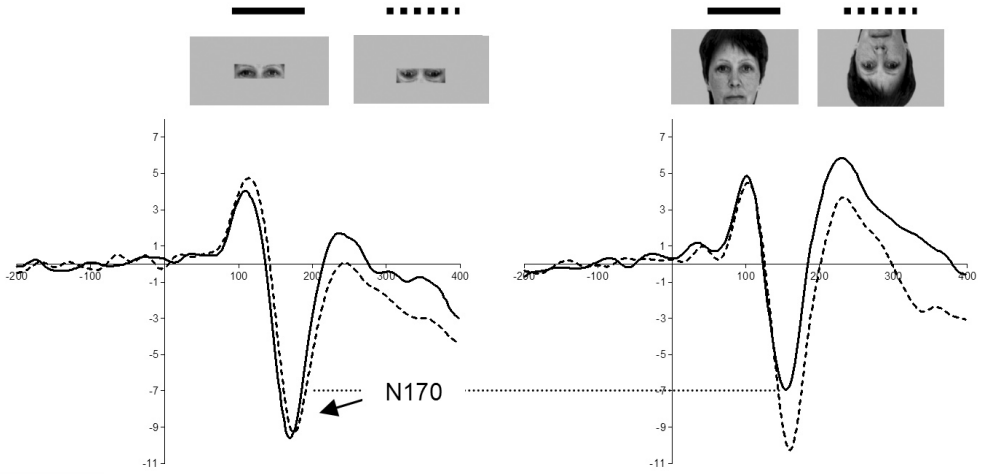

Figure 54: Figure showing both the larger N170 to isolated eyes than to full faces (compare left and right) and the larger N170 to inverted compared to upright faces (right). The N170 to upright and inverted eyes is of equal amplitude. Modified from Itier et al., 2007.

As a second evidence for a representation of features during the N170, it has been observed that rearranging all or part of the inner features of a face in an abnormal configuration also leads to face-like N170. Similarly than to isolated eyes, this manipulation usually enhances the N170 amplitude (George et al., 1996; Bentin et al., 1996; Caharel et al., 2006; Machi-cassia et al., 2006, see Figure 55; but see Zion-Golumbic & Bentin, 2007), or

[73] "Bubbles" (Gosselin & Schyns, 2001) is a method used to derive information about the visual cues in an image that are useful to resolve a given task (e.g. recognize a face). In each of several thousands of trials, randomly located small regions of the stimulus are presented to the subject through a number of round apertures (the bubbles), and the subject is asked to perform the task only based on the revealed image regions. By sorting trials in which the regions revealed through the apertures allowed the subject to correctly perform the task from trials in which the information revealed did not allowed for a correct response, the technique can reveal the cues that are necessary for the subject to perform the task. It has also been applied to EEG recording to determine the information in an image that drives the EEG amplitude at a given time point (Schyns et al., 2003; Smith et al., 2004; 2007; Schyns et al., in press).

reduces the M170 amplitude[74] (in MEG: Halgren et al., 2000; Harris & Nakayama, 2007), which nevertheless remains larger than non-faces (see Figure 56).

A third evidence comes from the observation that progressively adding visual noise onto a face image – a manipulation that mostly affects the visibility of inner facial features – progressively reduces the N170/M170 amplitude (Jemel et al., 2003a; Tarkiainen et al., 2003; Tanskanen et al., 2005).

In summary, there is evidence clearly pointing toward the fact that during the time-window of the N170, facial features, even when not organized in a facial configuration, are categorized as physiognomic features.

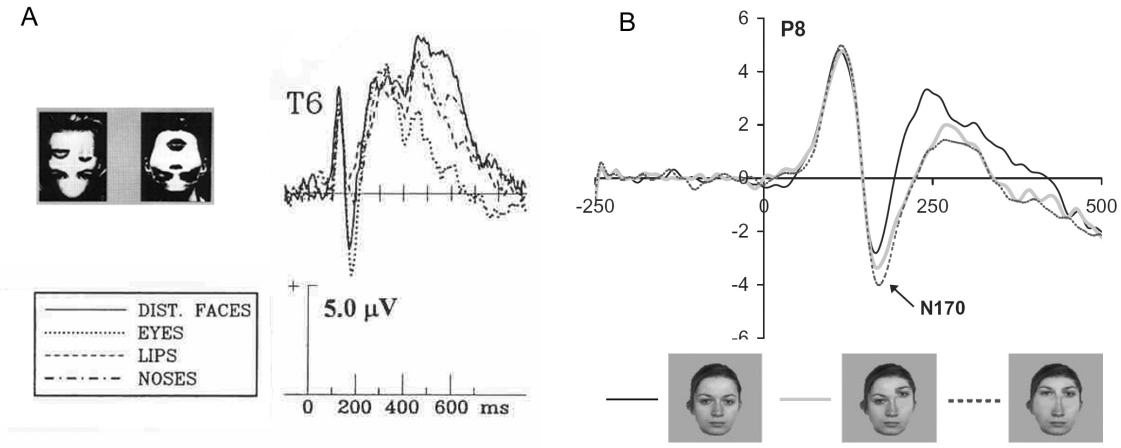

Figure 55: A. N170 response in the right hemisphere on a inferior temporal electrode to isolated eyes, lips or nose and faces with jumbled features (dist. Faces). The figure shows an example of jumbled face used in the experiment. B. The N170 is larger for faces with one or two eye displaced. Modified from Caharel et al., 2006.

[74] Note that because these EEG and MEG studies differ both in the amount of feature jumbling (e.g. all features vs. only the eyes) and in the number of different feature configurations used (e.g. Zion-Golumbic & Bentin, 2007 uses 80 different feature jumbling configurations while Machi-cassia et al., 2007 uses only two), it is difficult to derive a clear information of the coding of face configuration from these data. Moreover, the fact that MEG studies found a smaller M170 compared to normal faces while the opposite is generally found for the N170, may well be accounted for by these methodological differences.

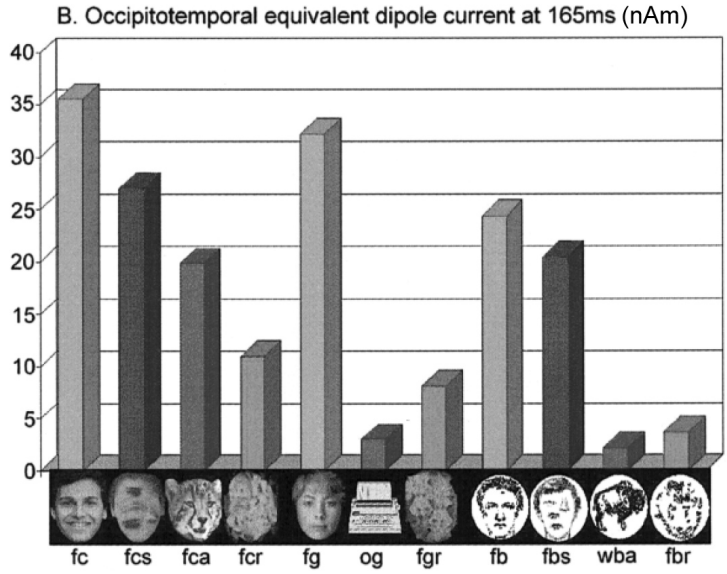

Figure 56: Amplitude of the occipito-temporal dipole corresponding to the M170 (equivalent to N170 is MEG) as a function of the stimulation (fc: face in colour; fcs: face in colour with jumbled features; fca: animal face in color; fcr: randomized colour face; fg: face in greyscale; og: object in greyscale; fgr: randomized greyscale face; fb: face in black and white; fbs: face in black and white with jumbled features; wba: whole-body black and white animal; fbr: randomized black and white face). Modified from Halgren et al., 2000.

c. The representation of face configuration during the N170 time-window

Numerous evidence also support the view that the configuration of a face is represented during the N170. The vast majority of studies that have investigated the processing of configural information at the N170 have relied on stimulus manipulation that alters the normal configuration of a face. Other studies have also more generally investigate how holistic processing allows for face detection during the N170.

First, presenting a face upside-down - a manipulations that is known to more strongly affect holistic processing, and the perception of metric distances between features than the processing of facial features (Tanaka & Farah, 1993; Rhodes, 1993; Freire et al., 2000; Goffaux & Rossion, 2006; 2007) - generally leads to a larger and delayed N170 as compared to upright faces (e.g. Rossion et al., 1999a; Linkenkear-Hansen et al., 1998; Itier & Taylor, 2002; Figure 54). A related effect is the observation that while an upright "Thatcherized" face[75] elicit a slightly larger N170 than the normal version of the same face (Millivozevich et

[75] If the eyes and mouth of a picture of a face are inverted, the face appears grotesque, unless it is presented upside-down. The normal appearance of the distorted face is known as the 'Thatcher illusion' because it was first demonstrated with the face of the former British prime minister, Margaret Thatcher (Thompson, 1980). This orientation dependent illusion has been linked to an interaction between configural/holistic processing and featural processing (Leder & Bruce, 2000; Carbon et al., 2006).

al., 2003; Carbon et al., 2005), this effect disappears (Millivozevich et al., 2003) or is reversed (Carbon et al., 2005) when these stimuli are presented upside-down (Figure 57).

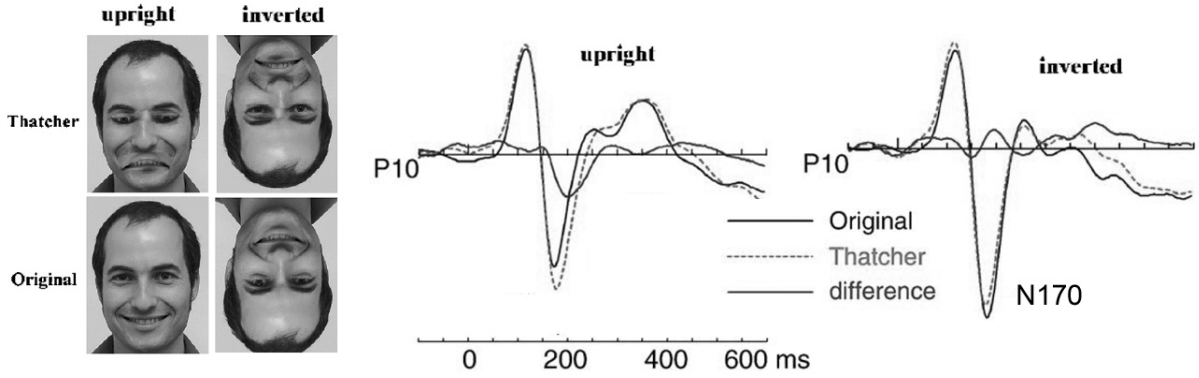

Figure 57: N170 response to normal and "Thatcherized" faces in the upright and inverted orientation. The N170 is larger for Thatcherized faces in the upright orientation but smaller in the inverted orientation. Modified from Carbon et al;, 2005.

Similar to face inversion, reducing configural/holistic processing by reversing the contrast of a face (Lewis & Johnston, 1997), also enhances and delays the N170 (Itier & Taylor, 2002; 2004c; Itier et al., 2006a; 2007).

Second, we have seen that rearranging the inner features of a face in an abnormal configuration usually leads to a larger N170 (George et al., 1996; Caharel et al., 2006; Machi-cassia et al., 2006) or smaller M170 (Halgren et al., 2000; Harris & Nakayama, 2007) than to a normal face. While the fact that the N170 displays a larger amplitude to a face with an abnormal configuration compared to non-face object can be taken as evidence that features are represented during the N170, the amplitude difference compared to normal faces is an evidence that face configuration also matters during this time-window. The same holds true for the larger N170 to eyes presented in isolation, even though in this case one can always argue that the amplitude difference between isolated eyes and whole faces could be due to differences in the amount of face information, independently of the configuration.

Third, schematic faces, in which the facial features are very crude/simple but are organised in a facial configuration (first order), elicit an N170 of similar amplitude as for photographic pictures of faces (Sagiv & Bentin, 2001; Latinus et al., 2005). Similarly, two-tone mooney faces (in which the features are poorly defined) or Arcimboldo "vegetable" faces that are perceived as faces mainly on the basis of holistic processing, elicit higher N170 amplitude than mooney/arcimboldo faces which are not perceived as faces (see Figure 49).

A last compelling evidence that a face is processed as an integrated whole during the N170 comes from one of our recent ERP study in which we used a stimulus manipulation similar to what is used in the face composite illusion paradigm (Young et al., 1987). That is, we manipulated the possibility to integrate – or not - the face into a holistic representation by comparing the ERP response to top and bottom parts of a face which were either aligned (i.e. allowing for an integration of the top and bottom face parts) or misaligned (i.e. making it impossible to integrate the top and bottom face parts) (see Figure 58). The N170 was much larger when the top and bottom part were misaligned than when they were aligned (Figure 58 –top row). Moreover, this difference was much stronger when both top and bottom parts of the stimulus were face parts compared to visual noise (Figure 58 – second to last rows). Interestingly however, when only the top part of the stimulus was a face part, there was still a large amplitude increase when the bottom part was misaligned, even though this bottom part was only made of visual noise (Figure 58 – second row). Additionally, in the aligned condition the amplitude of the N170 was equal whether the bottom part was a face part or visual noise, even though the latency of the N170 was longer in the latter case. Altogether, these observations suggest that the face representation activated during the N170 includes information about the general vertically organized shape of the face, even if it is very poorly defined such as a face top part with a bottom part composed of visual noise. Other results from the literature are compatible with this proposal. More precisely, the general contour or layout of the face seems crucial in determining the N170 amplitude. For instance, facial features (eyes, nose mouth) presented on their own elicit a much larger N170 than when presented with a simple line contout (Shibata et al., 2002). Moreover, when global contour of the face and the hair are included in the face stimulus, removing or jumbling the inner features may have only little effect on the N170 amplitude (although the latency is delayed) (Eimer, 1998; 2000b; Zion-Gollumbic & Bentin, 2007 – but see Caharel et al., 2006; George et al., 1996; Bentin et al., 1996).

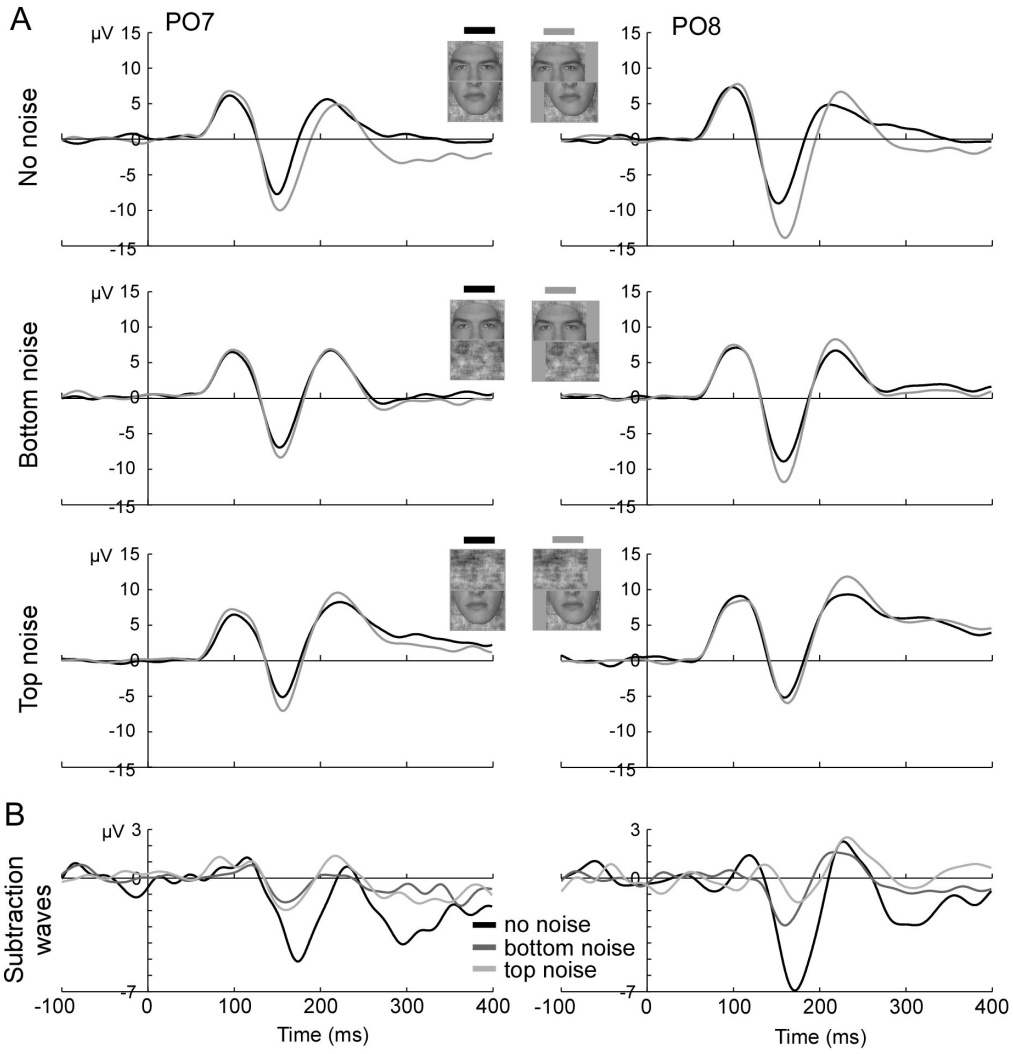

Figure 58: The N170 is strongly sensitive to the vertical alignment of the face, even when the bottom part is only made out of visual noise. That is, the N170 is strongly increased when the top and bottom parts of a face are not aligned (top row). A similar (though weaker) effect is observed when only the top part contains face information (the bottom being visual noise). These data are from one of our recent experiment not included in the empirical part of the thesis (Jacques & Rossion, in preparation).

d. Featural and configural information are integrated during the N170

We have seen in the two previous sections that there is clear evidence that both facial features and face configuration are represented during the N170 time window. Moreover, these informations seem to be integrated into a holistic representation (at least at the coarse level of the face category; the integration into an individual representation having not been tested) during this time-window. So far, attempts to isolate the effects of *featural* versus *configural* stimulus manipulations have been rather unsuccessful. In addition, there is no clear evidence to determine which of featural or configural information is processed first. Even though there is evidence that the processing of face parts and face configuration may be

distinct or are dissociable experimentally, it is more likely that these information are processed in parallel and in a strongly interactive way. This is most obvious when considering the strong interdependence between all the facial cues that make up a face; an interdependence that is subserved by holistic processing (e.g. Young et al., 1987; Tanaka & Farah, 1993). If this is the case that features and configuration are processed interactively, this may jut be part of the reason why it has proven so difficult to asses the time point at which each of these information is selectively represented based on ERP recording. Note that this question is nevertheless relevant given that some authors defend the idea that facial features are first represented and then assembled in a whole face configuration (e.g. Zion-Golumbic & Bentin, 2007; Harris & Nakayama, 2007), while other suggest that visual face representations are holistic in nature, the coding of local features taking place after or simultaneously as the coding of the whole face (e.g. Sergent, 1986; Tanaka & Farah, 1993).

3.2.3. Evidence for the time course of face categorization and the encoding individual faces.

In this section we will first examine the empirical work about the time course at which faces are detected or categorized in the visual system. This will lead us to consider the possibility that the evidence about whether a face is *present* or not in a fixated visual scene may accumulate before the point at which faces are *selectively* represented in the visual cortex (i.e. at the level of face-neurons). In a second section we will review the evidence about the temporal course at which information about individual faces starts to be represented in the visual system.

We would like to remind the reader that the questions of the temporal dynamics of face generic categorization and the encoding of an individual face representation are directly related to the experimental work of this thesis.

a. The categorization of a face as a face occurs during the N170.

In the previous sections, we have seen that there is a very large amount of evidence that faces are categorized as such during the N170 time window. These evidence are based on the fact that the N170 is strongly sensitive to the physiognomic value of stimuli; these stimuli eliciting a larger N170 than non-physiognomic stimuli. Thus, the onset time of the larger N170 to faces relative to non-face objects can be taken as the onset latency at which faces are

categorized as faces. Importantly, the term categorization is used here to refer to the fact that faces activate *specific* representations/processes.

Obviously, among the many type of information that constitute a *face,* the facial features and the global face configuration are probably the most diagnostic elements of the face category. That is, if either the configuration or the features are present –even in an incomplete way- in a stimulus will this stimulus be categorized as a face. It is therefore not surprising that, as we have seen already, both the N170 recorded in humans and the face-cells recorded in the IT cortex of the monkey are strongly sensitive to both face configuration and facial features. This therefore indicates that these elements are indeed used to detect a face in the visual environment. The fact that a face can still be detected even with poorly defined features (e.g. schematic, Mooney or Arcimboldo faces, low spatial frequency filtered faces) or configuration (e.g. jumbled inner facial features, inverted faces), is compatible with the properties of the N170 and with the view that this component reflects face categorization based on high-level (integrated) representations of facial features and configuration.

Importantly however, faces generally differ from non-face objects on a number of low-level visual properties such as spatial frequency amplitude spectrum, color distribution or organisation of local contrast. For instance, face stimuli may convey diagnostic information in different spatial frequency bands than object stimuli or visual scenes in general (Bosworth et al., 2006). In addition, we have already mentioned that photographs of human faces can be searched for efficiently among non-face objects (i.e., "pop out") partly based on the unique characteristics of spatial frequency amplitude spectrum of face images (VanRullen, 2006; see also Lewis & Edmonds, 2005). Such results strongly suggest that these low-level factors – which should typically be processed in relatively lower-level visual areas (e.g. V1, V2, V4) – may play a role in the categorization of the stimulus as a face *before* the onset of face-selective activity in the occipito-temporal cortex. The possibility that faces may actually be categorized before the N170 will be examined in the next section.

b. How early in the visual processing stream are face preferential responses observed: are faces categorized before the N170?

The question of whether faces may actually be categorized before the N170 mostly arise because some studies find that the P1 component (or the similar M1 visual component recorded in MEG), peaking around 100-110 ms and preceding the N170, is larger for faces than other non-face objects in a number of studies (e.g. Eimer, 1998; 2000c; d; Goffaux et al., 2003; Itier and Taylor, 2004a; Liu et al., 2002; Herrmann et al., 2005a; b; see Figure 59) and is larger and slightly later for inverted compared to upright faces[76] (Itier & Taylor, 2004a; c; Likenkeart-hansen, 1998; Schneider et al., 2007). However, such increased P1 amplitude for faces over objects is not consistently observed (e.g. Boutsen et al., 2006; Rossion et al., 2003a; Rousselet et al., 2005; 2007a). Moreover, most studies, even those reporting P1 effects, find a much more robust difference between faces and objects at the level of the N170 (e.g. Eimer, 1998; 2000c; d; Goffaux et al., 2003; Itier & Taylor, 2004a; c; Liu et al., 2002).

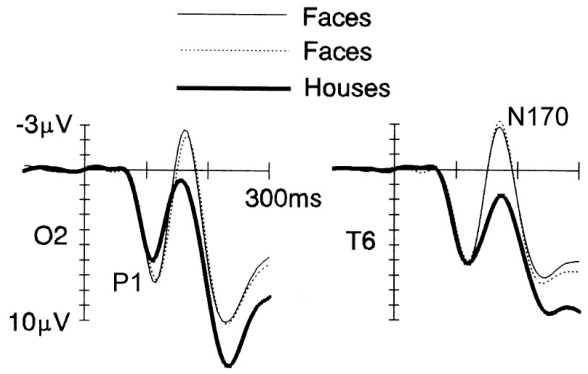

Figure 59: ERP waveform showing a larger P1 in response to face than house pictures, but only on an occipital electrode (O2, left); the effect being absent on an occipito-temporal electrode (T6, right). Modified from Eimer, 2000d.

Do P1 effects truly reflect face categorization processes? Because previous studies have indeed reported larger P1 in response to faces than to objects, it is fair to admit that there are probably electrophysiological differences between faces and non-face object categories taking place earlier than the N170 onset (around 130 ms), at the level of the P1. However, one must remain cautious in over-interpreting these P1 effects, at least for two reasons. First, as indicated earlier, the P1 is an early visual ERP thought to originate from striate and

[76]Here we will only address the larger amplitude of the P1 for faces than objects. The fact that the P1 may be larger for inverted faces will be addressed at a later point, see discussion in the article presented in chapter 7 (second article).

extrastriate visual areas (e.g. Clark et al., 1995; Di Russo et al., 2002) and is highly sensitive to low-level visual parameters such as luminance, contrast or spatial frequencies of the stimulus (see Regan, 1989). Small P1/M1 face effects, inconsistently found in the literature, may be due to differences in the global or local low-level parameters (e.g. energy in different spatial frequency bands, contrast, luminance, colour distribution,…) between stimulus categories generally not controlled in ERP and MEG studies. For instance, a MEG study showed that an M1 amplitude difference between photographs of faces and other categories was cancelled when removing surface information (color and texture) of the face stimuli (Halgren et al., 2000), prompting the authors to attribute this early difference to low-level properties (the M170 effect remained consistent). Other studies in which the faces and object stimuli were carefully controlled do not report P1 differences between categories (e.g. Rousselet et al., 2005). A second reason for remaining cautious about early P1 effects is that, as noted above, unlike N170 effects, these categorical differences at the level of the P1 are simply not consistently found.

Other evidence that the P1 is not particularly sensitive to faces *per se* come from studies in which a variable amount of visual noise is added to a face image. For instance, the P1/M1 is strongly correlated with the amount of noise in an image (Tarkiainen et al., 2002) or the noise's spatial frequency (Tanskanen et al., 2005), but is not correlated with the amount of face information in the image (Jemel et al., 2003a; Tanskanen et al., 2005; Tarkiainen et al., 2002). By contrast, in the same studies, the amplitude and latency of the N170/M170 is strongly correlated with the perception of a face stimulus.

In short, many observations indicate that fundamental face processes appear to take place primarily during the N170 time window rather than the P1.

To go back to the question we want to address in this section *(how early in the visual processing stream are faces categorized as faces?),* given the finding of a larger P1 (~ 80 ms onset, maximal at ~110 ms) for faces in some studies[77], is the N170 time-window (~ 130 ms onset) too late? The question is not an easy one to answer, and in fact it is a question that may not have a single and simple answer. For comparison, neurons recorded in the monkey inferior-temporal cortex start firing selectively to faces at around 80-90 ms (e.g. Baylis et al.,

[77] Some ERP studies have found even earlier differences between ERP conditions (usually around 50 ms after stimulus onset) under various experimental conditions (stimulus repetition, processing of face gender,…) (e.g. Mouchetan-Rostaing et al., 2000; Seek et al., 1997; George et al., 1997; Braeutigam et al., 2001). However, these effects can be explained by uncontrolled experimental designs or low-level visual differences between stimuli.

1987; Matsumoto et al., 2005; Rolls et al., 2006; Kiani et al., 2005), with the earliest and mean onset latencies being observed at around 70 and 100 ms respectively (e.g. Baylis et al., 1987; Kiani et al., 2005). Considering that these neurons will fire slightly later in the larger human brain (see Foxe and Schroeder, 2005; Schroeder et al., 1998; 2004), selective response to faces starting at around 120-130 ms, at the level of the N170 onset, seems to be a reasonable time frame for face visual categorization (with a maximal discriminability between faces and object categories at the 160-170 ms peak).

However, as already mentioned faces differ from other object categories on multiple low-level visual properties and are likely to elicit differential responses earlier in the hierarchy of cortical areas of the visual stream. Hence, in many situations there may probably be sufficient evidence accumulated in lower level visual areas to detect a face above chance level *before* the N170 onset latency. For instance, ERP and forced-choice saccadic eye movement studies indicate that categorization of animal or human faces in visual scenes can take place within 100-150 ms following stimulus onset, *including the perceptual decision* (Kirchner and Thorpe, 2006; Thorpe et al., 1996; VanRullen and Thorpe, 2001a; Thorpe et al., 2007). These response latencies suggest that visual stimuli may be, under certain circumstances, categorized reliably as faces before the N170 onset, perhaps within the time-window of the P1. However, while face categorization reflected on the N170 can be largely independent of low-level cues, rather reflecting what the observer is seeing (e.g. Bentin and Golland, 2002; Bentin et al., 2002; George et al., 2005; Jeffreys, 1996; Sagiv and Bentin, 2001), earlier effects may rather arise because of the presence of low-level cues that are statistically more frequently associated with faces. Supporting this low-level visual account for face detection, the fastest behavioral responses to detect a face in visual scenes in which the location and size of faces is largely unpredictable, are largely unaffected by stimulus vertical inversion (Rousselet et al., 2003).

More generally, the speed at which faces are categorized by the visual system, whether it is slightly before or at the N170 onset, is likely to depend on the nature of other stimuli, either presented simultaneously (in a visual scene) or interleaved in a face categorization task. That is, there is no single answer to the question of the speed at which faces can be categorized by the visual system: it depends on the rate of accumulation of evidence in multiple distributed populations of neurons in the visual system, and is dependent on the categorization context[78] (distracters, visual context, subjects expectations,…). Under certain

[78] Note that in natural viewing conditions, only the concept of information accumulation may actually be meaningful. Specifically, in natural environment, if faces are most of the time associated with a set of unique

circumstances, visual stimuli may indeed be categorized as faces before the N170 onset, but this is likely to be due to low-level properties that are statistically associated with face photographs rather than the perception of a face *per se*. Moreover, given the early onset latency of the P1 (~80 ms) as compared to the mean onset latency of face-selective neurons (100 ms in the monkey brain, Kiani et al., 2005, probably slightly later in the human brain), it is unlikely that this type of categorization based on image statistics involves 'face specific' neural processes (i.e. the processes carried out by these face neurons).

The question of the type of information that is most readily used for face detection, or more generally the type of information that makes up early face representations, is still a matter of debate (e.g. see Goffaux & Rossion, 2006; Sinha et al., 2006; Sergent, 1986).

c. How early is information about individual faces represented in the brain?

When using neuroimaging techniques such as fMRI and EGG that reveal the activity of large clusters of neurons rather than isolated neurons, there are mainly two paradigms than can be used to investigate neural correlates of face individual recognition or discrimination[79]: The first one is to look for differences between the response to familiar and to unfamiliar faces, and the second one is to look for differences between face that are presented repeatedly (either immediate or delayed repetition) compared to faces whose identity is not repeated.

---

visual characteristics and the visual system is able to accumulate and therefore make use of these characteristics in low-level visual areas to help face detection or categorization, then this accumulation should actually be considered as meaningful as categorization based on the activity from face-selective cells. The fact that these cells are able to selectively respond to faces almost as soon as they start responding is probably a consequence of this earlier accumulation of information in earlier visual areas. However, what we would like to stress here is that in these areas, the accumulation about the stimulus encountered will probably be only inferential. That is, if a non-face stimulus appears in the visual field that possess visual characteristics similar to a face, the activity accumulated in early visual areas will erroneously point toward the presence of a face. Then, it is probable that only at the level of face-neurons will this error be corrected; even though the "correction" may actually take some time to be performed. For instance, a recent study has shown that the FFA, which is considered a high-level visual area preferentially activated by faces, is more sensitive to oval patterns with more elements in the *upper* part as opposed to the same oval pattern with more elements in the *lower* part or to square stimuli with either more elements in the upper or lower part (Caldara et al., 2006). This suggests that patterns that contain some basic properties of faces (i.e. oval shape and more elements in the upper part), even though they do not strongly resemble a face, can partly trigger face-specific processes (See also Wilkinson et al. 2000 who showed that the FFA is more activated by concentric patterns than radial patterns or vertical gratings).

[79] The different terms used in face processing to refer to identity require some clarification. **Recognition** generally refers to the matching of an encountered face with a stored representation in memory. It is generaly associated with a conscious feeling of familiarity with a face, without neceseraly being able to retrive any additional information about the person. The term **identification** is used to refer to the retrieval or recollection of semantic information about a person's identity. Face **discrimination** is the fact of telling apart different face identities without a need to activate stored face representations. Discrimination can thus be performed on unfamiliar faces.

While contrasting familiar and unfamiliar faces will provide information *only* about face recognition and identification, the use of a repetition paradigm can be used to investigate face discrimination as well as recognition and identification.

Only a small number of studies have directly compared the ERP responses to familiar and unfamiliar faces. The vast majority of these studies have found that the N170 does not discriminate between unfamiliar faces and famous faces (e.g. politicians, movie stars, … ; Bentin & Deouel, 2000; Eimer, 2000d; Jemel et al., 2003b; c; 2005; Schweinberger et al., 2002b; Henson et al., 2003; Anaki et al., 2007; Caharel et al., 2005) or learned faces (Rossion et al, 1999b). In contrast to the absence of N170 difference observed in these studies, famous faces usually elicit an enhanced negativity between 300 and 500 ms (Bentin & Deouell, 2000; Eimer, 2000d; Jemel et al., 2003b; c) over central or fronto-central electrodes and an increased positivity between 500 and 700 ms over central or centro-parietal sites (Bentin & Deouell, 2000; Eimer, 2000d; Henson et al., 2003) when compared with unfamiliar faces. In these studies, the earliest effect of face familiarity -between 300 and 500 ms- was interpreted as reflecting face recognition and face identification mechanisms involved in the retrieval of semantic information about familiar faces (Bentin & Deouell, 2000; Eimer, 2000d). This interpretation was mostly derived from the similarity of this negativity with the N400 potential (with a fronto-central distribution) first described as reflecting semantic processing of words or semantic integration (Kutas & Hillyard, 1980). This classical N400 effect in language studies is elicited when a word within a sentence does not match the semantic context of the sentence and is thus thought to reflect the detection of a semantic incongruity during semantic integration. In this context, it is interesting to note that an N400 effect has also been found when familiar or learned faces contain internal structural incongruity (e.g. the face of Julia Roberts with the eyes of another person) suggesting this effect may also reflects contextual integration in long-term representation of familiar faces (Jemel et al., 1999; 2003b; Olivares et al., 1999; 2003). As we will see below, a similar N400 effect is elicited by faces that have been shown previously, usually within the same EEG recording session.

More recently, a series of studies actually found a larger N170 for personally familiar faces (subject's own face or mother's face, friends) or very famous face (French president Jacques Chirac) compared to unknown faces (Caharel et al., 2002; 2005; 2006). Similarly, Kloth et al. (Kloth et al., 2006) found a larger M170 (in MEG) to personally familiar faces (fellow university students) compared to unfamiliar faces. However, as for the findings of familiarity effects on the N170 (Caharel et al., 2002; 2005; 2006), the used of a very small

number of different stimuli repeated a large number of times prevents determining whether the observed effects reflect familiarity per se or if it rather reflects a familiarity effect that arise in interaction with a repetition effect[80]. Finally, one other study apparently found a small effect of personal familiarity on the N170 but did not further describe it, rather concentrating on later effects of familiarity (Herzmann et al., 2004).

In sum, the vast majority of studies found that the earliest effect of face familiarity occurs at around 300 ms after face onset. As already stated above, the finding that the N170 is not modulated by face familiarity has been taken as evidence that the neural processes taking place during the time-window of this component reflects face detection and are not involved in representing face at an individual level (Bentin & Deouell, 2000; Eimer, 2000d).

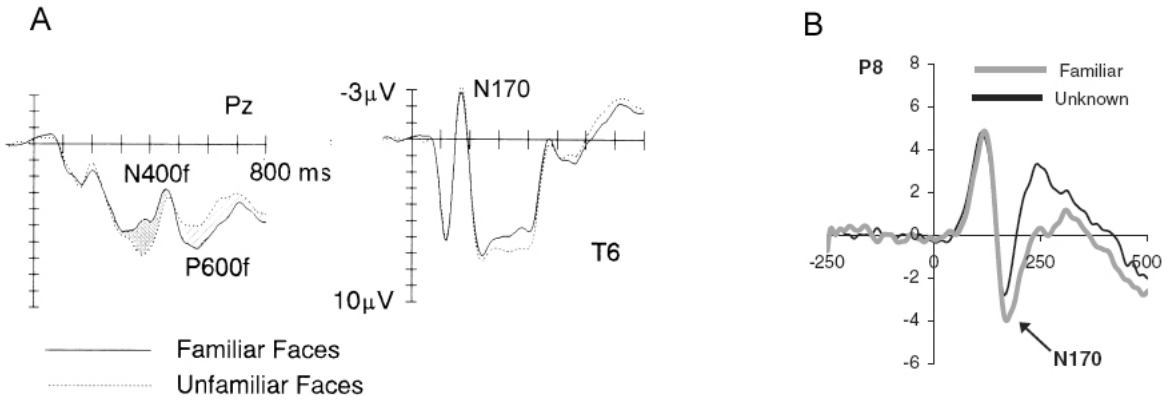

Figure 60: A. Figure showing an N170 of equal amplitude for familiar and unfamiliar faces at an inferior temporal electrode (T6), as well as an effect of familiarity in the N400 time range and between 500 and 700 ms ('P600f') on a midline parietal electrode. Note that negative polarity is plotted upward. Modified from Eimer, 2000d. B. Figure showing a difference between unknown and personally familiar faces at the N170 level on a similar electrode as in panel A. Modified from Caharel et al., 2006.

More evidence as to the time course of face identity processing has been gathered using face repetition paradigms, using either repetition after a delay during which other stimuli are presented, or immediate repetition.

[80] Recent preliminary evidence from our laboratory suggest that while the repetition (vs. non-repetition) of an unfamiliar face may lead to a reduction of the N170, no such reduction of N170 is observed for familiar faces if semantic information is associated with the familiar face. If this is the case, then, the apparently larger N170 for familiar faces in studies that use numerous faces repetition (Caharel et al., 2002; 2005; 2006) may be better described as an absence of N170 repetition effect (i.e. amplitude reduction) for familiar faces while the N170 to unfamiliar faces is reduced by repetition. Whether the familiarity effect arise due to familiarity *per se* or is due to repetition, it would remain that familiarity with faces probably modulates the activation of face representation during the N170. However, if the N170 amplitude difference between familiar and unfamiliar faces is due to repetition, this would suggest that the encoding of familiar faces during the N170 is not functionally different from the encoding of unfamiliar faces.

Most studies that have used **delayed repetition** include a phase during which faces are learned (either only visually or by association with a name and/or short semantic information such as an occupation) (e.g. Paller et al., 1999; 2000; 2003; Yovel & Paller, 2004; Joyce & Kutas, 2005; Curran & Hancock, 2007; MacKenzie & Donaldson, 2007). Most of these studies were actually inspired by earlier studies that investigated memory encoding and recollection using language material (i.e. words; Rugg & Allan, 2000). Other studies compared the ERP response to the first presentation of different faces with the ERP response to the same faces presented in a subsequent block of trial (e.g. Henson et al., 2003; Schweinberger et al., 2002b; Tanaka et al., 2006). Thus, all of these studies include a variable number of intervening face stimuli between the first and subsequent presentations.

Notably, *none* of the above cited studies that used delayed repetition of face identity reported a modulation of the N170 component when comparing repeated to unrepeated face identities. Rather, the most robust finding in these studies is that ERP response to repeated faces is more positive (or less negative) than to new/unrepeated faces during a time window ranging from around 300 to 500 ms. This effect is generally located over central or centro-parietal regions of the scalp but the location of this effect may vary across studies probably because of the different methodology used (variables manipulated, task, type of stimuli,…) and different location/computation of the reference electrode[81] (Itier & Taylor, 2002). This effect has been termed the "ERP repetition effect" or the "old/new ERP effect" (see Paller et al., 2000) and, as in studies comparing familiar to unfamiliar faces[82], many authors relate this effect to the N400 effect observed with verbal material. However, it is still hazardous at this stage of knowledge to establish a direct link between the N400 observed in language studies (either in semantic incongruity designs or in word repetition designs) and that elicited during face processing given that these effects are generally found in very different experimental setups, generating large variations in the scalp distribution and latency of this effect (see Caldara et al., 2004 for a direct comparison between N400 to faces and to words).

It is interesting to further note that this the old/new ERP effect around 400 ms after stimulus onset is generally of equal amplitude whether or not the subjects are able to recollect semantic information or name previously associated with the face, suggesting it reflects a process involved in *familiarity*-based recognition (i.e. the feeling of knowing that a person has

[81] For a component maximal over central electrode sites, the choice of the reference may strongly affect its topography.

[82] It is generally agreed that the repetition effect occurring during the 300-500 ms is probably similar to the effect observed during the same latency when comparing familiar to unfamiliar faces (e.g. Bentin & Deouel, 2000; Eimer, 2000). Familiar faces might actually be considered as "previously presented" faces, just as in the delayed repetition studies.

been encountered before), rather than the retrieval of semantic information *per se*. Supporting this view, strong N400 effect in delayed repetition designs can be observed for repeated visually learned faces for which no semantic information has been associated (e.g. Paller et al., 1999; 2000; Joyce & Kutas, 2005). The *recollection* of identity semantic information has rather been linked to a later ERP response over parietal regions (around 500 – 700 ms) that distinguish between repeated faces for which associated semantic information or name is correctly recollected from repeated faces for which the subjects did not recollect this information (Curran & Hancock, 2007; see also MacKenzie & Donaldson, 2007). Note that this dissociation has been more robustly described in studies using verbal material (Curran et al., 2006; see Rugg & Allan, 2000 for a review).

In any event, both the "N400" effect and the later (500-700 ms) "recollection effect" are most likely related to recognition processes that are *post*-perceptual and not visually (or modality) specific. Indeed, these effects can be observed irrespective of whether stimuli are presented in the visual (e.g. faces, objects, words) or in the auditory (words) modality. Moreover, the N400 effect is similar when stimuli presented during the study phase (e.g. lists of words) are presented in the visual or auditory modality (Curran & Dien, 2003).

As a final note, several evidence from ERP source localisation, intracranial recordings and fMRI, suggest that a large network of brain regions located in the frontal cortex (Caldara et al., 2004) and medial temporal lobe (Johnson & Hamm, 2000) participate in generating the N400 effects elicited by verbal or face/objects repetition. Interestingly, recent single neuron recordings in the human medial temporal lobe (amygdale, hippocampus, entorhinal cortex) revealed the existence of neurons that selectively respond to many different instances of a given familiar face (e.g. Bill Clinton; Jennifer Aniston) including printed name of the person (Quiroga et al., 2005). These neurons generally fire between 300 and 600 ms, compatible with the timing of the N400 component. While, these neurons may be taken as evidence for the existence of grand-mother cells responding only to a single individual (Konorski, 1967; Barlow, 1972), computation of the theoretical sparseness[83] of the representation (Waydo et al., 2006), rather suggest each of these neuron is involved in coding several individual and that each individual is represented by a small group of neurons (i.e. a sparse representation).

[83] The sparseness can be roughly defined as the number of stimuli that a given neuron responds to. It can also be seen as the number of neurons that are activated (i.e. code) by a given stimulus (see Foldiak, 2002). Sparseness thus defines how much a representation is distributed or not and can be quantified by a "sparseness coefficient", ranging from 1 (fully distributed representation – all neurons respond to all stimuli) to 0 (the representation is local – each neuron only codes for a single stimulus).

A larger number of ERP studies have investigated individual face repetition effects using **immediate repetition**. That is, these studies compared the ERP response to a face immediately preceded by the same face (either the same photo or a different photo of the same person) or by another face (i.e. a different identity). In these studies, the effect of different factors on ERP repetition effects were usually tested, such as *face familiarity* (Jemel et al., 2003b; c; 2005; Bentin & McCarthy, 1994), *masking and prime-to-target stimulus onset asynchrony* (Martens et al., 2006; Trenner et al., 2004), *face inversion* (Itier & Taylor, 2002; 2004c; Schweinberger et al., 2004); *luminance contrast reversal* (Itier et al., 2002; 2004c), *task* (Martens et al., 2006; Trenner et al., 2004; Huddy et al., 2003), *repetition of the same vs. different pictures* (Shweinberger et al., 1995; 2002a; Jemel et al., 2003c), *prime and target in the same vs. different visual field positions* (Heisz et al., 2006a; b), *number of face repetition* (Itier & Taylor, 2004; Heisz et al., 2006a; b), *within- vs. cross-domain priming* (Jemel et al., 2005).

In the pool of studies that used immediate face repetition, some studies found that the N170 is reduced in response to a face preceded by a face of the same identity compared to a face preceded by a face of a different identity (Jemel et al., 2003c; 2005; Martens et al., 2006; Trener et al., 2004; Guillaume & Tibergien, 2001; Heisz et al., 2006a; b; Campanella et al., 2000; Itier & Taylor, 2002; 2004c), while other studies did not find any difference between repeated and unrepeated faces at the N170 level (Zhang et al., 2001; Schweinberger et al., 1995; 2002a; 2004; Huddy et al., 2003; Bentin & MacCarthy, 1994; Mnatsakanian & Tarkka, 2004; Jemel et al., 2003b; Potter & Parker, 1997).

One of the challenges to interpret these data is that there do not seem to be many commonalities between studies that did or did not find individual face repetition effects at the N170. That is, of all the above mentioned factors that have been manipulated, the only one that more-or-less distinguish between the presence vs. absence of an effect is the use of the same vs. different pictures of the same person (different view point, illumination, expression,…) in "repetition" trials. More precisely, most studies that used fairly different pictures for repeated trials did not find a repetition effect at the N170 (Schweinberger et al., 1995; 2002a; Huddy et al., 2003; Mnatsakanian & Tarkka, 2004; Jemel et al., 2003b; c). However, this view is complicated by the fact that some of these latter studies that compared the repetition of same vs. different pictures did not find a repetition effect even when repeating the very same pictures (Schweinberger et al., 1995; 2002a; Jemel et al., 2003b). Moreover, other studies that only use the same face picture for repetition did not find an effect at the N170 either (Zang et al., 2001; Bentin & MacCarthy, 1994).

A further difficulty comes from the fact that there are also inconsistencies among the studies that did find that the N170 is reduced for repeated vs. unrepeated individual faces. For instance, Martens et al. (2006) found a repetition effect at the N170 for masked and unmasked presentation of the prime in a task where subjects had to judge whether the target face (i.e. the second face of the trial) was a singer or a politician, whereas no such effect was found when subjects performed a matching task between prime and target face. In the very same setting, the same group (Trenner et al., 2004) mostly did not find any N170 repetition effect, or these were confounded with low-level visual effects (see next paragraph). In another study, Itier & Taylor, (2002) found that the N170 is reduced for an immediately repeated face, but not when the repetition occurs after one intervening stimulus. In a related study by the same authors (Itier & Taylor, 2004c), the N170 was very slightly reduced to repeated faces but this time only when the same face was repeated more than 6 times. Moreover, in these two studies, the repetition effect at the N170 was observed whether the faces were presented upright, inverted or in a contrast-reversed format, even though these manipulation are though to strongly affect *perceptual* encoding of faces (i.e. those that should take place at the N170) (e.g. Freire et al., 2000; Tanaka & Farah, 1993). These are just a few examples, but they illustrate well the lack of consistency in the repetition effects manifested at the N170.

Even though an individual face repetition effect at a given time point may be indicative that the neural processes at work at this time point are involved in encoding an individual representation, other factors not reflecting identity coding per se and that may lead to such a repetition effect have also to be controlled for, or least taken into account. Specifically, in a least some studies that used the very same face pictures for repetition, one cannot fully determine whether the N170 repetition effect found is truly related to the encoding of an individual face representation per se or if rather reflects a mere low-level repetition effect that occurs because the very same picture is repeated in a relatively short interval (i.e. it may thus not be the identity that is coded but rather the distribution of the pixels in the image). This is particularly true for studies that actually found repetition effects at the N170 but also much earlier (Trener et al., 2004; Martens et al., 2006[84]; Guillaume & Tiberghien, 2001; see also Heisz et al., 2006a; b), at the P1 level (~100 ms) which mostly reflect low-level visual processing and not yet the encoding of an individual face representation. In addition, as noted above, the fact that neither face inversion nor contrast-

[84] In Trenner et al. (2004) and Martens et al. (2006)'s studies, the early (~100 ms) and N170 repetition effects more strongly occur when very short SOA (67 ms) between prime and target is used. The authors therefore suggested early (and maybe N170) effects may reflect a repetition effect occurring at the level of a "fast decaying iconic memory".

reversal reduced the identity repetition effect in Itier & Taylor (2002; 2004c), may also suggest these repetition effects partly involve lower-level effects. In these two studies, an interaction between identity repetition and the stimulus factor (i.e. reduced repetition effect for inverted and/or contrast-reversed faces compared to upright) was only found on later repetition effects (see also Schweinberger et al., 2004).

It thus largely remains to be determined whether these repetition effects observed on the N170 reflect the *perception* of facial identity or whether it only reflects lower-level visual processing of image attributes. If the former interpretation holds true, the type of face neural representations that subserve this effect remains to be investigated. Observations from previous studies indicate these representations are visual, fairly image-based, probably fast decaying but also entailing rapid learning (present for both familiar and unfamiliar). Note that a number of experiments presented in the empirical part of this thesis specifically aimed at clarifying the time course with which individual face representations are coded, through the use of repetition effects.

In addition to the rather inconsistent N170 repetition effect found in the literature, a much more robust individual face repetition effect is usually observed starting at around 220-250 ms after stimulus onset, thus in between the N170 between and the N400-like effect observed in delayed repetition and familiarity studies. This repetition effect generally arises in the form of a more negative ERP for repeated compared to unrepeated faces around 250 ms over temporal-parietal region, and has accordingly been termed the "N250" or "N250r" (Begleiter et al., 1995; Schweinberger et al., 1995; 2004). The N250 is sensitive to immediate (Schweinberger et al., 1995; 2002a; 2004; Itier & Taylor, 2004c; Trener et al., 2004, Martens et al., 2006; Herzmann et al., 2004) and short lag face identity repetitions (Pfütze et al., 2002; Itier & Taylor, 2004c), but not longer lag repetition (Schweinberger et al., 2002a; Huddy et al., 2003; Tanaka et al., 2006; see however Joyce & Kutas, 2005 that did find an N250 repetition effect after more than 1 hour) unless the face is repeated a large number of times or is highly familiar (Tanaka et al., 2006). Because the N250 repetition effect is larger for familiar faces (Herzmann et al., 2004; Schweinberger et al., 1995; Tanaka et al., 2006) and when the task requires the matching of identity between prime and target face (Trener et al., 2004; Martens et al., 2006) or the detection of a specific identity among other non-target faces (Tanaka et al., 2006), it has been suggested to reflect the activation of stored individual face representations in areas around the fusiform gyrus (Schweinberger et al., 2002a). Given the sensitivity of the N250 to the lag or the presence of intervening stimuli between prime and

target face and the fact that it is larger and earlier when the same picture is repeated compared to different pictures of the same person (Schweinberger et al., 2002a), this effect probably involves, at least partly, modality specific visual representations. This would contrast with the repetition and familiarity effects taking place at later time points (e.g. N400) that reflect more abstract modality-general representations and are mostly insensitive to the format of picture repetition (Schweinberger et al., 2002a; Curran & Hancock, 2007). Finally, even though the N250 was first suggested to reflect a face specific activation -being absent or reduced for repetition of cars and inverted faces (Schweinberger et al., 2004; Itier & Taylor, 2004c) -, recent evidence clearly showed that repetition of non-face objects for which the observer has acquired visual expertise elicit strong effects at the N250 (Scott et al., 2006).

Apart from the N170 and N250 effects, most of immediate repetition studies also find later repetition effects both at the central N400 and later (>500 ms) parietal effect similar, to what is found in delayed repetition and familiarity studies. However, as noted above, unlike these effects, the N170 and perhaps the N250 effects take place at the level of visual representations in visual cortex.

To conclude this section, focusing on repetition effects taking place at the level of visual representations (i.e. N170 and N250), what information can we derive from this review of the literature about the question we had in the first place: *how early is information about individual faces represented in the brain?*

Firstly, the lack of consistent familiarity or individual face repetition effects at the N170 has prompted many researchers to define the N170 as reflecting a mere "face detection stage", reflecting "*the operation of a neural mechanism tuned to detect (as opposed to identify) human faces, similar to the 'structural encoder' suggested by Bruce and Young (1986).*" (Bentin et al., 1996, pp. 551). In addition, the literature using delayed individual face repetition also found an absence of repetition effect during the N170 time window. Last, immediate repetition paradigms have revealed rather mixed results as to whether or not the ERP response during the N170 is sensitive to individual face repetition. Moreover, in at least some cases, the N170 repetition effect found may potentially be confounded with lower-level effects.

Secondly, much more consistent individual face repetition or familiarity effects are observed when later (>250 ms) ERP responses are considered. However, this timing does not fit with the known temporal dynamics of identity coding in non-human primate brain. Indeed, given the probable onset time of face-neurons in humans (i.e. around the onset of the N170, at

130 ms), these results would suggest that the activation of an individual face representation is delayed by more than 100 ms relative to face detection. This view is clearly incompatible with the finding that face-neurons in the monkey brain start to accumulate information about both face category and face identity simultaneously (Tovee & Rolls, 1995), with identity information being significantly represented in the neurons' responses not more than 40-50 ms after information about the global category (Sugase et al., 1999; Matsumoto et al., 2005). These "late" repetition effects might not therefore reflect the earliest visual encoding of individual faces in the human visual system. As we will try to demonstrate in the experimental work of the present dissertation, the time-window of the N170 may be a more compatible time-frame during which both face detection and the encoding of individual representations take place in the occipito-temporal cortex.

Finally, although we have been mostly referring to distinct ERP components in our review of the ERP literature on face processing, we would like to reinstate that ERP components should not be interpreted as discreet information processing stages in the brain. Therefore, analyses that are restricted to measurements of specific ERP components can only lead to confirmation of such an interpretation. As we will see in the experimental work, performing analyses on the entire ERP response within a given time window (e.g. from stimulus onset to 400 ms after stimulus onset), rather than on specific components, allows identifying the exact time point at which the effect of stimulus or task manipulations arise. These analyses may generally reveal significant differences between specific ERP components, which would have been missed by restricting analyses to ERP components.

3.3. Summary and conclusions

In this section on electrophysiological studies of face perception, we have reviewed a large amount of data- mostly about the N170- that we will here briefly summarize in order to derive a clearer picture of what information about face may be processed during the time-window of the N70.

First, the N170 is an ERP component taking place in the 130-200 ms after stimulus onset over inferior occipito-temporal regions of the scalp and has been described in numerous studies as larger in amplitude for faces compared to non-face object categories. The N170 is temporally coincident with a positivity (VPP) observed over central scalp regions and

described in earlier studies. These two components show identical response properties (Joyce and Rossion, 2005), in particular the larger response to faces, indicating that the VPP reflects the positive counterpart of the equivalent dipoles generating the N170 in the occipito-temporal cortex. The N170 is maximal in the the 5- to 15-Hz frequency band (Rousselet et al., 2007a).

Second, the larger N170 for faces is probably due to the recruitment of additional functionally defined areas in the visual cortex showing preferential response to face stimuli. These areas and the N170 do not only discriminate between faces and non-faces, but can also show some sensitivity to non-face categories (Rossion et al., 2000; Grill-spector et al., 2006; Avidan et al., 2002), especially for categories for which the observer is an expert (Tanaka & Curran, 2001; Rossion et al., 2002; 2004; Gauthier et al., 2003; Busey & Vanderkolk, 2005; Scott et al., 2006; Gauthier et al., 1999).

Third, the neural processes underlying the N170 are more implicated in high level visual categorization and can be largely independent of the low-level visual properties of the input stimulus. The N170 is thus more correlated with the *perception* of a face, and stimulus manipulations that reduce the ability to perceive the face (e.g. adding visual noise) may have large influence on this component.

Fourth, both facial features and face configuration (i.e. first order relation, global shape, shape-from-shading) are represented during the N170 time window. There is also evidence that these facial cues are integrated in a holistic template that can be used for face detection during this time-window. However, the type of cues that are initially used for face categorization and that constitute the "earliest" building blocks of the face representation are currently unknown.

Fifth, there is considerable evidence that faces are categorized as such during the N170 time-window. However, in many cases, information about faces may start to accumulate before the N170, mostly based on lower level visual cues statistically associated with the face category, such as color distribution, spatial frequency amplitude spectrum or general "ovalness" of the shape. In addition, while some authors defend the view that the neural processes taking place at this level only reflects face categorization/detection, some inconsistent evidence suggest information about face identity as well may be present during this time-window.

Last, and to conclude, multiple functional face processes take place during the N170 time-window and, as such, this component constitutes a critical time-window to study the time course of visual encoding of faces.

II. Contribution empirique

Chapter 4. Presentation of experimental work

In the first two chapters we have examined current evidence about the time course of face processing. Behavioural studies using RSVP or masking paradigms indicate that faces can be detected and recognized (or discriminated from each other) very rapidly and with a minimal amount of visual information. These proposals are entirely supported by data collected by single-neuron recordings in the monkey brain. These studies on the monkey have precisely temporally located the time course at which information about faces starts to accumulate; that is at around 80-100 ms after stimulus onset, both at a coarse *categorical* level and at a finer *individual* level. These studies also indicate that these information about face may accumulate at different rate, thus generating a decision about the category faster than a decision about face identity. However, as for humans, even though behavioural experiments reviewed in Chapter 2 provide some clues as to the time needed to recognize or discriminate individual faces, the question of the precise time course of early visual face processing (within the first 200-300 ms after stimulus onset) is still largely unclear.

The general goal of the experimental work presented next was therefore to attempt to clarify the general speed at which coarser and finer information about faces is represented in the human visual cortex.

In the remainder of the present Chapter, we briefly present and motivate each experiment that are included in our empirical work for the present dissertation.

Concurrent processing reveals competition between visual representations of faces (Chapter 5)

We have seen that single-cell recordings in inferior temporal cortex have suggested that faces are coded by overlapping populations of face-selective neurons; each neuron modulating its firing rate to more than one face identity. In this first experiment we thus reasoned that if the face-N170 originates from the activity of these overlapping populations of face-neurons coding for different individual faces, it should be reduced in amplitude in response to a face stimulus, when another face stimulus is being processed concurrently. More precisely, we hypothesized that if a face appears in the visual field while another face is already being visually processed, these face would compete for neural representation given

their partly overlapping representations. If this is the case, the N170 to the second face stimulation should be reduced in amplitude, because the neurons involved in processing the first face would be only partially available to encode the second face. To test this hypothesis, we measured the scalp ERP elicited by a face appearing in the near peripheral visual field, while the subject was already visually fixating another stimulus; either another face or a control meaningless stimulus.

By revealing interactions between visual face representations in the cortical regions that participate in generating the N170, this experiment indirectly provides information about the time-point at which face-selective neurons start to respond in the occipito-temporal cortex, encoding information about stimulus category (i.e. face) and face identity.

The speed of individual face categorization (Chapter 6)

We have seen that the ERP literature on face identity processing generally indicates that the identity of a face is encoded in the visual system at latencies around 250 ms at the level of the N250 component. Although there is some evidence from immediate repetition studies that information about face identity may be present during the N170 time-window, these evidence are not only weak and sometimes potentially confounded with low-level visual effects, but considerable inconsistencies are also found across studies, some finding a repetition effect at the N170 level, some not finding any effect. However, the finding that information about face identity arise around 230-250 ms in humans – whereas the face-selective neurons probably start to respond at around 120-130 ms at the onset of the N170 – is at odd with the fact that face-neurons in the monkey's IT convey significant information about face identity within the first 20 or 50 ms after response onset. Thus, given that the mean response time of these neurons in the monkey is about 100 ms, it would be reasonable to hypothesize than in human, information about face identity should be reflected in the neural response within a 120-200 ms time-window. Given the apparent inadequacy of the face repetition method to reveal consistent identity processing within the N170 time window, in our second experiment we used an original stimulus presentation method (based on continuous stimulus presentation,) to further qualify the time course of face identity coding in the human brain.

Using the face inversion effect to assess the time course of face categorization and discrimination (Chapter 7)

When pictures of faces are presented upside down (i.e. vertical inversion), their recognition at the individual level is dramatically impaired, in particular when compared with the same transformation applied to other visual object categories. In addition, there is a large body of evidence that suggest that face inversion also impairs and delays the detection or the categorization of a face, for example when encountered in visual scenes. Most interestingly, this very simple stimulus manipulation strongly impairing face categorization and face recognition while preserving most of the global low-level visual properties of the face image. It is therefore not surprising that face inversion has been used as a invaluable tool to investigate many functional aspects of "normal" high-level face processing.

In this chapter, we report three original studies in which we used the face inversion effect to try to precisely characterize the time course of face categorization and face identity discrimination using ERP. More precisely, we have attempted to relate the effects that vertical inversion has on face categorization and face individual discrimination, with the modulations of the ERP response induced by this stimulus manipulation.

An additional goal for these studies was to clarify the time course of the inversion effect as revealed through electrophysiological responses.

1. **The time course of the face inversion effect during individual face discrimination.**

In this third study we investigated the time course with which face inversion influences the coding of facial identity. We addressed this question by using an adaptation paradigm in ERPs. More specifically we measured the ERP response to a test face that was preceded by an adapting face presented for a duration of 3 seconds. The identity of the test face could be either the same identity or a different identity as the adapting face. Adapting and test faces were either presented in an upright or in an inverted orientation. The use of a long presentation time of the adapting face (3 secs) was inspired by a recent ERP study (Kovacs et al., 2005) in which the adapting stimulus was presented for 5 seconds and that found consistent face adaptation effects on the N170 at the category (not identity) level. Indeed, one possible reason for the lack of consistent face identity repetition effect at the N170 level in previous studies is the use of short stimulus presentation time (typically < 1 s).

By comparing the ERP responses to the test face when it was the same vs. a different identity as the adapting face, we could precisely locate the onset time at which information about face identity is present in the ERP response. The originality of this experiment is that we defined the onset time of the effect of inversion on the coding of facial identity as the earliest time point at which an ERP identity adaptation effect was observed for upright but not for inverted faces.

2. **Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans.**

This fourth experiment aimed at further investigating how face inversion affects face visual processing using ERP recordings. Here, rather than contrasting only upright and inverted faces as in previous studies, we recorded ERPs in response to pictures of faces presented at 12 different orientations from 0 to 360 degrees in 30 degrees steps, while subjects performed a face identity matching task. Presenting face stimuli at multiple orientations allowed us to precisely characterize the patterns of modulation of the P1 and N170 parameters (i.e. amplitude and latency) as well as the pattern of modulation of behavioral responses as a function of face orientation. Moreover, in order to find brain responses related to our face discrimination task, we identified the time points at which the pattern of ERP responses to face stimuli at multiple orientations was correlated with the behavioral performance at the individual face matching task. To achieve this we correlated the ERP signal amplitude across face orientations with behavioral data at each time point of the ERP response and each electrode.

3. **Dissociating the time course of face categorization and the coding of individual faces.**

In the two previous experiments presented (1. and 2.), we have used a manipulation of face vertical orientation to derive information about the time course of face categorization and face identity coding in the visual system. In the first on these experiments we used an identity adaptation design and in the second experiment we manipulated face orientation parametrically. Altogether, these studies support the view that the visual system accumulates category-level information about faces faster than information about face identity. More specifically, while faces may be first categorized at the onset of the N170, information about

face identity is significantly represented in the ERP response at a slightly later time-point but still during the N170 time-window.

In the present experiment, we aimed at replicating the findings of the two previous experiments by combining the two methodologies, namely an identity adaptation design and a parametric manipulation of face orientation. That is, we wanted to replicate these findings in a single experiment, which would make a stronger case for this fine temporal dissociation between face categorization and face identity encoding. Note that the results of this experiment are largely preliminary.

Individual faces are represented holistically during the N170 component (Chapter 8)

It is generally acknowledge that facial features are not perceived independently from each other but are rather integrated into a gestalt, or a so-called holistic representation (Galton, 1883). Although it is widely assumed that the integration of a face into a holistic representation occurs at a perceptual level, that is, as the face is perceptually encoded (Farah et al., 1998; Hole, 1994; Le Grand et al., 2004), there is no direct evidence for this yet.

Therefore, this experiment aimed at directly investigate the time course of holistic individual face encoding using one of the most commonly used paradigm to investigate holistic face processing, namely the "face composite effect" (Young et al., 1987). The face composite effect is a strong perceptual illusion in which two identical top parts of a face are perceived as being different if their respective bottom parts belong to two different identities. Notably, the perceptual illusion disappears if the top and bottom parts are laterally offset, so that they appeared misaligned.

Here we used a slightly modified version of the composite task in an ERP adaptation experiment to record both electrophysiological and behavioural measures of the face composite effect.

Chapter 5

Concurrent processing reveals competition between visual representations of faces

1. Paper 1:

Concurrent processing reveals competition between visual representations of faces

Corentin Jacques & Bruno Rossion

ABSTRACT

Scalp electrophysiological recordings in humans indicate that the processing of faces differs from other categories between 100 and 200 ms after stimulus onset, peaking at the N170. To test the effect of the addition of a second face stimulus in the visual field on the face-related N170, we measured this component in response to a laterally presented face stimulus while subjects were processing either a central face or a control stimulus. As early as 130 ms following the lateralized face stimulus, there was a strong (~40% of signal) and specific reduction of the N170 amplitude when subjects were concurrently processing a central face. This observation suggests that the early stages of the N170 reflect the activation of individual faces having overlapping and competing neural representations in the human occipito-temporal cortex.

For Copyright reasons, only the abstract of this published article is reproduced here.

2. Complementary results and discussion

Response to first stimulus: faces versus Fourier randomized faces

In this and subsequent articles using a concurrent presentation paradigm (Jacques & Rossion, 2006a; 2007a), we only presented figures and results performed on the ERP response to the second face because only this response is relevant when analyzing competition effects. However, when being interested in the time course of face visual processing in a more general way, as in this thesis, the ERP response to the first stimulus (face vs. Fourier phase-randomized face) may provide some additional interesting information. Indeed, since these stimuli are similar in terms of most global low-level visual properties[85], except for the presence of a structured shape information, the comparison of the ERP response to face and Fourier scrambled faces may provide an information as to when structured shape information starts to be coded in the human visual system. There are actually very few report of comparison between faces and Fourier scrambled faces in the ERP literature (e.g. Taylor et al., 2001; Rousselet et al., 2007a; in MEG: Tanskanen et al., 2007). Furthermore, the only study that precisely measured the onset time of the difference between the ERP to these stimuli presented data from a few individual subjects and used grey-scale degraded versions of face stimuli which may yield an overestimated time point for a differential ERP signal to these stimuli.

The comparison between the ERP to face and face noise is shown in Figure 61 for 2 posterior electrodes in each hemisphere. Strikingly, while we have seen earlier that the ERP to face and non-face objects usually starts after the P1 at around 130 ms (when low-level properties between categories are controlled), the ERP waveforms to faces and scrambled faces starts to separate at 100 ms, at the P1 level. Importantly this effect cannot be attributed to differences in the global low-level properties between these two types of stimuli, unlike the P1 category effect found in some ERP studies. Rather, this indicates that face stimuli are segregated from scrambled face at an earlier level in the visual hierarchy than non-face objects. That is, faces and scrambled faces probably generate an equally strong response in early visual areas at a global level, as is observed in Neuroimaging studies (Grill-spector et al., 2001). However, when the visual activity related to stimulus onset reaches a cortical level that shows sensitivity to simple or more complex shapes, such as V4 or areas in the lateral

[85] The phase randomization however randomly redistributes the phase information across spatial frequencies, which may slightly alter the spatial frequency amplitude spectrum.

occipital complex (LOC; e.g. Grill-spector et al., 2001), the global neural response from these areas is likely to be higher for stimuli with structured shapes such as faces and objects compared to visual noise. In contrast, faces and non-face objects, when controlled for low-level visual properties, probably generate more-or-less an equal global level of activity even in visual areas that differentiate between face and noise stimuli. Therefore, faces and non-face objects are likely to be segregated at a slightly later time point, when reaching areas that contain large proportions of neurons selective for faces at around 130 ms, compatible with the timing of our competition effect. Indeed, it is probable that it is only when these areas start to be activated by a visual input, that higher ERP amplitude for faces (during the N170) is observed.

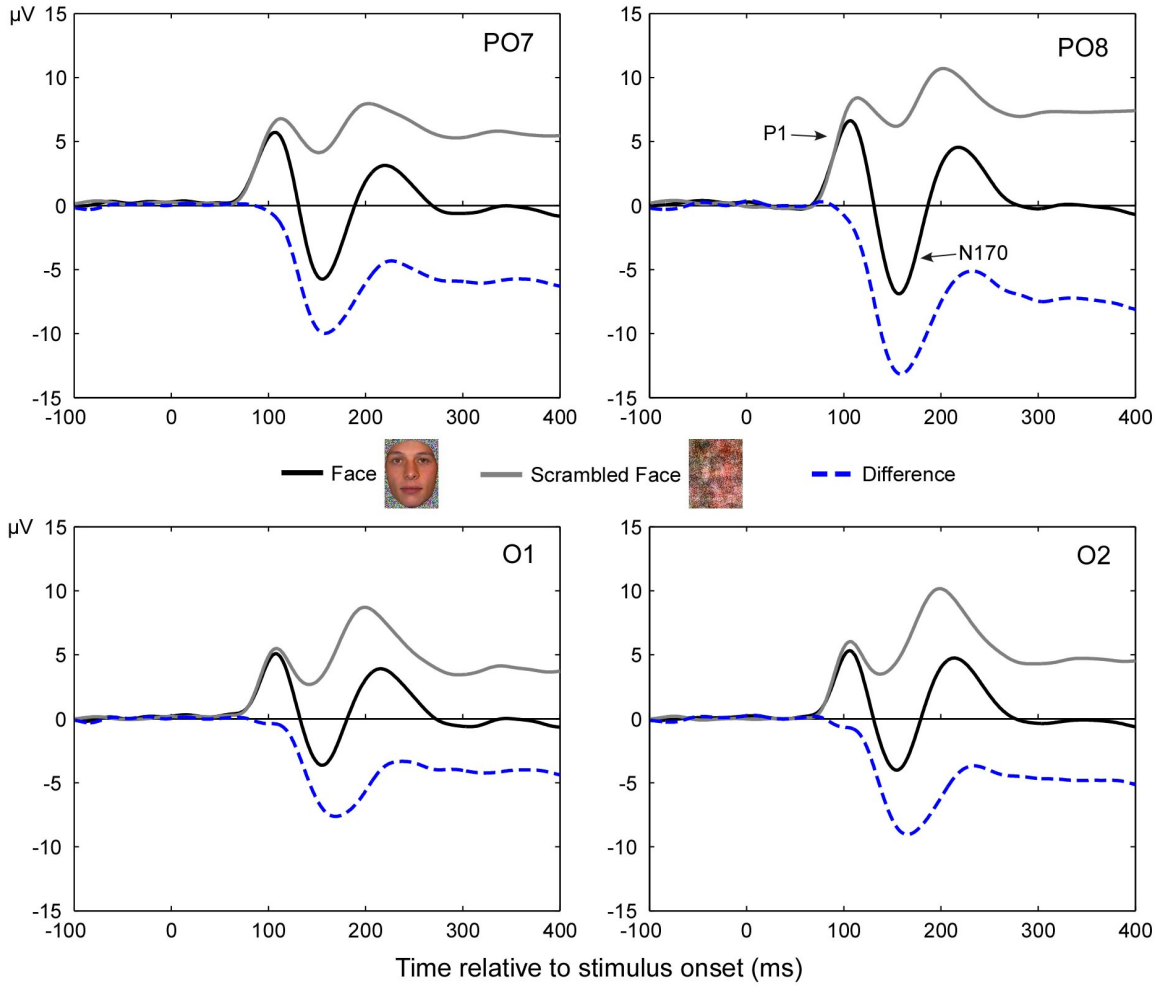

Figure 61: Grand average ERP response to the first stimulus (face or Fourier scrambled face presented at fixation) in the concurrent processing paradigm (Jacques & Rossion, 2004; 2007a) at two posterior (occipital and occipito-temporal) electrodes in each hemisphere.

Dissociating spatial attention from competition effects

In this concurrent processing paradigm, both with faces or object of expertise (Rossion et al., 2004; 2007) presented at fixation, there is at least one alternative explanation for this reduction of the N170 elicited by face presented peripherally. That is, subjects may allocate less attentional resources to the visual field location where the lateralized face stimulus appears, when another behaviorally relevant stimulus such as a face (or a car picture in car experts) is present at fixation. In other words, the reduction of the N170 recorded to a face stimulus may result from a general spatially based attentional modulation rather than from an underlying competition taking place between neural representations of faces, or of faces and objects of expertise.

In order to address this issue we conducted an experiment aiming at disentangling the effect of sensory competition between concurrently presented stimuli and the putative visual spatial attention effects that may act on the processing of faces presented in the periphery (Jacques & Rossion, 2007a). To do so, we manipulated both sensory stimulation and attentional load on the fixated stimulus in the same EEG recording session. Scalp ERPs elicited by the onset of a lateralized face were recorded while subjects were fixating a central stimulus, either a face or a control Fourier scrambled face. Attentional load at fixation was varied using two different tasks that were performed in different blocks, on the exact same stimulus sequence. The first task was the same as in previous experiments (Jacques and Rossion, 2004; Rossion et al., 2007; Rossion et al., 2004), and was considered to involve "low-attentional load" on the central fixated stimulus. The second task ("high-attentional load") was designed to increase and control the level of attention allocated to the central fixated stimulus.

In short, the same subjects were submitted to the exact same stimuli while the factors spatial attention (high load vs. low load at center) and sensory competition (face vs. scrambled-face at fixation) were crossed in a 2 x 2 design.

The results of this study can be briefly summarized as follow (see Figure 62):

1. The N170 to lateralized face was reduced starting at 130 ms when subjects were fixating another face relative to a Fourier scrambled face, replicating previous results.

2. The N170 to the lateralized face was reduced when subjects were engaged in the high-load task at fixation compared to the low-load task, revealing that spatial attention may strongly affects face visual processing.

3. The effects of attention and competition on the N170 were mostly orthogonal to each other, supporting the view that the N170 decrease for lateralized faces presented in the context of another face stimulus is independent of spatial attention.

4. Whereas sensory competition effects started at about 130 ms after stimulus onset, spatial attention modulated visual processes as early as 80 ms, in keeping with the time course of spatial attention effects on visual ERPs (Luck et al., 2000; Mangun, 1995).

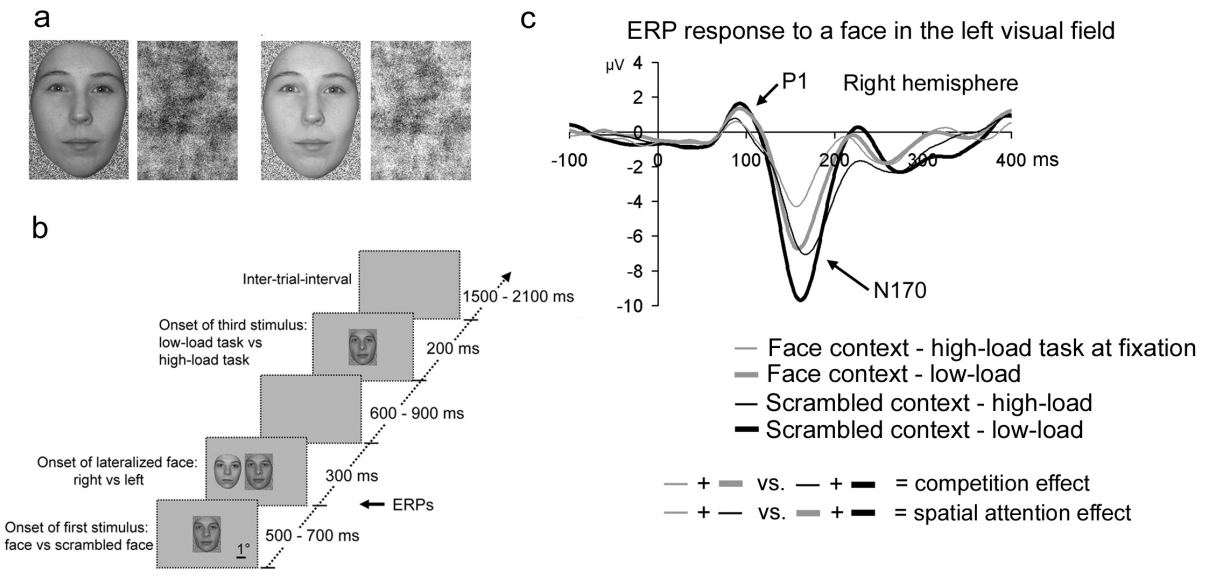

Figure 62: Dissociating spatial attention from competition effects (modified from Jacques & Rossion, 2007a). (a) Examples of faces and Fourier scrambled faces used in the experiment. In the high-load task at fixation, subjects had to match the luminance of the first and third stimulus of each trial (see (b)). Examples of "low" and "high" luminance stimuli are shown on the left and right respectively. (b) Time-line of the stimulus sequence used in the experiment. (c) Grand-average ERP recorded on a right hemisphere occipito-temporal electrode (PO8) in response to a face presented in the left visual field. ERP waveforms are presented for each of the 4 conditions (2 'stimulus at fixation' x 2 'task at fixation'). Note the competition effects starts at 130 ms and is independent of the spatial attention effect. The spatial ettention effect starts at 80 ms.

Thus, these results confirm that the previously observed N170 reduction to a laterally presented face when subject are fixating another face or an object of expertise is not due to spatial attention being differentially allocated to peripheral visual field in the different "competition" conditions (e.g. faces vs. scrambled faces). They also provide strong evidence that the N170 in response to faces is modulated by spatial attention, and that spatial attention and sensory competition do not reflect the same mechanisms of early selection of visual information in the extrastriate cortex.

Competition effects in foveal versus peripheral presentation

In this paradigm the first stimulus appears at fixation and the competition effect is measured in the ERP response to the second stimulus that appears in the peripheral visual field. It may therefore, be that presenting the second face in the peripheral visual field has delayed or biased the competition effect compared to the case where the second face would appear at the fovea. In an additional experiment (Jacques & Rossion, 2006a) we tested the effect of having the second face appearing at the fovea while the first stimuli (face or scrambled face) appeared in the peripheral visual field (Figure 63A). In this experiment we replicated the reduction of the N170 starting at around 130 ms when the first stimuli were faces compared to Fourier scrambled faces (Figure 63B). However, in these conditions, the magnitude of the competition effect was about half the size of the competition effect measured when the second face appears at the fovea (Figure 63C). This suggests that a stimulus that appears at the fovea has a competing advantage (i.e. is processed preferentially) over stimuli appearing at the periphery[86]. This bias in the competition effect as a function of the position of the stimuli is likely to be due to differences in the cortical magnification factor between the fovea and the peripheral visual field. More precisely, the competing advantage for stimuli appearing at the fovea may be due to the larger proportion of cortical surface dedicated to process visual information appearing at the fovea (See Rousselet et al., 2005; Jacques & Rossion, 2006a; Levy et al., 2001; Hasson et al., 2002 for additional discussion).

Complementary discussion

In this experiment we suggest that when two faces are presented concurrently, they activate visual representations that compete for neural representation in the occipito-temporal cortex. This competition probably arise either because of partly overlapping representations, or because the onset of the first face inhibits other neurons by means of lateral inhibitory connections.

We therefore suggest that the earliest time point at which we observed an ERP face competition effect in this concurrent processing experiment (~125-130 ms), reveals the time point at which face selective neurons are activated in the human occipito-temporal cortex. At this point we have to admit however that it may be argued that the measured latency of this ERP face competition effect does not truly reflect the time at which face neurons are activated

[86] Note that this is probably not true for moving stimuli to which the peripheral visual field is particularly sensitive

in the occipito-temporal cortex (i.e. it may be too late). Indeed, the onset of the competition effect may actually be slightly delayed with respect to the activation of face neural representations. While the data from this competition experiment cannot directly address this point, the exhibited latency generally fits well with the latency found in other experiments of this thesis, as will be seen later on (see second and experiments from Chapter 7).

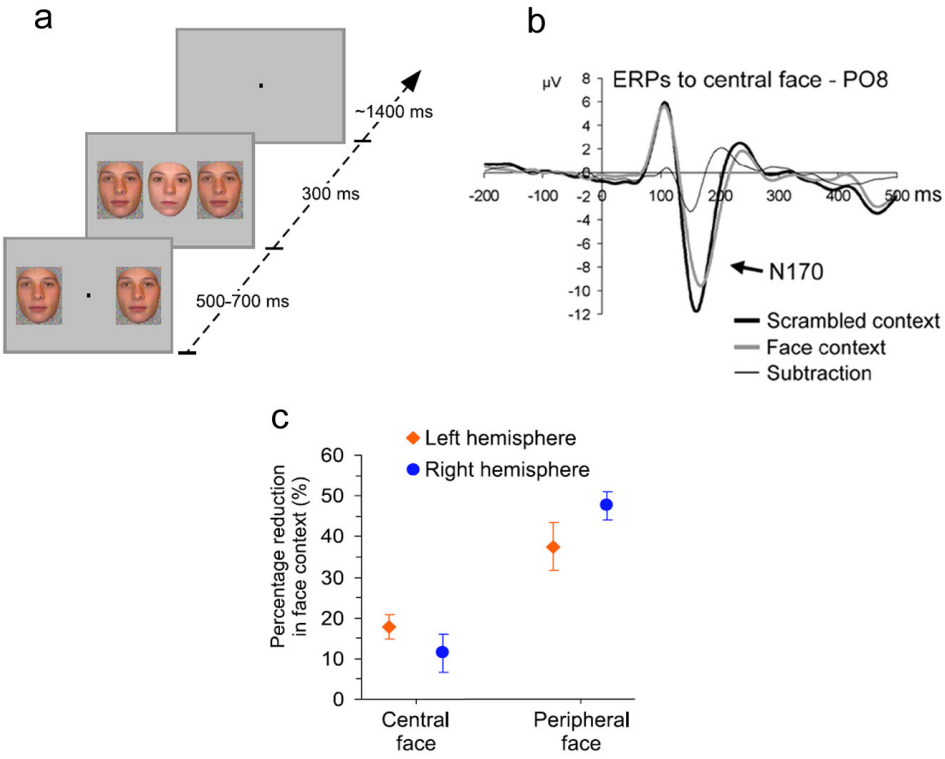

Figure 63: Design and summary of results of ERP data presented in Jacques & Rossion (2006a). A. Time line of stimulus sequence. Two identical stimuli (either faces or Fourier scrambled faces) appear on both sides of a fixation point. About 600 ms later a face appears at fixation. B. Grand average ERP response to the central face at occipito-temporal electrode PO8 (right hemisphere) in the face and Fourier scrambled face context. Thin line is the difference waveforms. C. Percentage of N170 amplitude reduction in the face context compared to scrambled face context as a function of whether the face appears at fixation (with peripheral "competitor" stimuli) or in the periphery (with a fixated competitor stimulus). The reduction of N170 amplitude is about twice smaller when the face appears at fixation ("central face" on the graph).

Another potential limitation of our competition approach to reveal the latency at which face representation are activated comes from the fact that competitive interactions occur primarily at the level of the neurons' receptive field. Because the size of the receptive field increases at each stage of cortical processing, the level of the visual hierarchy at which competitive suppression is observed varies as a function of the distance between the competing stimuli (Kastner et al., 2001; see also Schwartz et al., 2005). Therefore, in our study, the onset time of the ERP competition effect may thus rather reflect the time taken to

reach cortical regions in which neurons have sufficiently large receptive fields to encompass both face stimuli (~6.5°). Neurons which have sufficiently large receptive fields in the ventral stream are probably located in cortical areas V4 and beyond (TEO, TE, STS). In the monkey brain, V4 neurons generaly start firing at around 60 – 80 ms after stimulus onset, whereas the latency of face selective response in TE and STS face-neurons is observed slightly later, at around 80-90 ms. Therefore, taking into account the overal slower neural responses in the human brain[87], an ERP effect observed at 130 ms probably roughly translates into a 80 ms latency in monkeys; a latency that best fits with the onset latencies of neurons in IT (area TE) and in the STS containing face-neurons. As a last argument, the fact that the scalp distribution of the competition effect is strongly similar to the N170 distribution also supports the hypothesis that the competition effect takes place –at least partly- in those areas that are generating the N170.

A last point we want to make here is that while we suggested in this article that the large N170 reduction to the second face may be related to the fact that faces have overlapping neural representations *at the individual level*, unpublished preliminary results suggest that face identity only plays a minor role in this competition effect. These preliminary evidence suggest that the competition mainly arises at the face category level rather than at the level of individual face representations. We conducted a concurrent processing experiment in which the first stimuli appeared at the periphery and the second stimuli appeared at the fovea (similar to Jacques & Rossion, 2006a – see previous point), and we manipulated visual competition both at the face category level (i.e. face vs. scrambled face context) and the individual face level (i.e. in the face context, the lateral (first) and central (second) faces were either the same or different identities). In this experiment we replicate the competition effect on the second face at the N170 for face vs. scrambled face context, but the comparison of same face identity versus different identities only yield a very small difference (statistical significance not tested) at the N170, a stronger effect being observed later starting at 240 ms. Moreover, this small difference was restricted to the N170 peak, around 160-170 ms. The fact that the large difference starting at 130 ms rather reflects a category-level competition actually fits well with more recent evidence that the encoding of face identity rather starts around 160 ms after stimulus onset.

[87] It has been estimated that the neural response latencies in humans would be about 5/3 longer than in macaque monkey (Schroeder et al., 1998; Foxe & Schroeder, 2005).

Chapter 6
The Speed of Individual Face Categorization

Corentin Jacques & Bruno Rossion

ABSTRACT

How fast does the human visual system discriminate individual faces? To address this question, we used a continuous-stimulation paradigm in which event-related potentials (ERPs) to a face stimulus are recorded with respect to another face stimulus, rather than to a preceding blank-screen baseline epoch. Following the shift between two face stimuli, posterior sites showed an early negative ERP deflection that started at 130 ms and peaked at 160 ms, the latency of the N170, an ERP component associated with discriminating faces from objects. The ERP we recorded was larger in amplitude when the preceding stimulus was perceived as a different individual face rather than the same individual face, although face pairs were of equal physical distance in the two conditions. These findings provide direct evidence that individual face discrimination in humans can take place as early as 130 ms following stimulus onset, during the same time window as face detection.

For Copyright reasons, only the abstract of this published article is reproduced here.

Chapter 7

Using the face inversion effect to assess the time course of face categorization and discrimination

1. The time course of the inversion effect during individual face discrimination

1.1. Paper 3:

The time course of the inversion effect during individual face discrimination

Corentin Jacques, Olivier d'Arripe, Bruno Rossion

ABSTRACT

Human faces look more similar to each other when they are presented upside down, leading to an increase in error rate and response time during individual face discrimination tasks. This face inversion effect (FIE) is one of the most robust findings in the face processing literature. Recent neuroimaging studies using adaptation to face identity have shown that the "fusiform face area" was the primary neural source of the behavioral FIE. However, the time course of the FIE, that is, when inversion affects the coding of facial identity in the human brain, remains unclear. Here, we addressed this question by recording event-related potentials (ERPs) on the scalp during an adaptation paradigm with upright and inverted faces. Subjects were presented first with an adapting face stimulus for 3,000 ms, followed by a second face of either the same identity or a different identity. Starting at about 160 ms after stimulus onset, the ERP response to the second face stimulus was markedly reduced over occipitotemporal electrode sites when it was identical to the adapting face, during the N170 time window. When the exact same stimuli were presented upside down, the reduction of signal was smaller and took place about 30 ms later, in line with the behavioral effect of inversion. This result shows that face inversion affects the early encoding of face identity in the occipitotemporal cortex at about 160 ms. Because inversion is known to disrupt massively the integration of facial features, these observations provide indirect evidence that individual faces are processed holistically as early as 160 ms after stimulus onset.

For Copyright reasons, only the abstract of this published article is reproduced here.

The original article is freely available at http://www.journalofvision.org/7/8/3/

1. 2. Complementary results and discussion

Distribution of response times in the matching task for upright and inverted faces

In order to have a better idea of minimal time needed to produce a behavioural response in the face discrimination task, we computed the distribution of response time (independently of "same" and "different" conditions) separately for upright and inverted trials. Figure 64 depicts the resulting response time distributions (i.e. number of responses in successive 20 ms time-bins) for correct and incorrect trials, and Figure 65 shows the distribution difference between correct and incorrect responses for upright and inverted trials. These figures reveal that the earliest responses take place around 330-350 ms after stimulus onset, and that correct responses start to outnumber incorrect responses at around the same latencies. Thus, this analyses shows that compared to the minimal time needed to detect a face in pictures of visual scenes (~260 ms; Rousselet et al;, 2003), an additional 70-80 ms is required to discriminate between 2 faces presented in isolation. Interestingly, similarly to what is observed for face detection (Rousselet et al., 2003), the earliest correct responses are not affected by face inversion, the two distribution starting to separate at around 400 ms (Figure 65 - difference curve). This suggests that the earliest responses in this discrimination task may have been largely based on lower-level visual cues that distinguish between pictures of faces (e.g. colour, overall shape, contrast) and that are not affected by inversion.

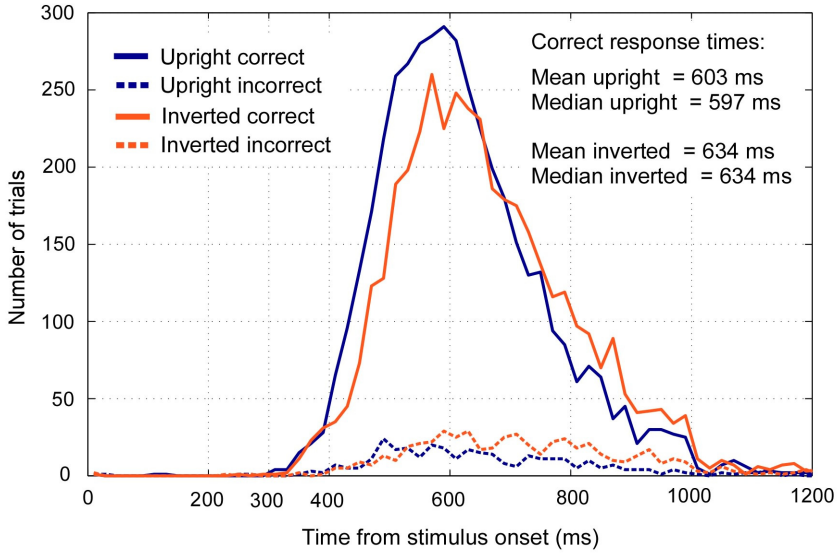

Figure 64: Distribution of response times for upright and inverted trials separatelly for correct and incorrect responses. The number of trials in successive 20 ms time-bins is plotted as a function of time from stimulus onset.

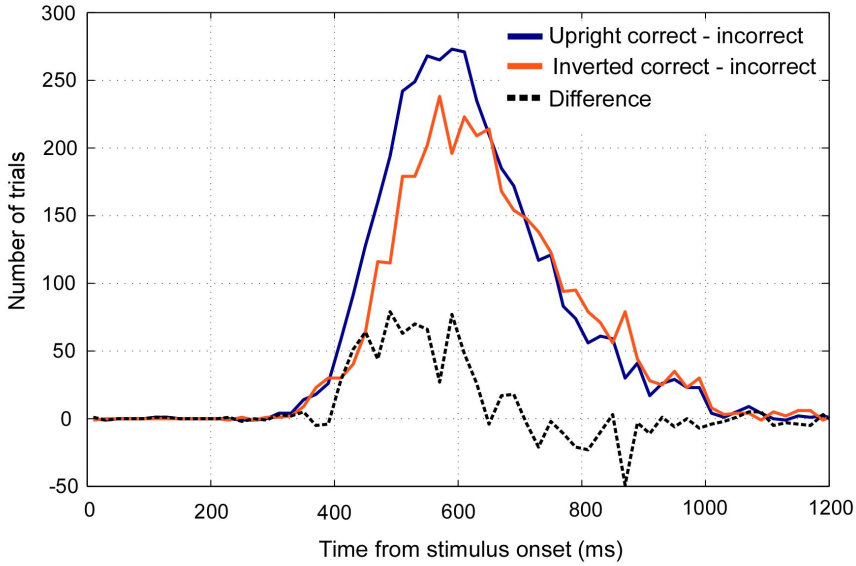

Figure 65: Distribution difference between correct and incorrect responses (see Figure 64) for upright and inverted trials. The difference between the resulting two curves is plotted in dashed line.

Finally, it is important to note that in this experiment, slightly different pictures of the same person were shown as adapting and test faces, and pictures were of slightly different size. As we will see later, both the mean and minimal response times are usually shorter when the very same images are used.

Large P1 inversion effect in response to the test face

In the results of this adaptation study, we report a particularly large P1 effect of inversion, the P1 amplitude of the test face being 1.5 µV larger for inverted compared to upright faces (see figure 3 of the article or Figure 66). We were surprised by this finding since we did not usually observe such an effect in our investigations of the face inversion effect. However, such findings have been reported in the ERP literature (e.g. Itier & Taylor, 2004a; Linkenkaer-Hansen et al., 1998; Schneider et al., 2007).

To further investigate this effect we collapsed the ERP waveforms to the test face across "same" and "different" trials separately for upright and inverted conditions. The resulting waveforms are shown in Figure 66 (upper row). Next we computed the difference waveforms between upright and inverted conditions, and based on the individual subject's data we plotted a 99% confidence interval (with 20 degrees of freedom) around the upright-inverted subtraction waveform (Figure 66, lower row). The confidence interval allows to determine the time point at which the two conditions start to significantly diverge (i.e. when the "zero" µV line is not included within the confidence interval). This analysis revealed that, for the test face, the ERP response to inverted face starts to be significantly larger than to the

upright face at around 65-70 ms after stimulus onset in the right hemisphere, and at about 100 ms in the left hemisphere. The difference in the right hemisphere was very robust and remained significant for about 150 ms.

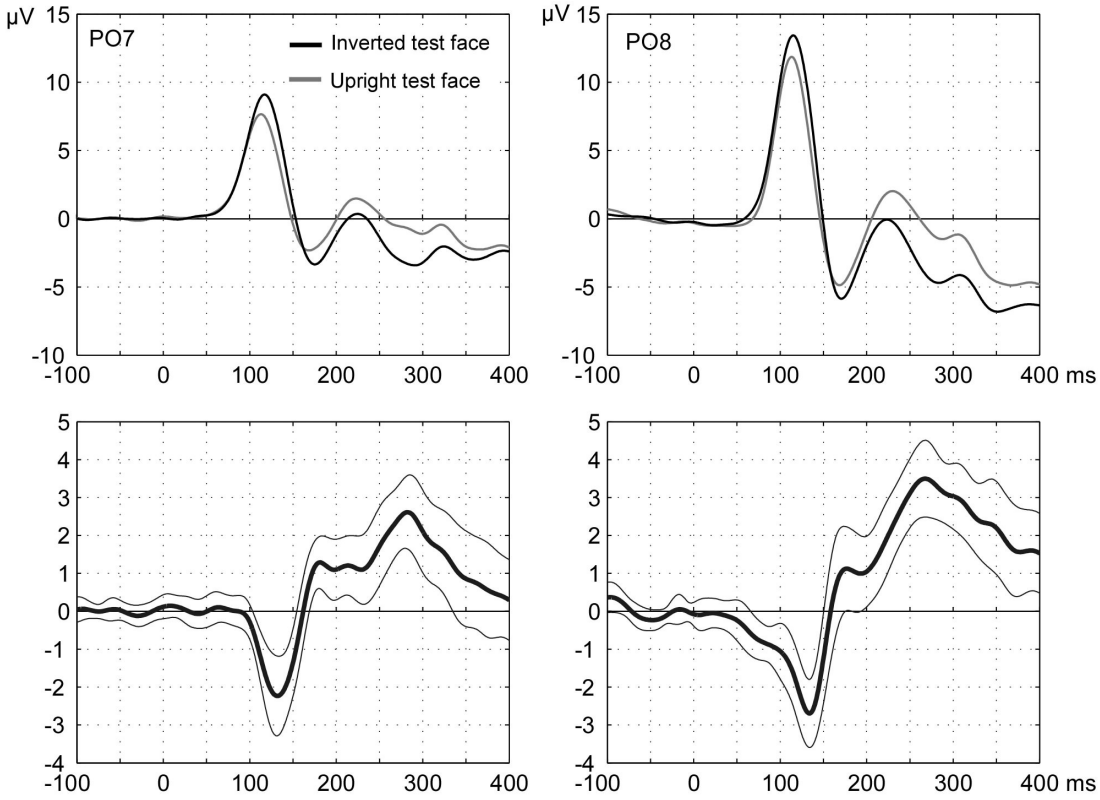

Figure 66: Upper row: Grand average ERP response to the test face collapsed across "same" and "different" conditions at electrodes PO7 and PO8. Bottom row: Subtraction waveforms representing the difference between the ERP response to the test face in the upright orientation (grey curve of upper row) and the response to the inverted orientation (black curve of upper row). Note the large difference starting to be significantly different from zero at about 65 ms after stimulus onset.

Since the ERP to the adapting face mostly did not show such an effect on the P1, we attribute this effect to the fact that the face on which the effect is strong (i.e. the test face) is preceded by another face. If this is the case, there are mostly two plausible interpretations of this very early difference.

First, given the very early onset of this effect, it could be that using an upright- but not an inverted- adapting face actually adapts the response of neurons throughout the visual hierarchy (even of neurons that are not face-selective), from higher-level visual areas until areas such as V1 or V2. This would result in a reduced response to the test face only when the adapting face is an upright face. We would therefore predict that the early modulation of the P1 to the test face is dependent on the orientation of the adapting face but *independent* of the

orientation of the test face. However, confirmation for this hypothesis would require an additional experiment since in the present experiment the test face is always preceded by a face of the same orientation, making it difficult to assess the effect of the orientation of the adapting and test faces independently.

As a second and simpler explanation, this effect may have nothing to do with the orientation of the faces per se, but may rather be due to the position of the test face with respect to the fixated location. More precisely, when the adapting face was presented upright, the subjects probably fixated the eye region (i.e. upper part of the stimulus), as this area of the face generally receives the longest fixation (Williams & Henderson, in press) and is the most informative for face discrimination. On the contrary, when the adapting face was presented inverted, subjects were probably fixating the lower part of the stimulus, that is the eye region as well (Williams & Henderson, in press). Therefore, given that the orientation of the test face was predictable from the orientation of the adapting face together with the fact that we used a very short inter-stimulus interval, it is likely that the test face appeared at a different location relative to fixation in the upright and inverted conditions. Since the amplitude of the early visual evoked potentials such as the C1 (main sources in V1) or the P1 are strongly influenced by the visual field location of a stimulus, this fixation bias is a plausible explanation –though not a very interesting one- for this strong inversion effect on the test face.

In any case, additional evidence is needed to disentangle these two interpretations.

Conclusions

In this study we thus replicate our finding that the encoding of face identity in face discrimination first occurs during the N170 time-window at around 160 ms. Moreover, given that we used slightly different pictures for the adapting and test faces in "same" trials, this suggests that the individual face representation at the N170 may not just be a pure image-based representation. This is also supported by the fact that face inversion delays and reduces this effect even though the exact same images were shown in both orientations.

2. Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans

2.1 Paper 4:

Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans

Corentin Jacques & Bruno Rossion

ABSTRACT

Picture-plane inversion dramatically impairs face recognition. Behavioral and event-related potential (ERP) studies suggest that this effect takes place during perceptual encoding of the face stimulus. However, the relationship between early electrophysiological responses to upright and inverted faces and the behavioral face inversion effect remains unclear. To address this question, we recorded ERPs while presenting 10 subjects with face photographs at 12 different orientations around the clock (30 degrees steps) during an individual face matching task. Using the variability in the neurophysiological responses introduced by the multiple orientations of the target face, we found a correlation between the ERP signal at 130-170 ms on occipito-temporal channels, and the behavioral performance measured on the probe stimulus. Correlations between electrophysiological signal and behavioural responses started about 10 ms earlier in the right hemisphere. Significant effects of orientation were observed already at the level of the visual P1 (peaking at 100 ms), but electrophysiological responses were not correlated with behavior until the face-sensitive N170 time-window. Overall, these observations indicate that plane-inversion affects the perceptual encoding of faces as early as 130 ms in occipito-temporal regions, leading directly to an increase in error rates and RTs during individual face recognition.

For Copyright reasons, only the abstract of this published article is reproduced here.

2.2. Complementary results and discussion

Distribution of response times in the matching task

Similarly to what we did for the complementary data analyses in the previous experiment, we computed the distribution of response time in all subjects collapsed across all orientations (Figure 67), or separately for upright and inverted trials (Figure 68). Figure 67 shows that correct responses clearly started to outnumber incorrect responses slightly before 300 ms after stimulus onset, thus slightly faster that what we observed in our previous experiment, in which the earliest correct responses were produced at around 330-350 ms.

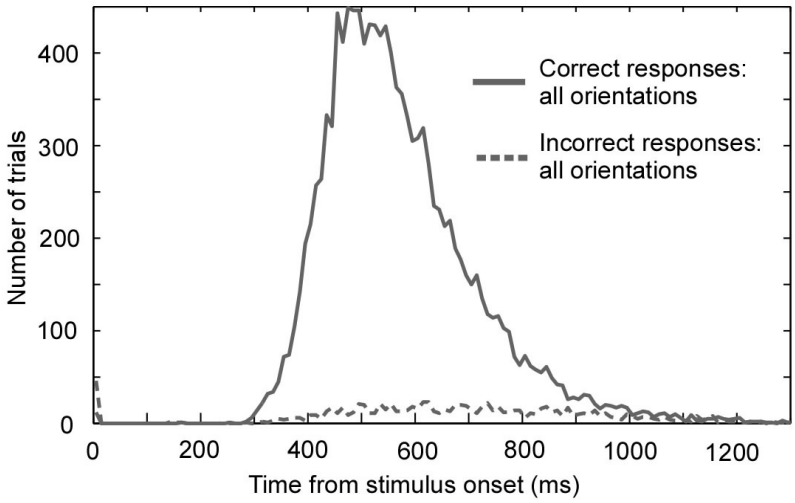

Figure 67: Distribution of response times across all orientations, separately for correct and incorrect responses. The number of trials in successive 10 ms time-bins is plotted as a function of time from stimulus onset.

We suggest this may partly be due to the fact that in the present study the same pictures were used in *same* identity condition (although the target face was 10% larger than the probe face), as opposed to slightly different pictures of the same person in our previous adaptation experiment. This would also be consistent will overall larger mean response times in the adaptation experiment (around 600 ms) as compared to the present experiment (around 550 ms). In other words, given the slight increase in difficulty due to the use of different pictures in the adaptation experiment, slightly more information has to be accumulated before reaching an accurate decision for the discrimination task.

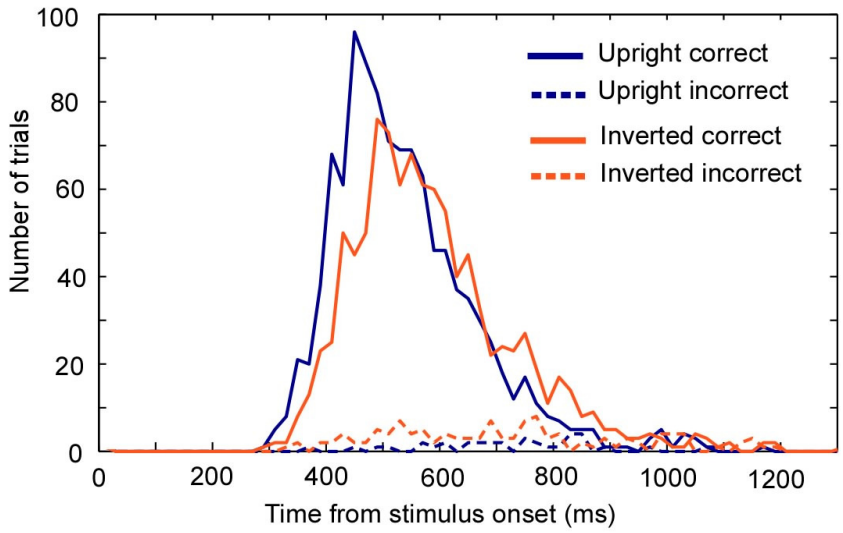

Figure 68: Distribution of response times for upright and inverted trials separately for correct and incorrect responses. The number of trials in successive 20 ms time-bins is plotted as a function of time from stimulus onset.

Surprisingly, in contrast to what was observed in the previous adaptation experiment, the earliest responses were readily influenced by face orientation (see Figure 68; the left tail of the distribution for inverted faces being translated to the right by about 30 ms compared to upright faces) despite the use of the same pictures in the same identity condition and the corresponding increased possibility to rely on lower-level image-based cues to perform the task. Several methodological differences between this and the previous adaptation experiment, such as presentation duration of the target face, the use of a visual mask and longer interstimulus interval between target and probe faces in the present study, or even the use of multiple vs. only two orientations, may explain this difference of results.

To conclude, these observations are be perfectly in line with the view that the accumulation of information about a stimulus depends both on the distance between the target and probe stimuli (the larger the distance, the more time is needed to match/discriminate between target and probe) and on the internal representations (the face representation is tuned for upright faces) (Perrett et al., 1998), and that therefore there is no unique answer as two the time point at which a given information will be sufficiently available in the neural response to perform the task at hand.

ERP response to the second (probe) face

Even though investigating identity repetition effects on ERPs was not the first aim of the present study, the design of the experiment allowed to search for such repetition effects. Thus, we analysed the response to the second (probe) face of each trial as a function of whether it was preceded by a face of the *same* identity or by a face of a *different* identity. The ERP responses to the probe face in the "same" and "different" identity conditions are plotted in

Figure **69** for upright (upper row) and inverted (lower row) orientations. At first sight there doesn't seem to be any repetition effect (i.e. amplitude difference between "same and "different" conditions) on the N170. However, one should note that for this second face, the N170 was maximal at about 145 ms in the upright orientation, thus considerably earlier than in response to the first stimulus, and actually earlier than the identity adaptation effect found in our previous study (Jacques et al., 2007). Therefore, to better assess whether there was any repetition effect at a later latency than the N170 peak and whether this putative repetition effect would be modulated by face orientation, we computed subtraction waveforms between the ERPs obtained "same identity" and "different identity" conditions separately for each orientation. To increase the signal-to-noise ratio, we collapsed the data for clockwise and counter-clockwise orientations (e.g. 30° and 330° were collapsed); except for 0° and 180° for which there was no corresponding clockwise and counter-clockwise orientations. Plotting these difference waveforms (Figure 70 – upper row) actually revealed that a difference starts to emerge at around 150-160 ms for the upright orientation (0°), both in the left and right hemispheres. In addition, the repetition effect was smaller and delayed for inverted faces. The onset latency of the repetition effect is thus in line with our previous finding of an adaptation effect occurring at about 160 ms for the upright orientation and slightly later for inverted faces. Interestingly, there was only little differences between the repetition effect for orientations 90° to 180°, both at the level of the N170 (early difference) and at the level of the later N250r (see Figure 70 – upper row, negative-going potential around 300 ms). In contrast, repetition effects at orientations 30° and 60° had similar amplitude in the N170 range as for upright faces, and were of intermediate amplitude in the N250r range. Thus, the patterns of modulation of repetition effects at the N170 and N250 partly match the pattern observed for behavioural modulation due to face orientation (see Figure 2 of the paper).

However, as shown in Figure 70 (middle row), the repetition effect at the N170 for upright faces started to be (moderately) significant (p<0.05) only at around 200 ms, and was

actually not significant in the N250 window. This was mainly owing to the fact that only a few subjects seemed to display a reliable repetition effect (see Figure 70 – bottom row and see below). There are several possible reasons why we did not observe reliable repetition effects in this study. First, in this experiment the first face was presented for only 400 ms, followed by a mask and an interstimulus interval of 800 ms. As mentioned earlier, "early" repetition effects are probably stronger when the first face is presented for a longer duration and a relatively short ISI between first and second face is used.

Second, relatively few subjects were included in this experiment, being sufficient for strong ERP and behavioural effects (e.g. effects of orientation on the amplitude of the N170) but insufficient for more subtle effects such as these repetition effects at the N170. Third, given the large number of face orientations used (12 orientations), a relatively limited number of trials were included in the *same* and *different* conditions (50-60 trials per conditions, resulting in about 1320 (=55*2*12) trials per subject. This is especially true for the upright and inverted orientations for which we could not collapse clockwise and counter-clockwise orientations.

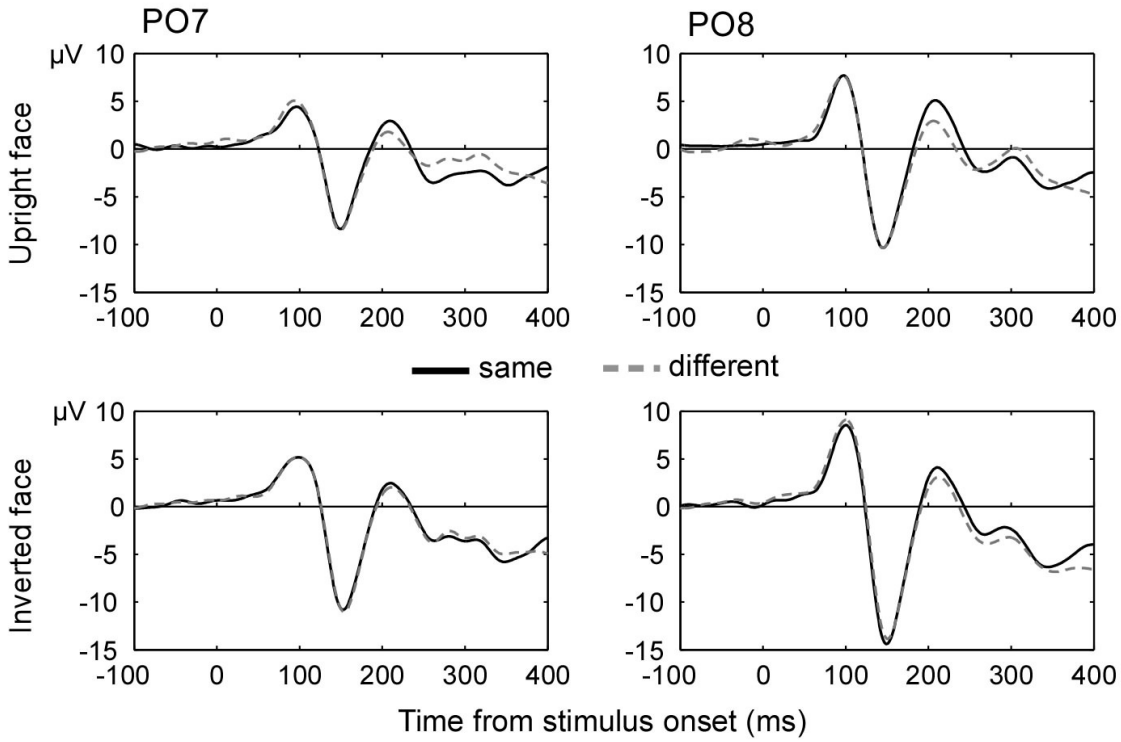

Figure 69: Grand average ERP response to the second face separately for "same" and "different" trials in the upright (upper row) and inverted (lower row) face orientations.

Conclusions

As mentioned in the present paper and earlier in this thesis, some have taken category (i.e. face vs. non-face) and face inversion effect at the P1 as an evidence that face-specific neural activity starts at the P1 level, thus earlier than the N170 time-window. However, in the present study we found that the earliest significant correlations between behavioral responses and electrophysiological signal only started after the P1 -around 110 ms- and was maximal at around 130 ms. Together with the fact that the P1 amplitude seemed to co-vary with the asymmetry of contrast in the face (due to the presence of the eye region) in the left vs. right and upper vs. lower visual field, this clearly further indicates that these effects of category and inversion at the P1 level are mostly driven by differences in the global or local low-level visual properties between stimuli. In a recent study we have conduxted (Jacques & Rossion, in preparation; not reported here), subjects were presented with upright and inverted faces, and were instructed to fixate either the *upper part* of the stimulus (i.e. the eye region for an upright face and the mouth region for an inverted face) or the *lower part* of the stimulus (i.e. the mouth region for an upright face and the eye region for an inverted face) in different blocks of trials. Most interestingly for the present discussion, we observed that while the P1 was larger for inverted faces when subjects fixated the upper part of the stimuli, the opposite result (i.e. larger P1 for upright faces) was found when subjects fixated the lower part of the stimuli. This suggests that the P1 is larger when the high contrast region of the eyes (or hairline when present) is located away from direct fixation. Given that in ERP experiments most subjects will naturally tend to fixate the upper part of the stimulus (as long as the orientation of the face is not predictable from trial to trial), this may partly explain why some studies find inversion effects at the P1; inverted faces having the eyes located the further away from fixation.

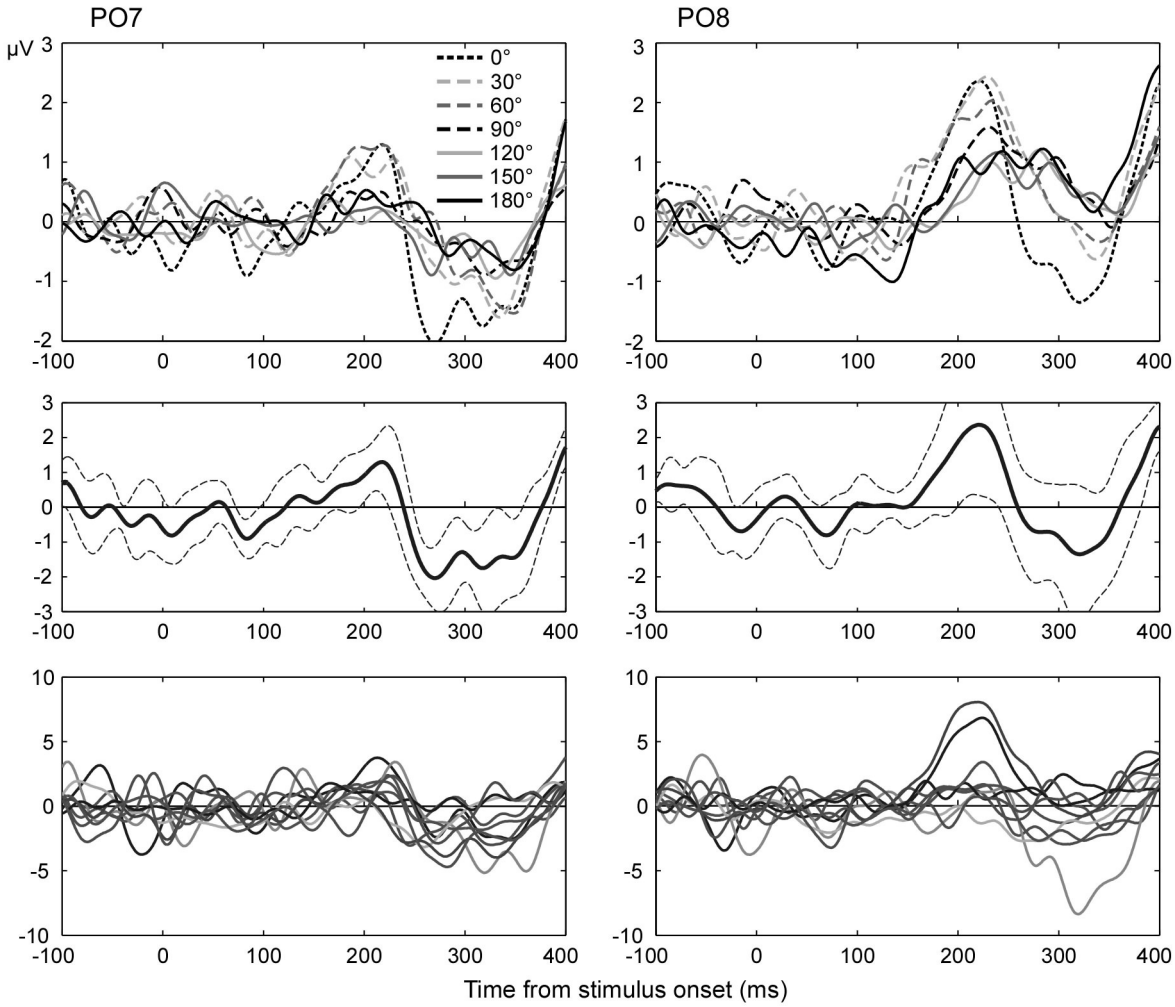

Figure 70: Subtraction waveforms between "same and "different" conditions for the second (probe) face of each trial. Upper row: subtraction waveforms at each face orientation between 0 and 180°. For orientations between 30° and 150°, clockwise and counter-clockwise orientations were averaged together (e.g. 30° averaged with 330°). Middle row: subtraction waveform (same-different) for the upright orientation together with a 95% confidence interval (2 tailed = +/- 2.1 x standard error of the mean). The waveform is statistically different from zero when the confidence interval does not include the "zero" line. Bottom row: subtraction waveform (same-different) for the upright orientation in each subject (N=10) showing that the same-different repetition effect is driven mostly by two subjects in the right hemisphere.

While we found orientation effects during the P1 time window in the present paper, these differences were not correlated with behaviour, suggesting they represent lower-level visual differences not functionally related to face categorization. What we rather suggested is that the effect of face orientation on the face categorization (i.e. the categorization of a face as a face) starts during the early correlation time window at around 115 ms and maximal at 130 ms. In this first correlation time window the effect of orientation was mainly reflected as a latency shift of the ERP waveform at the onset of the N170 (Figure 71), suggesting delayed face categorization due to abnormal face orientation. At a neural level, this may either correspond to a delay in the response onset of face-neurons (e.g. Tsao et al., 2006) or to a

slower rate of response accumulation for rotated faces despite a similar response onset time for upright and inverted faces (see Perrett et al., 1998).

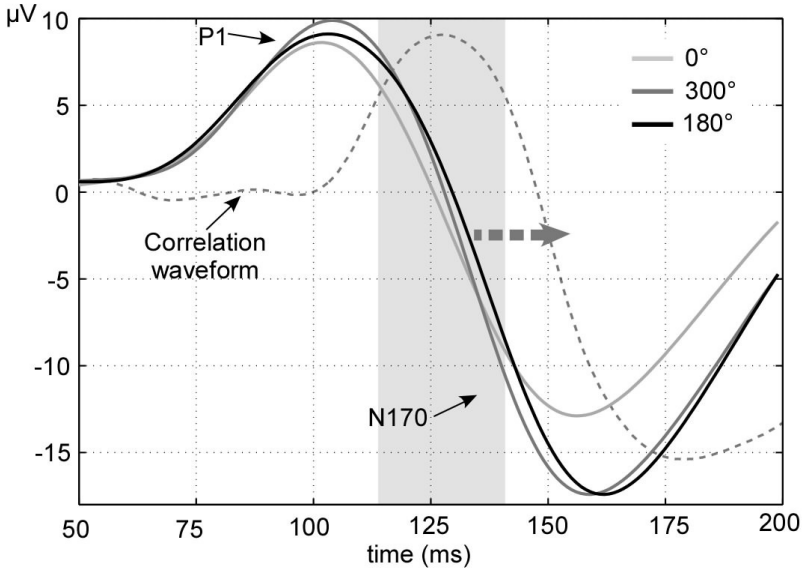

Figure 71: The effect of face orientation (here shown for 3 orientations) at around 130 ms is best described as a latency shift in the onset of the N170 component. The dotted line is the (scaled) correlation curve between ERP response at 12 orientations and behavioural performance at the matching task.

We further suggested that the second correlation time window, starting around 150 ms, may have more to do with representing information about face identity, a proposal compatible with the timing of identity adaptation effects found in our previous study (Jacques et al., 2007).

3. Dissociating the time course of face categorization and the coding of individual faces

Corentin Jacques & Bruno Rossion (in preparation)

Introduction

Based on the finding of our two previous studies manipulating face orientation, we have suggested that the visual system accumulates category-level information about faces faster than information about face identity, but that both of these information are represented during the N170. These suggestions are based on the following evidence. First, we found that varying face orientation – a manipulation that supposedly impairs both face categorization and face discrimination- affects the ERP response in a similar way as behavioural response at the onset of the N170 (120-130 ms). Second, using an identity adaptation design, we found that face orientation affects the coding of face identity at around 160 ms.

In this study, we aimed at assessing the time course of face categorization -as indexed by the early effects of face orientation[88]- and the time course of face identity coding -as indexed by identity adaptation- within the same experiment. To achieve this, we combined the designs of the two previous studies, namely a parametric manipulation of face orientation and an identity adaptation paradigm.

Because the data and results of the present study are largely preliminary, we will provide a minimal amount of methodological details and statistics, rather focusing on the description of the results. Despite the preliminary nature of these results we nevertheless felt that important information regarding our aim in this thesis could be gathered from this experiment.

[88] Note that we are aware that manipulating face orientation is an indirect way to assess the time course of face categorization. However, as already mentioned, numerous studies have shown that face inversion delays the detection of faces in visual scenes (Rousselet et al., 2003; Lewis & Edmonds, 2003; Purcell & Steward, 1988). Moreover, single-cell recordings in the monkey have shown that face-neurons are sensitive to face orientation from the very onset of their response. It is therefore most reasonable to use the effect of face orientation as a marker to reveal the time course of face categorization.

Methods

Participants

So far 9 paid volunteers (8 males, all right-handed, mean age 22.3 +/- 1.8 years) participated in this study. All subjects had normal or corrected-to-normal vision.

Stimuli

The same 30 front view face pictures (15 males) as in Jacques & Rossion (2007b) were used in this study. Briefly, these faces were without glasses, facial hair and make-up, and with neutral expression. All face photographs were edited to remove backgrounds, hair, and everything below the chin, and were equated for mean luminance within the same face-gender. All stimuli subtended approximately 2.8 by 3.7 degrees of visual angle.

Procedure

The design was a mix between our adaptation experiment (Jacques et al., 2007) and the multiple face orientation experiment (Jacques & Rossion, 2007b). In each trial, two faces (adapting and test) were presented sequentially, both in one of seven[89] picture-plane counter clock-wise orientation (0º, 330º, 330°, 270°, 240°, 210° or 180º)[90]. The test face was always presented at the same orientation as the adapting face. A trial started with a fixation point displayed at the center of the screen for 200 ms (Figure 72). Two hundred milliseconds after the offset of the fixation point, the adapting face appeared for about 2800 ms (2600 to 3000 ms), followed by an interval of random duration (150 to 350 ms), and then the second face for 200 ms. The offset of the second face was followed by an intertrial interval of approximately 1400 ms (1300 to 1500 ms). In half of the trials, the second face stimulus was of the same identity (same picture) as the first face. The same pairs of faces were presented at each orientation. Subjects performed an identity matching task between the adapting and test faces at each trial, and gave their response by pressing on one of two keys with their right hand (keys counterbalanced across subjects). Subjects performed 120 trials per orientation condition (60 "same" and 60 "different") for a total of 840 trials in 20 experimental blocks. The order of all the conditions was randomized within each block.

[89] We only used seven orientations in the present experiment mainly because the experiment would have been too long with twelve orientations given that each trial lasts for approximately 5 seconds.
[90] These orientation will be referred to in clock-wise values (i.e. 0°, 30°, 60°,…) in the following sections.

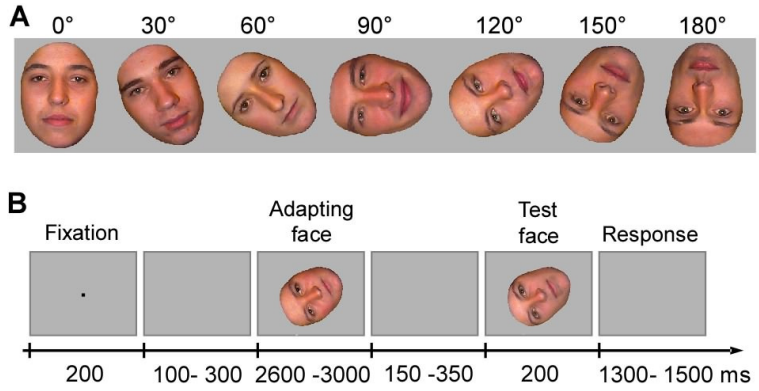

Figure 72: A: Examples of face stimuli at 7 counter clockwise orientations. B: Time-line of a trial sequence.

EEG recording

Scalp EEG was recorded from 58 tin electrodes mounted in an electrode cap (Quik-cap, Neuroscan Inc.), with a left earlobe reference, and electrode impedances kept below 10 kΩ. EEG analog signal was digitized at a 1,000-Hz sampling rate, and a digital anti-aliasing filter of 0.27 * sampling rate was applied at recording (at 1,000 Hz sampling rate, the usable bandwidth is 0 to approximately 270 Hz). EEG data were analyzed using EEprobe 3.2 (ANT, Inc.) and custom made Matlab codes. After filtering of the EEG with a digital 30-Hz low-pass filter, time windows in which the standard deviation of the EEG on any electrode within a sliding 200-ms time window exceeded 35 μV were marked as either EEG artifacts or blink artifacts. Blink artifacts were corrected by subtraction of a vertical electrooculogram (EOG) propagation factors based on EOG components derived from principal component analyses. Trials containing EEG artifacts were rejected. Subjects' averages were baseline corrected using the 100-ms prestimulus epoch and then re-referenced to a common average reference.

Analyses, results and discussion

In this section we mainly focus on the analyses of ERP waveforms rather than specific ERP components.

The raw effect of face orientation was measured on the ERP response to the adapting face only, mostly because we have seen in our previous adaptation experiment that face orientation strongly modulates very early (~65 ms) ERP response to the test face.

The identity adaptation effect was measured on the test face by computing subtraction waveforms between the ERPs obtained in the "same" condition and the ERPs obtained in the "different" condition.

Behavioural results

The relationship between face orientation and behavioral responses is plotted in Figure 73 for mean accuracy and response times (upper row). Normalized[91] accuracy and response times for each individual subject are show in the lower row of Figure 73. This figure illustrates that in this experiment, the patterns of average accuracy and response times are somewhat consistent with our previous result of an increase/decreased in response time/accuracy starting at around 90° - 120° (Jacques & Rossion, 2007b), and with the finding of a larger performance shift around 90° in numerous studies investigating various face manipulations (e.g. blurring, manipulation of local features, thatcherization, …) at multiple orientations (Collishaw & Hole, 2002; Murray et al., 2000; McKone et al., 2001). However, what this figure also shows is the highly variable patterns of accuracy and response times in individual subjects, something that wasn't observed in our previous study with multiple face orientations. This observation is at odd with the fairly large amount of trials used in this experiment (120 per orientation), and may thus reflect a partly inadequate experimental design to measure consistent behavioural performances. Specifically, the use of a small number of faces (30 faces) repeated a large number of time, the use of the same picture (and no change in size) in same identity condition coupled with a long unmasked presentation time of the adapting stimulus (~2800 ms) and short adapting-test interstimulus interval (~250 ms), may have rendered the task too easy (96% on average with only 2% difference between upright and inverted faces while there was 12% difference in Jacques & Rossion, 2007b) and may have triggered abnormal response strategies. Thus, while there clearly is an inversion effect both in accuracy and response times, this ceiling effect may have caused a lack of precise sensitivity of these behavioural measures to the parametric manipulation of face orientation. This was actually supported by the fact that the earliest correct response times in this study were produced around 260-270 ms for upright faces and around 300 ms for inverted faces (Figure 74). These response times are even faster than in the previous study (~300 ms

[91] Normalization was performed by subtracting independently for each subject, the mean accuracy/response time over all orientation. This was done mainly for display purposes to centre around zero the accuracy and response times of each subjects, and plot the resulting data in a single plot.

and ~330 ms for upright and inverted face respectively), and suggest the identity matching task was very (too) easy to perform.

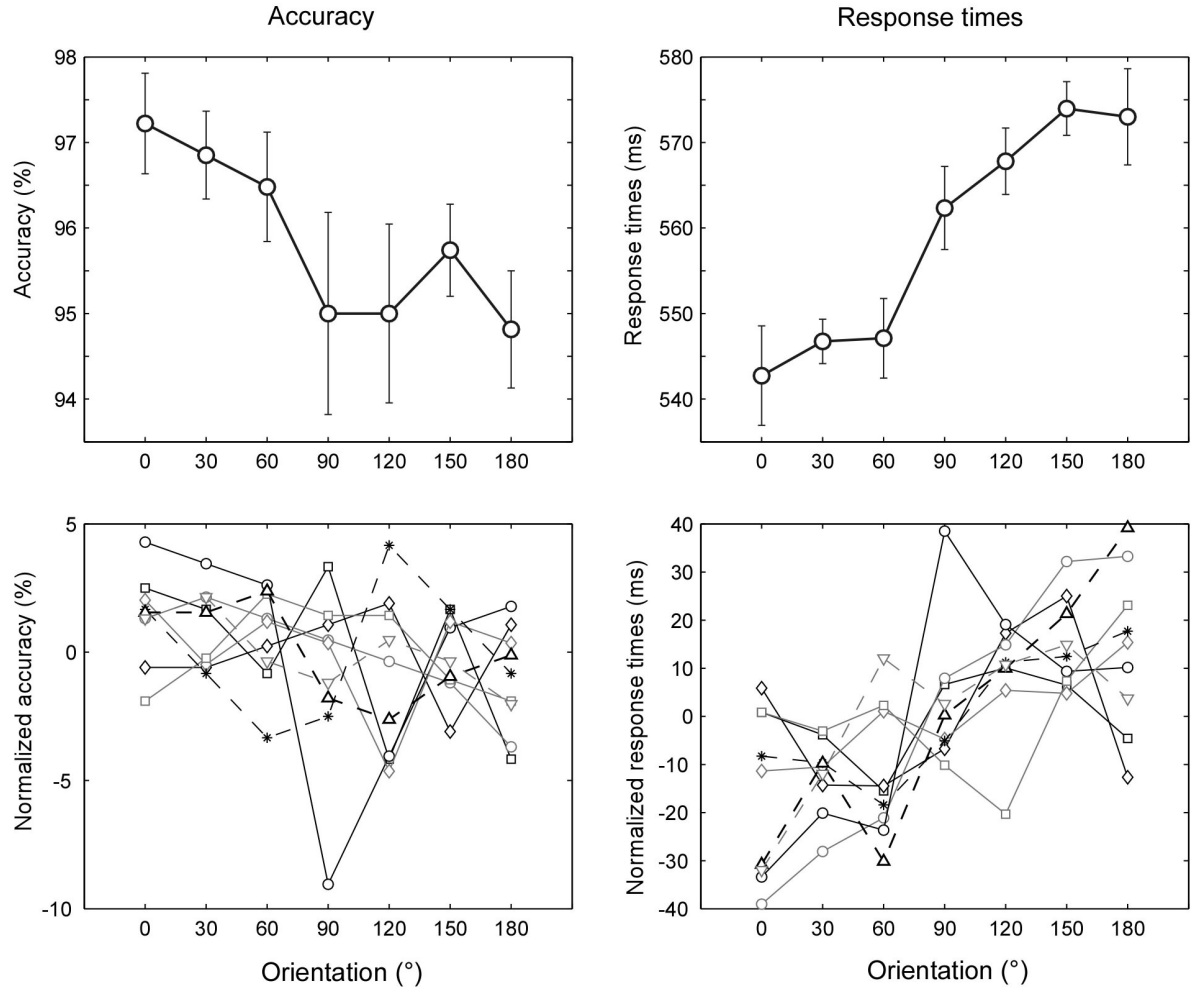

Figure 73: Behavioural results. Accuracy (left) and response times (right) as a function of face orientation. Upper row displays behavioural data averaged over all 9 subjects. Error bars represent standard errors of the mean computed after normalizing the data for each individual subject (see bottom row). Bottom row: Normalized accuracy and response time are show for each subject. Normalization was performed by subtracting independently for each subject, the mean accuracy/response time over all orientation.

Unfortunately, given this lack of consistent results in the individual subject's behavioural data, we will not be able to carry out the point-by-point correlation analyses between behaviour and ERP responses as in our previous study (see Jacques & Rossion, 2007b). Instead, we will rely on other measures to try to characterize the time course of face categorization and individual face encoding.

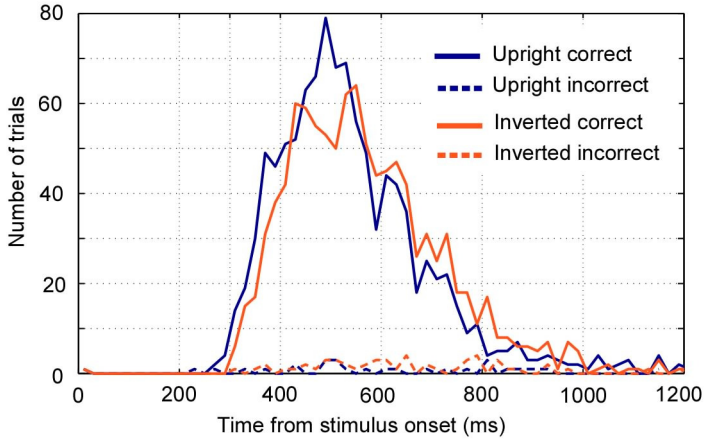

Figure 74: Distribution of response times for upright and inverted trials separately for correct and incorrect responses. The number of trials in successive 20 ms time-bins is plotted as a function of time from stimulus onset.

ERP response to the adapting face

The grand average ERP waveforms obtained at each orientation in response to the adapting face are displayed in Figure 75. Notably, the pattern of amplitude modulation of the N170 as a function of orientation was very close to that found in our previous experiment (Jacques & Rossion, 2007b). Moreover, similarly to what we observed previously, the effects of orientation on the ERP waveforms were most consistent right after the P1 (Figure 76), being maximal at around 130 ms, later at around 180 ms. A third, longer lasting, difference was also found starting at about 320 ms. The results observed for the adapting face therefore mostly replicate our previous observations and does not add further information.

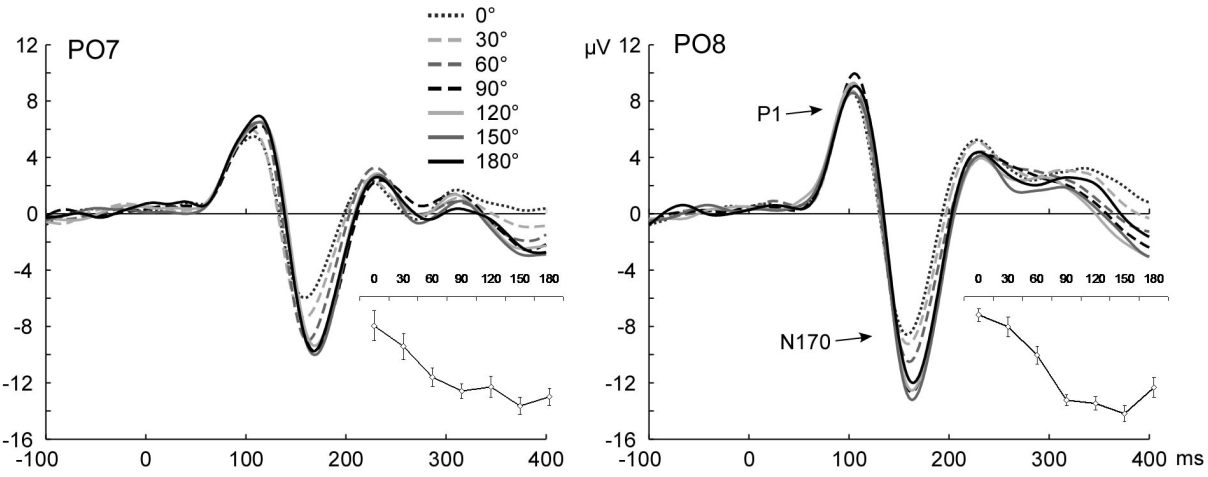

Figure 75: Grand average ERP response to the adapting face at each orientation for 2 occipito-temporal electrodes (PO7:left hemisphere; PO8: right hemisphere). The line graphs in the insets represent the N170 amplitudes in all orientations.

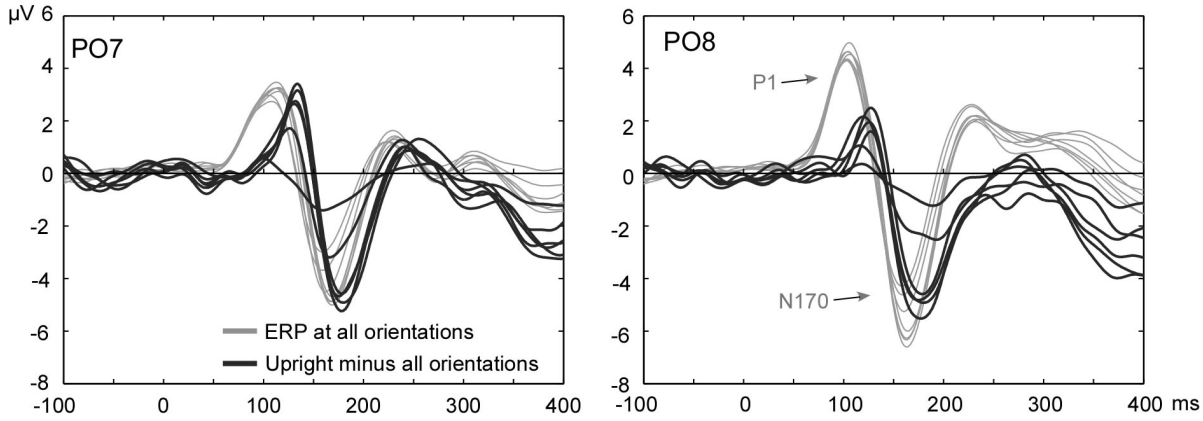

Figure 76: Difference waves between ERP elicited by a face at 0° and other face orientations are shown in dark grey. Note the peaks in the amplitude of the difference waveforms at around 130 ms, 180 ms and beyond 300 ms after stimulus onset. ERP waves (scaled) to all face orientations are shown for comparison in light grey.

ERP response to the test face

The second aim of this experiment was to measure the onset time of identity adaptation effect, and how this would be modulated by face orientation. The ERP responses to the test face obtained in the *same* and *different* identity conditions are shown for 0°, 60° and 180° in Figure 77. Again, as expected, a strong identity adaptation effect was observed in the N170 time window for the upright orientation, while a reduced effect and no effect were found for 60° and inverted orientations respectively. When plotting the ERP difference between same and different identity conditions (Figure 78; see PO8-right hemisphere), the largest difference in the N170 time-window was found for upright faces followed by the 30° orientation. However, given the small number of subjects, the order of amplitude differences for subsequent orientations is more difficult to interpret. The onset latency of these differences seemed to be slightly more consistent, with the earliest differences being observed for 0°, 30°, 90° and 120° followed by 60° and then 150° and 180°. However, while the difference in the adaptation effect for upright and inverted faces is quite strong, these preliminary results do not yet provide reliable information about the more fine-grained subtle differences in identity adaptation between successive orientations[92]. Thus, additional subjects

[92] It may actually be that different orientations influence differently the amplitude and latency of the N170 adaptation effect. If this is the case, an adaptation index taking into account amplitude within a given time-window and onset latency (e.g. the adaptation effect at 150° is quite large but has a late onset time) may provide a better account for this early difference as a function of face orientation. This index could be or instance: adaptation index = (mean amplitude of adaptation in 150-250 ms) / (onset latency of adaptation).

are needed to more precisely map the pattern of identity adaptation as a function of face orientation.

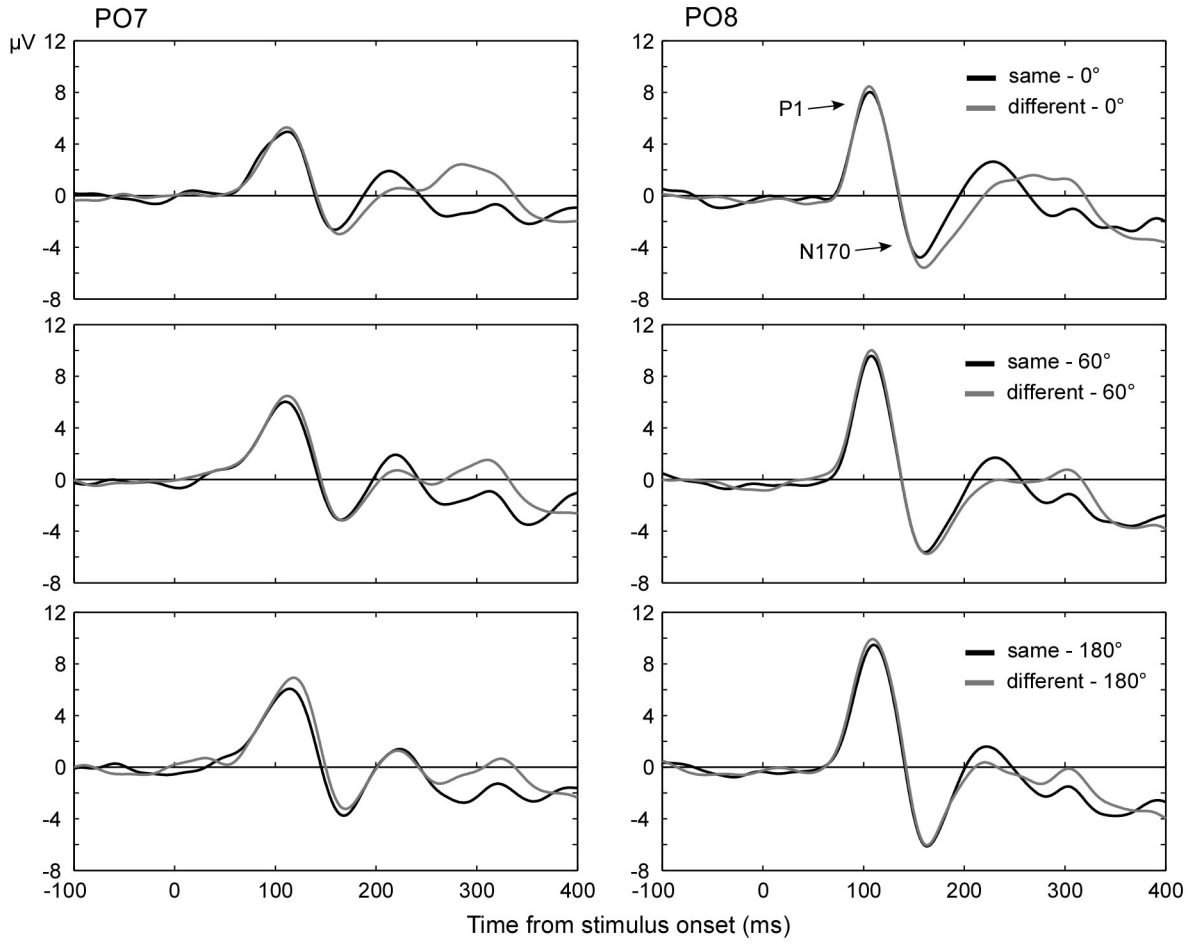

Figure 77: Grand average ERPs to the test face in the same and different identity conditions for upright (upper row), 60° (middle row) and inverted (bottom row) orientations. Waveforms are shown for two occipito-temporal electrodes (PO7, PO8).

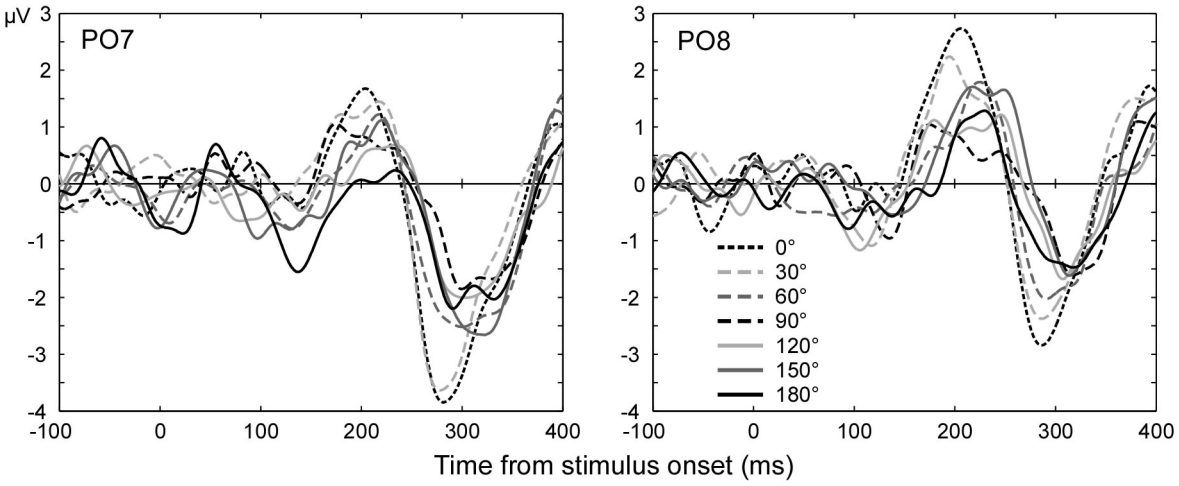

Figure 78: Subtraction waveforms (different – same) to the test face in all seven orientations at electrodes PO7 and PO8.

The second strong adaptation effect occurs during the N250 time-window, starting around 250 ms after stimulus onset. In keeping with the much more robust repetition/adaptation effects found on this component, the ordering of the adaptation effect as a function of orientation was more consistent. That is, the strongest adaptation effect was found for upright faces, followed by 30°, then 60°. In general, subsequent orientations (from 90° to 180°) generated similarly lower adaptation effects at the N250, as was observed in the previous study of parametric face orientation manipulation (see complementary analyses for the previous paper).

Dissociating the time course of face categorization and the coding of individual faces

Since we were not able compute point-by-point correlations between ERPs and behavioural data because of inconsistent single-subject data, we relied on a different method to try dissociating the time course of face categorization and the coding of individual faces. We simply compared the effect of **face orientation** measured on the *adapting* face with the **identity adaptation** effect measured on the *test* face in the upright orientation.

First, the *effect of orientation* on the first face was quantified by computing the standard deviation between the ERP waveforms for all orientations. We reasoned that an effect of orientation on ERP waveforms should increase the standard deviation. Conversely, when there is no effect of orientation (i.e. no systematic difference between ERP waveforms across orientations) the standard deviation should be minimal. Therefore, computing the standard deviation between conditions is a good way to determine the time points the maximally discriminate between the different orientations. The resulting curves are depicted for all 58 electrodes in Figure 79 (upper row). This figure highlights the waveforms obtained for the two occipito-temporal electrodes (PO7, PO8) that best capture this orientation effect. As can be seen in this figure, the standard deviation between orientations showed peaks at around 130 ms, 180 ms and at a later time point (>350 ms), comparable to what is observed when computing the difference between upright orientation and all other orientations (see Figure 76). To determine the earliest time point at which the standard deviation separates from zero, we computed a 95% confidence interval based on individual subjects data (with 8

degrees of freedom). This confidence interval is plotted around the baseline-corrected[93] standard deviation in Figure 80 (upper row) for the two occipito-temporal channels. This figure indicates that the standard deviation start to reliably differ from zero at about **115-120 ms** after stimulus onset (marked by a vertical line in Figure 80; see also Figure 79), in line with our previous results. As indicated in our previous study (Jacques & Rossion, 2007b), this effect may represent the well-described latency increase of the N170 as a result of face inversion (Rossion et al, 1999a; 2000; Eimer, 2000; Itier & Taylor, 2002; 2004a; c). The next consistent orientation effect started around **150 ms** in both hemisphere and thus largely overlapped the N170 time-window. This second effect mostly represents the well known face inversion effect resulting in an amplitude increase of the N170 (see Figure 75) (e.g. Rossion et al, 1999a; 2000; Eimer, 2000; Itier & Taylor, 2002; 2004a; c).

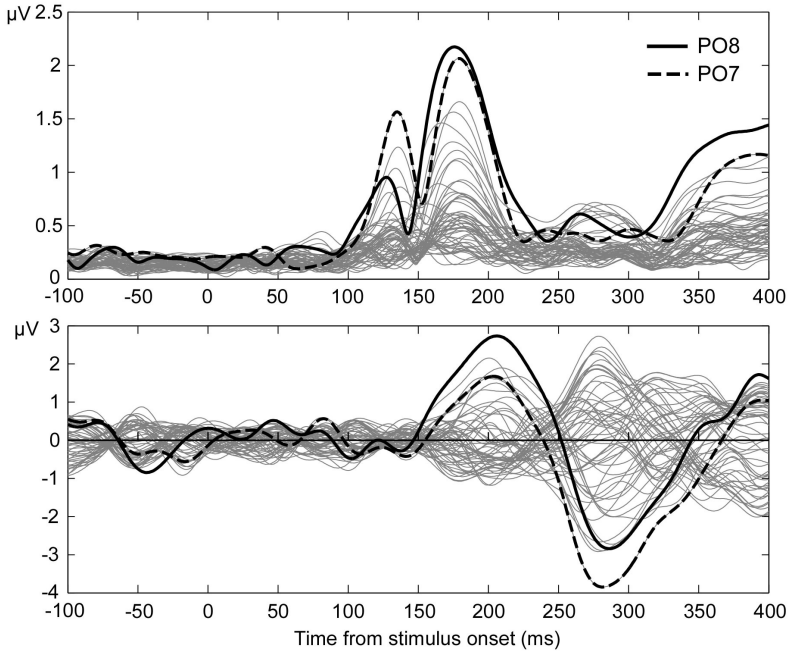

Figure 79: Effects of face orientation and identity adaptation. Upper row: orientation effects on the ERP response to the adapting face. Standard deviation between the ERP waveforms for all orientations is plotted for the 58 electrodes. Occipito-temporal electrodes PO7 (left hemisphere - dashed line) and PO8 (right hemisphere - solid line) are highlighted. Bottom row: Identity adaptation effect (same – different) measured on the ERP response to the test face for the 58 electrodes. PO7 and PO8 are highlighted similarly as in upper row.

Second, the latency at which information about face identity is represented in the ERP signal was determined by measuring the onset time of the *identity adaptation* (i.e. subtracting the ERP waveforms in the same vs. different identity conditions) in the upright orientation.

[93] The standard deviation was baseline-corrected by subtracting the mean standard deviation value in the interval from -100 to 0 ms relative to stimulus onset. This was done to be able to assess the onset time of the orientation effect relative to the baseline period.

Subtraction waveforms of the identity adaptation effect for all 58 electrodes are shown in Figure 79 (bottom row), again with electrodes PO7 and PO8 highlighted. The time point at which the identity adaptation takes place was also assessed by computing a 95% confidence interval (df = 8) around the subtraction waveforms (Figure 80 middle row). This analysis revealed that the identity adaptation started to be significant at about 160 ms after stimulus onset in the right hemisphere and slightly later (170 ms) in the left hemisphere (see vertical lines in Figure 80). A second difference in the subtraction waveforms corresponded to the N250 component previously described as consistently indexing identity repetition effects (e.g. Schweinberger et al., 2002a; 2004).

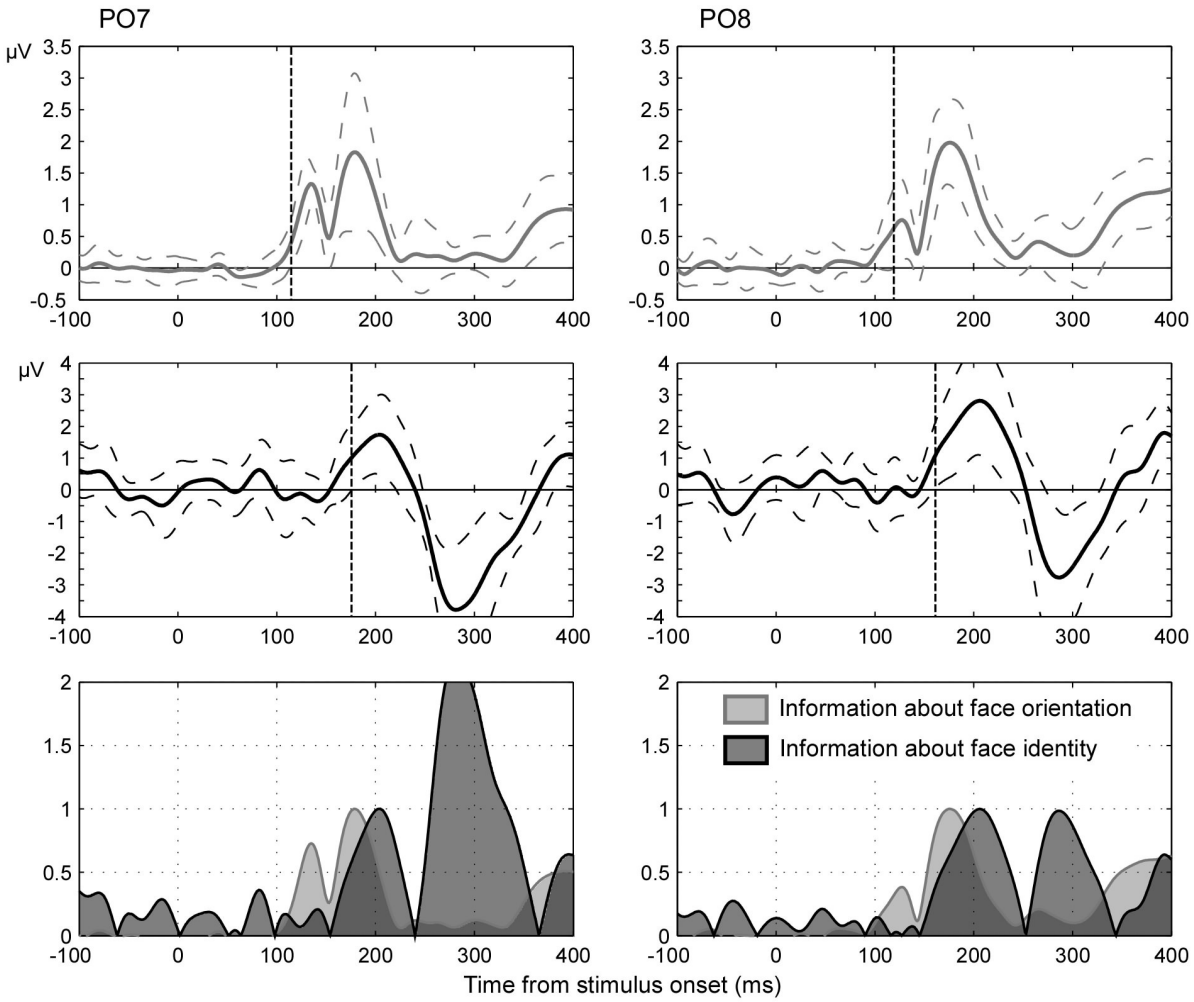

Figure 80: Comparison between orientation effect and identity adaptation effect at two occipito-temporal electrodes (PO7/PO8). Upper row: Baseline corrected standard deviation between the ERP to all orientations in response to the adapting face. Dashed line represents a 95% confidence interval around the standard deviation. Middle row: subtraction waves between the ERP response to the test face in the same vs. different identity conditions for the upright orientation. Dashed lines are the 95% confidence interval. Bottom row: waveforms for the effect of orientation (normalized values of upper row) and the effect of identity adaptation (normalized absolute value of middle row) are superimposed for comparison. Note the earlier onset time of the orientation effect (~115 ms) compared to the effect of face identity adaptation (~160 ms).

Thus, as expected, a direct comparison between the effect of face orientation and the effect of upright face identity adaptation (see Figure 80, bottom row) reveals that the adaptation effect starts on average about 40 to 50 ms later than the effect of face orientation. This difference can also be well apprehended when plotting the accumulation of ERP amplitude[94] across time for the orientation and the adaptation effects (Figure 81). This figure shows that information (reflected by ERP amplitude) about face orientation accumulated slightly faster in the left hemisphere (PO7) than in the right hemisphere (see also Figure 79 and Figure 80). We attribute this hemisphere effect to the fact that we only presented faces in counter-clockwise orientations, with the eye regions in the ipsi-lateral visual field relative to the left hemisphere. This is supported by the fact that comparing the ERP to upright vs. inverted face yielded an orientation effect at 130 ms of equal amplitude in the left and right hemisphere, while this early effect was much larger in the left hemisphere when comparing upright with the 90° face orientation.

Cumulative plots (Figure 81) also clearly show the faster and earlier accumulation of information about face identity in the right compared to the left hemisphere. This observation is perfectly in line with the large amount of evidence from behaviour (e.g. Hillger & Koenig, 1991), neuroimaging (e.g. Haxby et al., 2000), electrophysiology (e.g. Rossion et al., 2003a) and neurophsychology (Bouvier & Engel, 2006) revealing a strong right hemispheric dominance in face processing, especially when the processing of face identity is considered. Because ERP amplitude is continuously accumulated on these cumulative plots, they may in some circumstances better render the continuous evolution of information processing, thus allowing to more closely achieving the goal of not referring to distinct ERP components.

[94] The cumulative plots were created by computing for each time point between 0 and 400 ms the cumulative sum of absolute ERP amplitude. The plots obtained for the orientation and adaptation effects were then normalized by dividing each plot by its respective maximum cumulative response (i.e. corresponding to the cumulative response at 400 ms).

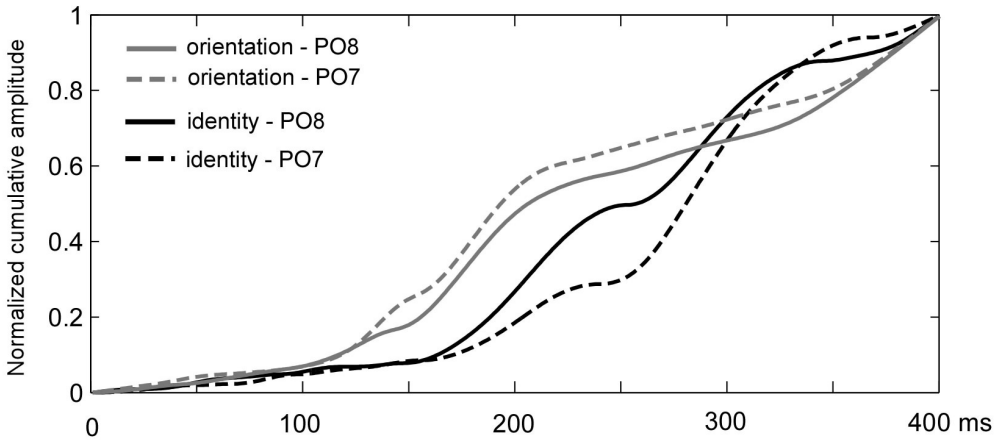

Figure 81: Comparison between orientation effect and identity adaptation effect with cumulative ERP amplitude. The curves represent the cumulation of ERP amplitude across time for the effect of orientation (based on the standard deviation response) and the effect of identity adaptaiton (based on same-different subtraction waveforms). The cumulative curves were normalized so as to be equal to 1 at 400 ms after stimulus onset. Curves for electrode PO7 are shown in dashed lines; solid lines for PO8.

General discussion

First we replicated our previous results (Jacques & Rossion, 2007b), showing that face orientation modulates the ERP response to a face at multiple time-scales. Indeed, we found a first modulation at about 115-120 ms (maximal at ~130 ms) at the onset of the N170 component and a second modulation starting at about 150 ms (maximal at 180 ms) overlapping the N170 peak. We also found a third strong effect of orientation starting at about 320-350 ms.

Second, we replicate our results on identity adaptation, again finding an effect starting at about 160 ms after the onset of the test face, in the time-window of the N170 component. This adaptation effect was clearly smaller and delayed in response to inverted faces, again replicating our previous results (Jacques et al., 2007). Given the finding of a behavioural effect of inversion in the identity matching task, the delayed and reduced adaptation effect at the N170 may at least partly reflects the delayed and "poorer" coding of inverted faces in the occipito-temporal cortex. However, we note that too few subjects have been so far included in our analyses to precisely characterize the relationship between face orientation and the pattern of modulation of adaptation effects (both at the N170 and later time-points). Determining the time course of this relationship using averaged ERP in individual subjects and single trial

ERP responses may help to clarify how neural responses are used to generate a behavioural response in an identity discrimination task.

One of the most interesting outputs of this experiment is the temporal dissociation between the effect of face orientation and identity adaptation effects which was revealed in the same experiment. As stated above, modulation of ERP response by face orientation was used as a marker to determine the time course of face categorization[95], and identity adaptation allowed tagging when in the ERP response is information about face identity significantly represented. Importantly, the finding that significant identity adaptation effects are found later than category-related effects does not necessarily imply that the visual processes involved in these different levels of categorization evolve in a serial way; with face identity being represented only *after* global face categorization has been performed. Indeed, single-cell recording studies have shown that information about stimulus category and facial identity may both be present in the first few spikes of a neuron's response (Tovee & Rolls; 1995; Rolls et al., 1999). Moreover, a large proportion of these face-neurons are involved both in global face categorization and in encoding face identity (Matsumoto et al., 2005; Sugase et al., 1999). However, given that information about facial identity accumulates slower than information about stimulus category, the information about these levels of categorization is significantly represented in the neural response at different time-window. Therefore, the fast accumulation of neural activity about the stimulus category may generate sufficient activity to be recorded in the ERP response at scalp level near the onset of the neural response (~120-130 ms). However, the differential neural response to same vs. different identities due to adaptation may not be large enough to be measured at scalp level before 160 ms.

[95] Even though face categorization can be assessed by manipulating face orientation, further experiments manipulating stimulus category and adaptation may provide complementary information about the time course of face and object visual categorization.

Chapter 8
Individual faces are represented holistically during the N170 component

Corentin Jacques & Bruno Rossion (in preparation)

Introduction

Human face recognition is based on the processing of multiple cues such as local facial features (e.g. eyes, mouth, hair), the relative distances between these features (e.g. inter-occular distance), properties of the skin surface (e.g. texture, color, reflectance) and general shape apprehended by the interaction of an oriented light-source and the bone structure of the face (e.g. shape-from-shading). It is generally accepted that these facial cues are not perceived independently from each other but are rather integrated into a gestalt, or a so-called holistic representation (refs). Evidence that faces are processed holistically come from the finding that the identification of facial features is strongly influenced by the other face parts or the whole face context (Sergent, 1984; Young et al., 1987; Tanaka and Farah, 1993).

One of the most widely used paradigm to investigate holistic face processing is the so-called "face composite effect" initially created by Young and colleagues (1987). The face composite effect is a strong perceptual illusion in which two identical top parts of a face are perceived as being different if their respective bottom parts belong to two different identities. Notably, the perceptual illusion disappears if the top and bottom parts are laterally offset, so that they appeared misaligned. Although it is widely assumed that the integration of a face into a holistic representation[96] occurs at a perceptual level, that is, as the face is perceptually encoded (Farah et al., 1998; Hole, 1994), there is no direct evidence for this yet. In contrast to this perceptual account for holistic processing, recent studies using signal detection theory have rather suggested that holistic effects observed in the face composite paradigm (i.e.

[96] The notion of holistic representation in the present case refers to the fact that face information is integrated into an *individual* representation. We have seen earlier that holistic processing can also be defined as a process that integrate facial information to help face categorization/detection. For instance, as already stated, holistic processing is used at a basic-level categorization to detect faces in two-tone mooney images or vegetable Arcimboldo faces. Furthermore, there is no doubt that that this type of holistic processing is used to categorize faces during early face visual processing indexed by the N170. It is however yet unclear whether holistic processing involved in face individualization is just a "higher resolution" version of holistic processing used to categorize/detect faces (see Sergent, 1986), or if these are functionally dissociable processes. One evidence for different holistic face representations comes from the finding that in some prosopagnosic patients (e.g. patient PS, Rossion et al., 2003b) holistic processing can readily be used to detect faces in mooney-images (Busigny et al., in preparation), but not to generate individual representations (Ramon & Rossion, 2007).

considering to identical top parts to be different when aligned to two different bottom parts) occurs primarily at a decisional level, resulting from strong response biases (e.g. Wenger & Ingvalson, 2002, 2003).

It is only very recently that functional neuroimaging (fMRI) has provided direct evidence that neurons in the cortical areas that respond preferentially to faces, both in the fusiform (FFA) and inferior occipital (OFA) gyri, integrate the top and bottom parts of a composite face into a holistic representation (Schiltz & Rossion, 2006). However, owing to the poor temporal resolution of fMRI, this finding cannot disentangle between the hypotheses that holistic processing occurs during the initial perceptual encoding of individual faces as originally proposed, or whether the holistic effect observed in these areas reflects re-intrant signal from post-perceptual or decisional processes. Determining the locus of holistic processing is crucial because, whether it is found that this effect has a perceptual or a decisional locus, it will introduce strong constrains for models of face recognition.

The recording of scalp event-related potentials (ERP) allows investigating brain events related to perceptual processes with a millisecond time resolution. To directly tackle the issue of the time course of holistic processing, we recorded ERPs in a face composite paradigm. Specifically, we used an adaptation paradigm in which an adapting face composed of separate top and bottom parts (i.e. a composite face) was presented for 3 seconds and rapidly followed by a second (test) composite face. The adapting and test composite faces could either (1) have identical top and bottom face parts (*same*), (2) differ only in their bottom parts (*bottom different*), or (3) different in both their top and bottom parts (*top+bottom different*). In addition, adapting and test faces were either presented both in an aligned or both in a misaligned format. The composite illusion was therefore measured by comparing the ERP waveforms elicited by the test face in the *same* vs. *bottom-different* conditions. We hypothesized that, for the aligned format, the composite illusion due to holistic processing in the *bottom-different* condition would lead to a recovery from adaptation when compared to the *same* condition. In contrast, because the composite illusion is not observed for misaligned faces, we mostly should not find any release from adaptation when comparing misaligned *same* to misaligned *bottom different* conditions. Moreover, because the *top+bottom different* condition can be considered as a full identity change from the adapting face to the test face (i.e. both top and bottom parts are of a different identiy), the comparison of this condition with the *same* condition would provide additional information about the time course of face identity encoding in the visual system.

Methods

Participants

Seventeen paid volunteers (5 men, 1 left-handed, mean age 21.6 +/- 1.7 years) participated in the experiment. All subjects had normal or corrected-to-normal vision.

Stimuli

Stimuli used in this experiment were constructed from 30 top parts of faces (17 males & 13 females) (see Figure 82) and 43 bottom parts of faces (23 males & 20 females). Face tops and bottoms were taken from whole faces cropped to remove background, hair and everything below the chin and horizontally cut in two halves above the nostril. All resulting top and bottom parts were converted to greyscale and equated for mean pixel luminance.

Composite face stimuli were then created by pairing each face top with 2 different bottom parts (some bottom parts being used for more than one top), resulting in 30 pairs of stimuli; each pair with an identical top part but different bottom parts (see Figure 82). The created pairs of composite faces were either formed by top and bottom face parts precisely aligned to each other ("aligned" format) or were formed by the very same top and bottom parts with the bottom slightly offset to the right (30 pixels = 0.6° of visual angle). These stimuli with offset bottom parts are commonly called "misaligned" composite faces. Aligned faces subtended 2.5° in width x 3.4 in height of visual angles, while misaligned faces subtended 3.1° in width x 3.4 in height. The gap between top and bottom parts was 3 pixels in height (<0.1°).

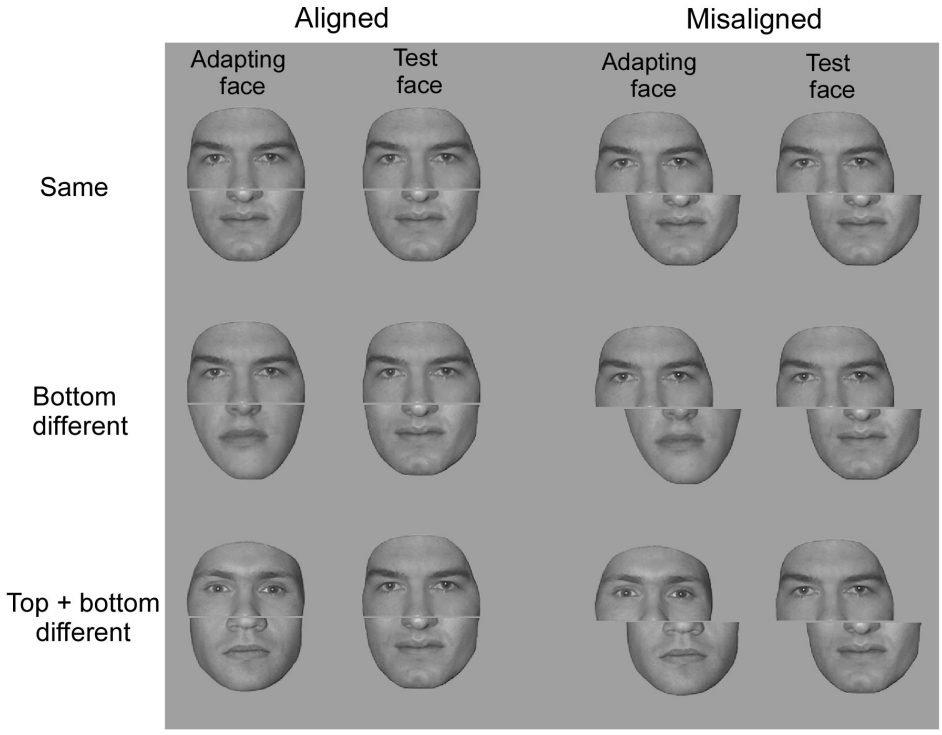

Figure 82: Examples of composite faces and pairs of composite faces (adapting and test) for the 6 conditions used in the experiment. Note that the top parts of the composite faces in middle row are perceived as being slightly different in the aligned face format (left) despite the fact that they are identical and that only the bottom parts of the two composite faces are different (= composite face illusion). Notably, this illusion does not occur in the misaligned format (right).

Procedure

Stimuli were displayed against a light grey background using E-prime 1.1 (PST) at 110 cm viewing distance. In each trial, two face stimuli were presented sequentially, either both in the aligned or both in the misaligned format. A trial started with a fixation point displayed at the centre of the screen for 200 ms (Figure 83). Two-hundred milliseconds after the offset of the fixation point (randomized between 100 and 300 ms), the first face (adapting stimulus) appeared for about 2800 ms (2600 to 3000ms) with the top part being presented at fixation. The offset of the adapting face was followed by an interval of random duration (150 to 350 ms), and then the second face (test face) for 200 ms. An inter-trial-interval of about 1600 ms (1500 to 1700 ms) separated the offset of the test face from the next trial. Identity of the top and the bottom parts were manipulated between the adapting and the test face in three different equally likely conditions. In one third of the trials, both the top and bottom parts of the test face stimulus were different from those of the adapting face (Figure 82 –bottom row – "top+bottom different" condition). In a second third of the trials, only the bottom part of the test face stimulus was different from that of the adapting face, the top part being identical (Figure 82–middle row – "bottom different" condition). This condition was critical as it leads

to the perception of a new identity in the aligned and not in the misaligned format. In the last third of the trials, both the top and bottom parts of the test face stimulus were *identical* to those of the adapting face (Figure 82–top row – "same" condition). Importantly, the same pairs of faces were presented in the aligned and misaligned orientations, so that any effect cannot be attributed to the particular pairing of adapting and test faces. Subjects were instructed to attend only to the top part of the face (appearing at fixation) and to press one of two response keys (keys counterbalanced across subjects) corresponding to whether the adapting and test top parts were the same or different pictures. Note that, according to these instructions, both the "same" and "bottom different" conditions required a "same" response and only the "top+bottom different" condition required a "different" response. Subjects performed 90 trials per condition, resulting in 540 trials. The order of all the conditions was randomized within each block.

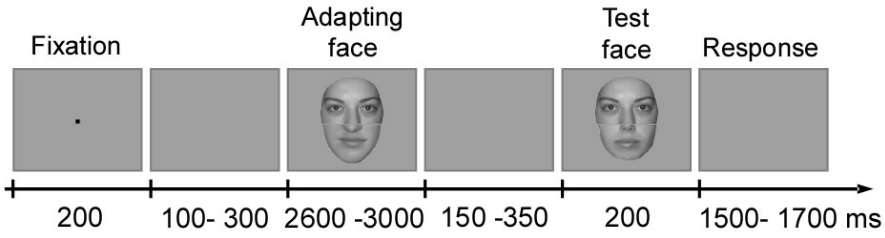

Figure 83: Time of the stimulus sequence in each trial.

EEG recording

Scalp EEG was recorded from 58 tin electrodes mounted in an electrode cap (Quik-cap, Neuroscan Inc.), with a left earlobe reference. Two pairs of bipolar electrodes were used to record vertical and horizontal eye movements. Electrode impedances were kept below 10 kΩ. EEG analog signal was digitized at a 1,000-Hz sampling rate, and a digital anti-aliasing filter of 0.27 * sampling rate was applied at recording (at 1,000 Hz sampling rate, the usable bandwidth is 0 to approximately 270 Hz). EEG data were analyzed using EEprobe 3.2 (ANT, Inc.) and custom made Matlab codes. After filtering of the EEG with a digital 30-Hz low-pass filter, time windows in which the standard deviation of the EEG on any electrode within a sliding 200-ms time window exceeded 35 μV were marked as either EEG artifacts or blink artifacts. Blink artifacts were corrected by subtraction of a vertical electrooculogram (EOG)

propagation factors based on EOG components derived from principal component analyses. Trials containing EEG artifacts were rejected. Subjects' averages were baseline corrected using the 100-ms prestimulus epoch and then re-referenced to a common average reference.

ERP component analyses

ERP amplitude analyses were performed on the visual components P1 (maximal at approximately 105 ms) and the face-sensitive N170 (maximal at approximately 165 ms) elicited by the second face of the trial sequence (test face). Amplitude values of the P1 and the N170 were measured at five pairs of occipito-temporal electrodes in the left and right hemisphere where both components were the most prominent (P7/8, P5/6, PO7/8, PO5/6, O1/2). Amplitudes were quantified for each condition as the mean voltage measured within 30-ms windows centred on the grand average peak latencies of the components' maximum. The amplitude values of each component were then submitted to separate repeated measures analyses of variance (ANOVA) with the factors alignment (aligned, misaligned), identity (*same*, *bottom different*, *top+bottom different*), hemisphere (left, right) and electrode (five levels).

Results and discussion

Behavioural results

Error rates and response times for experiment 1 are displayed in Figure 84. Error rates and response times at the face top part matching task were submitted to repeated measures ANOVAs with the factors identity (3 levels) and alignment (2 levels).

Error rates

The ANOVA performed on response times revealed a significant effect of *identity* ($F(2, 32) = 9.1$, $p < 0.001$), *alignment* ($F(1, 16) = 12.6$, $p < 0.005$) and a significant interaction between the two factors ($F(2, 32) = 22.1$, $p < 0.0001$). As expected, subjects made much more errors with aligned compared to misaligned faces in the *bottom different* condition ($p < 0.0001$). Moreover, Subjects made also more errors in misaligned compared to saligned faces in the *top+bottom different* condition ($p < 0.05$). Last, subjects made more error in the *top+bottom different* compared to *same* ($p < 0.05$), probably due to the oddity of the correct response in this condition ("different").

Response times

The ANOVA performed on response times also revealed a significant effect of *identity* ($F(2, 32) = 14.6$, $p < 0.0001$), *alignment* ($F(1, 16) = 32.3$, $p < 0.0001$) and a significant interaction between the two factors ($F(2, 32) = 23.6$, $p < 0.0001$). The interaction revealed that subjects were much slower for aligned compared to misaligned faces in the *bottom different* condition ($p < 0.0001$). Subjects were also slightly slower in aligned compared to misaligned faces in the *same* condition ($p < 0.05$) and the opposite was observed in the *top+bottom different* ($p < 0.05$). In addition, subjects were generally slower in the *top+bottom different* condition ($p < 0.005$), again probably due to the smaller number of "different" response compared to "same" response.

In summary, we observed a strong face composite effect (i.e. increased errors and response times for aligned compared to misaligned in the *bottom different* condition), replicating numerous previous studies.

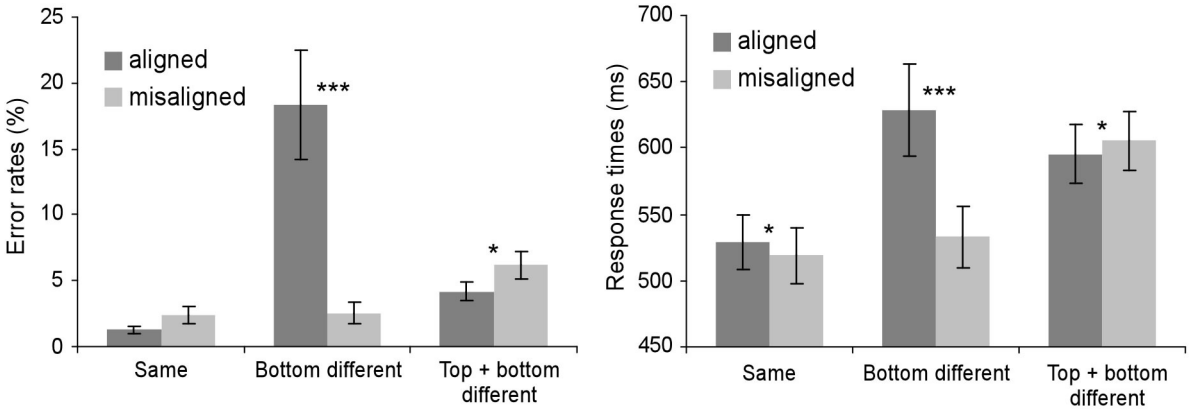

Figure 84: Behavioural results for the composite task. Error rates are shown on the left and response times on the right side of the figure.

Electrophysiological results

ERP response to adapting face

The ERP waveforms to the adapting face in the aligned an misaligned conditions (collapsed across the 3 identity conditions) are shown in Figure 85 for 2 occipito-temporal electrodes in the left (PO7) and right (PO8) hemisphere. These are shown for illustration purposes and no statistics are reported. However, as can be very clearly seen in the figure, the

N170 component elicited by misaligned faces was much larger than the N170 elicited by aligned faces. This result is similar to the above described observation (see chapter 3), and most probably reflects the disruption in configural/holistic face processing.

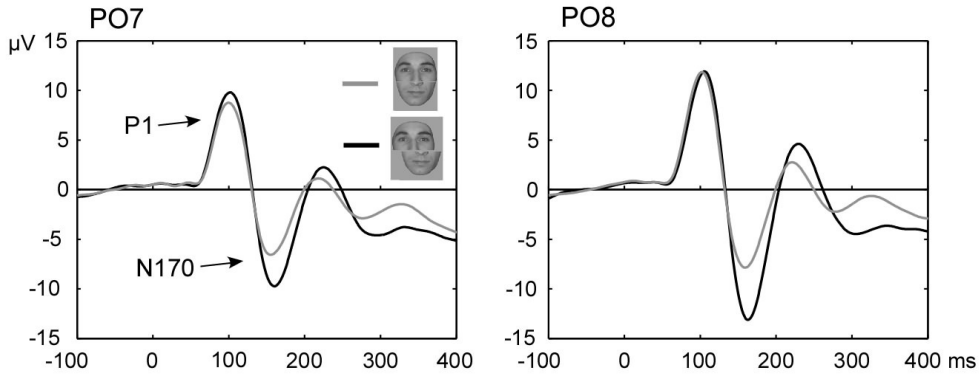

Figure 85: Grand average ERP response to the adapting face in aligned and misaligned format for two occipito-temporal electrodes (PO7- left hemisphere and PO8 – right hemisphere).

ERP response to test face

Grand average ERP waveforms elicited by the test faces in the different conditions at two ocippito-temporal electrode (PO7/PO8) are depicted in Figure 86.

P1 amplitude

A significant effect of alignment ($F(1, 16) = 13.1$, $p < 0.005$), qualified by a significant interaction between alignment and hemisphere ($F(1, 16) = 16.6$, $p < 0.001$) revealed that the P1 was larger in the misaligned condition, but only in the left hemisphere ($p<0.001$), not in the right hemisphere ($p=0.43$). There was also a significant effect of identity ($F(2, 32) = 8.8$, $p < 0.005$), indicating a larger P1 in the *top+bottom different* condition relative to *same* ($p<0.005$) and *bottom different* ($p<0.005$) conditions. The P1 amplitude for these two later conditions did no differ ($p=0.57$). Importantly, the identity x alignment interaction was not significant ($F(2, 32) = 0.8$, $p = 0.45$), indicating that no differential processing of identity between aligned and misaligned faces occurred at the P1 level. Finally, the effect of identity was slightly larger on the most posterior electrodes (PO7/8, PO5/6, O1/2) as revealed by a significant identity x electrode interaction ($F(8, 128) = 3.3$, $p < 0.05$). The identity effect was however significant on all 5 pairs of electrodes analysed.

N170 amplitude

As observed for the adapting stimulus, misaligned faces elicited a larger N170 ($F(1, 16) = 24.5$, $p < 0.005$). The main effect of identity was not significant ($F(2, 32) = 2.5$, $p = 0.1$) but, importantly, a significant interaction between alignment and identity ($F(2, 32) = 4.1$, $p < 0.05$) revealed that the effect of identity was significant in the aligned format ($p<0.01$) and not in the misaligned format ($p=0.52$) (Figure 86).

An additional ANOVA performed only on the N170 recorded in the aligned format revealed a significant interaction between identity and hemisphere ($F(2, 32) = 4.2$, $p < 0.05$). In the right hemisphere, the N170 was larger for bottom different compared to same ($p<0.01$), and for *top+bottom different* compared to same ($p<0.005$). The N170 elicited by *top+bottom different* was only marginally larger than for bottom different ($p=0.07$). In the left hemisphere, the only marginally significant effect was a larger N170 for *top+bottom different* compared to same ($p=0.09$), the two other comparisons being more clearly non-significant ($p's>0.11$).

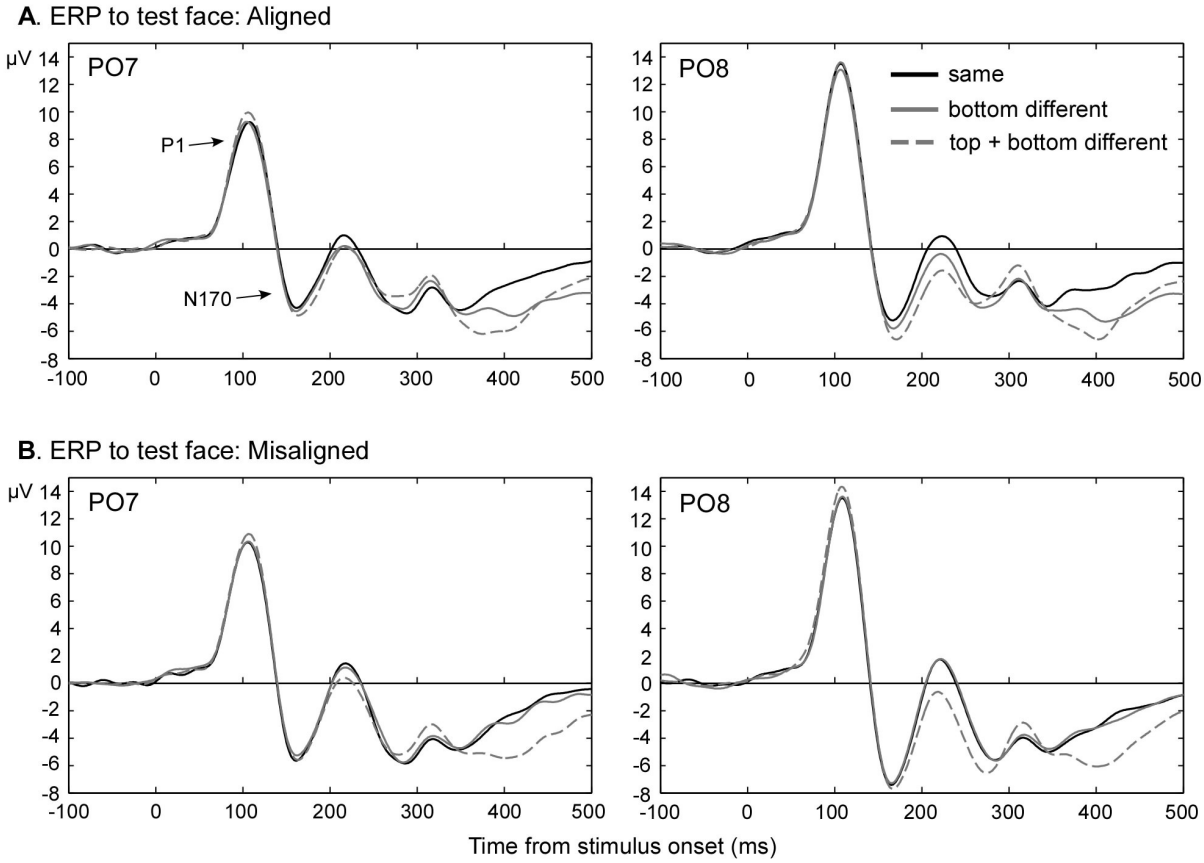

Figure 86: Grand average ERP waveforms elicited by the test face in the aligned (top row) and misaligned (bottom row) face format at electrodes PO7 and PO8. The ERP to the *same*, *bottom different* and *top+bottom different* conditions are displayed for each format.

To summarize, the N170 exhibited a graded amplitude increase from *same* to *top+bottom different* condition, the N170 to bottom different being of intermediate amplitude. Most importantly, this effect was only observed in the aligned condition, not in the misaligned, in line with the composite effect observed on behavioural data. In other words, this pattern of ERP response at the N170 is consistent with the perceptual illusion that occurs only in the aligned format that top parts of adapting and test faces are different when actually only the bottom parts of these faces are different.

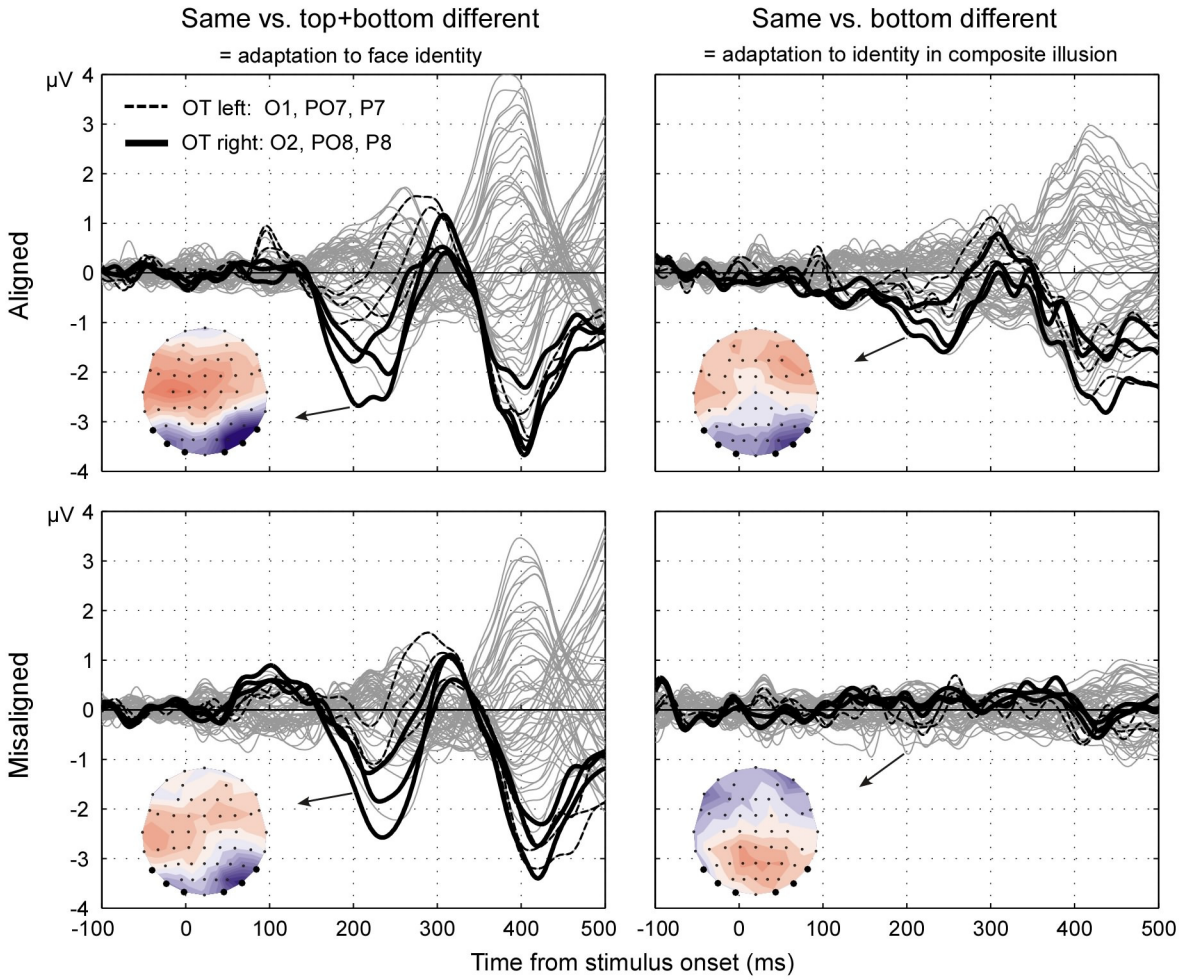

Figure 87: Subtraction waveforms comparing same to top+bottom different conditions (left) and comparing same to bottom different conditions (right) in the aligned (top row) and misaligned (bottom row) format for 58 electrodes. Highlighted electrodes are three pairs of occipito-temporal electrodes in the left (dotted black lines) and right (full black lines) hemispheres. The location of these electrodes is shown on the topographical maps in each plot. Each map shows the distribution of ERP difference at around 200 ms after stimulus onset.

The time course of individual face holistic encoding

In order to precisely characterize the timing of the adaptation effects on the ERP responses to the test face, we performed point-by-point analyses on the difference waveforms computed between the different conditions. More precisely, we first computed the difference between the ERP elicited in the *same* condition and the ERP elicited in the *top+bottom different* condition, both for aligned and misaligned faces (Figure 87, left part). Given that these subtraction waveforms involve a comparison between a condition where both the top and bottom of the face are repeated (*same*) and a condition where both the top and bottom of the face are *not* repeated (*top+bottom different*), we considered this comparison as reflecting an identity adaptation (adaptation to face identity). The resulting difference waves are plotted in Figure 87 (left part) for 58 electrodes, with 3 occipito-temporal electrodes in each hemisphere highlighted. In a second comparison, we computed the difference between the *same* condition and the *bottom different* condition, again both for aligned and misaligned faces (Figure 87, right part). These subtraction waveforms involved a comparison between the condition where both the top and the bottom of the face are repeated (*same*) and the condition in which the identity of the top of the test face was perceived as different in the aligned condition only, as the result of the composite illusion (i.e. triggered by the different bottom part (adaptation to identity in composite illusion- Figure 87).

To statistically assess the onset time at which difference waves depart from zero, we used a permutation test based on the method from Blair and Karniski (1993). This method was already used to assess the time course of adaptation effects in our previous ERP adaptation study (Jacques et al., 2007). The *p*-values (two-tailed) were computed based on the distribution of 5000 permutations at each time-point and electrode. The results of this analysis are shown in the form of topographical maps plotted every 20 ms between 0 and 300 ms (Figure 88) and between 300 and 600 ms (Figure 89) for the 4 difference waveforms computed. Only significant differences at the $p < 0.01$ threshold are represented on the topographical maps. These significant differences are color-coded according to the amplitude (in μV) of the ERP difference, with blue representing differences of negative polarity and red representing difference of positive polarity.

In the following sections we describe significant differences that lasted for more than 20 ms (to account for the temporal auto-correlation in ERP data) and included clusters of electrodes rather than single electrodes.

Interval from 0 to 300 ms

The topographical maps of the significant identity adaptations (i.e. *same* vs. *top+bottom different*) from 0 to 300 ms are shown in the first two rows of Figure 88. The earliest consistent identity adaptation effects were observed at around 160 ms in the aligned format and slightly later (180 ms) and smaller in the misaligned format. Interestingly, these differences were both restricted to occipito-temporal electrodes in the right hemisphere, with concomitant difference of opposite polarity at anterior temporal sites in the left hemisphere. This early adaptation effect lasted until about 280 ms both for the aligned and misaligned conditions.

In addition to these robust differences, earlier and smaller differences were observed in the misalign format within the P1 range (~100 ms) on a few occipito-temporal electrodes. This was due to the *top+bottom different* condition eliciting a slightly larger P1 than in the *same* condition. A very early difference (20-40 ms) was also observed on a few centro-parietal electrodes, and most likely reflects an effect due to the offset of the adapting stimulus only in misaligned format (the same difference is observed when comparing same to bottom different conditions in misaligned format). Indeed, this difference is too early to reflect cortical processing of the test stimulus, given that the earliest responses in the human primary visual cortex (V1) take place around 50–60 ms after stimulus onset (Foxe & Simpson, 2002; Clark & Hillyard, 1996).

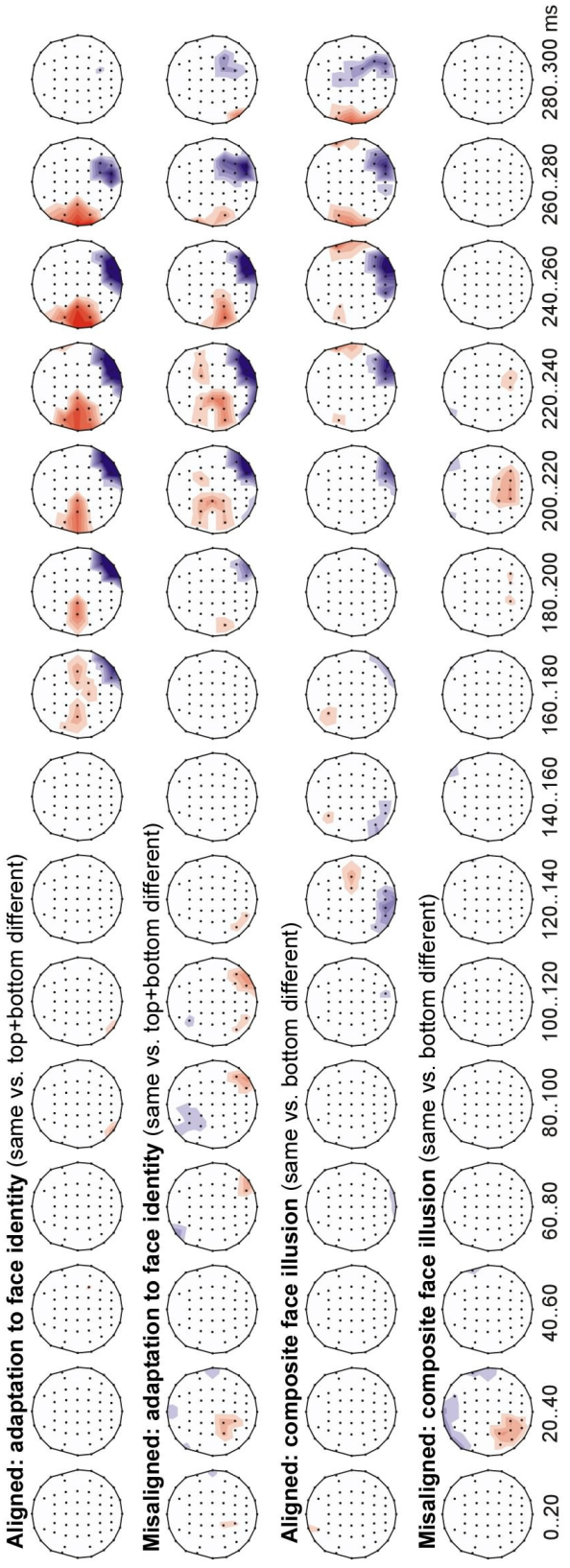

Figure 88: Topographical maps of the significant differences (p<0.01) between *same* and *top+bottom different* conditions in the aligned (first row) and misaligned (second row) format and between *same* and *bottom different* in the aligned (third row) and misaligned (fourth row) format for the interval between 0 and 300 ms after the onset of the test face. Each map shows an average of 20 ms of the respective ERP difference. The significant differences are color-coded according to the amplitude (in µV) of the ERP difference, with blue representing negative values and red representing positive values.

The topographical maps of the significant identity adaptation effect in the composite illusion (i.e. *same* vs. *bottom different*) are shown for the interval from 0 to 300 ms in the last two rows of Figure 88. In the aligned format, a significant difference (p<0.01) started at 160 ms on 3 right hemisphere low occipito-temporal electrodes (O2, PO8, P8; see Figure 87). This difference then spread to more electrodes at around 200 ms after the onset of the test face, but was still restricted to the right hemisphere. This early difference was also accompanied by an effect of opposite polarity at anterior temporal sites in the left hemisphere although of smaller magnitude to what is observed in the identity adaptation effect. The spatio-temporal properties of the adaptation effect under the face composite illusion were thus strikingly similar (although its amplitude was smaller) to those observed in the identity adaptation measured when comparing *same* to *top+bottom different* (compare left and right part of the top row of Figure 87 and compare the top row and third row of Figure 88). In stark contrast to the differences observed in the aligned format, almost no difference was observed when the same conditions are compared (*same* vs *bottom different*) in the misalign format. More precisely, the only effect observed was a small difference at around 200 ms over parietal and centro-parietal electrodes and this very early (20-40 ms) difference already described above. Thus, contrary to the aligned format that generated qualitatively comparable effects in the identity adaptation and composite illusion, the near-absence of effect in the misaligned format under the composite illusion was in strong contrast to the strong identity adaptation effect in the same format (compare left and right part of the bottom row of Figure 87 and compare the second and fourth rows of Figure 88).

In addition to the effect observed starting at 160 ms in the aligned condition in the composite illusion, an earlier difference occurred around the onset of the N170 in the 120-140 ms time window mostly over parieto-occipital electrodes (e.g. POZ, PO3) in the left hemisphere. Surprizingly, no such effect was observed when comparing *same* to *top+bottom different* conditions. A potential interpretation for this effect will be discussed later.

Interval from 300 to 600 ms

When considering later ERP responses, large differences were found between 350 ms and 450 ms after stimulus onset, distributed over bilateral occipito-temporal regions with a polarity reversal over fronto-central electrodes. These differences were observed on all comparisons except in the misaligned format when comparing *same* to *bottom different* conditions (composite illusion, last row of Figure 89). In this last comparison, the only significant difference was observed from 440 to 520 ms over centro-parietal electrodes. A last

strong difference started a little before 500 ms only when comparing *same* to *top+bottom different* conditions. This effect is due to the presence of a large ERP component, maximal over parietal regions, in the ERP response to the *top+bottom different* condition (Figure 90). This component is most probably a P300 component generated due to the fact that these conditions (aligned and misaligned) required a "different" response which had a smaller probability than "same" response. Moreover, the relatively long latency of this component may come from the fact that it may take some time to sort out the fact that in these conditions the top parts between adapting and test faces are physically different, whereas they are only perceptually different in the *bottom different* aligned condition.

Is the composite effect at the N170 due to physical changes in the bottom part?

Some may argue that the composite effect in the N170 time-window is due merely to the fact that the bottom parts of the face were physically different between the adapting and the test face, rather than to holistic integration of top and bottom face parts. However, several arguments support the latter (holistic) proposal.

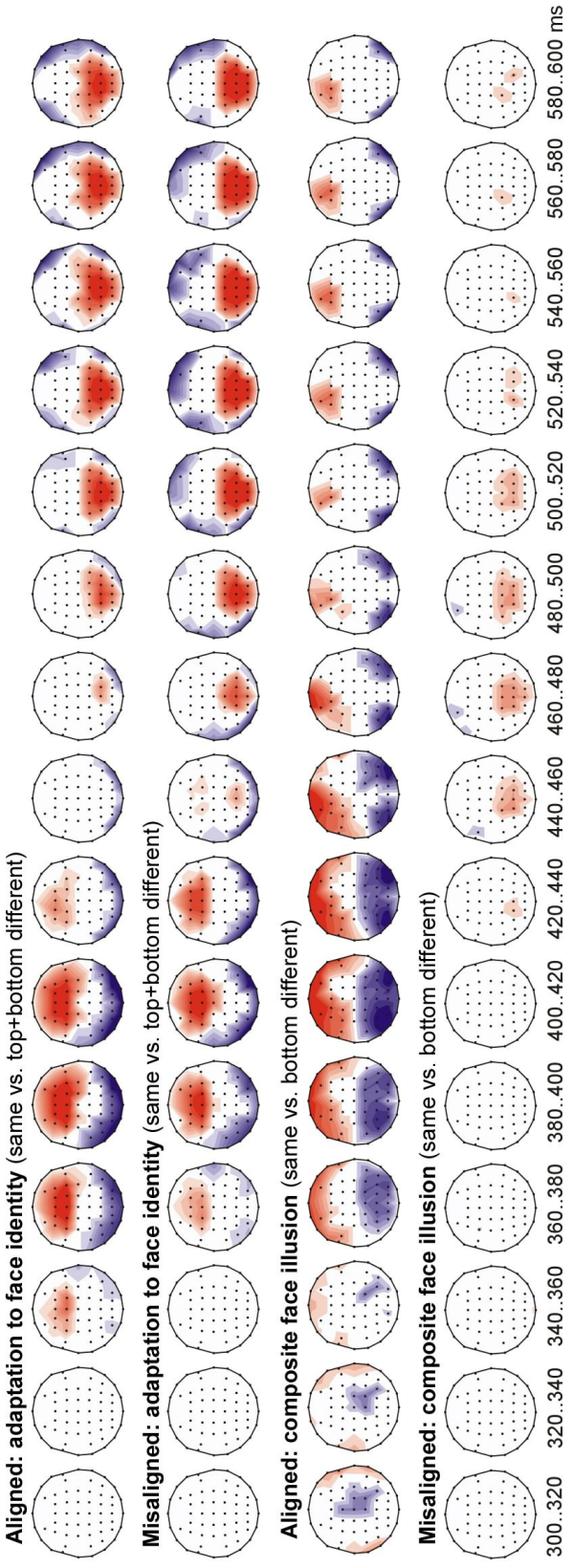

Figure 89: Identical to Figure 88 except the interval is from 300 to 600 ms. Note also that the color scale is different than in Figure 88.

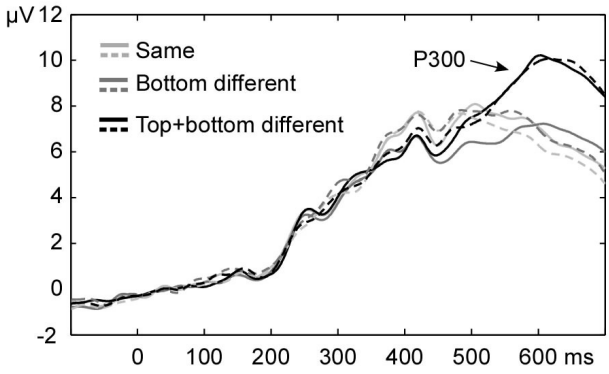

Figure 90: Grand averaged ERP waveforms at parietal electrode PZ showing the P300 for the *top+bottom different* conditions both in aligned (solid lines) and misaligned (dashed lines) face format.

First, the composite effect in the N170 time-window was observed in interaction with alignment. That is, it was observed in the aligned format, while there were very few differences even at later time points in the misaligned format when comparing *same* to *bottom-different* (Figure 87 and Figure 88). Given that the bottom part in the misaligned format was only slightly offset (~0.6° of visual angle) compared to the aligned format (see Figure 91), it is unlikely that the physical change in the bottom part would have strongly influenced ERP response in the aligned format and not in the misaligned format. As can be seen in Figure 91, the average distance between the eye region which is fixated during composite task (De Heering et al., in press), and the mouth region was about 1.6° in the aligned and 1.7° in the misaligned format. In these conditions, the mouth region in the bottom part of the composite face appears at the outer edge of the foveal region of the visual field (i.e. including a region of about 1-1.5° of radius around fixation (plotted is a dashed-line circle in Figure 91), both in aligned and misaligned formats.

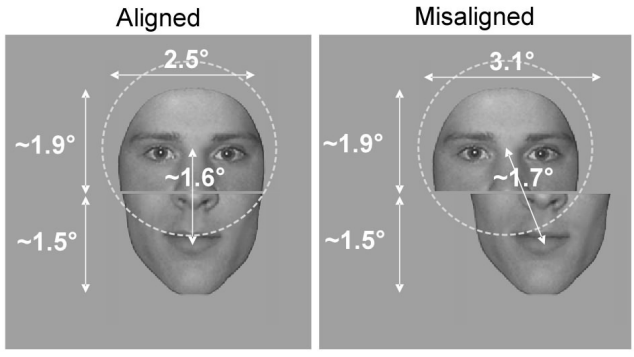

Figure 91: Size measures of composite faces in the aligned and misaligned format. The dashed circle represents a circle of 3° of diameters (roughly the size of macular visual field) centred on the putative fixation location in the composite task (see De Heering et al. in press).

Second, to strengthen our view that that the ERP composite effect at the N170 is not due to physical change in the bottom part, we correlated the amount of physical change between bottom parts of each stimulus pair used in the experiment, with the ERP amplitude in the *bottom different* condition at each time-point and each electrode. We first computed an index of image similarity between the bottom parts of each stimulus pair based on inter-stimulus pixel-wise correlations[97]. Then, for each time point and electrode, a vector composed of the z-scored ERP amplitude value for the *bottom different* condition in single-trials from all subjects (i.e. about 1400 values) was Pearson correlated with a vector composed of the bottom part inter-stimulus correlation associated with each trial in each subject. This was repeated for all electrodes and time points to obtain a map of correlation between ERP amplitude and amount of physical change in the bottom part of the composite face. Significant correlations (p<0.05) as a function of time and electrode are shown in Figure 92. Left part of the figure indicates color-coded significant correlation coefficients (R) and right part indicates the corresponding p-values.

This analysis reveals an early (~40 ms) correlation pattern in a large number of electrodes probably due to some random ERP fluctuations[98], given it very early onset time (~20 ms). Very weak correlation (p-values > 0.01) can also be seen in the P1 range on parietal, centro-parietal and pre-frontal electrodes. A correlation between physical change in the bottom face part and ERP amplitude can also be observed in the N170 time window (~150-200 ms), but only on midline fronto-central and frontal electrodes at which no ERP composite effect is actually observed (Figure 87 and Figure 88). This analysis thus further suggest that the ERP composite effect at the N170 in the aligned condition is not due to physical changes in the bottom parts between adapting and test faces. What we rather claim is that the increase of N170 amplitude in the *bottom different* relative to *same* condition, is related to the perceptual illusion of a different face identity due to holistic processing.

[97] Interstimulus pixel-wise correlation is computed between two given images using the pixel intensity values. The pixel values of each image are first vectorized (e.g. a 10 pixels x 10 pixels image becomes a 100 x 1 vector) and the two resulting vectors are Pearson correlated.

[98] The fact that this early correlation is observed on many electrodes may merely be due to the use of an average reference to analyse ERP responses.

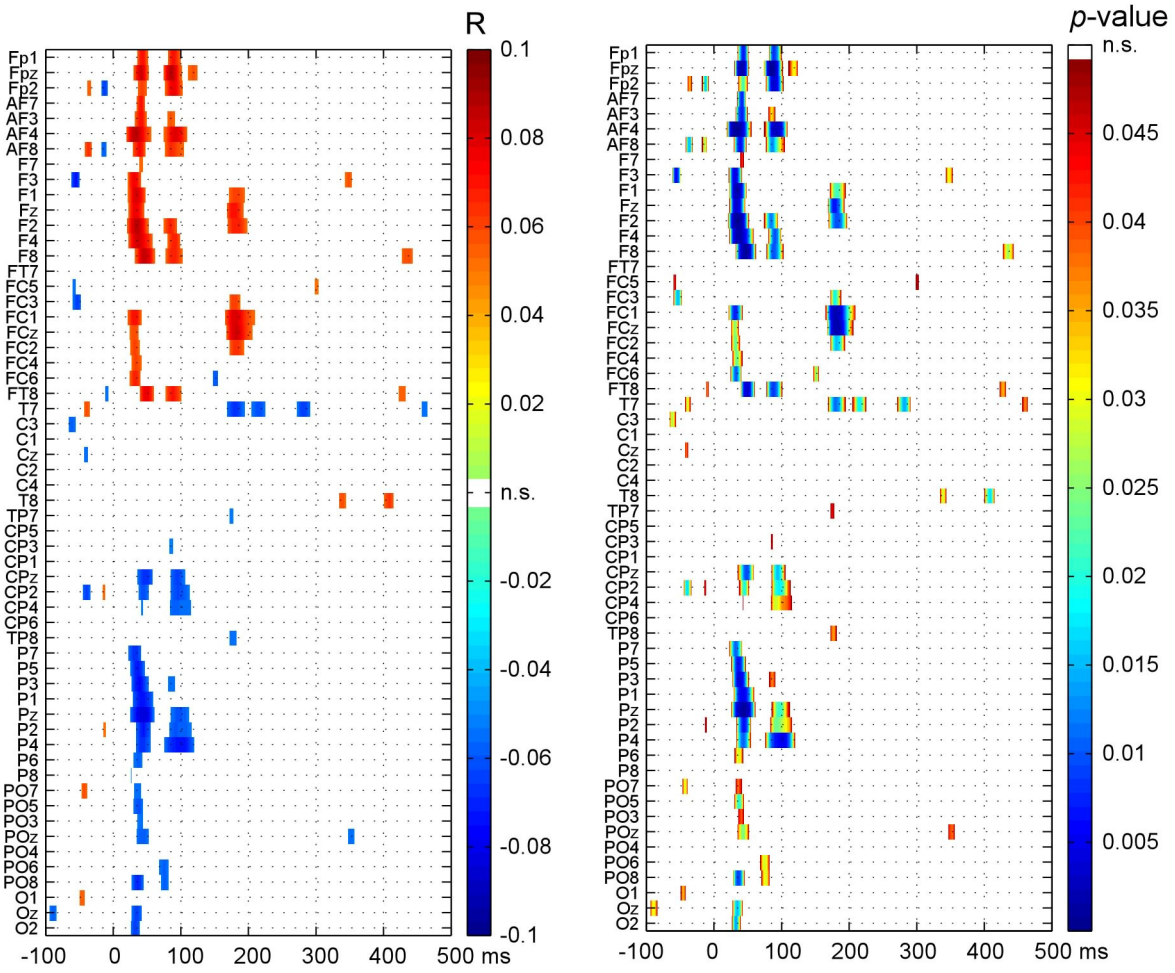

Figure 92: Results of the correlation analysis between single-trial ERP amplitudes and the amount of physical changes in the bottom parts between adapting and test faces in the aligned *bottom different* condition. Left: the map (time x electrodes) shows significant correlations ($p<0.05$) color-coded as a function of the correlation coefficient. Right: Map of the *p*-values associated with significant correlation coefficients shown in the left-hand map.

As a final point, we want to briefly comment the observation that comparing the ERP responses in the *bottom different* vs. *same* conditions revealed relatively early differences at the onset of the N170 component (120-140 ms) on a number of posterior electrodes (Figure 88; third row). This observation may seem inconsistent with the absence of such a difference when comparing *same* to *top+bottom different* conditions and with our proposal the information about facial identity becomes available in the ERP response only after 160 ms. One possible interpretation for this early difference is that it is partly due to the change in the bottom part of the face; the amplitude of this difference however not being correlated with the amount of *physical* change in the bottom (hence the absence of correlation in this time window – Figure 92). Moreover, this difference is not observed in the *top+bottom different*

condition (even though the bottom changes as well in this condition) probably because the change in the top part of the face occurring at fixation has "saturated" the neural response or is processed preferentially over the change in the bottom part[99]. In other words, in the *top+bottom different* condition, because there is already a large change in the fixated top part between adapting and test faces, an additional change of the bottom part makes little difference to the visual system. This effect may actually bare some resemblance to the phenomenon of *change blindness*, that is, the fact that an observer generally fails to detect a change between two images if a visual event (e.g. a saccade or a flickering of the image) occurs simultaneously to the change in the image (e.g. see Simons & Rensink, 2005 for a review). In this case, the visual event obscuring the change would be the fact that a different top part appears at fixation. In contrast, the absence of any physical change in the fixated top part of the face during the *bottom different* condition would allow for the change of the bottom part to be perceived and to influence the ERP response. We have however actually no explanation as to why such an effect at 120-140 ms is not observed in the misaligned condition and additional experiments may be needed to clarify this observation.

Conclusions

In this experiment we used a face composite task during ERP recording to investigate the time course with which holistic processing influence the coding of individual faces in the visual cortex. Results can be summarized as follows. First, we replicated the well-known composite effect observed in behavioural studies, namely an increase in error rate and reaction times to discriminate two identical top parts when these are aligned with two different bottom parts. Second, the test face yielded a larger ERP response, starting at around 160 ms in the N170 time-window, when its bottom part was different from the adapting face compared when it was identical to the adapting face. Most importantly, this ERP difference was observed only when the top and bottom parts were aligned, mirroring the behavioural face composite illusion. Third, this N170 composite effect and the N170 increase when both the top and bottom parts of the test face were different from the adapting face were similarly lateralized to the right hemisphere. This observation further underlines the dominant role of this hemisphere in face (holistic) processing (refs).

[99] As a matter of fact, visual information that appears at the fovea generally has a large competing advantage, being preferentially processed over information appearing in the periphery. We have seen in the complementary analysis of our first paper that such a competition bias toward the fovea is also evident in ERPs (Jacques & Rossion, 2006a).

Therefore, the present finding bring direct support to the view that holistic processing of individual faces occurs at a perceptual level, in contrast to the recent suggestions that holistic effect in the face composite paradigm occurs primarily at a decisional level (e.g. Wenger & Ingvalson, 2002; 2003). While our results do not question the fact that *part* of the behavioural composite effect is probably due to decisional response bias, they nevertheless directly point to the perceptual nature of holistic individual face representations in the occipito-temporal cortex. Furthermore, the onset time of the ERP composite effect over occipito-temporal regions in the right hemisphere (~160 ms) was similar to the identity adaptation found here and in our previous ERP adaptation study (Jacques et al., 2007), suggesting holistic integration directly participates in the activation of individual face representations. This result is also compatible with the finding that brain areas that respond preferentially to faces in the visual cortex (FFA and OFA) are involved in representing faces at an individual level (Gauthier et al., 2000), through holistic processing (Schiltz & Rossion, 2006).

In conclusion, even though the role of holistic processing in face recognition is well-known from behavioural studies, to our knowledge, the present study constitutes the first direct evidence that face holistic processing occurs during early perceptual encoding of individual faces in the visual cortex.

General discussion and perspectives

The general aim of the present thesis was to clarify the time course with which face information at different levels of resolution (e.g. category or individual level) are represented in the visual system. We relied on an electrophysiological approach allowing to measure brain responses involved in visual face processing in the human brain with a very high temporal resolution.

A tentative summary of the time course of early face information processing is presented as a time-line in Figure 93. This summary is based both on previous findings as well as on results from the empirical part of this thesis. This figure represents the visual information about face identity processed in the visual cortex within the first 400 milliseconds, focusing mostly on the time-interval between 100 and 200 ms which appears to be critical for early face visual processing. The figure is organized as follows:

- The first and second rows represent the likely sequence of the ventral stream cortical areas through which face processing probably evolves from categorization to the coding of identity. These areas are shown for the human (first row) and for some corresponding areas in the monkey brain (second row - truncated after IT for clarity).
- The third row provides some examples of the visual-face information represented as a function of time and possible visual area(s) (see first and second rows). This row also illustrates some of the possible perceptual decisions that can be made about faces based on the information accumulated at a particular time point (written in blue ink on the figure).
- The fourth row schematically represents how information about face accumulates over time and how processing this information participate in generating perceptual face representation allowing perceptual decisions at different level of resolution. Note the progressively slower rate of information integration for progressively finer-grained representations.
- The fifth row depicts how the encoding of information about face maps onto scalp onto successive scalp ERP components (C1, P1, N170, N250, N400) as measured in our and previous ERP experiments.
-The sixth row is a time-line representing the onset latency at which the ERP response significantly indexes the coding by the visual system of the information represented in third and fourth rows.

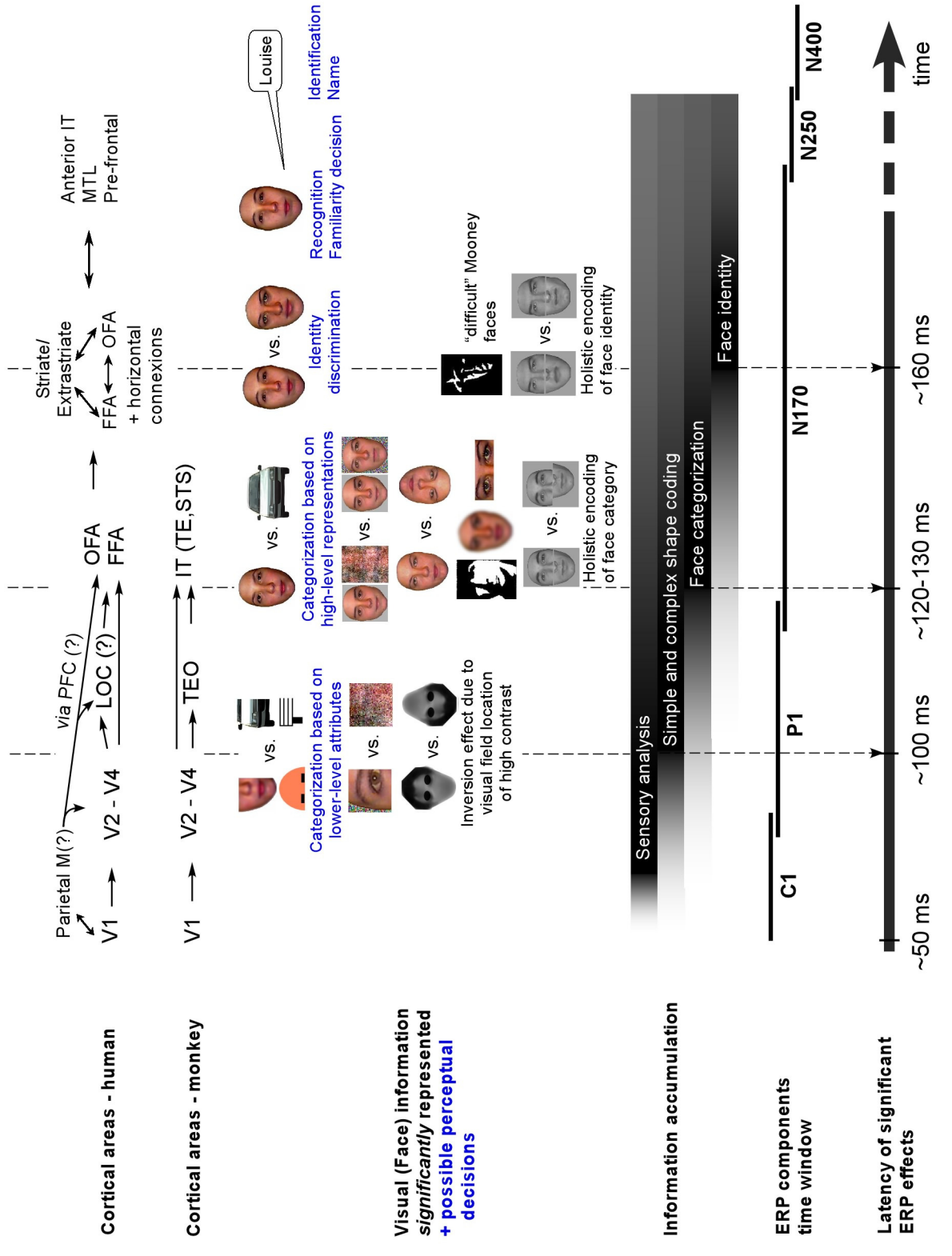

Figure 93: Time line summarizing the time course of face processing, focussing mostly on the 100-200 ms after stimulus onset time interval. Question marks in the first row indicate brain pathways for which there is an uncertainty as to whether or not they are involved in early face processing. See main text for details.

In the next paragraphs we provide a more detailed explanation about the time course of face information processing as represented in this figure.

Before 120 ms: face categorization based on low-level visual features.

In the human brain, the earliest activation of V1 neurons, which is considered the first cortical visual stage, is observed at around 50 ms after the onset of a static stimuli (e.g. a face) appearing at the fovea. We thus consider here that the cortical sensory/visual analysis starts in V1 at around this latency (see 4th and 6th rows in Figure 93). Following V1 activation, the information for object perception is transmitted to V2 and V4 in the ventral stream[100]. As can be seen in the first row of the figure, we also added a potential role of the dorsal stream areas of the parietal cortex (e.g. MT, MST) in shaping neural selectivity of the ventral stream areas. More precisely, a large number of areas of the dorsal stream are activated almost simultaneously as V1 and generally before the earliest activation in V2 or V4 (see Bullier, 2001; Lamme & Roelfsema, 2000; Schmolesky et al., 1998). This fast activation is mostly driven by magnocellular inputs from the lateral geniculate nucleus reaching V1 about 15-20 ms before parvocellular inputs (Nowak & Bullier, 1995; Maunsell et al., 1999; Schmolesky et al., 1998). Bullier and colleagues (see for example Bullier & Nowak, 1995; Bullier, 2001; Bullier et al., 2001) suggested that the result of this first coarse analysis of a visual stimulus may be fed back to V1 or other areas of the ventral stream in time to influence the earliest responses of parvocellular neurons in these areas. This early and fast visual analysis may thus serve as a header to guide visual processing in the areas of the ventral stream (Bullier, 2001). A related proposal (Bar, 2003; see also Ullman, 1995) indicates that the fast magnocellular activation of the parietal cortex may rapidly activate coarse representations in the pre-frontal cortex acting as "initial guesses" of what the encountered object may be. These "initial guesses" may then rapidly be fed back to inferior-temporal cortex to guide visual categorization. This possible route is also depicted in the first row of Figure 93 (see arrow from parietal to LOC or FFA-OFA).

Even though lower level areas of the ventral stream (represented as V2, V4, LOC or TEO in the monkey) would probably not contain face-selective neurons, information about

[100] Note that, as already discussed in chapter 2, some V2, V3 or even V4 neurons may be activated before other neurons in V1 have been activated. At this point we would like to remind the reader that this "model" certainly does not pretend to account for the complexity of the neural circuitry involved in early face visual coding.

the visual stimulus being encountered may nevertheless already starts to accumulate in these areas. This is probably especially true for areas V4 and LOC (homologous of area TEO in the monkey?) which contain neurons that are strongly sensitive to shape or colour information (e.g. Pasupathy & Connor, 1999) but not yet sensitive to stimulus category. We have indicated at various places in this thesis that faces generally differ from non-face objects on a number of low-level visual properties such as spatial frequency amplitude spectrum, color distribution, organisation of local contrast or global shape. Given these low-level visual differences, faces and other objects are likely to elicit differential responses in regions of the cortical hierarchy (e.g. V2, V4, LOC) that are not sensitive to faces *per se (i.e. features such as eyes, nose and mouth and/or their basic first-order organization)*. As also mentioned in chapter 3 of this thesis, these low-level differences are also likely to generate the amplitude difference between faces and non-face objects observed in some studies starting at around 100 ms, at the level of the P1 ERP component (e.g. Itier & Taylor, 2004a; Hermann et al., 2005). As depicted in the 3rd row of Figure 93, the lower-level visual information accumulated in these regions might actually be used to detect a face with above chance level performances. This proposal would be compatible with the finding that subjects are able to make a saccade toward a visual scene containing a face in 100-120 ms (Thorpe et al., 2007), a timing which is probably too fast to rely on the activity of face-selective neurons in occipito-temporal cortex (OTC).

Supporting the view that areas such as V4 or LOC are activated during the P1 component (~100 ms), comparing the ERP response to faces and Fourier scrambled faces indicate that stimuli containing or not a structured shape elicit strong differential brain activity at a global level at or slightly before 100 ms after stimulus onset (see complementary analysis presented in chapter 5 of the present document). Based on the fact that neurons' response latencies in humans have been estimated to roughly correspond to 5/3 of the latencies measured in the monkey brain, a human ERP effect that start at 100 ms translate roughly into a 60 ms latency in monkeys, which fits with the onset latency of neurons in V4 or TEO (~60-90 ms). Supporting this proposal, single-cell and human neuroimaging studies have indeed revealed that strong differential responses between stimuli containing structured shape compared to visual noise are generally observed starting around visual areas such as V4 or LOC in the ventral stream hierarchy.

Last, along with the finding that faces sometimes elicit a larger P1 component (see previous paragraph), the effect of face inversion observed on the P1 in some previous studies has been taken as an evidence that faces may be categorized during this time window (e.g.

Itier & Taylor, 2004a; c; see also Herrmann et al., 2005b; Liu et al., 2002). Given that in the second experiment of chapter 7 we found no correlation between the amplitude modulation of the P1 component and behavioural responses, we believe a more parsimonious account for this inversion effect at this latency is to consider that it is merely due to the differential location of high contrast regions (mostly the eyes and hair line – see 3rd row in Figure 93) with respect to fixation, between upright and inverted faces.

Around 120-130 ms: generic face categorization in the occipito-temporal cortex

Based in the results of our first experiment (Chapter 5), we have suggested that when multiple faces are presented concurrently in the visual field, they compete for overlapping visual representation in the OTC. Such competition between visual representations has been most thoroughly described in the framework of the biased competition model for visual attention[101] developed by Desimone and colleagues (e.g. Desimone, 1998). This model has received much empirical support (see Desimone, 1998; Reynolds & Chelazzi, 2004; Kastner & Ungerleider, 2000 for reviews) which we won't describe here. Importantly however, this model posits that competition for neural representation between multiple objects should take place to the extent that these objects normally recruit common neural populations (Desimone, 1998). Because faces are coded by overlapping neural populations in the monkey's IT cortex (Young & Yamane, 1992; Rolls & Tovee, 1995b), we hypothesized that the earliest time point at which we observed an ERP competition effect in our concurrent processing experiment (~125-130 ms; see 3rd row of Figure 93, second cluster of pictures) reveals the time point at which face neural representations (localized in humans in the FFA and OFA – see 1st row in Figure 93) are activated in the human OTC.

In the second and third studies of chapter 7, we measured the effect of parametrically varying face orientation on the ERP response recorded at scalp level. Both the computation of

[101] The biased competition hypothesis (Desimone and Duncan, 1995) suggests that when multiple stimuli are presented in nearby locations, they activate populations of neurons that interact in a mutually suppressive way. Such competitive interactions are though to index competition for neural representation, thus reflecting the limited capacity of single neurons to represent and selectively respond to multiple stimuli present simultaneously in their receptive field. According to this biased-competition model for attentional selection, selective or spatial attention mechanisms give a processing advantage to an attended stimulus or attended location when several competing stimuli are present in the visual field (e.g. Luck *et al.*, 1997; Chelazzi *et al.*, 1998; Reynolds *et al.*, 1999). Hence, when multiple stimuli – e.g. two faces – compete for a neuron's response, if one of the stimuli is attended, the response of the neuron is biased toward that elicited when this attended stimulus is presented alone, thereby partly eliminating the influence of the unattended stimulus on the neuron's response.

point-by-point correlation between the electrophysiological and behavioural responses and the computation of the standard deviation between the ERP responses to different face orientations revealed that this parametric manipulation starts to modulate face processing at around 115-120 ms (maximal at 130 ms), in the slope between the P1 and the N170 peaks. We note that this range of latencies is very similar to the ERP face competition effect that we tentatively linked with the onset of face neural representations in OTC (see above). Compatible with this proposal, given that face inversion impairs both face detection and face recognition, we hypothesized that this early ERP effect, which takes the form of a delay in the onset of the N170, **reflects the delayed activation of face neural representations** (i.e. delayed face categorization) in OTC due to abnormal face orientation. This proposal is perfectly in line with the findings of slightly higher response times to detect inverted relative to upright faces (Lewis & Edmonds, 2003; 2005; Rousselet et al., 2003; see also Purcell & Stewart, 1988). The proposal that face orientation modulates early face categorization is perfectly compatible with the finding that face-neurons modulate their firing rate to different face orientations from the very onset of their response (see Tsao et al., 2006, supplementary material; see also Perrett et al., 1982; 1985a). We also note that two-tone Mooney images or Arcimboldo paintings, in which a face is easily perceivable based on holistic processing[102] (see Mooney face in the 3rd row of Figure 93, second cluster of pictures), elicit a larger N170 than similar images which are not perceived as faces[103] (George et al., 2005; Latinus & Taylor, 2005).

To conclude, the latencies observed for early ERP effects both in the competition paradigm and with the manipulation of face orientation (~120-130 ms) are compatible with the onset latencies of the earliest face-selective responses in face-cells of the inferior-temporal cortex of the monkey. Therefore, we suggest that this latency reflect the earliest time point at

[102] In the first row of Figure 93, we have included a potential direct projection from extrastriate areas such as V2 or V4 to the FFA, thereby by-passing the LOC and OFA. This direct pathway has been proposed based on the observation that prosopagnosic patients with lesions of the OFA can have a normal activation of the FFA (Rossion et al., 2003b; Steeves et al., 2007) and the recent finding from diffusion tensor imaging (DTI) in the human brain that areas V3 and V4v are connected to the FFA. Therefore, one possible role for these connexions would be to participate in the rapid activation of a coarse holistic face representation in the FFA where neurons probably have receptive fields sufficiently large to represent the whole face (Rossion, in press; see also Schiltz & Rossion, 2006). Refinement of the face representation needed to generate an individual representation would then call on the OFA and other lower-level visual areas which contain neurons with smaller receptive fields.

[103] Note however that for some Mooney images, the face is not as readily perceivable as in other images and it may take some additional time to generate a full face percept. Indeed, for these Mooney faces, holistic processing may have to rely strongly on interactions with low-level visual areas for illusory contour completion to take place, in order to bind image parts into a face representation. Therefore, these more difficultly perceived Mooney faces are unlikely to elicit a larger N170 than Mooney images that do not contain a face. The additional time that may be needed to detect these Mooney faces is represented in the 3rd row of Figure 93.

which faces are categorized as such based on high-level visual representations (see 3rd row of Figure 93, second cluster of pictures). This latency is fully compatible with the observation that the ERP response to faces generally separate from the ERP response to non-face objects at around 120-130 ms (at the onset of the N170), even though the precise latency has been vary rarely measured (see Rousselet et al., 2005 for an exception[104]). Finaly, these ERP latencies are in keeping with the view that fast visual categorization (such as revealed in behavioural studies with RSVP, masking or go/no-go paradigms) relies on the global information rapidly transmitted through the visual cortical hierarchy. This fast information transmission about the global category may rely mostly on the earliest neural responses at each cortical stage and transmitted through feed-forward connections (Thorpe et al., 1996; VanRullen & Thorpe, 2001b; see also Lamme & Roelfsema, 2000).

After 150 ms: representation of individual faces within and beyond the occipito-temporal cortex

As a final main result of our empirical work in this thesis, we found that measures of the time course of face identity coding, through continuous face stimulation (chapter 6), identity adaptation (first and third experiments of chapter 7, chapter 8) or computations of correlation between ERP signal and face discrimination (second experiment of chapter 7), consistently reveal that information about face identity starts to be significantly represented in the ERP response during the N170 time-window starting at around 160 ms after face onset (see 3rd row in Figure 93, third cluster of pictures). Together with our proposal that face neural representations starts to be activated at around 120-130 ms, these findings suggest that the finer-grain information needed to represent face identity is accumulated at a slower rate than the coarser information allowing for face categorization (see 4th row in Figure 93). The finding of a 30 to 40 ms delay between the significant coding of fine and coarse information about faces actually fits well with the observation that a large proportion of face-neurons in the monkey IT cortex are involved in representing face information at different levels of resolution with information about face category and face identity being significantly represented at different time windows of the neurons' responses (Matsumoto et al., 2005; Sugase et al., 1999). More precisely, at a neural population level an average delay of about 50

[104] It seems however that the latency of the difference between faces and houses in Rousselet et al. (2005) is found slightly later than 130 ms. This may be due to the fact that the stimuli in the two categories used in this experiment were well controlled for low-level visual properties but were consequently strongly degraded and noisy.

ms is found between the availability of coarse (category level) information (90-140 ms) and fine (individual level) information (140 – 190 ms) (Matsumoto et al., 2005). Interestingly, if at least partly similar populations of neurons are involved in both face categorization and encoding of face identity, our findings would suggest that while face categorization is mostly based on information provided by the rapid feed-forward sweep of activation through the ventral stream (see previous section), the accumulation of finer-grained information needed to represent face identity relies much more strongly on the interactions between neurons within areas of OTC (through horizontal connexions), on interactions between different regions preferentially activated for faces (i.e. OFA and FFA) and interactions with lower-level visual areas (through recurrent bi-directional connexions) (see 1st row in Figure 93). This proposal stems from the hypothesis that the refinement of visual representations in high-level visual areas strongly depends on recurrent interactions with lower-level visual areas (e.g. V1-V2) that are able to represent more local and precise information given their smaller receptive fields and their selectivity for simple features (see Lee et al., 1998; Lamme & Roelfsema, 2000; Bullier, 2001; Hochstein & Ahissar, 2002). In addition, the finding that a lesion of the OFA in a prosopagnosic patient leads to an absence of sensitivity to face identity in the FFA, suggests a critical role of the recurrent interaction between FFA and OFA in the coding of face identity (Rossion et al., 2003b; Schiltz et al., 2006). Moreover, horizontal connexions have also been linked to the fine tuning of neurons selectivity in IT (see Wang et al., 2000).

In addition and complementary to the proposal exposed in the previous paragraph, the delay between the representation of face category and face identity in IT may also be related to the differential speed at which information useful for these different levels of categorization are integrated by the visual system (Sergent, 1986; see also Bullier, 2001). More precisely the information from different spatial frequency bands are not represented simultaneously in the visual system, at least in low-level structures such as LGN or V1. Information in the low spatial frequency bands (LSF), transmitted by both magnocellular and parvocellular pathways, is processed faster than high spatial frequencies (HSF) which are transmitted mostly by the parvocellular pathway (see Maunsell et al., 1999; Schmolesky et al., 1998; Bullier, 2001; Bullier & Nowak, 1995). This may thus partly explain why generic face categorization, which can be performed based on rapidly transmitted LSF representations, is performed faster than individual face processing which is based on the accumulation of both LSF and slower HSF information (see Sergent, 1986; Schyns & Oliva, 1994; Loftus & Harley, 2004). In other words, in this framework the slower representation of fine-grained compared to coarse visual information stems from the fact that LSF information reaches high-

level visual areas while HSF information is still being processed in lower-visual areas. This view that fine (HSF) information about faces reaches high-level visual cortex with a delay compared to LSF information (Sergent, 1986; Goffaux & Rossion, 2006) is compatible with the proposal that visual perception evolves from *coarse-to-fine* representations about the visual environment (Sergent, 1986; Schyns & Oliva, 1994; Loftus & Harley, 2004). Moreover, this view is also in agreement with the proposal that a first-pass analysis of information from the magnocellular pathway may serve as a header to guide visual processing throughout the ventral stream (Bullier, 2001; Bar, 2003 – see above).

The latency difference between the ERP effects that we attribute to face categorization (~120-130 ms) and individual face coding (~160 ms) is also in agreement with the differences between the minimal response times observed in a behavioural face detection task (~250 ms; Rousselet et al., 2003) and in our face discrimination tasks (290-300 ms; see complementary results in Chapter 7) for upright faces. Interestingly, this 40-50 ms difference observed with *minimal* response times transforms into a much larger difference when taking the *mean* response times in these studies (380-400 for face detection vs. 550-600 ms for face discrimination, thus 150-200 ms difference). This is due to the fact that there is much more inter-trial variability in discrimination than in detection response times, compatible with the view that the variance in the rate of information integration increases as a function of the resolution of the information needed for the perceptual decision (see Ratcliff, 2006).

Whereas our analysis of reaction times distributions in individual discrimination tasks reveals that the earliest correct responses are produced at around 300 ms, even faster reaction times may probably be expected using a more sensitive experimental design and apparatus, such as used in go/no-go experiments (VanRullen & Thorpe, 2001b). If we consider that a minimum of 100 ms is needed to generate a motor command (see VanRullen & Thorpe, 2001b), our reaction time analyses also points to a coding of face identity before 200 ms, compatible with our ERP findings.

Also compatible with the view that face perception proceeds in a continuous accumulation of information, we found that vertical inversion both reduced the ERP identity adaptation effect and delayed the onset of a significant face adaptation effect by about 30 ms. This means that the time needed for the visual system to discriminate between 2 faces increases when face orientation deviates from the upright orientation, mirroring behavioural responses. Preliminary results from the last experiment of Chapter 7 partly suggest that intermediate orientations between upright and inverted face orientations lead to intermediate latency/amplitude of identity adaptation. However, further testing is needed to determine

whether or not the pattern of latency modulation of the adaptation effect at the N170 as a function of face orientation parallels the pattern of behavioural modulation in a face discrimination task. Such a finding would provide a stronger case for the view that the perceptual decision in face discrimination is –at least partly- based on the early neural activity in OTC around the N170 time-window.

In a last experiment (chapter 8), results obtained using an adaptation design with the composite face illusion strongly suggest that the early coding of face identity in OTC relies on holistic face representations, in line with recent findings that face identity is represented holistically in the FFA and to a lesser extent in the OFA (Schiltz & Rossion, 2006). This result, obtained with a technique providing fine temporal resolution, is actually the first evidence that face identity is processed holistically *during* perceptual encoding, in contrast to the recently proposed view that holistic effects in face processing are due solely to decisional response bias (Wenger & Ingvalson, 2002; 2003). This finding is crucial in our opinion because it brings a clear answer on the nature of the individual face representations in the OTC. Hence, if this result is confirmed, any description of how face identity is coded in the visual cortex should be able to account for the holistic nature of individual face representations.

In addition to the identity adaptation starting during the N170 time window, similar effects were also found at later time points during the time-window of the so-called N250 component (Schweinberger et al., 1995; 2004). The sensitivity to identity repetition of the N250 has been consistently described previously (e.g. Schweinberger et al., 2002a; 2004), in contrast to the N170 effect (see Chapter 3), and probably reflects the fact that the information accumulated within OTC at this latency allows for a more "robust" representation of individual faces. In our experiments, we found that the N170 and the N250 show relatively similar sensitivity to face orientation (see chapter 7), suggesting that the effects in the N250 and N170 time ranges involve largely overlapping face neural representations and should not be considered as distinct stages of face processing. However, the finding that the N170 and N250 adaptation/repetition effect have slightly different dipolar neural sources (Schweinberger et al., 2002) and scalp topographies (see first experiment in chapter 7; Jacques et al., 2007) suggest additional visual or non-visual regions (such as medial temporal lobe – MTL, see Quiroga et al., 2005) may be implicated in the N250 effect.

Finally, we have seen in Chapter 3 that the presentation of familiar vs. unfamiliar faces or the delayed repetition of a previously seen face generally elicits a differential ERP response at around 300-500 ms over central or frontal regions, during the time window of the N400

(e.g. Paller et al., 2000; Bentin & Deouel, 2000) and at a later time-point (500-700 ms) over parietal regions. As briefly mentioned in chapter 3, these ERP differences are likely to be generated by large cortical networks including visual and non-visual areas (such as MTL, AIT, pre-frontal cortex – PFC; see 1st row) involved both in the decision about face familiarity and in the recollection of semantic information and name of familiar individuals.

Conclusion: both face detection and the coding of face identity occur during the time window of the N170

As noted in Chapter 3, it is largely agreed among ERP researchers investigating face processing, that the N170 merely represents a visual stage during which faces are categorized as faces (i.e. face detection; e.g. Bentin et al. 1996; Bentin & Deouell, 2000; Eimer, 2000d; Schweinberger et al., 2002a; Zion-Gollumbic & Bentin, 2007). According to these studies, the access to face individual representations, or the recognition of familiar faces would take place at later time points, during the N250 or the N400 time windows (e.g. Bentin & Deouell, 2000; Eimer, 2000d; Schweinberger et al., 2002a; 2004). As we have seen throughout this dissertation, this proposal is incompatible with findings both from the monkey neurophysiology and psychophysical studies in humans. Therefore, based on the results of our empirical work, we propose that *both* face detection and the activation of individual face representations take place during the time window of the N170. This proposal does not however imply that such a fast coding of individual faces should occur in every situation. The speed at which an individual representation about an encountered face is extracted will depend on numerous external and internal factors. Moreover, as already noted, the processing of face identity does not stop after the time point at which just enough evidence has been gathered to generate a motor response. Additional time is certainly needed to reach a fully detailed representation that will allow the extraction of fine-grained information about the individual face. Additional time is also required, relative to this early individual face coding, to consciously recollect various semantic information about the person encountered.

The proposal that both face detection and face individualization take place at the N170 is clearly at odd with the conventional "*ERP-component-as-functional-stage*" view generally advocated in psychophysiology. Through our work in this dissertation, we hope to have illustrated that ERP components, such as the N170, should not be described as functional stages. Therefore, we propose that rather than concentrating on describing what neural or

cognitive process a particular ERP component reflects, future ERP research on face processing should focus on methodologies that use such electrophysiological responses as a tag in order to reveal temporal as well as functional properties of face perception.

This proposal probably calls for a revision of the approach classically used in face ERP research. This approach generally consists in manipulating a few dimensions in a stimulation to modulate the N170 response, without any precise knowledge as to what this manipulation entails in terms of face processing or face representation. In our empirical work, we have, as much as possible, attempted to avoid this approach. Instead, we mostly relied on the combination of two other main methods. The first one implies measuring the response to the same face while varying the context parameters in which this face appears (e.g. in the competition, identity reversal and adaptation experiments). This approach allows obtaining a relatively pure measure of the manipulated parameters, minimizing potential confounds with low-level visual factors that necessarily have to be taken into account when directly manipulating the stimulus. The second approach relies on a parametric stimulus manipulation to generate variability in the ERP response. Combined with a measure of behavioural responses, this second approach may be a powerful tool to address many unresolved questions in the field of face processing.

More generally, we believe that a way to avoid focusing on specific ERP components is to perform precise temporal analysis such as point-by-point statistics as used in our experiments. These analyses allows revealing the exact latency at which an effect is reflected on the ERP response, providing a better understanding of the temporal dynamics of perceptual processes, than analyses that rely purely on the measure of specific ERP components.

Obviously, we are not the first to claim that ERP components should not be apprehended as if they were functional stages. Also the type of approach and analyses we have used in our empirical work have been used previously in other fields of research, such as the field of visual attention (e.g. Luck et al., 2000; Mangun, 1995), or researches carried out to determine the time course of visual scene categorization (VanRullen & Thorpe, 2001; Rousselet et al., 2007b). These have relied on approaches manipulating the instruction given to the subjects, rather than directly manipulating the stimulation.

The goal of this dissertation was to enhance our knowledge of the precise time frame for face categorization, at the basic and individual levels. With this information in hand, future work should concentrate on determining the nature of the face information that is represented during these time windows. We have seen in Chapter 3 that there is a large

amount of evidence suggesting that both local facial features and face configural information are represented during the N170. Other type of facial cues that appear crucial in determining ones identity, such as skin texture or colour or general three dimensional shape of the face, have to date received little interest in ERP studies of face processing, probably owing to the difficulties in manipulating these cues separately (but see Hill et al., 1996; Russell et al., 2007) and the lack of knowledge of psychophysical data by the ERP community of researchers. Thus, it remains to be determined precisely how and when these multiple facial cues combine and participate in face detection and face individualization. We believe the empirical approach used and developed in our empirical work may be particularly well suited to address these critical questions in the field of face processing.

Bibliographie générale

Afraz, S. R., Kiani, R., Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. Nature, 442, 692-695.

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

Allison, T., Puce, A., McCarthy, G. (2002). Category-sensitive excitatory and inhibitory processes in human extrastriate cortex. J Neurophysiol, 88, 2864-2868.

Allman, J., Miezin, F., Mcguinness, E. (1985). Stimulus Specific Responses from Beyond the Classical Receptive-Field - Neurophysiological Mechanisms for Local Global Comparisons in Visual Neurons. Annu Rev Neurosci, 8, 407-430.

Anaki, D., Zion-Golumbic, E., Bentin, S. (2007). Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. Neuroimage, 37, 1407-1416.

Andrews, T. J., Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. Neuroimage, 21, 91-98.

Andrews, T. J., Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage, 23, 905-13.

Ashbridge, E., Perrett, D. I., Oram, M. W., Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. Cognitive Neuropsych, 17, 13-34.

Avidan, G., Behrmann, M. (2002). Correlations between the fMRI BOLD signal and visual perception. Neuron, 34, 495-497.

Avidan, G., Hasson, U., Hendler, T., Zohary, E., Malach, R. (2002). Analysis of the neuronal selectivity underlying low fMRI signals. Curr Biol, 12, 964-72.

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-Psych Fo, 69, 11-21.

Bacon-Mace, N., Mace, M. J. M., Fabre-Thorpe, M., Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. Vision Res, 45, 1459-1469.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cognitive Neurosci, 15, 600-609.

Bar, M. (2004). Visual objects in context. Nat Rev Neurosci, 5, 617-629.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? Perception, 1, 371-394.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? Perception, 1, 371-394.

Baylis, G. C., Rolls, E. T., Leonard, C. M. (1985). Selectivity between Faces in the Responses of a Population of Neurons in the Cortex in the Superior Temporal Sulcus of the Monkey. Brain Research, 342, 91-102.

Baylis, G. C., Rolls, E. T., Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. Journal of Neuroscience, 7, 330-42.

Bentin, S., McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. Journal of Experimental Psychology: Learning Memory and Cognition, 20, 130-149.

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

Bentin, S., Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cognitive Neuropsych, 17, 35-54.

Bentin, S., Golland, Y. (2002). Meaningful processing of meaningless stimuli: The influence of perceptual experience on early visual processing of faces. Cognition, 86, B1-B14.

Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., von Cramon, Y. D. (2002). Priming visual face-processing mechanisms: Electrophysiological evidence. Psychol Sci, 13, 190-193.

Bentin, S., DeGutis, J. M., D'Esposito, M., Robertson, L. C. (2007a). Too many trees to see the forest:

Performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. J Cognitive Neurosci, 19, 132-146.
Bentin, S., Taylor, M. J., Rousselet, G. A., Itier, R. J., Caldara, R., Schyns, P. G., Jacques, C., Rossion, B. (2007b). Much ado about nothing: controlling for interstimulus perceptual variance does not abolish N170 face sensitivity. Nat Neurosci, 10, 801-802.
Biederman, I., Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philos Trans R Soc Lond B Biol Sci, 352, 1203-19.
Blanz, V., Tarr, M. J., Bulthoff, H. H. (1999). What object attributes determine canonical views? Perception, 28, 575-599.
Bodamer, J. (1947). Die-Prosop-agnosie. Archives Psychiatrie Nervenkrankh, 179, 6-54.
Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. Trends Cogn Sci, 11, 118-125.
Bosworth, R. G., Bartlett, M. S., Dobkins, K. R. (2006). Image statistics of American Sign Language: comparison with faces and natural scenes. J Opt Soc Am A Opt Image Sci Vis, 23, 2085-96.
Botzel, K., Grusser, O. J. (1989). Electric Brain Potentials-Evoked by Pictures of Faces and Non-Faces - a Search for Face-Specific Eeg-Potentials. Exp Brain Res, 77, 349-360.
Botzel, K., Schulze, S., Stodieck, S. R. G. (1995). Scalp Topography and Analysis of Intracranial Sources of Face-Evoked Potentials. Exp Brain Res, 104, 135-143.
Boutsen, L., Humphreys, G. W., Praamstra, P., Warbrick, T. (2006). Comparing neural correlates of configural processing in faces and objects: An ERP study of the Thatcher illusion. Neuroimage, 32, 352-367.
Bouvier, S. E., Engel, S. A. (2006). Behavioral deficits and cortical damage loci in cerebral achromatopsia. Cereb Cortex, 16, 183-191.
Braeutigam, S., Bailey, A. J., Swithenby, S. J. (2001). Task-dependent early latency (30-60 ms) visual processing of human faces and other objects. Neuroreport, 12, 1531-6.
Breitmeyer, B. G., Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. Perception & Psychophysics, 62, 1572-1595.
Bruce, C., Desimone, R., Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol, 46, 369-384.
Bruce, V., Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77 (Pt 3), 305-327.
Bruce, V., Humphreys, G. W. (1994). Recognizing objects and faces. Vis Cog, 141-180,
Bruyer, R., Galvez, C., Prairial, C. (1993). Effect of Disorientation on Visual Analysis, Familiarity Decision and Semantic Decision on Faces. British Journal of Psychology, 84, 433-441.
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, 497-503.
Bullier, J. (2001). Integrated model of visual processing. Brain Res Rev, 36, 96-107.
Bullier, J., Hupe, J. M., James, A. C., Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. Prog Brain Res, 134, 193-204.
Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., Herrmann, C. S. (2004). Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. Clin Neurophysiol, 115, 1810-1820.
Busey, T. A., Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. Vision Res, 45, 431-48.
Caharel, S., Poiroux, S., Bernard, C., Thibaut, F., Lalonde, R., Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. Int J Neurosci, 112, 1499-1512.
Caharel, S., Courtay, N., Bernard, C., Lalonde, R., Rebai, M. (2005). Familiarity and emotional expression influence an early stage of face processing: An electrophysiological study. Brain Cognition, 59, 96-100.
Caharel, S., Fiori, N., Bernard, C., Lalonde, R., Rebai, M. (2006). The effects of inversion and eye displacements of familiar and unknown faces on early and late-stage ERPs. Int J Psychophysiol, 62, 141-151.
Caharel, S., Lalonde, R., Bernard, C., Leleu, A., Rebaï, M. (in preparation). The N170: an electrophysiological index of visual categorization dependent on the level of familiarity with objects and face formats.

Caldara, R., Jermann, F., Arango, G. L., Van der Linden, M. (2004). Is the N400 category-specific? A face and language processing study. Neuroreport, 15, 2589-93.
Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., Hauert, C. A. (2006). The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. Neuroimage, 31, 313-9.
Calder, A. J., Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci, 6, 641-651.
Campanella, S., Hanoteau, C., Depy, D., Rossion, B., Bruyer, R., Crommelinck, M., Guerit, J. M. (2000). Right N170 modulation in a face discrimination task: an account for categorical perception of familiar faces. Psychophysiology, 37, 796-806.
Carmel, D., Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition, 83, 1-29.
Cauquil, A. S., Edmonds, G. E., Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention? Neuroreport, 11, 2167-71.
Cheal, M. L., Lyon, D. R. (1991). Central and Peripheral Precuing of Forced-Choice Discrimination. Q J Exp Psychol-A, 43, 859-880.
Chelazzi, L., Duncan, J., Miller, E. K., Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol, 80, 2918-40.
Chun, M. M., Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. Cognitive Psycholology, 36, 28-71.
Clark, V. P., Fan, S., Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum Brain Mapp, 2, 170-187.
Collin, C. A., Liu, C. H., Troje, N. F., McMullen, P. A., Chaudhuri, A. (2004). Face recognition is affected by similarity in spatial frequency range to a greater degree than within-category object recognition. J Exp Psychol Human, 30, 975-987.
Collishaw, S. M., Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? Perception, 31, 287-296.
Crouzet, S., Thorpe, S., Kirchner, H. (2007). Category-dependent variations in visual processing time. J Vision, /7/9/922/, J. Vision abstract.
Curran, T., Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: an ERP study. Psychophysiology, 40, 979-88.
Curran, T., Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. Neuroimage, 36, 464-471.
Damasio, A. R., Damasio, H., Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. Neurology, 32, 331-341.
Damasio, A. R., Tranel, D., Damasio, H. (1990). Face Agnosia and the Neural Substrates of Memory. Annu Rev Neurosci, 13, 89-109.
Dawson, G. D. (1954). A summation technique for the detection of small evoked potentials. Electroencephalogr Clin Neurophysiol, 6, 65-84.
De Baene, W., Premereur, E., Vogels, R. (2007). Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J Neurophysiol, 97, 2900-2916.
de Haan, M., Pascalis, O., Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. J Cognitive Neurosci, 14, 199-209.
DeAngelis, G. C., Ghose, G. M., Ohzawa, I., Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. Journal of Neuroscience, 19, 4046-4064.
Deco, G., Rolls, E. T. (2004). A Neurodynamical cortical model of visual attention and invariant object recognition. Vision Res, 44, 621-642.
Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. Neuroimage, 35, 1495-1501.
Dehaene, S., Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition, 79, 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 Res, 40, 2187-2200.
Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., Orban, G. A. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: A functional magnetic resonance imaging study. Journal of Neuroscience, 24, 2551-2565.
Desimone, R., Albright, T. D., Gross, C. G., Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience, 4, 2051-2062.
Desimone, R. (1991). Face-Selective Cells in the Temporal Cortex of Monkeys. J Cognitive Neurosci, 3, 1-8.
Desimone, R., Duncan, J. (1995). Neural Mechanisms of Selective Visual-Attention. Annu Rev Neurosci, 18, 193-222.
Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. Philos T Roy Soc B, 353, 1245-1255.
Di Russo, F., Spinelli, D. (1999). Spatial attention has different effects on the magno- and parvocellular pathways. Neuroreport, 10, 2755-62.
Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. Hum Brain Mapp, 15, 95-111.
Di Russo, F., Martinez, A., Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex, 13, 486-99.
Di Russo, F., Pitzalis, S., Aprile, T., Spitoni, G., Patria, F., Stella, A., Spinelli, D., Hillyard, S. A. (2007). Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. Hum Brain Mapp, 28, 323-34.
Diamond, R., Carey, S. (1986). Why faces are and are not special: an effect of expertise. J Exp Psychol Gen, 115, 107-117.
Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J Vision, 3, 586-598.
Downing, P. E., Chan, A. W. Y., Peelen, M. V., Dodds, C. M., Kanwisher, N. (2006). Domain specificity in visual cortex. Cereb Cortex, 16, 1453-1461.
Edwards, R., Xiao, D. K., Keysers, C., Foldiak, P., Perrett, D. (2003). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. J Neurophysiol, 90, 1245-1256.
Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H., Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. J Neurophysiol, 91, 358-371.
Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? Neuroreport, 9, 2945-2948.
Eimer, M., McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: evidence from event-related potentials. Neuroreport, 10, 255-259.
Eimer, M. (2000a). Effects of face inversion on the structural encoding and recognition of faces - Evidence from event-related brain potentials. Cognitive Brain Res, 10, 145-158.
Eimer, M. (2000b). The face-specific N170 component reflects late stages in the structural encoding of faces. Neuroreport, 11, 2319-2324.
Eimer, M. (2000c). Attentional modulations of event-related brain potentials sensitive to faces. Cognitive Neuropsych, 17, 103-116.
Eimer, M. (2000d). Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clin Neurophysiol, 111, 694-705.
Einhäuser, W., Mundhenk, T. N., Baldi, P., Koch, C., Itti, L. (2007). A bottom–up model of spatial attention predicts human error patterns in rapid scene recognition. Journal of Vision, /7/10/6/, 1-13.
Engel, A. K., Moll, C. K., Fried, I., Ojemann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. Nat Rev Neurosci, 6, 35-47.
Fabre-Thorpe, M., Richard, G., Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. Neuroreport, 9, 303-308.
Fairhall, S. L., Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. Cereb Cortex, 17, 2400-6.
Farah, M. J., Wilson, K. D., Drain, M., Tanaka, J. N. (1998). What is "special" about face perception? Psychol Rev, 105, 482-498.

Faust, C. (1955). [Evaluation of posttraumatic organic personality disorders.]. Dtsch Med Wochenschr, 80, 1237-9.
Fei-Fei, L., Iyer, A., Koch, C., Perona, P. (2007). What do we perceive in a glance of a real-world scene? J Vision, 7/1/,
Felleman, D. J., Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex, 1, 1-47.
Ferrera, V. P., Nealey, T. A., Maunsell, J. H. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. Journal of Neuroscience, 14, 2080-2088.
Foldiak, P. (2002). Sparse coding in the primate cortex. In The Handbook of Brain Theory and Neural Networks, Second Edition (M. Arbib, Eds.). Cambridge, MA: MIT Press, pp. 1064 - 1068.
Foldiak, P., Xiao, D. K., Keysers, C., Edwards, R., Perrett, D. I. (2003). Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog Brain Res, 144, 107-116.
Formisano, E., Goebel, R. (2003). Tracking cognitive processes with functional MRI mental chronometry. Curr Opin Neurobiol, 13, 174-81.
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, 139-150.
Foxe, J. J., Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. Neuroreport, 16, 419-23.
Freedman, D. J., Riesenhuber, M., Poggio, T., Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science, 291, 312-316.
Freedman, D. J., Riesenhuber, M., Poggio, T., Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. Journal of Neuroscience, 23, 5235-5246.
Freedman, D. J., Riesenhuber, M., Poggio, T., Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb Cortex, 16, 1631-1644.
Freire, A., Lee, K., Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. Perception, 29, 159-170.
Fujita, I., Tanaka, K., Ito, M., Cheng, K. (1992). Columns for Visual Features of Objects in Monkey Inferotemporal Cortex. Nature, 360, 343-346.
Galton, F. (1883). Inquiries into human faculty and its development. London: Macmillan.
Gauthier, I., Behrmann, M., Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? J Cognitive Neurosci, 11, 349-370.
Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. J Cogn Neurosci, 12, 495-504.
Gauthier, I., Curran, T., Curby, K. M., Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. Nat Neurosci, 6, 428-32.
George, N., Evans, J., Fiori, N., Davidoff, J., Renault, B. (1996). Brain events related to normal and moderately scrambled faces. Cognitive Brain Res, 4, 65-76.
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.
George, N., Jemel, B., Fiori, N., Chaby, L., Renault, B. (2005). Electrophysiological correlates of facial decision: Insights from upright and upside-down Mooney-face perception. Cognitive Brain Res, 24, 663-673.
Ghose, G. M., Maunsell, J. (1999). Specialized representations in visual cortex: A role for binding? Neuron, 24, 79-85.
Girard, P., Hupe, J. M., Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. J Neurophysiol, 85, 1328-1331.
Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. European Journal of Neuroscience, 10, 1563-1573.
Goffaux, V., Gauthier, I., Rossion, B. (2003). Spatial scale contribution to early visual differences

between face and object processing. Cognitive Brain Res, 16, 416-424.
Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., Rossion, B. (2005a). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. Vis Cog, 12, 878-892.
Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., Rossion, B. (2005b). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. Perception, 34, 77-86.
Goffaux, V., Rossion, B. (2006). Faces are "spatial" - Holistic face perception is supported by low spatial frequencies. J Exp Psychol Human, 32, 1023-1039.
Goffaux, V., Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. J Exp Psychol Hum Percept Perform, 33, 995-1002.
Gold, J. I., Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. Trends Cogn Sci, 5, 10-16.
Gomez Gonzalez, C. M., Clark, V. P., Fan, S., Luck, S. J., Hillyard, S. A. (1994). Sources of attention-sensitive visual event-related potentials. Brain Topogr, 7, 41-51.
Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. Neuron, 24, 31-47.
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron, 24, 187-203.
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, 837-843.
Grill-Spector, K., Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica, 107, 293-321.
Grill-Spector, K., Kourtzi, Z., Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision Res, 41, 1409-22.
Grill-Spector, K., Malach, R. (2004). The human visual cortex. Annu Rev Neurosci, 27, 649-677.
Grill-Spector, K., Knouf, N., Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci, 7, 555-62.
Grill-Spector, K., Kanwisher, N. (2005). Visual recognition - As soon as you know it is there, you know what it is. Psychol Sci, 16, 152-160.
Grill-Spector, K., Sayres, R., Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat Neurosci, 9, 1177-1185.
Gross, C. G., Bender, D. B., Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science, 166, 1303-6.
Gross, C. G., Rocha-Miranda, C. E., Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol, 35, 96-111.
Gross, C. G., Bruce, C. J., Desimone, R., Fleming, J., Gattass, R. (1981). Cortical visual areas of the temporal lobe: three areas in the macaque. In Cortical Sensory Organization, Vol. 2, Multiple Visual Areas (C. N. Woolsey, Eds.). New York: Humana Press, pp. 187-216.
Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. Philos T Roy Soc B, 335, 3-10.
Gross, C. G. (1998.). Brain, Vision, Memory: Tales in the History of Neuroscience. Cambridge: MIT Press.
Gross, C. G. (2002). Genealogy of the "grandmother cell". Neuroscientist, 8, 512-8.
Guérit, J.-M. (1998). Les potentiels évoqués. Paris: Masson.
Guillaume, F., Tiberghien, G. (2001). An event-related potential study of contextual modifications in a face recognition task. Neuroreport, 12, 1209-16.
Guyonneau, R., Kirchner, H., Thorpe, S. J. (2005). Animals roll around the clock: The rotation invariance of ultra-rapid visual processing. Perception, 34, 131-132.
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.
Halit, H., de Haan, M., Schyns, P. G., Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? Brain Research, 1117, 154-161.

Harries, M. H., Perrett, D. I. (1991). Visual Processing of Faces in Temporal Cortex - Physiological Evidence for a Modular Organization and Possible Anatomical Correlates. J Cognitive Neurosci, 3, 9-24.
Harris, A. M., Duchaine, B. C., Nakayama, K. (2005). Normal and abnormal face selectivity of the M170 response in developmental prosopagnosics. Neuropsychologia, 43, 2125-2136.
Harris, A., Nakayama, K. (2007). Rapid Adaptation of the M170 Response: Importance of Face Parts. Cereb Cortex,
Hasselmo, M. E., Rolls, E. T., Baylis, G. C. (1989). The Role of Expression and Identity in the Face-Selective Responses of Neurons in the Temporal Visual-Cortex of the Monkey. Behav Brain Res, 32, 203-218.
Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron, 22, 189-99.
Haxby, J. V., Hoffman, E. A., Gobbini, M. I. (2000). The distributed human neural system for face perception. Trends Cogn Sci, 4, 223-233.
Heisz, J. J., Watter, S., Shedden, J. A. (2006a). Automatic face identity encoding at the N170. Vision Res, 46, 4604-4614.
Heisz, J. J., Watter, S., Shedden, J. M. (2006b). Progressive N170 habituation to unattended repeated faces. Vision Res, 46, 47-56.
Henderson, J. M. (2003). Human gaze control during real-world scene perception. Trends Cogn Sci, 7, 498-504.
Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cereb Cortex, 13, 793-805.
Henson, R. N., Kiebel, S., Kilner , J., Friston, K., Hillebrand, A., Barnes, G. R., Singh, K. D. (2005). Time-frequency SPMs for MEG data on face perception: Power changes and phase-locking. HBM abstract,
Henson, R. N., Mattout, J., Singh, K. D., Barnes, G. R., Hillebrand, A., Friston, K. (in press). Population-level inferences for distributed MEG source localisation under multiple constraints: Application to face-evoked fields. Neuroimage,
Herrmann, M. J., Ehlis, A. C., Muehlberger, A., Fallgatter, A. J. (2005a). Source localization of early stages of face processing. Brain Topogr, 18, 77-85.
Herrmann, M. J., Ehlis, A. C., Ellgring, H., Fallgatter, A. J. (2005b). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). Journal of Neural Transmission, 112, 1073-1081.
Hershler, O., Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. Vision Res, 45, 1707-1724.
Herzmann, G., Schweinberger, S. R., Sommer, W., Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. Psychophysiology, 41, 688-701.
Hill, H., Schyns, P. G., Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. Cognition, 62, 201-222.
Hillyard, S. A., Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. P Natl Acad Sci USA, 95, 781-787.
Hochstein, S., Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron, 36, 791-804.
Hoffman, E. A., Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci, 3, 80-4.
Hoffmann, M. B., Straube, S., Bach, M. (2003). Pattern-onset stimulation boosts central multifocal VEP responses. J Vision, 3, 432-439.
Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. Perception, 23, 65-74.
Hopfinger, J. B., Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. Psychol Sci, 9, 441-447.
Horton, J. C., Adams, D. L. (2005). The cortical column: a structure without a function. Philos Trans R Soc Lond B Biol Sci, 360, 837-62.
Hsu, R.-L., Abdel-Mottaleb, M., Jain, A. K. (2002). Face detection in color images. Transactions on

Pattern Analysis and Machine Intelligence, 24, 696 - 706.
Hubel, D. H., Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology, 195, 215-243.
Hubel, D. H., Wiesel, T. N. (1969). Anatomical demonstration of columns in the monkey striate cortex. Nature, 221, 747-750.
Huddy, V., Schweinberger, S. R., Jentzsch, I., Burton, A. M. (2003). Matching faces for semantic information and names: an event-related brain potentials study. Cognitive Brain Res, 17, 314-326.
Hung, C. P., Kreiman, G., Poggio, T., DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science, 310, 863-866.
Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol, 85, 134-145.
Ishai, A., Ungerleider, L. G., Martin, A., Schouten, H. L., Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. P Natl Acad Sci USA, 96, 9379-9384.
Itier, R. J., Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: A repetition study using ERPs. Neuroimage, 15, 353-372.
Itier, R. J., Taylor, M. J. (2004a). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex, 14, 132-142.
Itier, R. J., Taylor, M. J. (2004b). Source analysis of the N170 to faces and objects. Neuroreport, 15, 1261-1265.
Itier, R. J., Taylor, M. J. (2004c). Effects of repetition learning on upright, inverted and contrast-reversed face processing using ERPs. Neuroimage, 21, 1518-1532.
Itier, R. J., Latinus, M., Taylor, M. J. (2006a). Face, eye and object early processing: What is the face specificity? Neuroimage, 29, 667-676.
Itier, R. J., Herdman, A. T., George, N., Cheyne, D., Taylor, M. J. (2006b). Inversion and contrast-reversal effects on face processing assessed by MEG. Brain Research, 1115, 108-120.
Itier, R. J., Alain, C., Sedore, K., McIntosh, A. R. (2007). Early Face Processing Specificity: It's in the Eyes! J Cogn Neurosci, 19, 1815-1826.
Ito, M., Fujita, I., Tamura, H., Tanaka, K. (1994). Processing of contrast polarity of visual images in inferotemporal cortex of the macaque monkey. Cereb Cortex, 4, 499-508.
Ito, M., Tamura, H., Fujita, I., Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol, 73, 218-226.
Jacques, C., Rossion, B. (2004). Concurrent processing reveals competition between visual representations of faces. Neuroreport, 15, 2417-2421.
Jacques, C., Rossion, B. (2006a). The time course of visual competition to the presentation of centrally fixated faces. J Vision, 6, 154-162.
Jacques, C., Rossion, B. (2006b). The speed of individual face categorization. Psychol Sci, 17, 485-492.
Jacques, C., Rossion, B. (2007a). Electrophysiological evidence for temporal dissociation between spatial attention and sensory competition during human face processing. Cereb Cortex, 17, 1055-1065.
Jacques, C., Rossion, B. (2007b). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. Neuroimage, 36, 863-876.
Jacques, C., d'Arripe, O., Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. J Vision, 7(8):3, 1-9.
Jasper, H. H. (1958). The ten–twenty system of the International Federation. Electroen Clin Neuro, 10, 371-375.
Jeffery, L., Rhodes, G., Busey, T. (2006). View-specific coding of face shape. Psychol Sci, 17, 501-505.
Jeffreys, D. A., Axford, J. G. (1972). Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. Exp Brain Res, 16, 1-21.
Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. Exp Brain Res, 78,

193-202.
Jeffreys, D. A., Tukmachi, E. S. A., Rockley, G. (1992). Evoked-Potential Evidence for Human Brain Mechanisms That Respond to Single, Fixated Faces. Exp Brain Res, 91, 351-362.
Jeffreys, D. A., Tukmachi, E. S. (1992). The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res, 91, 340-50.
Jeffreys, D. A. (1993). The influence of stimulus orientation on the vertex positive scalp potential evoked by faces. Exp Brain Res, 96, 163-172.
Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. Vis Cog, 3, 1-38.
Jellema, T., Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? Neuropsychologia, 41, 1728-1737.
Jellema, T., Maassen, G., Perrett, D. I. (2004). Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. Cereb Cortex, 14, 781-790.
Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., Bruyer, R. (2003a). Stepwise emergence of the face-sensitive N170 event-related potential component. Neuroreport, 14, 2035-2039.
Jemel, B., Calabria, M., Delvenne, J. F., Crommelinck, M., Bruyer, R. (2003b). Differential involvement of episodic and face representations in ERP repetition effects. Neuroreport, 14, 525-530.
Jemel, B., Pisani, M., Calabria, M., Crommelinck, M., Bruyer, R. (2003c). Is the N170 for faces cognitively penetrable? Evidence from repetition priming of Mooney faces of familiar and unfamiliar persons. Brain Res Cogn Brain Res, 17, 431-46.
Jemel, B., Pisani, M., Rousselle, L., Crommelinck, M., Bruyer, R. (2005). Exploring the functional architecture of person recognition system with event-related potentials in a within- and cross-domain self-priming of faces. Neuropsychologia, 43, 2024-2040.
Jiang, F., Blanz, V., O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation - A view from the other side of the mean. Psychol Sci, 17, 493-500.
Johnson, B. W., Hamm, J. P. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. Clin Neurophysiol, 111, 532-45.
Johnson, J. S., Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. J Vision, 3, 499-512.
Joyce, C., Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. Clin Neurophysiol, 116, 2613-2631.
Joyce, C. A., Kutas, M. (2005). Event-related potential correlates of long-term memory for briefly presented faces. J Cogn Neurosci, 17, 757-67.
Joyce, C. A., Schyns, P. G., Gosselin, F., Cottrell, G. W., Rossion, B. (2006). Early selection of diagnostic facial information in the human visual cortex. Vision Res, 46, 800-813.
Kanwisher, N., McDermott, J., Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17, 4302-4311.
Kersten, D., Mamassian, P., Yuille, A. (2004). Object perception as Bayesian inference. Annu Rev Psychol, 55, 271-304.
Keysers, C., Xiao, D. K., Foldiak, P., Perrett, D. I. (2001). The speed of sight. J Cognitive Neurosci, 13, 90-101.
Keysers, C., Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. Trends Cogn Sci, 6, 120-125.
Kiani, R., Esteky, H., Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J Neurophysiol, 94, 1587-1596.
Kim, M., Ducros, M., Carlson, T., Ronen, I., He, S., Ugurbil, K., Kim, D. S. (2006). Anatomical correlates of the functional organization in the human occipitotemporal cortex. Magn Reson Imaging, 24, 583-590.
Kirchner, H., Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Res, 46, 1762-1776.
Kleinschmidt, A., Cohen, L. (2006). The neural bases of prosopagnosia and pure alexia: recent insights from functional neuroimaging. Curr Opin Neurol, 19, 386-91.
Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory

performance. Cognitive Brain Res, 19, 302-316.
Kloth, N., Dobel, C., Schweinberger, S. R., Zwitserlood, P., Bolte, J., Junghofer, M. (2006). Effects of personal familiarity on early neuromagnetic correlates of face perception. European Journal of Neuroscience, 24, 3317-3321.
Kobatake, E., Tanaka, K. (1994). Neuronal Selectivities to Complex Object Features in the Ventral Visual Pathway of the Macaque Cerebral-Cortex. J Neurophysiol, 71, 856-867.
Kobatake, E., Wang, G., Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J Neurophysiol, 80, 324-330.
Konorski, J. (1967). Integrative activity of the brain; an interdisciplinary approach. Chicago: University of Chicago Press.
Kourtzi, Z., Erb, M., Grodd, W., Bulthoff, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb Cortex, 13, 911-20.
Kourtzi, Z., Betts, L. R., Sarkheil, P., Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol, 3, e204.
Kovacs, G., Vogels, R., Orban, G. A. (1995). Cortical correlate of pattern backward masking. Proc Natl Acad Sci U S A, 92, 5587-91.
Kovacs, G., Zimmer, M., Banko, E., Harza, I., Antal, A., Vidnyanszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. Cereb Cortex, 16, 742-53.
Kreiman, G., Koch, C., Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci, 3, 946-953.
Kutas, M., Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207, 203-5.
Lamme, V. A. F., Super, H., Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. Current Opinion in Neurobiology, 8, 529-535.
Lamme, V. A. F., Rodriguez-Rodriguez, V., Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb Cortex, 9, 406-413.
Lamme, V. A. F., Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci, 23, 571-579.
Latinus, M., Taylor, M. J. (2005). Holistic processing of faces: Learning effects with Mooney faces. J Cognitive Neurosci, 17, 1316-1327.
Latinus, M., Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. Brain Research, 1123, 179-187.
Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., Chabriat, H. (2001). Diffusion tensor imaging: concepts and applications. Journal of Magnetic Resonance Imaging, 13, 534-546.
Leder, H., Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. Q J Exp Psychol-A, 53, 513-536.
Lee, K. J., Perrett, D. (1997). Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces. Perception, 26, 733-752.
Lee, T. S., Mumford, D., Romero, R., Lamme, V. A. F. (1998). The role of the primary visual cortex in higher level vision. Vision Res, 38, 2429-2454.
Lee, D. K., Itti, L., Koch, C., Braun, J. (1999). Attention activates winner-take-all competition among visual filters. Nat Neurosci, 2, 375-381.
Lee, K. J., Perrett, D. I. (2000). Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. Perception, 29, 1291-1312.
Lee, T. S., Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. P Natl Acad Sci USA, 98, 1907-1911.
Lehky, S. R. (2000). Fine discrimination of faces can be performed rapidly. J Cognitive Neurosci, 12, 848-855.
Lewis, M. B., Johnston, R. A. (1997). The Thatcher illusion as a test of configural disruption. Perception, 26, 225-7.
Lewis, M. B., Edmonds, A. J. (2003). Face detection: Mapping human performance. Perception, 32, 903-920.

Lewis, M. B., Edmonds, A. J. (2005). Searching for faces in scrambled scenes. Vis Cog, 12, 1309-1336.
Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. Neurosci Lett, 253, 147-150.
Liu, J., Higuchi, M., Marantz, A., Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. Neuroreport, 11, 337-341.
Liu, C. H., Collin, C. A., Rainville, S. J. M., Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. J Exp Psychol Human, 26, 956-979.
Liu, C. H., Chaudhuri, A. (2002). Reassessing the 3/4 view effect in face recognition. Cognition, 83, 31-48.
Liu, J., Harris, A., Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. Nat Neurosci, 5, 910-916.
Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. Vision Res, 45, 2287-2297.
Loftus, G. R., Harley, E. M. (2004). How different spatial-frequency components contribute to visual information acquisition. J Exp Psychol Hum Percept Perform, 30, 104-18.
Logothetis, N. K., Pauls, J., Bulthoff, H. H., Poggio, T. (1994). View-Dependent Object Recognition by Monkeys. Curr Biol, 4, 401-414.
Logothetis, N. K., Pauls, J., Poggio, T. (1995). Shape Representation in the Inferior Temporal Cortex of Monkeys. Curr Biol, 5, 552-563.
Logothetis, N. K., Sheinberg, D. L. (1996). Visual object recognition. Annu Rev Neurosci, 19, 577-621.
Logothetis, N. K., Guggenberger, H., Peled, S., Pauls, J. (1999). Functional imaging of the monkey brain. Nat Neurosci, 2, 555-562.
Lu, Z. L., Dosher, B. A. (1998). External noise distinguishes attention mechanisms. Vision Res, 38, 1183-1198.
Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., Hawkins, H. L. (1994). Effects of Spatial Cueing on Luminance Detectability - Psychophysical and Electrophysiological Evidence for Early Selection. J Exp Psychol Human, 20, 887-904.
Luck, S. J., Chelazzi, L., Hillyard, S. A., Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol, 77, 24-42.
Luck, S. J., Woodman, G. F., Vogel, E. K. (2000). Event-related potential studies of attention. Trends Cogn Sci, 4, 432-440.
Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge, MA: MIT Press.
Lueschow, A., Miller, E. K., Desimone, R. (1994). Inferior Temporal Mechanisms for Invariant Object Recognition. Cereb Cortex, 4, 523-531.
Macchi Cassia, V., Kuefner, D., Westerlund, A., Nelson, C. A. (2006). Modulation of face-sensitive event-related potentials by canonical and distorted human faces: the role of vertical symmetry and up-down featural arrangement. J Cognitive Neurosci, 18, 1343-1358.
Mace, M. J. M., Richard, G., Delorme, A., Fabre-Thorpe, M. (2005a). Rapid categorization of natural scenes in monkeys: target predictability and processing speed. Neuroreport, 16, 349-354.
Mace, M. J. M., Thorpe, S. J., Fabre-Thorpe, M. (2005b). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? European Journal of Neuroscience, 21, 2007-2018.
Mack, M. L., Gauthier, I., Sadr, J., Palmeri, T. J. (in press). Object detection and basic-level categorization; Sometimes you know it is there before you know what it is. Psychonomic Bulletin & Review,
MacKenzie, G., Donaldson, D. I. (2007). Dissociating recollection from familiarity: electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect. Neuroimage, 36, 454-63.
Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. Science, 295, 690-694.
Mangun, G. R. (1995). Neural Mechanisms of Visual Selective Attention. Psychophysiology, 32, 4-18.
Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and

Processing of Visual Information. San Francisco, CA: W. H. Freeman.
Martens, U., Schweinberger, S. R., Kiefer, M., Burton, A. M. (2006). Masked and unmasked electrophysiological repetition effects of famous faces. Brain Research, 1109, 146-157.
Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci, 2, 364-369.
Martinez, A., DiRusso, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., Hillyard, S. A. (2001). Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Res, 41, 1437-1457.
Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., Kawano, K. (2005). Population dynamics of face-responsive neurons in the inferior temporal cortex. Cereb Cortex, 15, 1103-12.
Maunsell, J. H. R., Gibson, J. R. (1992). Visual Response Latencies in Striate Cortex of the Macaque Monkey. J Neurophysiol, 68, 1332-1344.
Maunsell, J. H. R., Ghose, G. M., Assad, J. A., McAdams, C. J., Boudreau, C. E., Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. Visual Neurosci, 16, 1-14.
Maurer, D., Grand, R. L., Mondloch, C. J. (2002). The many faces of configural processing. Trends Cogn Sci, 6, 255-260.
Mayer, E., Rossion, B. (2007). Prosopagnosia. In The Behavioral and Cognitive Neurology of Stroke (O. B. Godefroy, J., Eds.). Cambridge: Cambridge University Press, pp. 315-334.
Mazard, A., Schiltz, C., Rossion, B. (2006). Recovery from adaptation to facial identity is larger for upright than inverted faces in the human occipito-temporal cortex. Neuropsychologia, 44, 912-922.
McCarthy, G., Puce, A., Belger, A., Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. Cereb Cortex, 9, 431-444.
McKeeff, T. J., Remus, D. A., Tong, F. (2007). Temporal limitations in object processing across the human ventral visual pathway. J Neurophysiol, 98, 382-393.
McKone, E. (2004). Isolating the special component of face recognition: Peripheral identification and a Mooney face. J Exp Psychol Learn, 30, 181-197.
Merigan, W. H., Maunsell, J. H. R. (1993). How Parallel Are the Primate Visual Pathways. Annu Rev Neurosci, 16, 369-402.
Michel, C. M., Thut, G., Morand, S., Khateb, A., Pegna, A. J., Grave de Peralta, R., Gonzalez, S., Seeck, M., Landis, T. (2001). Electric source imaging of human brain functions. Brain Res Brain Res Rev, 36, 108-18.
Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R. (2004). EEG source imaging. Clin Neurophysiol, 115, 2195-2222.
Michel, C., Rossion, B., Han, J., Chung, C. S., Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. Psychol Sci, 17, 608-615.
Mihaylova, M., Stomonyakov, V., Vassilev, A. (1999). Peripheral and central delay in processing high spatial frequencies: reaction time and VEP latency studies. Vision Res, 39, 699-705.
Milivojevic, B., Clapp, W. C., Johnson, B. W., Corballis, M. C. (2003). Turn that frown upside down: ERP effects of thatcherization of misorientated faces. Psychophysiology, 40, 967-78.
Minnebusch, D. A., Suchan, B., Ramon, M., Daum, I. (2007). Event-related potentials reflect heterogeneity of developmental prosopagnosia. European Journal of Neuroscience, 25, 2234-2247.
Mnatsakanian, E. V., Tarkka, I. M. (2004). Influence of the background on the recognition and comparison of targets: Behavioral and ERP study. Int J Psychophysiol, 54, 156-156.
Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. Eur J Neurosci, 12, 303-10.
Murray, J. E., Yong, E., Rhodes, G. (2000). Revisiting the perception of upside-down faces. Psychol Sci, 11, 492-496.
Musselwhite, M. J., Jeffreys, D. A. (1985). The influence of spatial frequency on the reaction times

and evoked potentials recorded to grating pattern stimuli. Vision Res, 25, 1545-1555.
Nakayama, K., He, Z. J., Shimojo, S. (1995). Visual surface representation: A critical link between lowerlevel and higher level vision. In An Invitation to Cognitive Science: Visual Cognition (S. M. Kosslyn and D. N. Osherson, Eds.). Cambridge, MA: M.I.T. Press, pp. 1-70.
Nasanen, R., Ojanpaa, H., Tanskanen, T., Paallysaho, J. (2006). Estimation of temporal resolution of object identification in human vision. Exp Brain Res, 172, 464-471.
Nederhouser, M., Yue, X., Biederman, I. (2006). Predicting psychophysical similarity of complex shapes from measures of physical similarity. J Vision, 6, 320.
Neville, H., Snyder, E., Woods, D., Galambos, R. (1982). Recognition and surprise alter the human visual evoked response. Proc Natl Acad Sci U S A, 79, 2121-3.
Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. Perception, 22, 1287-1298.
Nowak, L. G., Munk, M. H. J., Girard, P., Bullier, J. (1995). Visual Latencies in Areas V1 and V2 of the Macaque Monkey. Visual Neurosci, 12, 371-384.
Oliva, A., Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. Progress in Brain Research, 155, 23-36.
Olivares, E. I., Iglesias, J., Antonieta Bobes, M. (1999). Searching for face-specific long latency ERPs: a topographic study of effects associated with mismatching features. Brain Res Cogn Brain Res, 7, 343-56.
Olivares, E. I., Iglesias, J., Rodriguez-Holguin, S. (2003). Long-latency ERPs and recognition of facial identity. J Cogn Neurosci, 15, 136-51.
Op De Beeck, H., Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. Journal of Comparative Neurology, 426, 505-518.
Oram, M. W., Perrett, D. I. (1992). Time Course of Neural Responses Discriminating Different Views of the Face and Head. J Neurophysiol, 68, 70-84.
Oram, M. W., Xiao, D., Dritschel, B., Payne, K. R. (2002). The temporal resolution of neural codes: does response latency have a unique role? Philos T Roy Soc B, 357, 987-1001.
Orban, G. A., Van Essen, D., Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn Sci, 8, 315-324.
Paller, K. A., Bozic, V. S., Ranganath, C., Grabowecky, M., Yamada, S. (1999). Brain waves following remembered faces index conscious recollection. Cognitive Brain Res, 7, 519-531.
Paller, K. A., Gonsalves, B., Grabowecky, M., Bozic, V. S., Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. Neuroimage, 11, 98-110.
Paller, K. A., Hutson, C. A., Miller, B. B., Boehm, S. G. (2003). Neural manifestations of memory with and without awareness. Neuron, 38, 507-516.
Palmeri, T. J., Gauthier, I. (2004). Visual object understanding. Nat Rev Neurosci, 5, 291-U17.
Pasqual-Marqui, R. D., M., M. C., Lehman, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. IEEE Trans Biomed Eng, 42, 658–665.
Pasupathy, A., Connor, C. E. (1999). Responses to contour features in macaque area V4. J Neurophysiol, 82, 2490-502.
Perrett, D. I., Rolls, E. T., Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res, 47, 329-42.
Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., Jeeves, M. A. (1984). Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. Human Neurobiology, 3, 197-208.
Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., Jeeves, M. A. (1985a). Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. P Roy Soc Lond B Bio, 223, 293-317.
Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D., Jeeves, M. A. (1985b). Visual Analysis of Body Movements by Neurons in the Temporal Cortex of the Macaque Monkey - a Preliminary-Report. Behav Brain Res, 16, 153-170.
Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A. J., Potter, D. D., Broennimann, R., Harries, M. (1988). Specialized Face Processing and Hemispheric-Asymmetry in Man and Monkey - Evidence from Single Unit and Reaction-Time Studies. Behav Brain Res, 29, 245-258.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., Thomas, S. (1991). Viewer-Centered and Object-Centered Coding of Heads in the Macaque Temporal Cortex. Exp Brain Res, 86, 159-173.
Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J. (1992). Organization and Functions of Cells Responsive to Faces in the Temporal Cortex. Philos T Roy Soc B, 335, 23-30.
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, 111-145.
Pestilli, F., Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. Vision Res, 45, 1867-1875.
Peterson, M. A., Gibson, B. S. (1994). Object Recognition Contributions to Figure Ground Organization - Operations on Outlines and Subjective Contours. Perception & Psychophysics, 56, 551-564.
Pfutze, E. M., Sommer, W., Schweinberger, S. R. (2002). Age-related slowing in face and name recognition: evidence from event-related brain potentials. Psychol Aging, 17, 140-60.
Philiastides, M. G., Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex, 16, 509-518.
Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology, 37, 127-152.
Pilz, K. S., Thornton, I. M., Bulthoff, H. H. (2006). A search advantage for faces learned in motion. Exp Brain Res, 171, 436-447.
Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: A functional MRI study. P Natl Acad Sci USA, 102, 6996-7001.
Posner, M. I., Snyder, C. R., Davidson, B. J. (1980). Attention and the detection of signals. J Exp Psychol, 109, 160-74.
Potter, M. C., Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 81, 10-15.
Potter, M. C. (1975). Meaning in visual search. Science, 187, 965-966.
Potter, M. C. (1976). Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory, 2, 509-522.
Potter, D. D., Parker, D. M. (1997). Dissociation of event-related potential repetition effects in judgments of face identity and expression. J. Psychophysiol, 11, 287–303.
Pourtois, G., Grandjean, D., Sander, D., Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. Cereb Cortex, 14, 619-633.
Puce, A., Allison, T., Bentin, S., Gore, J. C., McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci, 18, 2188-99.
Purcell, D. G., Stewart, A. L., Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. Perception, 25, 1091-1108.
Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435, 1102-1107.
Ramon, M., Rossion, B. (2007). What's lost in prosopagnosia? An investigation of familiar face processing in a singlecase of pure prosopagnosia working in a kindergarten. J Vision, 9, 122.
Ratcliff, R. (1988). Continuous Versus Discrete Information-Processing - Modeling Accumulation of Partial Information. Psychol Rev, 95, 238-255.
Ratcliff, R. (2006). Modeling response signal and response time data. Cognit Psychol, 53, 195-237.
Rebai, M., Poiroux, S., Bernard, C., Lalonde, R. (2001). Event-related potentials for category-specific information during passive viewing of faces and objects. Int J Neurosci, 106, 209-226.
Reddi, B. A. J., Carpenter, R. H. S. (2000). The influence of urgency on decision time. Nat Neurosci, 3, 827-830.
Regan, D. (1989). Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. New York: Elsevier.
Reynolds, J. H., Desimone, R. (1999). The role of neural mechanisms of attention in solving the

binding problem. Neuron, 24, 19-29.
Reynolds, J. H., Chelazzi, L., Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. Journal of Neuroscience, 19, 1736-53.
Richmond, B. J., Optican, L. M., Spitzer, H. (1990). Temporal Encoding of 2-Dimensional Patterns by Single Units in Primate Primary Visual-Cortex .1. Stimulus-Response Relations. J Neurophysiol, 64, 351-369.
Riesenhuber, M., Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat Neurosci, 2, 1019-1025.
Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. Behav Brain Res, 76, 191-7.
Rolls, E. T., Judge, S. J., Sanghera, M. K. (1977). Activity of neurones in the inferotemporal cortex of the alert monkey. Brain Research, 130, 229-238.
Rolls, E. T., Perrett, D. I., Caan, A. W., Wilson, F. A. (1982). Neuronal responses related to visual recognition. Brain, 105 (Pt 4), 611-646.
Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Human Neurobiology, 3, 209-222.
Rolls, E. T., Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res, 65, 38-48.
Rolls, E. T. (1992). Neurophysiological Mechanisms Underlying Face Processing within and Beyond the Temporal Cortical Visual Areas. Philos T Roy Soc B, 335, 11-21.
Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., Azzopardi, P. (1994). The Responses of Neurons in the Temporal Cortex of Primates, and Face Identification and Detection. Exp Brain Res, 101, 473-484.
Rolls, E. T., Tovee, M. J. (1995a). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Exp Brain Res, 103, 409-20.
Rolls, E. T., Tovee, M. J. (1995b). Sparseness of the Neuronal Representation of Stimuli in the Primate Temporal Visual-Cortex. J Neurophysiol, 73, 713-726.
Rolls, E. T., Treves, A., Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. Exp Brain Res, 114, 149-162.
Rolls, E. T., Tovee, M. J., Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. J Cognitive Neurosci, 11, 300-11.
Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron, 27, 205-218.
Rolls, E. T., Aggelopoulos, N. C., Zheng, F. S. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. Journal of Neuroscience, 23, 339-348.
Rolls, E. T., Franco, L., Aggelopoulos, N. C., Jerez, J. M. (2006). Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. Vision Res, 46, 4193-4205.
Rolls, E. T. (2007). The representation of information about faces in the temporal and frontal lobes. Neuropsychologia, 45, 124-143.
Rosch, E., Mervis, C., Gray, W., Johnson, D., Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.
Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., Guerit, J. M. (1999a). Spatio-temporal localization of the face inversion effect: an event-related potentials study. Biol Psychol, 50, 173-189.
Rossion, B., Campanella, S., Gomez, C. M., Delinte, A., Debatisse, D., Liard, L., Dubois, S., Bruyer, R., Crommelinck, M., Guerit, J. M. (1999b). Task modulation of brain activity related to familiar and unfamiliar face processing: an ERP study. Clin Neurophysiol, 110, 449-462.
Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport, 11, 69-74.

Rossion, B., Gauthier, I. (2002). How does the brain process upright and inverted faces? Behav Cogn Neurosci Rev, 1, 62-74.
Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. Psychol Sci, 13, 250-257.
Rossion, B., Joyce, C. A., Cottrell, G. W., Tarr, M. J. (2003a). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. Neuroimage, 20, 1609-1624.
Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., Mayer, E. (2003b). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain, 126, 2381-2395.
Rossion, B., Kung, C. C., Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. P Natl Acad Sci USA, 101, 14521-14526.
Rossion, B., Collins, D., Goffaux, V., Curran, T. (2007). Long-term expertise with artificial objects increases visual competition with early face categorization processes. J Cognitive Neurosci, 19, 543-555.
Rossion, B., Jacques, C. (in preparation). The role of viewpoint and image symmetry in accounting for early face preferential responses in the human brain.
Rossion, B., Jacques, C. (in press). Does physical interstimulus variance account for early electrophysiological face preferential responses in the human brain? Ten lessons on the N170. Neuroimage,
Rossion, B. (in press). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. Neuroimage,
Rousselet, G. A., Mace, M. J. M., Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. J Vision, 3, 440-455.
Rousselet, G. A., Mace, M. J. M., Fabre-Thorpe, M. (2004a). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? J Vision, 4, 13-21.
Rousselet, G. A., Thorpe, S. J., Fabre-Thorpe, M. (2004b). How parallel is visual processing in the ventral pathway? Trends Cogn Sci, 8, 363-370.
Rousselet, G. A., Mace, M. J. M., Fabre-Thorpe, M. (2004c). Spatiotemporal analyses of the N170 for human faces, animal face's and objects in natural scenes. Neuroreport, 15, 2607-2611.
Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B. (2005). Spatial scaling factors explain eccentricity effects on face ERPs. J Vision, 5, 755-763.
Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B. (2007a). Single-trial EEG dynamics of object and face visual processing. Neuroimage, 36, 843-862.
Rousselet, G. A., Mace, M. J., Thorpe, S. J., Fabre-Thorpe, M. (2007b). Limits of event-related potential differences in tracking object processing speed. J Cognitive Neurosci, 19, 1241-1258.
Rugg, M. D., Allan, K. (2000). Event-related potential studies of long-term memory. In The Oxford Handbook of Memory (E. Tulving and F. I. M. Craik, Eds.). Oxford: Oxford University Press, pp.
Russell, R., Biederman, I., Nederhouser, M., Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. Vision Res, 47, 157-165.
Sagiv, N., Bentin, S. (2001). Structural encoding of human and schematic faces: Holistic and part-based processes. J Cognitive Neurosci, 13, 937-951.
Sary, G., Vogels, R., Orban, G. A. (1993). Cue-Invariant Shape Selectivity of Macaque Inferior Temporal Neurons. Science, 260, 995-997.
Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. Journal of Neuroscience, 25, 4294-4306.
Sayers, B. M., Beagley, H. A. (1974). Objective evaluation of auditory evoked EEG responses. Nature, 251, 608-609.
Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., Crommelinck, M., Orban, G. A. (1999). Neuronal mechanisms of perceptual learning: changes in human brain activity with training in orientation discrimination. Neuroimage, 9, 46-62.

Schiltz, C., Sorger, B., Caldara, R., Ahmed, F., Mayer, E., Goebel, R., Rossion, B. (2006). Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. Cereb Cortex, 16, 574-586.
Schiltz, C., Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. Neuroimage, 32, 1385-94.
Schmolesky, M. T., Wang, Y. C., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., Leventhal, A. G. (1998). Signal timing across the macaque visual system. J Neurophysiol, 79, 3272-3278.
Schneider, B. L., DeLong, J. E., Busey, T. A. (2007). Added noise affects the neural correlates of upright and inverted faces differently. J Vision, 7(4):4, 1-24.
Schroeder, C. E., Mehta, A. D., Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex, 8, 575-92.
Schroeder, C. E., Molholm, S., Lakatos, P., Ritter, W., Foxe, J. J. (2004). Human-Simian correspondence in the early cortical processing of multisensory cues. Cogn Process, 5, 140–151.
Schuller, A. M., Rossion, B. (2001). Spatial attention triggered by eye gaze increases and speeds up early visual activity. Neuroreport, 12, 2381-2386.
Schuller, A. M., Rossion, B. (2005). Spatial attention triggered by eye gaze enhances and speeds up visual processing in upper and lower visual fields beyond early striate visual processing. Clin Neurophysiol, 116, 2565-2576.
Schweinberger, S. R., Pfutze, E. M., Sommer, W. (1995). Repetition Priming and Associative Priming of Face Recognition - Evidence from Event-Related Potentials. J Exp Psychol Learn, 21, 722-736.
Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., Kaufmann, J. M. (2002a). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Cognitive Brain Res, 14, 398-409.
Schweinberger, S. R., Pickering, E. C., Burton, A. M., Kaufmann, J. M. (2002b). Human brain potential correlates of repetition priming in face and name recognition. Neuropsychologia, 40, 2057-2073.
Schweinberger, S. R., Huddy, V., Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. Neuroreport, 15, 1501-1505.
Schyns, P. G., Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. Psychol Sci, 5, 195–200.
Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., Gosselin, F. (2003). A principled method for determining the functionality of brain responses. Neuroreport, 14, 1665-1669.
Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. J Cognitive Neurosci, 18, 1453-1465.
Seeck, M., Grusser, O. J. (1992). Category-Related Components in Visual Evoked-Potentials - Photographs of Faces, Persons, Flowers and Tools as Stimuli. Exp Brain Res, 92, 338-349.
Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., Schomer, D. L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. Neuroreport, 8, 2749-54.
Sereno, M. E., Trinath, T., Augath, M., Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. Neuron, 33, 635-652.
Sergent, J. (1984). An investigation into component and configural processes underlying face perception. Br J Psychol, 75, 221-42.
Sergent, J. (1986). Microgenesis of face perception. In Aspects of face processing (H. D. Ellis, M. A. Jeeves, F. Newcombe and A. M. Young, Eds.). Dordrecht: Martinus Nijhoff, pp. 17–33.
Sergent, J., Ohta, S., Macdonald, B. (1992). Functional Neuroanatomy of Face and Object Processing - a Positron Emission Tomography Study. Brain, 115, 15-36.
Sergent, J., Signoret, J. L. (1992). Functional and Anatomical Decomposition of Face Processing - Evidence from Prosopagnosia and Pet Study of Normal Subjects. Philos T Roy Soc B, 335, 55-62.

Séverac-Cauquil, A., Edmonds, G. E., Taylor, M. J. (2000). Is the face-sensitive N 170 the only ERP not affected by selective attention? Neuroreport, 11, 2167-2171.
Shadlen, M. N., Newsome, W. T. (1996). Motion perception: Seeing and deciding. P Natl Acad Sci USA, 93, 628-633.
Sheinberg, D. L., Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. Journal of Neuroscience, 21, 1340-1350.
Shibata, T., Nishijo, H., Tamura, R., Miyamoto, K., Eifuku, S., Endo, S., Ono, T. (2002). Generators of visual evoked potentials for faces and eyes in the human brain as determined by dipole localization. Brain Topogr, 15, 51-63.
Shmuel, A., Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. J Neurosci, 16, 6945-64.
Sigala, N., Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature, 415, 318-320.
Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? Neuron, 24, 49-65.
Sinha, P., Balas, B. J., Ostrovsky, Y., Russell, R. (2006). Face recognition by humans: 19 results all computer vision researchers should know about. Proceedings of the IEEE, 94, 1948-1962.
Smith, M. L., Gosselin, F., Schyns, P. G. (2004). Receptive fields for flexible face categorizations. Psychol Sci, 15, 753-761.
Smith, M. L., Gosselin, F., Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. Neuroimage, 37, 974-984.
Sorger, B., Goebel, R., Schiltz, C., Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. Neuroimage, 35, 836-52.
Steeves, J. K., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., Humphrey, G. K., Milner, A. D., Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. Neuropsychologia, 44, 594-609.
Stringer, S. M., Trappenberg, T. P., Rolls, E. T., de Araujo, I. E. T. (2002). Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. Network-Comp Neural, 13, 217-242.
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.
Suzuki, W., Matsumoto, K., Tanaka, K. (2006). Neuronal responses to object images in the macaque inferotemporal cortex at different stimulus discrimination levels. Journal of Neuroscience, 26, 10524-35.
Swithenby, S. J., Bailey, A. J., Brautigam, S., Josephs, O. E., Jousmaki, V., Tesche, C. D. (1998). Neural processing of human faces: a magnetoencephalographic study. Exp Brain Res, 118, 501-510.
Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. Journal of Neuroscience, 16, 4240-4249.
Tamura, H., Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. Cereb Cortex, 11, 384-399.
Tamura, H., Kaneko, H., Kawasaki, K., Fujita, I. (2004). Presumed inhibitory neurons in the macaque inferior temporal cortex: Visual response properties and functional interactions with adjacent neurons. J Neurophysiol, 91, 2782-2796.
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.
Tanaka, K. (1993). Neuronal mechanisms of object recognition. Science, 262, 685-688.
Tanaka, J. W., Farah, M. J. (1993). Parts and Wholes in Face Recognition. Q J Exp Psychol-A, 46, 225-245.
Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu Rev Neurosci, 19, 109-139.
Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. Current Opinion in Neurobiology, 7, 523-529.
Tanaka, J. W., Curran, T. (2001). A neural basis for expert object recognition. Psychol Sci, 12, 43-47.

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, J. W., Curran, T., Porterfield, A. L., Collins, D. (2006). Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. J Cognitive Neurosci, 18, 1488-1497.
Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. Cereb Cortex, 15, 526-534.
Tanskanen, T., Nasanen, R., Ojanpaa, H., Hari, R. (2007). Face recognition and cortical responses: effect of stimulus duration. Neuroimage, 35, 1636-44.
Tarkiainen, A., Cornelissen, P. L., Salmelin, R. (2002). Dynamics of visual feature analysis and ob object-level processing in face versus letter-string perception. Brain, 125, 1125-1136.
Taylor, M. J., Itier, R. J., Allison, T., Edmonds, G. E. (2001). Direction of gaze effects on early face processing: eyes-only versus full faces. Cognitive Brain Res, 10, 333-340.
Thierry, G., Pegna, A. J., Dodds, C., Roberts, M., Basan, S., Downing, P. (2006). An event-related potential component sensitive to images of the human body. Neuroimage, 32, 871-879.
Thierry, G., Martin, C. D., Downing, P., Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nat Neurosci, 10, 505-511.
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. European Journal of Neuroscience, 14, 869-876.
Thorpe, S., Crouzet, S., Kirchner, H. (2007). Saliency maps and ultra-rapid choice saccade tasks. J Vision, /7/9/30/, J. Vision abstract.
Tjan, B. S. (2001). Adaptive object representation with hierarchically distributed memory sites. Adv. Neural Inf. Process. Syst., 13, 66–72.
Tong, F., Nakayama, K., Vaughan, J. T., Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. Neuron, 21, 753-9.
Tootell, R. B. H., Dale, A. M., Sereno, M. I., Malach, R. (1996). New images from human visual cortex. Trends Neurosci, 19, 481-489.
Tootell, R. B., Tsao, D., Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. J Neurosci, 23, 3981-9.
Torralba, A., Oliva, A. (2003). Statistics of natural image categories. Network-Comp Neural, 14, 391-412.
Tovee, M. J., Rolls, E. T., Treves, A., Bellis, R. P. (1993). Information Encoding and the Responses of Single Neurons in the Primate Temporal Visual-Cortex. J Neurophysiol, 70, 640-654.
Tovee, M. J., Rolls, E. T., Azzopardi, P. (1994). Translation-Invariance in the Responses to Faces of Single Neurons in the Temporal Visual Cortical Areas of the Alert Macaque. J Neurophysiol, 72, 1049-1060.
Tovee, M., Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. Vis Cog, 2, 35-58.
Tovee, M. J. (1998). Is face processing special? Neuron, 21, 1239-1242.
Treisman, A. M., Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12, 97-136.
Trenner, M. U., Schweinberger, S. R., Jentzsch, I., Sommer, W. (2004). Face repetition effects in direct and indirect tasks: an event-related brain potentials study. Cognitive Brain Res, 21, 388-400.
Ts'o, D. Y., Frostig, R. D., Lieke, E. E., Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. Science, 249, 417-420.
Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. Nat Neurosci, 6, 989-995.
Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. Science, 311, 670-674.
Tsunoda, K., Yamane, Y., Nishizaki, M., Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci, 4, 832-838.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. Cereb Cortex, 5, 1-11.
Ullman, S., Vidal-Naquet, M., Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nat Neurosci, 5, 682-687.
Ungerleider, L. G., Haxby, J. V. (1994). 'What' and 'where' in the human brain. Current Opinion in Neurobiology, 4, 157-165.
Ungerleider, L. G., Galkin, T. W., Desimone, R., Gattass, R. (2007). Cortical Connections of Area V4 in the Macaque. Cereb Cortex,
Valentine, T., Bruce, V. (1988). Mental Rotation of Faces. Mem Cognition, 16, 556-566.
Van Essen, D. C. (2003). Organization of Visual Areas in Macaque and Human Cerebral Cortex. In The Visual Neurosciences (C. L. M. W. J.S., Eds.). The MIT Press, pp. 507-521.
Vanni, S., Warnking, J., Dojat, M., Delon-Martin, C., Bullier, J., Segebarth, C. (2004). Sequence of pattern onset responses in the human visual areas: an fMRI constrained VEP source analysis. Neuroimage, 21, 801-17.
VanRullen, R., Thorpe, S. J. (2001a). The time course of visual processing: From early perception to decision-making. J Cognitive Neurosci, 13, 454-461.
VanRullen, R., Thorpe, S. J. (2001b). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. Perception, 30, 655-668.
VanRullen, R., Guyonneau, R., Thorpe, S. J. (2005). Spike times make sense. Trends Neurosci, 28, 1-4.
VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. Vision Res, 46, 3017-3027.
Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. European Journal of Neuroscience, 11, 1239-55.
Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. Trends in Cognitive Science, 9, 585-594.
Wang, G., Tanaka, K., Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. Science, 272, 1665-1668.
Wang, G., Tanifuji, M., Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. Neurosci Res, 32, 33-46.
Wang, Y., Fujita, I., Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. Nat Neurosci, 3, 807-813.
Wang, Y. C., Zhou, Y. F., Ma, Y. Y., Leventhal, A. G. (2005). Degradation of signal timing in cortical areas V1 and V2 of senescent monkeys. Cereb Cortex, 15, 403-408.
Watanabe, S., Kakigi, R., Puce, A. (2003). The spatiotemporal dynamics of the face inversion effect: A magneto- and electro-encephalographic study. Neuroscience, 116, 879-895.
Waydo, S., Kraskov, A., Quiroga, R. Q., Fried, I., Koch, C. (2006). Sparse representation in the human medial temporal lobe. Journal of Neuroscience, 26, 10232-10234.
Webster, M. J., Ungerleider, L. G., Bachevalier, J. (1991). Connections of Inferior Temporal Areas Te and Teo with Medial Temporal-Lobe Structures in Infant and Adult Monkeys. Journal of Neuroscience, 11, 1095-1116.
Webster, M. J., Bachevalier, J., Ungerleider, L. G. (1994). Connections of Inferior Temporal Areas Teo and Te with Parietal and Frontal-Cortex in Macaque Monkeys. Cereb Cortex, 4, 470-483.
Wenger, M. J., Ingvalson, E. M. (2002). A decisional component of holistic encoding. J Exp Psychol Learn Mem Cogn, 28, 872-92.
Wenger, M. J., Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. J Exp Psychol Learn Mem Cogn, 29, 1106-18.
Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., Goodale, M. A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. Curr Biol, 10, 1455-8.
Williams, C. C., Henderson, J. M. (in press). The face inversion effect is not a consequence of aberrant eye movements. Mem Cognition,
Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J

Neurophysiol, 92, 1830-1839.
Yamane, S., Kaji, S., Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp Brain Res, 73, 209-214.
Yin, R. K. (1969). Looking at upside-down faces. J Exp Psychol Human, 81, 141–145.
Young, A. W., Hay, D. C., McWeeny, K. H., Ellis, A. W., Barry, C. (1985). Familiarity decisions for faces presented to the left and right cerebral hemispheres. Brain Cogn, 4, 439-50.
Young, A. W., Hellawell, D., Hay, D. C. (1987). Configurational Information in Face Perception. Perception, 16, 747-759.
Young, M. P., Yamane, S. (1992). Sparse Population Coding of Faces in the Inferotemporal Cortex. Science, 256, 1327-1331.
Young, M. P. (2000). The architecture of visual cortex and inferential processes in vision. Spatial Vision, 13, 137-146.
Yovel, G., Paller, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered. Neuroimage, 21, 789-800.
Yovel, G., Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. Curr Biol, 15, 2256-2262.
Zeki, S. (1993). A Vision of the brain. Oxford: Blackwell Scientific Publication.
Zhang, Y., Wang, Y., Wang, H., Cui, L., Tian, S., Wang, D. (2001). Different processes are involved in human brain for shape and face comparisons. Neurosci Lett, 303, 157-60.
Zhong, Y. M., Rockland, K. S. (2003). Inferior parietal lobule projections to anterior inferotemporal cortex (area TE) in macaque monkey. Cereb Cortex, 13, 527-540.
Zion-Golumbic, E., Bentin, S. (2007). Dissociated neural mechanisms for face detection and configural encoding: evidence from N170 and induced gamma-band oscillation effects. Cereb Cortex, 17, 1741-9.