



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Cotutelle internationale avec :

Présentée et soutenue par :
Gladys Barragan-Jason

Le 18 février 2013

Titre :

La dynamique du traitement des visages : du percept à la familiarité

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

Centre de recherche Cerveau et Cognition (CerCo)

Directeur(s) de Thèse :

Emmanuel Barbeau (CerCo, Toulouse)

Rapporteurs :

Boutheina Jemel (LRNEC, Montréal, Québec)

Bruno Rossion (IPSY, IoNS, Louvain, Belgique)

Autre(s) membre(s) du jury :

Nathalie George (CRICM, UPMC, Paris)

Simon Thorpe (CerCo, Toulouse)

Jean-Christophe Sol (CHU, Toulouse)

Merci !

Je remercie *Bruno Rossion* et *Boutheina Jemel* pour avoir accepté d'évaluer ce travail.

Je remercie *Simon Thorpe*, *Nathalie George* et *Jean-Christophe Sol* pour avoir accepté d'être membre du jury.

Merci à *Emmanuel* pour son encadrement, ses encouragements, son écoute, sa compréhension. Merci pour sa patience et son aide qui ont permis de réaliser et de finaliser ce travail de thèse. Et d'une !

Merci à *Fabienne* pour son soutien, sa bonne humeur et son enthousiasme en toutes circonstances. Je n'oublie toujours pas la proposition de stage...

Merci à toute l'équipe médicale de Ranguel pour leur accueil et leur confiance.

Merci à *Luc Valton* pour ses précieux conseils, son implication et pour m'avoir permis dès mon « plus jeune âge » de participer à de nombreuses recherches passionnantes. Je suis vraiment une sacrée chanceuse, merci beaucoup !

Je tiens également à remercier *Marie Denuelle* pour sa gentillesse et sa confiance.

Merci à toute l'équipe paramédicale du laboratoire du sommeil et de l'épilepsie : *Joce*, *Martine*, *Mireille*, *Sylvie*, *Véro*, *Ginou*, *Céline*, *Sandrine*,... sans oublier *Bernard* pour simplement avoir été supers en toutes circonstances. Vous êtes vraiment géniaux, c'est toujours un plaisir de venir vous voir!

Merci à tous les volontaires et les patients pour avoir accepté de participer aux protocoles.

Merci à *Michèle* pour sa bienveillance et sa compréhension. Merci pour son écoute et son côté humain, indispensable, lorsqu'on se retrouve malencontreusement autour de la fameuse table ronde...

Merci à *Lolly Tyler* pour m'avoir accueillie en stage dans son laboratoire. Merci beaucoup à *Alex Clarke* pour son accueil, sa gentillesse, sa patience et son encadrement.

Merci à tous les chercheurs du CerCo et les ITAs pour leur gentillesse, leur disponibilité et les nombreuses discussions de couloirs.

Merci à tous les étudiants du CerCo pour la bonne ambiance, les soirées et pour m'avoir fait assez confiance pour vous représenter au conseil de labo.

Merci à *Mélanie* pour les fou-rires (c'est sympa l'éeG) et pour son aide pendant les révisions...

Merci à *Seb* pour m'avoir appris les rouages du CerCo et pour avoir été si cool à mon arrivée.

Merci à *Laure* pour sa bonne humeur et pour ne pas m'en avoir trop voulu pour avoir été enrôlée à la tête d'une certaine association.

Merci à *Rodi* pour son amitié, sa gentillesse et sa sincérité... heureusement qu'elle est là cette Rodi !

Je tiens à remercier mes joyeux acolytes, *Valérie* et *Gabriel* pour toutes les discussions (modèles au hasard,... Gab ? et les potins, Valérie ?), et pour leur soutien permanent.

Merci à *Gab* pour son amitié et sa gentillesse, pour nos collaborations présentes et futures qui seront grâce à lui toujours agréables et détendues.

Merci à *Marlène* pour les longues conversations café, les fou-rires et son franc parlé qui la rend si attachante. Merci d'avoir été là pour le meilleur et pour le pire...

Merci aux relecteurs de fautes d'orthographe, *Thomas*, *Marlène* et *Valérie*...il en reste peut-être quelques unes, c'est vraiment très long une thèse.

Merci à *Jonathan* pour nos fou-rires à l'hôpital autour du chariot de la honte.

Merci à *Romain* pour avoir été collègue, colloque et amie (avec un « e », parfaitement !), tout ça à la fois.

Merci à ma famille, en particulier à ma mère, pour ne pas m'avoir demandée chaque jour ce que j'allais faire après... pour leur soutien et leur amour. Merci beaucoup !

Merci à toi, cher Dr *Maximus*, pour ton aide sur le traitement des données EEG, pour les collaborations passées et futures, pour ce que l'on partage au quotidien et pour tous les trucs trop cool que l'on a prévu. Ça valait le coup de faire une thèse, ne serait-ce que pour faire ta rencontre...

A ma petite maman,

A ma famille,

A Maxime.

Sommaire

1	Chapitre 1. La reconnaissance des visages familiers.....	13
1.1	Traitement des visages et comportement	13
1.1.1	Un traitement performant et rapide: le débat de la spécificité	14
1.1.2	Modèle cognitif de la reconnaissance des visages.....	27
1.1.3	Dynamique temporelle de la reconnaissance des visages.....	31
1.2	Organisation cérébrale et perception des visages	38
1.2.1	Régions cérébrales impliquées dans la reconnaissance des visages	38
1.2.2	Modèles de la reconnaissance des visages	52
1.2.3	Potentiels évoqués et dynamique de la reconnaissance des visages.....	59
2	Chapitre 2. A la recherche du temps minimum	75
2.1	Déterminer la latence comportementale minimale pour reconnaître le visage d'une personne familière.....	75
2.1.1	Tâches de catégorisation visuelle rapide : tâche de Go/no-go et de Speed accuracy boosting (SAB).....	75
2.1.2	Utilisation de stimuli naturels en contexte : optimisation des performances comportementales.....	78
2.1.3	Calcul du temps de réaction minimum	81
2.2	Déterminer la latence neurophysiologique minimale pour reconnaître le visage d'une personne familière	83
2.2.1	Des techniques de haute précision temporelle	84
2.2.2	A la recherche de l'activité électrophysiologique la plus précoce : Moyennage vs essai par essai	90
2.3	Présentation des études.....	95
3	Chapitre 3. A quelle vitesse reconnaît-on un visage célèbre?	97
3.1	Article 1: How fast is famous face recognition? (publié)	101
3.2	Article 2: Fast and Famous: Looking for the fastest speed at which a face can be recognized (soumis).....	105
3.3	Article 3 : Top-down vs bottom-up face recognition (en préparation).....	109
4	Chapitre 4. Quelles sont les latences électrophysiologiques de la catégorisation rapide des visages célèbres ?	131
4.1	Article 4: The neural dynamics of human face detection in the wild (en préparation)	135
4.2	Article 5: The speed of face recognition (soumis).....	153
5	Discussion générale et perspectives.....	175

5.1	La reconnaissance des visages familiers, un traitement rapide ?	175
5.1.1	A quelle vitesse reconnaît-on un visage familier ?	175
5.1.2	A quelle latence notre cerveau reconnaît-il un visage familier ?	179
5.2	Reconnaissance vs catégorisation : un coût temporel révélateur	182
5.2.1	La reconnaissance d'un visage familier nécessite 100-200 ms de plus que sa catégorisation	182
5.2.2	Une différence électrophysiologique de 120-150 ms entre la catégorisation et la reconnaissance.....	183
5.3	Vers l'hypothèse d'un traitement massivement parallèle	185
5.3.1	Reconnaissance vs catégorisation : des mécanismes distincts ?.....	185
5.3.2	Qu'est-ce qui différencie la catégorisation de la reconnaissance bottom-up ?	186
5.3.3	Le modèle d' « emerging familiarity »	189
5.4	Latences électrophysiologiques intracrâniennes de la reconnaissance des visages célèbres : Etude préliminaire.....	193
5.5	Pour aller plus loin... Perspectives et conclusion générale.....	198
5.5.1	Le mot sur le bout de la langue : un accès direct à la méta-cognition	198
5.5.2	Etudier la personnes sous tous ces aspects : décodage du concept	202
5.5.3	Conclusion	203

Introduction Générale

Nous sommes le 21 septembre 2012, premier jour de l'automne. Une magnifique journée ensoleillée vient de commencer. Malheureusement, ce n'est pas vraiment le moment d'aller gambader dehors car selon mon Google agenda, cette date marque aussi la finalisation de ce mémoire de thèse ! J-28 pour le premier jet ! Une vague de stress m'envahit,... Je jette un coup d'œil par la fenêtre, j'aperçois au loin une silhouette que je connais bien, c'est la petite grand-mère d'à côté, qui passe tous les jours avec sa brouette afin de déposer ses déchets végétaux dans le compost de l'autre côté du champ. Elle m'a vu, elle aussi, à travers la baie vitrée et me salue d'un signe de la main avant de vaquer à ses occupations. J'aperçois aussi Mimi-chat et Miou-miou, les chats de la maison, qui se prélassent et profitent des premiers rayons de soleil, allongés sur le petit muret tiède qui nous sépare des voisins. Je détourne mon regard dans la direction opposée, côté rivière,... J'aperçois au loin un promeneur accompagné d'un chien je pense, un labrador, oui c'est ça ! Ils doivent sans doute aller pêcher à la rivière, les veinards!

La cloche de l'église retentit tout à coup,... Déjà 10h00 ! Assez divagué, je me remets un peu au travail. Je dois maintenant écrire quelques lignes sur mon sujet de thèse, le traitement des visages familiers,... La situation que je viens de décrire plus haut en illustre d'ailleurs fort bien les différents aspects. Je viens en effet de réaliser différents traitements cognitifs me permettant de détecter une présence humaine (le promeneur) dans une scène naturelle (route, rivières, végétation,...), d'identifier et d'interagir avec une personne familière (ma voisine), de différencier mon propre chat d'un autre animal et de retrouver le nom qui lui est associé (Mimi-chat), tout cela très rapidement, sans m'en rendre compte, sans aucun effort, de manière totalement automatique et irréprouvable. Vraiment, je m'étonne chaque jour des prouesses que le cerveau est capable d'accomplir !

Le cerveau traite en permanence des informations d'une formidable complexité provenant du monde extérieur. Ces informations très diverses (visuelles, tactiles, odorantes,...) et uniques (un même objet peut être présenté sous de nombreux points de vue, à différents niveaux d'éclairage,...) sont pourtant traitées et reconnues avec brio par notre cerveau, nous permettant d'interagir correctement avec notre environnement. Pour reconnaître, il faut

tout d'abord connaître, c'est-à-dire avoir formé une représentation de l'objet en mémoire et être capable d'y accéder en cas de besoin. Le cerveau ne va pas traiter chacune des ces informations comme si elle était unique, ce serait beaucoup trop coûteux ; il va les regrouper au sein de catégories (vivant, non-vivant, animal,...) (Rosch, 1975).

Une question reste en suspend,...Comment mon cerveau va-t-il réagir et traiter ce stimulus particulier que constitue « ma voisine »? Chaque personne que l'on connaît semble constituer un stimulus que nous traitons de manière unique (« je reconnais ma voisine, il n'en existe qu'un seul exemplaire ») même s'il est possible de regrouper plusieurs informations multimodales (voix, visage, démarche, traits de caractère,...) relatives à cette personne dans une *catégorie individuelle*. Le visage tient une place privilégiée dans la reconnaissance des personnes. Les visages véhiculent en effet de nombreuses d'informations telles que l'âge, les émotions, les intentions, le genre, l'attractivité qui sont cruciales pour nos relations sociales. C'est pourquoi, dans l'opinion publique mais également dans la littérature scientifique, la capacité à reconnaître et identifier les visages (ex : saluer un collègue, sauter au cou de son petit ami ou voir que son acteur préféré joue dans le dernier film de Woody Allen) semblent constituer une habileté triviale exécutée de manière efficace, très *rapidement* et sans effort.

Pourtant, les visages semblent être des stimuli de forte complexité visuelle : ils sont très similaires entre eux (deux yeux, un nez, une bouche), porteurs de nombreuses caractéristiques (distance entre les yeux, hauteur de la bouche, longueur du nez, texture,...) et sujets à de nombreux changements (luminosité, expression faciale, point de vue, âge,...). Reconnaître un visage familier parmi des centaines d'inconnus devrait donc au contraire nécessiter une puissance de calcul phénoménale qui devrait prendre du temps. Dans ce sens, l'hypothèse, selon laquelle un système de reconnaissance spécifique permettrait de traiter *rapidement* les informations complexes véhiculées par ces stimuli particuliers que sont les visages, a été suggérée. Effectivement, d'un point de vue évolutionniste, il semble crucial de pouvoir rapidement déterminer si la personne qui nous fait face est amie ou ennemie pour adapter son comportement en conséquence. *Cependant, est-ce réellement si rapide ? Ne serait-ce pas simplement une idée reçue ? Rapide, admettons, mais à quelle vitesse ?*

Ma thèse en 5 points

✓ De part leur rôle fondamental dans nos interactions sociales, les visages sont supposés être des stimuli particuliers reconnus rapidement et efficacement. Pourtant, la vitesse de cette reconnaissance reste inconnue, les latences liées à la familiarité étant très variables voire contradictoires à travers les études.

✓ A l'aide de tâches à forte contrainte temporelle (tâche de go/no-go et SAB procédure), nous voulons déterminer les latences comportementales et électrophysiologiques les plus précoces permettant de reconnaître un visage familier.

✓ Nos résultats révèlent que reconnaître un visage familier nécessite au minimum 360 ms au niveau comportemental (temps de réaction minimum) et 230 ms au niveau cérébral, voire 200 ms pour les réponses les plus rapides (analyse multi-variée essai par essai).

✓ Reconnaître un visage au niveau de la familiarité semble donc plutôt lent par rapport à sa catégorisation en tant que stimulus visage et nécessite la mise en jeu de processus additionnels aux mécanismes purement feedforward proposés pour la catégorisation.

✓ Trois modèles théoriques – (1) un modèle purement feedforward, (2) l'implication du « core network » ou (3) l'activation de l'ensemble de la voie ventrale en parallèle - peuvent être proposés, nos résultats supportant plutôt le troisième modèle. Nous suggérons que reconnaître un visage familier nécessiterait la mise en jeu de boucles récurrentes et des interactions à distance, impliquant l'ensemble de la voie visuelle ventrale, y compris les régions les plus antérieures comme le cortex périrhinal.

Le premier chapitre sera composé de deux parties théoriques : la première partie portera sur les études comportementales de la reconnaissance des visages et la seconde sur les études neurophysiologiques relatives à cette fonction, chacune comportant trois sous-parties sur (1) une présentation de l'état de l'art, (2) les modèles théoriques et (3) la dynamique temporelle.

Un second chapitre présentera la méthodologie particulière que nous avons utilisée : nous expliquerons comment et pourquoi, en adaptant le protocole de catégorisation visuelle rapide (tâche de go/no-go, développée initialement pour étudier la rapidité du système visuel) et en développant de nouveaux protocoles de catégorisation ultra-rapide (« Speed and Accuracy Boosting procedure »), nous déterminerons les latences comportementales et électrophysiologiques les plus rapides pour reconnaître explicitement des visages célèbres parmi des inconnus. Nous détaillerons les différentes analyses (évaluation du temps de réaction comportemental minimum et méthode de MVPA (Multi-Variate Pattern Analysis) appliquée à des données obtenues en EEG de surface).

Le chapitre 3 présentera les données comportementales (Article 1, 2 et 3) acquises au cours de cette thèse permettant de déterminer le temps minimal pour reconnaître un visage célèbre.

Le chapitre 4 étudiera la dynamique électrophysiologique de cette reconnaissance (EEG de surface) à travers deux articles (Article 4 et 5) ainsi que son possible lien avec la décision de familiarité lors d'une tâche de reconnaissance rapide de visages célèbres.

Enfin, l'ensemble de ces travaux sera discuté dans un dernier chapitre permettant l'élaboration d'un modèle, la présentation de données préliminaires en EEG intracérébral et proposera des pistes et projets permettant d'aller plus loin dans l'investigation.

1 Chapitre 1. La reconnaissance des visages familiers

Ce premier chapitre porte sur la reconnaissance des visages. Par « reconnaissance », j'entends son sens strict, c'est-à-dire les processus permettant de déterminer qu'un visage est connu ou non. Même si, globalement, peu d'études se sont intéressées à la reconnaissance des visages en ces termes, on observe cependant un intérêt grandissant pour les visages familiers depuis le début des années 80. Les principales études comportementales et neurophysiologiques de la reconnaissance des visages familiers seront donc présentées dans ce premier chapitre.

Traitement des visages et comportement

Les visages semblent constituer l'un des stimuli visuels les plus importants de notre environnement. En effet, nous traitons en permanence le visage de nos congénères afin d'en extraire les moindres détails et d'obtenir des informations sur l'état émotionnel ou même sur ce que pense ou perçoit notre interlocuteur. Notre propension à traiter les visages de manière compulsive nous amène même à en voir là où il n'y en a pas (par exemple : gravure d'un rocher, tache, nuages (Figure 1)). Ce phénomène appelé paréidolie a donné naissance à de nombreuses anecdotes. Par exemple, plusieurs apparitions supposées de la Vierge Marie ou de Jésus dans les nuages, sur des murs, ou encore sur des toasts ont été rapportées (voir Figure 1). Cette « obsession » pour les visages nous permettrait d'inférer à autrui des états mentaux afin d'en prédire le comportement et de pouvoir ainsi s'adapter et interagir correctement en société.

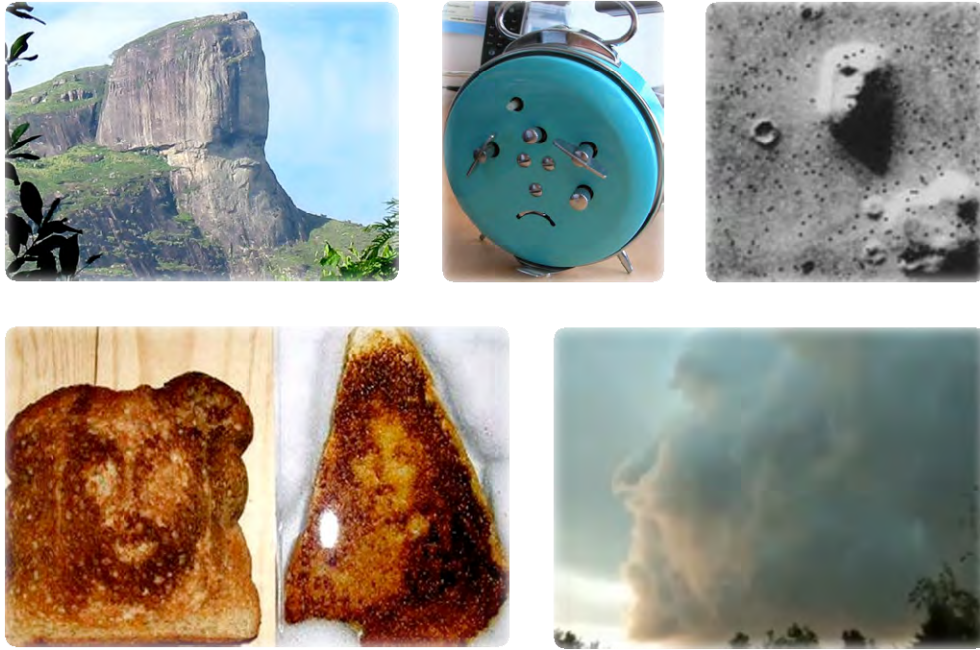


Figure 1 Exemples de paréidolie (source: Wikipedia). En haut : La falaise de Pedra da Gavea située au Brésil, évoquant un immense visage barbu à l'origine de nombreuses légendes amérindiennes sur l'existence d'un ancêtre représentant un dieu ou un roi, un réveil matin avec une triste mine, une photographie d'un rocher prise sur mars évoquant un visage. En bas : Apparitions supposées de Jésus et de Marie sur des toasts et profil de visage dans les nuages.

1.1.1 Un traitement performant et rapide: le débat de la spécificité

L'idée selon laquelle les visages seraient des stimuli particuliers provient de différentes observations : tout d'abord, les visages sont des stimuli visuels particuliers tant sur le plan social et affectif (Adolphs, 2003), développemental (Pascalis et al., 1995) que phylogénétique (Leopold et Rhodes, 2010). Ils nous renseignent sur l'identité, le genre, l'âge, le groupe ethnique ou encore les émotions de la personne à reconnaître (Squier et Mew, 1981). Les visages sont également des stimuli très complexes : Ils constituent en effet une catégorie visuelle dont les exemplaires sont tellement homogènes (ovale contenant deux yeux, un nez, une bouche) qu'ils nécessitent un traitement précis de nombreuses caractéristiques (distance entre les yeux, hauteur de la bouche, longueur du nez, texture, etc) pour être différenciés les uns des autres. Les visages sont également soumis à de nombreux changements tels que les variations de luminosité ou de point de vue par exemple. Enfin, contrairement aux autres objets, les visages sont des stimuli qu'il est nécessaire d'individualiser ; nous devons obligatoirement déterminer qu'il s'agit de « ce » visage en

particulier (traitement au niveau de la familiarité ou de l'identité) et non d'un visage parmi d'autres (traitement au niveau superordonné, voir Figure 2).

Tous ces facteurs laissent supposer que différencier un visage familier particulier de centaines d'inconnus devrait nécessiter une puissance de calcul phénoménale. Par conséquent, la reconnaissance du visage au niveau individuel devrait être un processus lent et fastidieux. Pourtant, la majorité des études suggèrent au contraire que la reconnaissance des visages serait rapide, automatique et même possible en quasi absence d'attention (Reddy et al., 2006). Dans ce sens, l'hypothèse, selon laquelle un système de reconnaissance spécifique permettrait de traiter rapidement les informations complexes véhiculées par ces stimuli particuliers que sont les visages, a été suggérée.

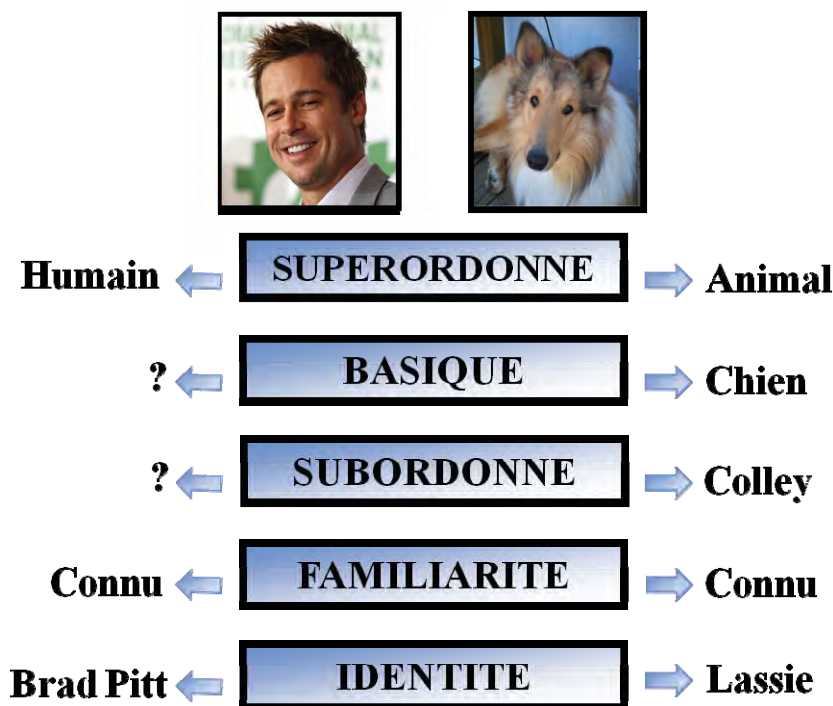


Figure 2 Illustration des différents niveaux de catégorisation adaptée des travaux de Rosch (1975) sur les objets. La photographie de gauche peut être catégorisée à différents niveaux d'abstraction. A un niveau superordonné (humain), familier (« je connais ce visage ») ou de l'identité (« il s'agit de Brad Pitt, un acteur célèbre »). La photographie de droite peut être catégorisée au niveau superordonné (animal), au niveau basique (chien) au un niveau subordonné (colley) et même au niveau de la familiarité (« je connais ce chien ») ou de l'identité (« c'est Lassie » ou « c'est mon propre chien, Nala »). Selon la théorie de Rosch, les objets seraient catégorisés au niveau basique. Les visages humains quant à eux nécessitent d'être traités au niveau de la familiarité ou de l'identité (Tanaka, 2001).

1.1.1.1 Un traitement spécifique des visages

Plusieurs arguments soutiennent l'hypothèse de l'existence d'un **traitement spécifique** des visages (e.g. Morton et Johnson, 1991 ; Diamond et Carey ; 1986 ; Yin, 1969 ; Kanwisher et al., 1997). Premièrement, il semblerait que la reconnaissance des personnes par leur visage apparaisse très tôt au cours du développement. Dès la naissance, le nouveau-né aurait tendance à regarder préférentiellement les stimuli ressemblant à des visages comparés à d'autres catégories d'objets (Johnson et al., 1991) (Figure 3).

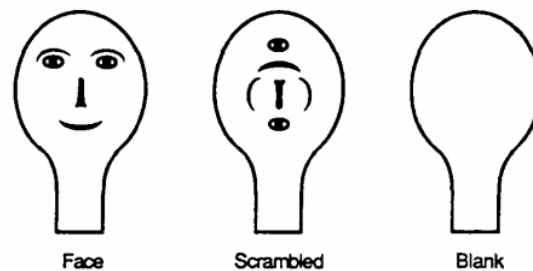


Figure 3 Exemples de stimuli utilisés dans Johnson et al. (1991). A gauche, un stimulus de type visage, au centre un stimulus comprenant les éléments du visage mais dont la configuration a été modifiée, à droite un stimulus de forme ovale sans élément du visage. Le nouveau-né aurait tendance à regarder plus longtemps les stimuli ayant une configuration de type visage comparé à des visages mélangés ou dépourvus d'éléments.

De plus, il semble qu'un nouveau né soit même capable de reconnaître le visage de sa mère (Pascalis et al., 1995). Par exemple, des études menées chez des nouveau-nés de 4 jours ont montré qu'ils fixaient plus longtemps leur mère lorsqu'elle était dans la pièce en comparaison à une personne étrangère (Field et al., 1984 ; Bushnell et Mullin, 1989). Une autre étude, utilisant une technique appelée la technique de « High Amplitude Sucking » permettant d'enregistrer le nombre de suctions effectuées par le nouveau-né face à une stimulation (le nombre de suctions est augmenté par les stimuli non-familiers) semble venir confirmer ces résultats (Walton et al., 1992). Il semble cependant que le nouveau-né utilise des informations non-spécifiques au visage telles que les cheveux ou la tenue vestimentaire (Pascalis et al., 1995).

Deuxièmement, certains effets comportementaux spécifiques aux visages ont été rapportés chez l'adulte suggérant que cette spécificité de traitement perdurerait tout au long de l'existence. Tout d'abord, les stimuli visages sont détectés très rapidement dans une scène

visuelle. En effet, une étude utilisant un protocole de choix saccadique a montré que les humains peuvent initier des saccades vers les visages comparés à d'autres catégories d'objets (2 stimuli présentés simultanément sur un écran) seulement 100 ms après la présentation de l'image (Crouzet et al., 2010) (Figure 4). De plus, lorsqu'il est demandé au sujet de faire une saccades vers les images ne contenant pas de visage, les toutes premières réponses (avant 140 ms) sont en faveur des visages comme si les sujets présentaient un biais vers cette catégorie de stimuli (Crouzet et al., 2010 ; Crouzet et Thorpe, 2011) (Figure 4).

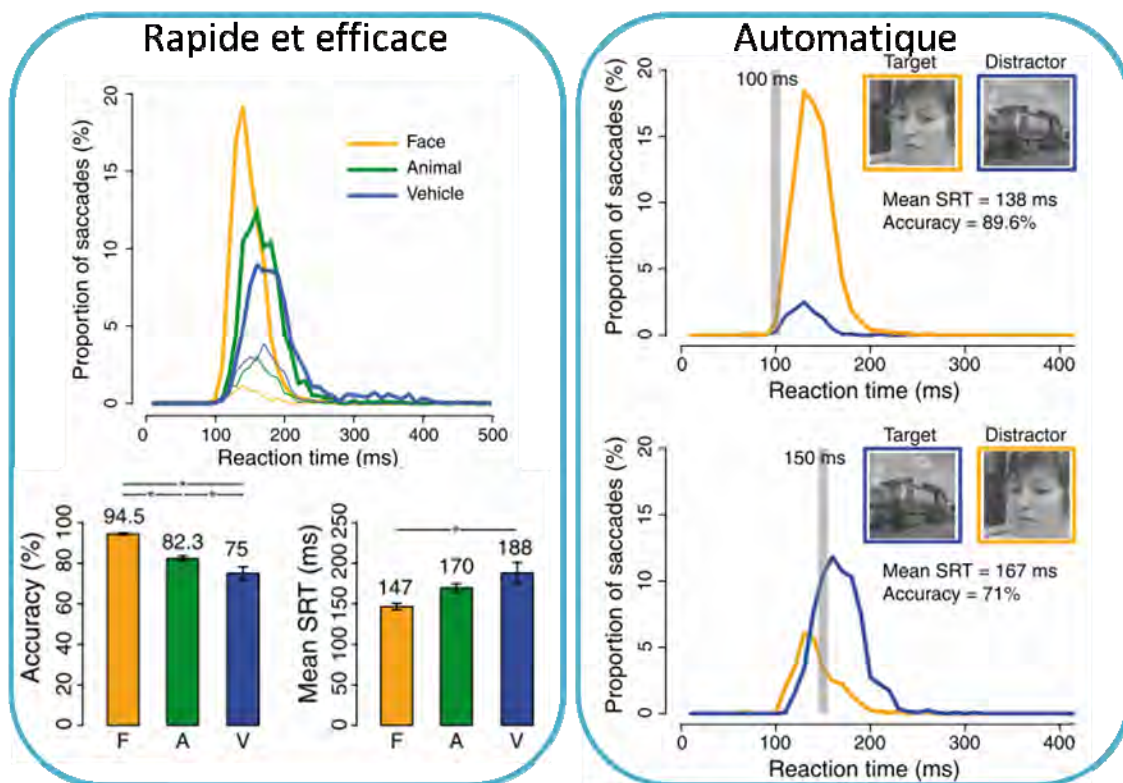


Figure 4 Figures extraites de Crouzet et al. (2010). A gauche : Distribution des temps de réaction en saccades pour trois catégories d'objets (visage en orange, animal en vert, véhicule en bleu) montrant que les réponses les plus rapides sont en faveur de la catégorie des visages. En bas, la précision et le temps de réaction moyen sont présentés pour chaque catégorie. Encore une fois, les visages sont reconnus avec la plus grande précision et plus rapidement que les autres catégories d'objets. A droite : le graphique présenté en haut montre que les saccades correctes vers les visages (en orange) sont significativement différentes des saccades incorrectes vers les véhicules (en bleu) dès 100 ms. En bas, le graphique montre que lorsque les véhicules sont cibles, les saccades sont initiées plus tardivement, 150 ms après la présentation de l'image. Une forte proportion de saccades vers les visages (ici distracteurs) est également observée pour les réponses les plus rapides témoignant d'un biais vers les visages.

D'autre part, l'inversion d'un stimulus visuel (rotation à 180°) entraîne une dégradation plus importante des performances dans des tâches de traitement de visages comparés à d'autres catégories d'objets (Yin, 1969 ; Diamond et Carey, 1986 ; Valentine et Bruce, 1986 ; Rhodes et al., 1993; Farah et al., 1995). Le fait d'inverser le visage perturberait un traitement dit configural ou holistique spécifique aux visages, c'est-à-dire un traitement global en fonction des relations spatiales (traitement comme un tout indissociable ; Galton, 1883 ; Tanaka et Farah, 1993 ; Rossion et al., 2000 ; Schiltz et Rossion, 2006). Les autres catégories d'objets sembleraient au contraire, plutôt traitées de manière analytique, c'est-à-dire, trait par trait, ce traitement étant moins perturbé par l'inversion (Diamond et Carey, 1986 ; Rhodes et al., 1993 ; Farah et al., 1995). Une illusion permettant d'illustrer ce traitement dit configural est l'illusion Thatcher (Thompson, 1980, Figure 5). En effet, lorsqu'un visage est présenté à l'envers, l'inversion de certains éléments du visage comme la bouche ou les yeux n'est pas perçue par l'observateur alors que cette transformation « saute » littéralement aux yeux lorsque le visage est présenté à l'endroit. Ce traitement configural semble donc perturbé lorsque l'on présente le visage à l'envers.



Figure 5 Illustration de l'illusion Thatcher (adapté de Thompson, 1980). A gauche le visage intact inversé et à l'endroit. A droite, les yeux et la bouche ont été inversés, et cette transformation passe inaperçue lorsque le visage est présenté à l'envers alors qu'elle est très visible lorsque le visage est présenté à l'endroit.

De plus, un autre argument de taille en faveur d'un système spécifique de traitement des visages est l'existence d'un trouble spécifique appelé la prosopagnosie (Bodamer, 1947). La prosopagnosie peut être définie comme « l'incapacité à reconnaître les visages familiers, sur la base de la perception visuelle, ne pouvant être expliquée ni par des déficits visuels de bas niveau, ni par une altération cognitive telle que la confusion mentale, une aphasie, une amnésie ou d'autres signes de détérioration intellectuelle » (définition extraite de Joubert et al., 2008). Ainsi, les personnes atteintes de prosopagnosie semblent incapables de reconnaître une personne *en se basant uniquement sur leur visage* alors que le traitement des autres informations véhiculées par le visage (âge, genre, ethnie) reste préservé (Sergent et Poncet, 1990). Il existe deux types de prosopagnosies, la prosopagnosie acquise où les troubles apparaissent suite à un traumatisme, une lésion ou une maladie neuro-dégénérative (Evans et al., 1995 ; Joubert et al., 2003) et la prosopagnosie développementale apparaissant au cours du développement et ayant pour origine des lésions néonatales ou des anomalies génétiques. Dans le cas de la prosopagnosie développementale, les patients ne réalisent souvent pas qu'ils sont incapables de reconnaître des visages et mettent en place des stratégies différentes basées sur des informations non-faciales, telles que les cheveux, la démarche, les vêtements, la voix ou encore un détail particulier comme des lunettes ou une moustache. Au contraire la prosopagnosie acquise est vécue comme un réel handicap par les personnes touchées qui mettent en place de manière consciente des stratégies similaires aux personnes atteintes de prosopagnosie développementale.

De manière intéressante, dans certains cas appelés prosopagnosie pure, l'aptitude pour reconnaître les objets est totalement préservée (Riddoch et al., 2008 ; Rossion et al., 2003 ; Caldara et al., 2005 ; Busigny et al., 2010). Par exemple, Busigny et al. (2010) présente le cas de GG, un patient atteint de prosopagnosie acquise. A l'aide d'un large panel d'expériences, ces auteurs montrent que GG présente un déficit spécifique du traitement holistique des visages au niveau individuel. Dans l'une de leur expérience, ils observent que les performances de GG dans des tâches de discrimination d'objets individuels sont équivalentes aux performances des contrôles alors qu'une diminution significative de ses performances est observée lorsque les stimuli utilisés sont des visages (Figure 6). De plus, même si cela reste discuté dans la littérature, l'absence d'effet d'inversion chez les patients

prosopagnosiques a été rapportée (Farah et al., 1995 ; Busigny et al., 2010). En effet, Busigny et al. (2010) montrent, à l'aide d'une tâche de matching, que GG ne présente pas d'effet d'inversion contrairement aux sujets contrôles (Figure 6). Ainsi, le cas de ce patient suggère qu'une atteinte spécifique du traitement individuel des visages peut être observée, argument fort en faveur de l'hypothèse d'un module spécifique de traitement.

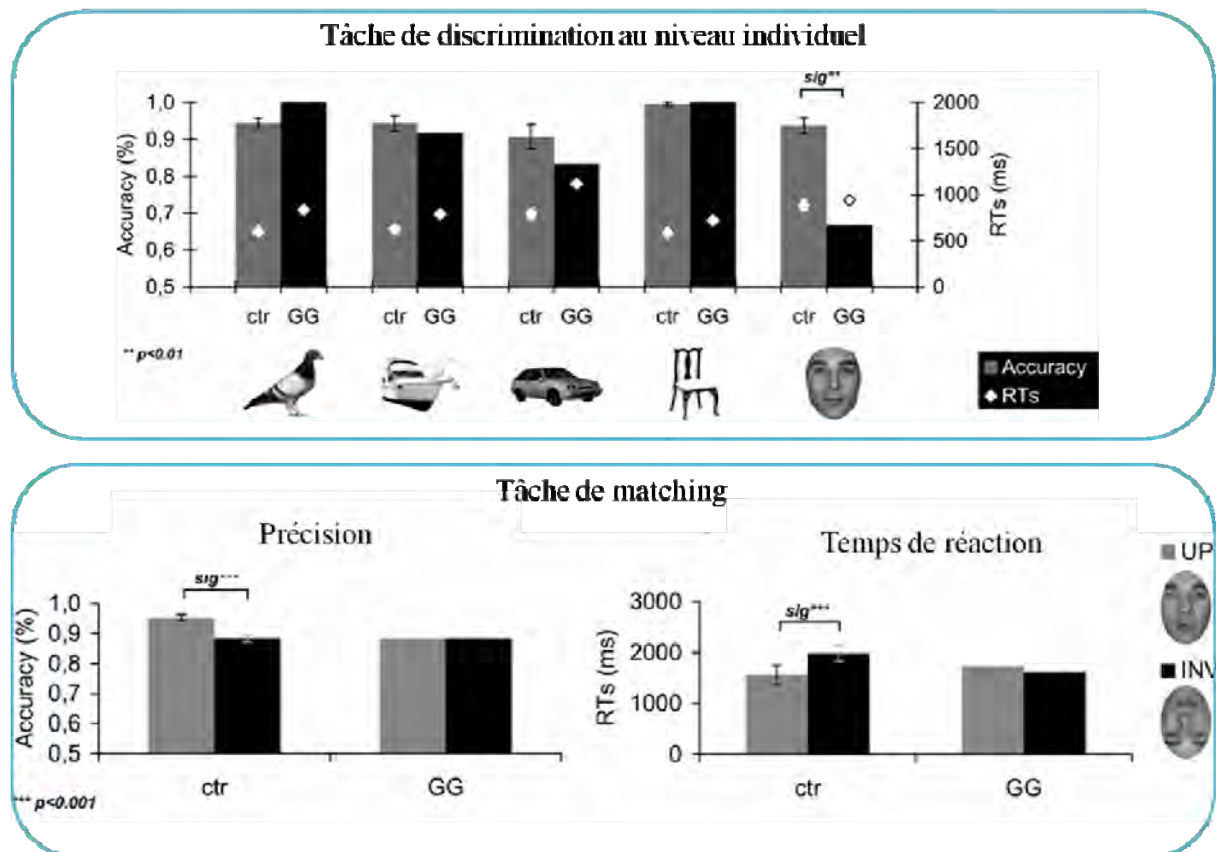


Figure 6 Figures extraites de Busigny et al. (2010) montrant en haut les performances (les losanges représentent les temps de réaction et les barres représentent la précision) des sujets contrôles en gris et du patient GG en noir dans une tâche de discrimination d'objets au niveau individuel. Aucune différence significative n'est observée pour les différentes catégories d'objets (oiseaux, bateaux, voitures, chaises) excepté pour la catégorie des visages. Ainsi, la prosopagnosie ne serait pas due à déficit global du traitement des objets au niveau individuel mais bien à un dysfonctionnement du traitement des visages au niveau individuel. En bas sont présentés les résultats d'une tâche de matching contrastant les performances (précision et temps de réaction) des contrôles et du patient lorsque les visages sont à l'endroit (en gris) vs à l'envers (en noir). Les sujets contrôles montrent un effet d'inversion contrairement au patient. Ces deux expériences sont des arguments en faveur d'un dysfonctionnement du traitement holistique des visages au niveau individuel dans la prosopagnosie.

Dernier point, l'effet Thatcher a également été retrouvé chez d'autres primates comme le macaque à qui l'on présentait des photographies de ses congénères (Adachi et al., 2009). En effet, la spécificité du traitement des visages semble retrouvée chez de nombreuses autres espèces animales. Afin de comprendre comment la perception des visages s'est développée au cours de l'évolution, une revue récente de Leopold et Rhodes (2010) a compilé différents travaux étudiant la manière dont plusieurs espèces animales traitaient les visages. Pour cela, les auteurs se sont intéressés aux différents aspects du traitement des visages (identité, expression émotionnelle, regard, attractivité, développement et spécialisation neurale) chez des espèces de primates (chimpanzés, macaques) et de non-primates (mammifères, vertébrés non-mammaliens et invertébrés) afin de déterminer quelles informations les animaux étaient capables d'extraire du visage de leurs congénères, ou des autres espèces animales. Ils ont ainsi apporté de nombreux arguments (Figure 7) en faveur d'une origine évolutive du traitement des visages, dont les différentes composantes se retrouvaient à travers les espèces.

Par exemple, les perruches seraient plus rapides pour discriminer une paire de visages d'autres perruches lorsqu'elles sont intactes comparées à des paires de visages mélangés, suggérant un rôle important de la configuration pour effectuer la tâche (Brown et Dooling, 1993) (Figure 7). Si l'on s'intéresse aux émotions faciales, les moutons sembleraient communiquer et reconnaître la peur en écarquillant les yeux et en rabattant leurs oreilles en arrière (Tate et al., 2006). De plus et de manière intéressante, les attributs faciaux semblent également intervenir dans la sélection du partenaire sexuel. En effet, les singes trouveraient plus attractifs un congénère au visage symétrique (Waitt et Little, 2006). Ils seraient également capables de reconnaître des similarités familiales à partir du visage de leurs congénères (entre une mère et ses petits par exemple) (Parr et de Waal, 1999). Enfin, de manière surprenante, il semblerait que les grands dauphins (Marino et al., 2007) et les éléphants (Plotnick et al., 2006) soient capables de reconnaître leur propre visage dans un miroir.

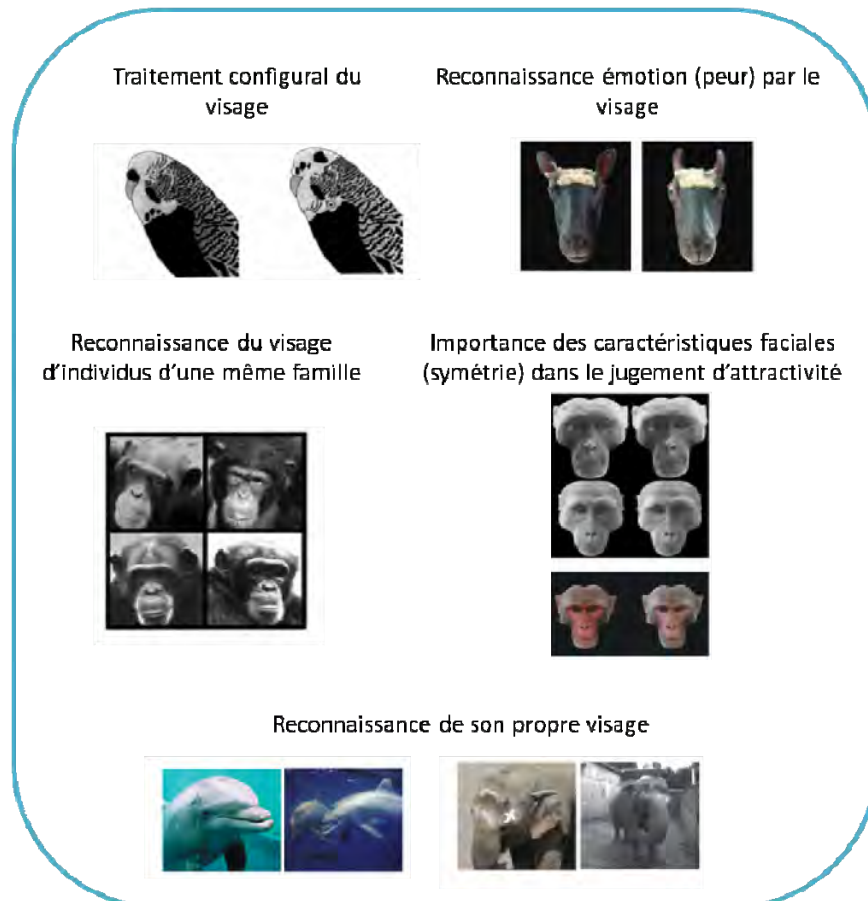


Figure 7 Illustration de différents comportements en faveur d'un traitement spécifique des visages chez d'autres espèces animales (adaptés de Rhodes et Leopold, 2010).

1.1.1.2 Arguments à l'encontre d'un traitement spécifique des visages

L'existence d'un traitement spécifique des visages reste controversée. Trois principaux arguments sont rapportés dans la littérature.

Un premier argument contre l'hypothèse d'un traitement spécifique des visages est le fait que *les visages soient généralement traités au niveau individuel* alors que les autres objets sont traités le plus souvent au niveau basique. En effet, dans les années 70, les études pionnières de Rosch (1975) portant sur le traitement des objets, proposent que les objets visuels soient catégorisés à différents niveaux d'abstraction, catégoriser un stimulus étant une activité cognitive consistant à regrouper des objets non identiques dans des catégories (Mervis et Rosch, 1981). Les catégories les plus abstraites correspondant aux catégories les plus « larges » sont appelées catégories superordonnées (animal, vivant,...) et les catégories les plus fines sont appelées subordonnées (colley, rouge-gorge,...). Entre ces deux niveaux de

catégorisation, le niveau basique (chien, lapin,...) correspondrait au niveau de catégorisation où les objets seraient directement catégorisés (Figure 2). Selon la théorie de Rosch, ce niveau constituerait le point d'entrée de la reconnaissance des objets typiques (Jolicoeur et al., 1986), le point d'entrée étant défini comme le premier point de contact entre le stimulus visuel et ses représentations en mémoire. Les visages, au contraire des autres objets nécessiteraient le plus souvent un traitement sous-ordonné (« il s'agit du visage de ma voisine et pas de la boulangère ») comparés aux autres stimuli visuels (« il s'agit d'un avion et non d'un oiseau »). L'apparente spécificité du traitement des visages à travers une atteinte spécifique de cette catégorie d'objet (ex : la prosopagnosie) pourrait refléter un trouble plus général de l'individualisation des objets. Dans ce sens, Damasio et al. (1982) rapportent plusieurs cas de patients présentant des troubles de l'individualisation de plusieurs catégories d'objets y compris des visages. Damasio postule de ce fait l'existence d'un système non-spécifique commun qui s'activerait lorsque tout type d'objets nécessiterait d'être individualisé.

Inspirée des travaux de Damasio, une alternative à ce premier argument est l'hypothèse de l'**expertise** qui suggère l'existence d'un système de traitement spécifique pour tous les objets dont nous sommes experts (experts canins, ornithologues, y compris les visages,...), c'est-à-dire que nous devons traiter à un niveau sous-ordonné (Diamond et Carey, 1986). En effet, plusieurs auteurs proposent que tous les stimuli dont nous sommes experts fassent l'objet d'un traitement recrutant des *routines perceptives spécialisées* (Tanaka et Taylor, 1991 ; Tanaka, 2001).

Une étude de Tanaka et Taylor (1991) montre que des experts, par exemple des éleveurs canins, catégorisent les chiens aussi spontanément et rapidement au niveau subordonné (labrador) qu'au niveau basique (chien). Dans ce sens, Tanaka (2001) fait l'hypothèse que les humains, en tant qu'experts des visages devraient présenter les mêmes effets que les experts canins et catégoriser les stimuli visages directement (et donc plus rapidement) au niveau subordonné. L'auteur teste cette hypothèse à travers quatre expériences (Figure 8). Dans la première expérience, les sujets doivent donner le nom de 26 personnes célèbres et 20 objets naturels ou manufacturés le plus rapidement possible. Leurs réponses sont classées en superordonnées (animal, être vivant), basiques (humain, personne) et

subordonnées (femme, homme ou nom propre). Les résultats montrent un avantage du niveau basique pour les objets et un avantage du niveau subordonné pour les visages. Tanaka suggère cependant que les sujets catégorisaient les visages au niveau subordonné par conventions sociales.

Ainsi, il développe une seconde étude utilisant la tâche de vérification de catégorie développée par Rosch (1975). Dans cette expérience, la catégorie (basique ou subordonnée) est présentée au sujet suivi de la photographie (8 chiens et 8 célébrités répétés 6 fois), le sujet ayant pour consigne de répondre le plus rapidement possible si l'image appartient à la catégorie présentée. Un avantage temporel du niveau basique est observé pour les images de chiens tandis que les temps de réaction sont similaires entre les catégories subordonnées et basiques pour les visages. Tanaka conclut donc que le point d'entrée pour les visages célèbres est bien le niveau subordonné de l'identité.

Il tente ensuite de préciser la nature de ces différents niveaux de catégorisation. Pour cela, dans une troisième expérience il souhaite déterminer si, comme le suggère Jolicoeur et al. (1984), la catégorisation subordonnée nécessite des processus perceptuels additionnels à la catégorie basique. Pour cela, il teste l'effet du masquage sur la tâche de vérification de catégorie. Ainsi, comme dans l'expérience précédente, la catégorie précède la présentation des stimuli (8 chiens, 8 oiseaux, 8 célébrités répétés 4 fois) qui sont ensuite masqués dès 75 ms après la présentation de l'image (courte exposition) ou plus tardivement, après 950 ms (longue exposition). Les résultats montrent en effet, que pour les objets dont nous ne sommes pas experts (chiens, oiseaux), une augmentation des temps de réaction et une diminution de la précision sont observées pour le niveau subordonné lors d'une courte exposition du stimulus. De manière intéressante aucun effet n'est observé pour les visages.

L'ensemble de ces travaux supportent l'hypothèse selon laquelle le traitement perceptif des objets d'expertise, et en particulier celui des visages, refléterait un mécanisme performant de traitement au niveau subordonné des exemplaires d'une catégorie (Tanaka et Curran, 2001 ; Rossion et al., 2002) plutôt qu'un système de traitement spécifique dédié aux visages.

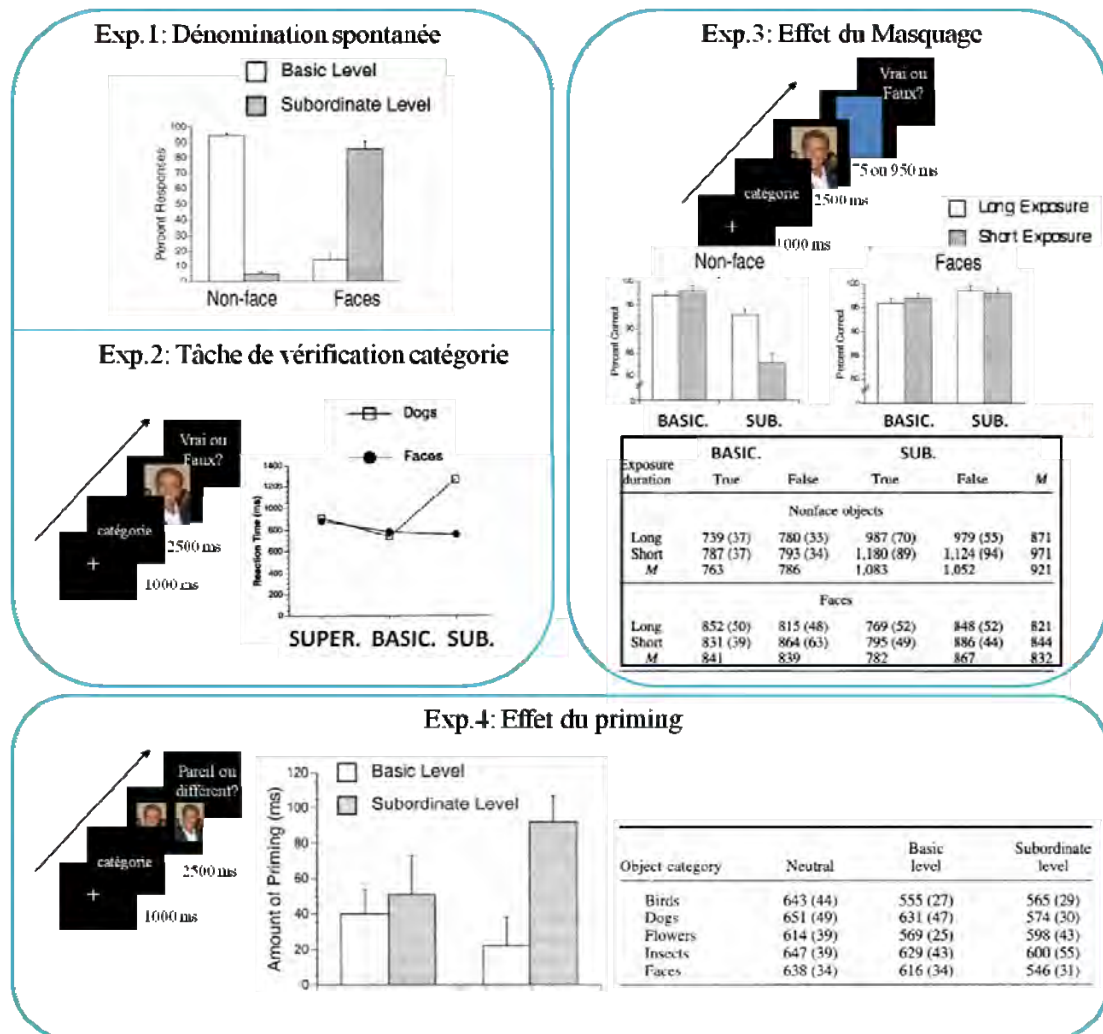


Figure 8 Synthèse des résultats de Tanaka (2001) en faveur d'une catégorisation des visages directement au niveau individuel en faveur de l'hypothèse selon laquelle les humains seraient experts des visages. Leur première expérience (en haut à gauche) montre que les sujets dénomment spontanément les visages au niveau individuel (nom propre) plutôt qu'au niveau basique (10% donne le genre contre 90% le nom de la personne) au contraire des objets qui sont en majorité dénommés au niveau basique. La seconde expérience montre que lors d'une tâche de vérification de catégorie comme développée par Rosch, les sujets sont aussi rapides pour répondre lorsque la catégorie est subordonnée comparée au niveau superordonné ou basique pour les stimuli visages contrairement aux objets pour lesquels un avantage basique est observé. Une troisième expérience montre dans une tâche de vérification de catégorie que le temps d'exposition (75 ms vs 950 ms) n'a pas d'effet sur les performances (précision et temps de réaction) des sujets pour les visages alors qu'une diminution des performances (diminution de la précision et augmentation des temps de réaction) est observée pour les autres objets. Enfin, la quatrième expérience teste l'effet du priming sur une tâche de « matching » (appuyer sur une touche si les deux images sont les mêmes, une autre si elles sont différentes) et montre un effet du priming par l'identité pour les visages contrairement aux autres objets qui ne montrent pas d'effet.

Enfin, un dernier argument à l'encontre d'un système spécifique des visages est le fait que certains objets, autre que les visages, semblent être traités de *manière configurale*. Par exemple, à la fin des années 80, une étude proposée par Diamond et Carey (1986) montre que sous certaines conditions, l'effet d'inversion, supposé spécifique aux visages, pouvait être observé pour des objets non-visages dont nous sommes experts (Figure 9). Dans leur seconde expérience, des photographies de chiens ou de visages sont présentées à l'envers ou à l'endroit de manière aléatoire (l'ordre étant contrebalancé à travers les sujets) à des experts canins et des novices ayant pour consigne de les mémoriser. Immédiatement après cette phase d'encodage, les sujets réalisent une tâche de choix forcé, les items étant présentés à l'endroit ou à l'envers. Les résultats montrent bien un effet d'inversion pour les stimuli visages chez les novices et les experts canins. Cependant cet effet est également observé pour les photographies de chiens chez les experts. Ces résultats supportent donc l'hypothèse d'un traitement configural de tous les objets visuels dont nous sommes experts, y compris les visages.

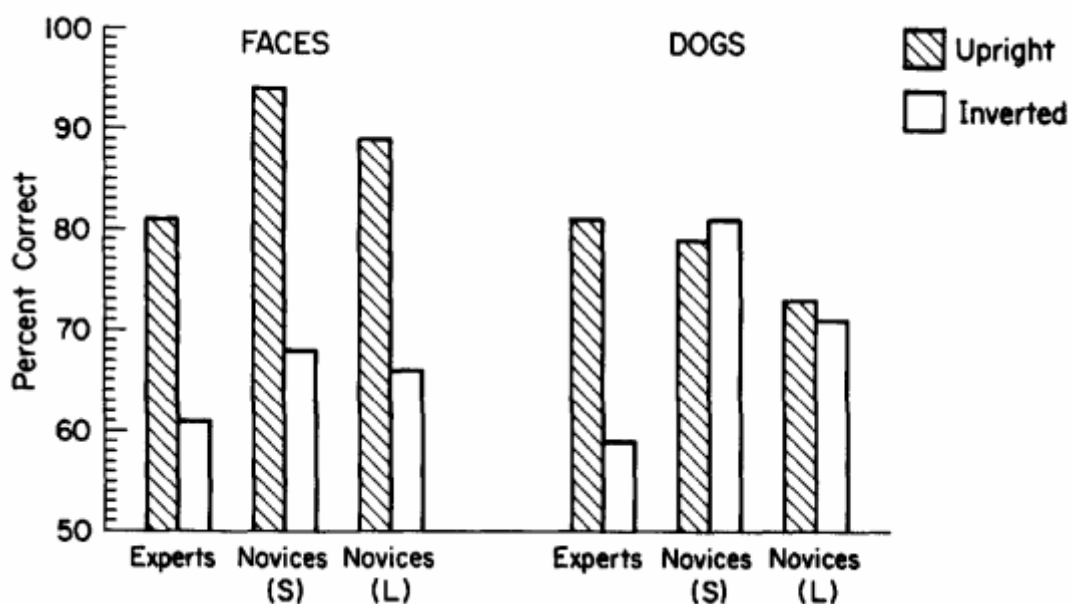


Figure 9 Figure extraite de Diamond et Carey (1986) montrant les performances d'experts et de novices dans des tâches de choix forcé. Les lettres S et L désignent le nombre de stimuli présenté au sujet, S désignant un faible nombre d'items et L un large nombre d'items. Un effet d'inversion est observé chez les novices et les experts canins pour les visages quelle que soit la taille de l'échantillon. Cet effet d'inversion est également observé chez les experts lorsque les objets de leur expertise (chiens) sont présentés. Ces résultats supportent donc l'hypothèse de l'expertise.

✓ Dans cette première partie, nous avons pu constater que malgré leur complexité, les visages sont traités de manière rapide et efficace. Une première hypothèse pour expliquer ce traitement performant suggère l'existence d'un traitement spécifique des stimuli visages. Cependant, plusieurs arguments (le traitement individuel, l'expertise et le traitement configural/holistique) vont à l'encontre de cette hypothèse et convergent plutôt vers l'existence de mécanismes communs permettant de traiter tous les objets nécessitant un traitement sous-ordonné. Nous avons également souligné l'existence d'une atteinte de ce système appelée la prosopagnosie. En effet, plusieurs études ont rapporté des cas de patients présentant une reconnaissance des visages perturbée à différents niveaux donnant naissance à des syndromes distincts. Ces observations ont permis l'élaboration d'un modèle célèbre, le modèle de Bruce et Young (1986) proposant un traitement hiérarchique séquentiel des personnes. Ce modèle sera présenté dans la partie suivante.

1.1.2 Modèle cognitif de la reconnaissance des visages familiaux

Plusieurs travaux ayant débuté dans les années 80 ont permis l'émergence d'une hypothèse selon laquelle la reconnaissance d'un visage familial impliquerait plusieurs étapes qui se succèderaient de manière hiérarchique, une étape ne pouvant être atteinte que lorsque les traitements liés à l'étape précédente étaient terminés (Hay et Young, 1982 ; Bruce et Young, 1986). Ainsi, un visage ne serait complètement identifié qu'après des étapes distinctes.

Le modèle de Bruce et Young (1986), largement rapporté dans la littérature, propose une organisation hiérarchique séquentielle (Figure 10).

Une première étape correspondrait à l'analyse visuelle perceptive réalisée sur le visage aboutissant à la formation d'une représentation structurale. Cette première étape nécessiterait tout d'abord, une première analyse picturale permettant d'extraire les informations relatives à l'âge, au sexe, à l'ethnie et à l'émotion aboutissant à l'élaboration d'une représentation picturale qui serait dépendante du point de vue. Une seconde étape d'analyse, extraite de la représentation picturale permettrait ensuite d'analyser les propriétés structurales et invariantes du visage, afin qu'il soit reconnu quels que soient ses changements (point de vue, expression faciale, niveau d'éclairage). Ainsi, cette étape aboutirait à la formation d'une représentation du visage perçu qui serait comparée aux

visages stockés en mémoire, les représentations de chaque visage familial étant supposées être stockées dans des Unités de Reconnaissance Faciales (URF) spécifiques.

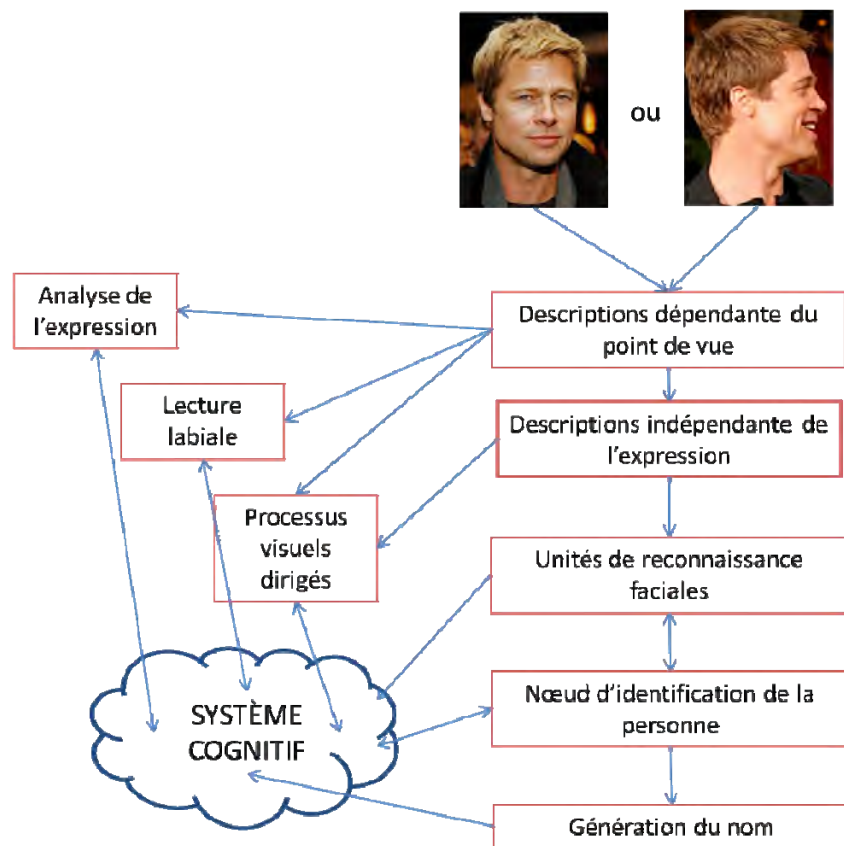


Figure 10 Adaptation du modèle séquentiel de Bruce et Young (1986) présentant les différentes étapes intervenant dans le traitement d'un visage familier. Après l'encodage structural, le visage est tout d'abord comparé aux visages stockés en mémoire (URF) puis identifié (NIP) et enfin dénommé.

La question de l'encodage et du stockage de ces représentations reste cependant débattue. En effet, les auteurs s'accordent à dire que chez l'adulte, les visages familiaux seraient représentés dans un espace multidimensionnel des visages ("face-space") dont chaque dimension correspondrait à une information critique pour discriminer les visages. Les visages moyens seraient situés au centre de cet espace et les visages atypiques seraient éloignés de ce centre (Rhodes, 1996; Rhodes et al., 1987; Valentine, 1991).

Deux modèles ont été proposés pour expliquer l'encodage et le stockage des visages familiaux dans cet espace. Une première hypothèse proposée par Valentine (1991, 1999), suggère que les visages seraient encodés sous forme de représentations individuelles

élaborées à partir des différentes expositions que le sujet aurait eues avec ce visage particulier (« exemplar-based coding»). Ainsi, selon cette hypothèse, les visages seraient codés en termes de valeur absolue dans chacune des dimensions de l'espace des visages (Figure 11).

Au contraire, une seconde hypothèse proposée par Rhodes et Jeffery (2006) suggère que les visages seraient encodés par rapport à un visage moyen (stock de normes perceptuelles) élaboré à partir de tous les visages perçus (« norm-based coding »). Ce second type de stockage permettrait au système visuel de s'abstraire de ce qui est commun à tous les visages afin de se focaliser sur ce qui est unique à l'individu. Selon cette hypothèse, le visage moyen aurait donc un statut particulier puisque l'identité des visages serait codée en termes de « déviation » par rapport à cette norme (vecteur). Ces deux hypothèses sont présentées en Figure 11.

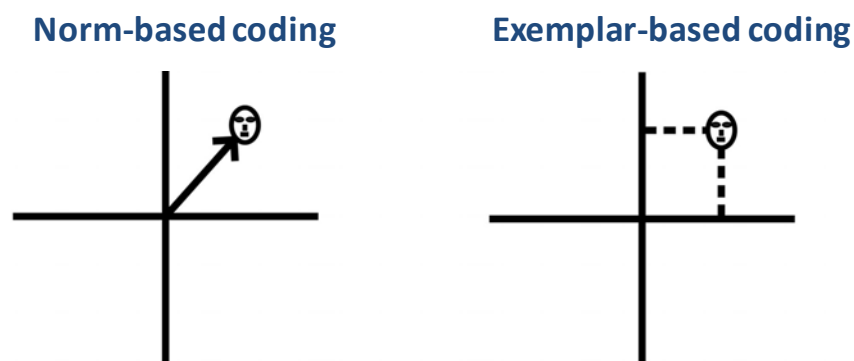


Figure 11 Illustration des deux hypothèses du codage des visages, adaptée de Jeffery et al. (2011), avec une représentation du codage par rapport à une norme (vecteur) à gauche et le codage à partir d'un exemplaire (valeur absolue moyenne) à droite.

Même si ces deux modèles divergent d'un point de vue théorique, ils sont tous deux compatibles avec l'idée d'une représentation de chaque visage qui nécessiterait d'être activée pour que le visage soit reconnu.

La seconde étape du modèle de Bruce et Young (1986) correspondrait donc à l'activation de l'URF spécifique au visage familier présenté. C'est lors de cette seconde étape que la *familiarité* avec le visage présenté serait possible.

Une troisième étape d'*identification* correspondant à l'activation du Nœud d'Identité de la Personne (NIP), chaque URF étant reliée à un NIP spécifique, est ensuite suggérée. Les NIPs seraient des nœuds d'entrée plurimodaux activés par toutes les modalités sensorielles (visage, voix, noms,...) et seraient spécifiques à chaque personne.

Enfin, une étape d'activation et d'accès aux informations sémantiques sur la personne serait possible, suivie d'une dernière étape correspondant à la récupération du nom.

Cette hiérarchie (détection → familiarité → identification → dénomination) a été mise en évidence via de nombreuses études. En effet, plusieurs études de cas cliniques montrent que ce système de reconnaissance peut être perturbé à différents niveaux donnant naissance à des syndromes distincts. Par exemple, les patients prosopagnosiques qui n'arrivent pas à traiter correctement le visage n'ont ni sentiment de familiarité, ni informations sémantiques, ni accès au nom de la personne à reconnaître lorsqu'ils doivent reconnaître une personne uniquement à partir de son visage (Young et al., 1990). D'autres cas de patients présentent une préservation du sentiment de familiarité mais avec atteinte des étapes suivantes (De Haan et al., 1991).

De plus, un autre trouble, l'agnosie des personnes définie comme l'incapacité à identifier les personnes à partir de tous types d'informations (physionomie, son de la voix, détails physiques,...) a été rapportée. Plus particulièrement, ce trouble est caractérisé par une perte du savoir sémantique sur la personne (Barbarotto et al., 1995 ; Evans et al., 1995 ; Gainotti et al., 2003 ; Joubert et al., 2003, 2006). La mémoire sémantique regroupe les connaissances générales que l'on a sur le monde, comme le savoir sur les objets ou sur les items uniques (qui n'existent qu'en un seul exemplaire comme les personnes familières ou les lieux familiers). L'agnosie des personnes n'est pas due à un trouble perceptif, le sujet étant capable de réaliser différentes tâches perceptives (appréciation de distances, orientations, copie de dessin, appariement objet réel-objet dessiné, traitement des visages au niveau superordonné). L'agnosie des personnes est une agnosie sémantique car ce trouble est présent quelle que soit la modalité et semble donc être associé à une perte totale du concept qui pourrait correspondre aux NIPs du modèle de Bruce et Young (1986).

Enfin, des cas de patients anomiques présentant seulement une atteinte de l'accès aux noms propres, dernière étape du modèle, ont été rapportés (Semenza et Zettin, 1989).

✓ *Dans cette partie, nous avons abordé différents troubles de la reconnaissance des personnes rapportées dans la littérature (prosopagnosie, agnosie des personnes, anomie) qui suggèrent l'existence d'une hiérarchie dans le traitement des personnes depuis leur détection jusqu'à la récupération du nom permettant l'élaboration du modèle de Bruce et Young. En effet, la prosopagnosie semble être due à une altération du traitement du visage alors que le concept de la personne est préservé. Au contraire, l'agnosie des personnes semble être accompagnée d'une perte du concept. Reconnaître un visage familier impliquerait donc différentes étapes distinctes organisées de manière hiérarchique. Cependant, une information importante qui n'est pas mise en avant par ces travaux est l'aspect temporel. En effet, si l'on part du principe qu'une étape en précède une autre, elle devrait également être plus rapide. La dynamique temporelle de ces différentes étapes de traitement a été investiguée par le biais de travaux s'intéressant au traitement des visages familiers ou non comparés à des visages non familiers ou d'autres catégories d'objets.*

1.1.3 Dynamique temporelle de la reconnaissance des visages

La chronométrie mentale constitue l'un des premiers outils utilisés en sciences cognitives permettant de décomposer et d'identifier précisément les mécanismes clés impliqués dans les différentes fonctions cognitives (Posner, 1978 ; Donders, 1869 ; Sternberg, 1966). Etudier les latences comportementales paraît donc être un bon outil pour caractériser les processus strictement nécessaires intervenant lors d'une tâche cognitive particulière.

1.1.3.1 Des latences comportementales variables pour reconnaître un visage familier

Si l'on s'intéresse aux études portant sur la dynamique temporelle de la reconnaissance des visages familiers, on constate rapidement que les temps de réaction comportementaux sont très variables à travers les études.

Si l'on considère seulement une demi-douzaine d'études utilisant des tâches de reconnaissance de visages célèbres (visages célèbres vs visages inconnus), les temps de

réaction varie d'environ 400 à plus de 900 ms en moyenne (Anaki et al., 2007 ; Lewis et Ellis, 2000 ; Kampf et al., 2002 ; Caharel et al., 2005 ; Tanaka et al., 2001 ; Baird and Burton, 2008 ; Anaki et Bentin, 2009 ; Herzmann et al., 2004) (Figure 12).

De même, une large variabilité est aussi observée pour les visages personnellement connus (491 à 875 ms) (Caharel et al., 2005 ; Ramon et al., 2011; Herzmann et al., 2004).

De plus, ces latences semblent très lentes comparées à des tâches dites rapides de détection de visages par exemple (140 à 489 ms, Crouzet et al., 2010 ; Eimer, 2000 ; Latinus et Taylor, 2006 ; Rousselet et al., 2003). Les résultats de ces études sont compilés en Figure 12 (détails présentés en annexe 1).

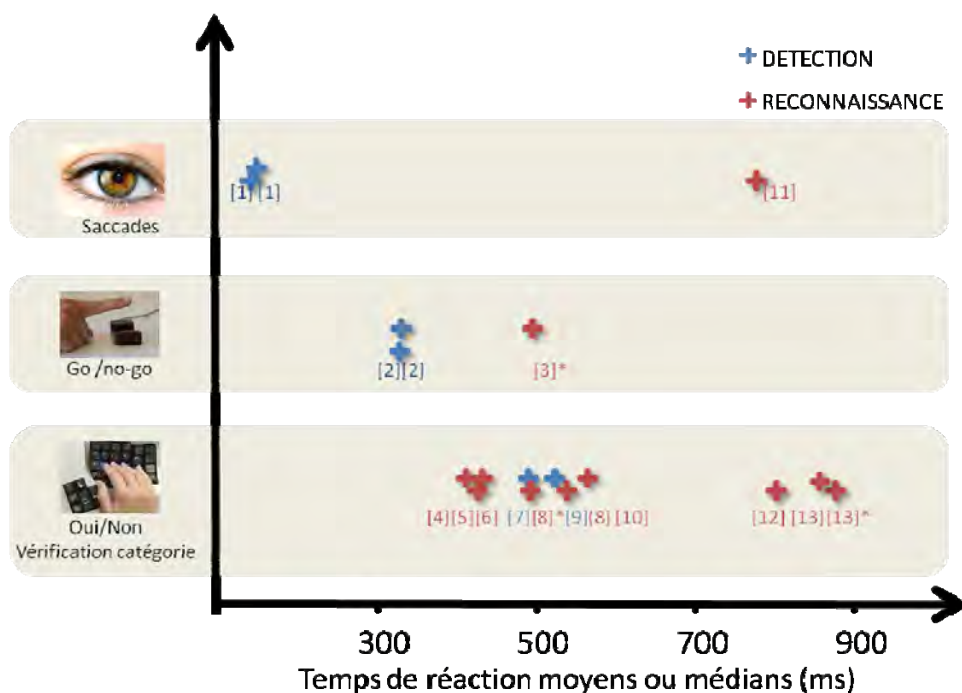


Figure 12 Compilation des temps de réaction moyens rapportés dans 13 études. Quatre études portent sur la détection des visages (en bleu) et 9 études portent sur la reconnaissance des visages (en rouge). Les études sont organisées en fonction du paradigme utilisé. Les protocoles utilisant des saccades oculaires sont en haut, puis les protocoles en go/no-go et enfin les études utilisant un appui bouton avec 2 choix possibles (cette dernière catégorie d'étude utilisant la tâche de vérification de catégorie ou un appui bouton lorsqu'une personne célèbre et présenté et un autre bouton lorsqu'il s'agit d'un inconnu). Les chiffres entre crochets correspondent aux différentes études suivantes : [1] Crouzet et al., 2010 ; [2] Rousselet et al., 2003 ; [3] Ramon et al., 2011 ; [4] Anaki et al., 2007 ; [5] Lewis et Ellis, 2000 ; [6] Kampf et al., 2002 ; [7] Eimer, 2000 ; [8] Caharel et al., 2005 ; [9] Latinus et Taylor , 2006 ; [10] Tanaka et al., 2001 ; [11] Baird and Burton, 2008 ; [12] Anaki et Bentin, 2009 ; [13] Herzmann et al., 2004. Les * signalent l'utilisation de visages personnellement connus rapportée dans 3 études, les autres utilisant des visages célèbres. Le détail de ces études est résumé dans le tableau 1 en annexe.

Cette variabilité peut être expliquée par de nombreux facteurs méthodologiques et conceptuels aboutissant à des données contradictoires et difficilement interprétables. En effet, plusieurs raisons semblent à l'origine d'une absence de consensus au sujet de la vitesse de la reconnaissance des visages.

1.1.3.2 Variabilité et facteurs confondants

Variabilité des stimuli et des modalités de réponses

Si l'on considère la littérature portant sur la vitesse de la reconnaissance des visages, les stimuli utilisés sont très variés : il s'agit le plus souvent de photographies couleurs ou noir et blanc qui peuvent être détournées (Caharel et al., 2005) ou non (Anaki et Bentin., 2009).

Les protocoles sont de nature très diverse, certains nécessitent la répétition des stimuli (1 seul stimulus personnellement connu répété 100 fois dans Caharel et al. (2005) contre 144 visages familiers différents dans Kampf et al. (2002)). En effet, plusieurs études utilisant des protocoles de priming (Lewis et Ellis, 2000) ainsi qu'une étude récente (Ramon et al., 2011) ont montré que les temps de réaction étaient fortement diminués dès la première répétition. Enfin, comme présenté en Figure 12, certains paradigmes nécessitent une réponse motrice (appuyer sur un bouton ou relever le doigt) et d'autres une réponse en saccades.

Reconnaissance ascendante ou descendante des visages

Un second point, lui, conceptuel, souvent occulté dans la majorité des études sur la reconnaissance des visages est le type de reconnaissance : reconnaissance de type top-down (descendante) vs reconnaissance de type bottom-up (ascendante).

En effet, il existe des traitements cognitifs de type « bottom-up » (ou « data-driven »), littéralement « du bas vers le haut ». Ce type de traitement est ainsi indépendant des hypothèses ou des attentes du sujet, il est purement perceptif. A l'inverse, il existerait également des traitements cognitifs de type « top-down » ou « Theory-driven » (du haut vers le bas), qui se réfèrent aux traitements basés sur les connaissances ou les attentes du sujet et pas seulement sur l'information perçue. Ces attentes viendraient moduler l'information sensorielle permettant ainsi d'identifier plus rapidement et facilement l'image perçue.

En se basant sur ces termes de bottom-up et top-down, il est possible de différencier deux types de reconnaissance. Une reconnaissance top-down est réalisée lorsque le sujet effectue la recherche d'une personne en particulier et a ainsi pu pré-activer des informations sur cette personne. Au contraire lors d'une reconnaissance bottom-up, le sujet ne sait pas quel visage va être présenté et doit donc reconnaître la personne sans pouvoir pré-activer d'informations qui auraient pu accélérer ses performances. Comme le montre la Figure 13, la reconnaissance bottom-up pourrait être comparée à la situation d'un sujet qui, lors d'une promenade croiserait subitement l'un de ses collègues. Au contraire, la reconnaissance top-down pourrait être comparée à la situation d'un sujet qui chercherait activement ce même collègue au milieu d'une salle pleine d'inconnus. Très clairement, le protocole de vérification de catégorie appliqué aux personnes célèbres par Tanaka (2001) (étude décrite en section 1.1.1.2) se place dans le cadre d'une reconnaissance top-down, le nom « Brad Pitt » présenté avant la photographie permettant d'activer fortement des informations visuelles diagnostiques sur cette personne.

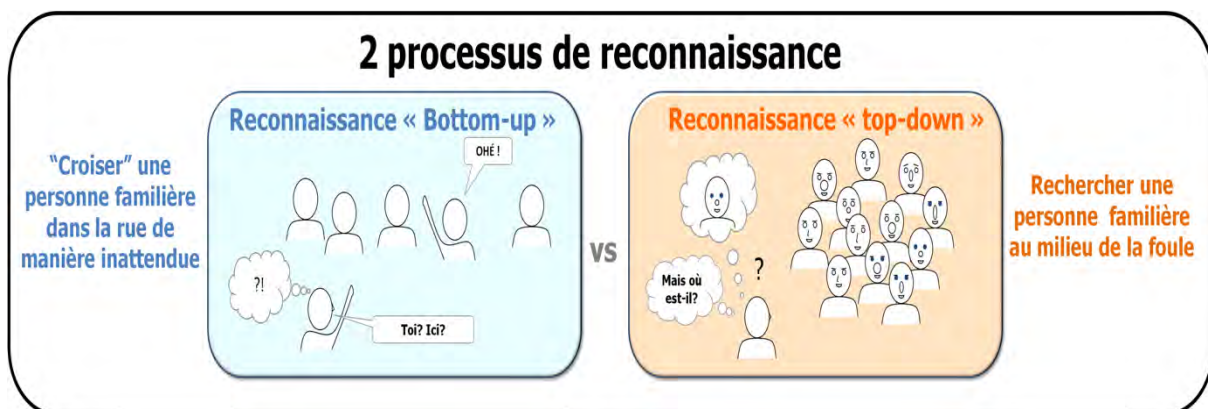


Figure 13 Illustration des deux processus de reconnaissance. A gauche la reconnaissance bottom-up, à droite la reconnaissance top-down.

Stratégies : Familiarité vs identification

Même si cela est encore débattu dans la littérature (Squire et Wixted, 2007 ; Montaldi et Mayes, 2010 ; Wixted et Squire, 2011 ; Montaldi et Mayes, 2011), la reconnaissance des objets serait basée sur deux mécanismes distincts : la familiarité et la recollection. Dans ce cadre, la familiarité peut être définie par le sentiment d'avoir déjà vu un stimulus, sans aucune autre information contextuelle au contraire de la recollection, qui s'accompagne d'informations contextuelles spatio-temporelles à propos du stimulus. Un exemple très

concret de cette distinction, décrit par Yovel et Paller (2004), est le « butcher-on-the-bus phenomenon » (phénomène du boucher dans le bus) qui fait référence à l'expérience de ressentir un fort sentiment de familiarité pour une personne sans pour autant savoir de qui il s'agit ou encore où on l'a rencontré. Ce phénomène se produit le plus fréquemment lorsque l'on croise une personne dans un contexte inhabituel. Le sentiment éprouvé est très fort (le sujet semble sûr de connaître cette personne) sans pour autant accéder aux informations concernant la personne (sans l'identifier).

Dans les études sur la reconnaissance des visages, il n'est pas toujours clairement mentionné si la tâche de reconnaissance requiert l'accès à l'identité de la personne ou un simple sentiment de familiarité. Cependant, les réponses basées sur la familiarité sont supposées être plus rapides que les réponses basées sur l'identité ou la recollection (Yovel et Paller, 2003 ; Bruce et Young, 1986 ; Besson et al., sous presse). L'absence de contraintes temporelles permettrait ainsi aux sujets d'utiliser différentes stratégies pour répondre entraînant une large variabilité dans les temps de réaction rapportés.

Une solution pour contraindre les sujets à répondre rapidement serait d'utiliser un protocole de catégorisation visuelle rapide. Ce protocole a été développé initialement pour étudier la rapidité du système visuel (Thorpe et al., 1996). De manière surprenante, il semblerait qu'en utilisant ce type de contrainte dans des tâches purement visuelles, des résultats contradictoires à la théorie de Rosch (1975) et à l'hypothèse d'expertise de Tanaka et Taylor (1991) soient rapportés pour le traitement des objets.

En effet, dans plusieurs études au cours desquelles les sujets devaient répondre le plus rapidement possible à une catégorie d'objets lors d'une tâche de go/no-go (voir description détaillée du protocole dans la section 2.1), les résultats montrent que les humains sont en fait capables de catégoriser des objets au niveau superordonné (animal vs non-animal, 340-380 ms en moyenne) plus rapidement qu'au niveau basique (chien vs autres animaux, 452 ms en moyenne) (Macé et al., 2009 ; Rousselet et al., 2003).

A notre connaissance, aucune étude n'a comparé le niveau superordonné et le niveau de la familiarité (Figure 14) en utilisant un protocole de catégorisation visuelle rapide de visages

célèbres. Seulement une étude utilisant ce type de protocole a étudié le temps de réaction des sujets pour reconnaître des visages personnellement connus parmi des visages non-familiers. Les latences rapportées sont de 530 ms en moyenne (370 ms au minimum) (Ramon et al., 2011).

Il semblerait donc intéressant de savoir comment se comporteraient des sujets face à un protocole de catégorisation visuelle rapide comme proposé par Macé et al. (2009) mais appliqué à la catégorisation de visages célèbres afin de comparer directement ces deux niveaux de catégorisation chez les mêmes sujets. L'utilisation du protocole de catégorisation visuelle rapide permettrait peut-être de contrôler un grand nombre de facteurs intervenant dans la variabilité des temps de réaction rapportés dans la littérature.

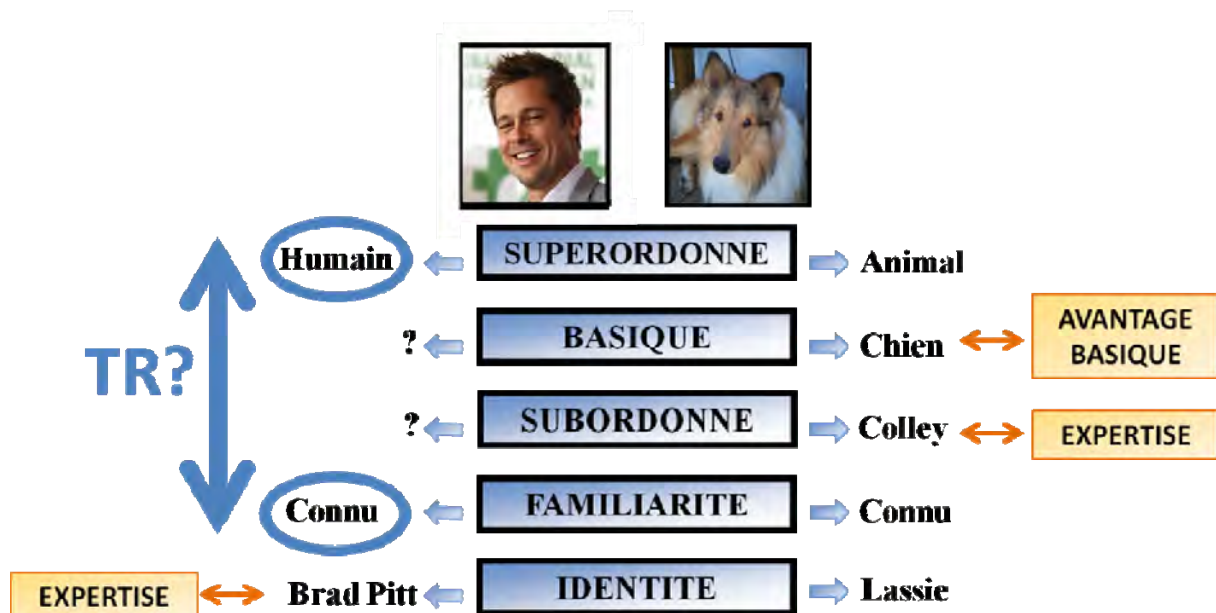


Figure 14 Adaptation du modèle de Rosch (1975). Selon la théorie de Rosch, les objets typiques seraient catégorisés au niveau basique tandis que les objets dont nous sommes experts (ornithologues, éleveurs canins) ou atypiques (pingouin) seraient catégorisés au niveau subordonné (Jolicoeur et al., 1984 ; Tanaka et Taylor, 1991). Selon l'hypothèse de l'expertise soutenue par Tanaka et Taylor (1991), en tant qu'experts en visages, les humains catégoriseraient les visages au niveau le plus subordonné possible qui est le niveau de l'identité. Cependant, les travaux de Macé et al. (2009) montrent qu'une catégorisation visuelle rapide au niveau superordonné précède la catégorisation basique. A gauche : Le visage de Brad Pitt peut être catégorisé à différents niveaux : une catégorisation superordonnée pourrait correspondre à une catégorisation humain/non-humain, une catégorisation au niveau de la familiarité pourrait être une catégorisation connu/inconnu. Le temps entre ces deux niveaux de catégorisation reste encore inconnu.

✓ Dans cette partie, nous avons abordé la question de la vitesse de la reconnaissance des visages célèbres. Nous avons constaté qu'un grand nombre de facteurs en particulier le type de reconnaissance (bottom-up vs top-down) ainsi que les stratégies utilisées par les sujets (familiarité vs identification) semblent contribuer à la forte variabilité des temps de réaction rapportés dans la littérature aboutissant à une absence de consensus sur cette vitesse. Pour palier à ces différents facteurs et déterminer la vitesse de la reconnaissance, nous proposons donc d'utiliser un protocole de catégorisation visuelle rapide qui permettrait de contraindre les sujets à répondre rapidement.

✓ Cette première partie comportementale suggère bien que la reconnaissance des visages est sous-tendue par un système performant et rapide mais révèle cependant une absence de consensus concernant cette vitesse de traitement. Qu'en est-il au niveau cérébral ? A travers la description des différentes régions identifiées dans la reconnaissance des visages, des modèles anatomo-fonctionnels proposés et de la dynamique temporelle de ces activations, la seconde partie abordera la question de la représentation des visages célèbres au niveau cérébral.

Organisation cérébrale et perception des visages

L'étude de la reconnaissance des visages a principalement été abordée à partir de la modalité visuelle, l'utilisation de photographies permettant de tester facilement cette fonction cognitive.

Dans ce chapitre, nous décrirons l'organisation du système visuel ainsi que les différents modèles computationnels qui ont été proposés pour rendre compte de la reconnaissance des objets visuels en général. Certains modèles anatomo-fonctionnels de la reconnaissance des visages, basées sur les données issues de la neuro-imagerie, seront ensuite présentés.

Enfin, nous verrons comment les données électrophysiologiques permettent d'appréhender la dynamique neurophysiologique de la reconnaissance des visages.

1.2.1 Régions cérébrales impliquées dans la reconnaissance des visages

1.2.1.1 Le système visuel

Percevoir un stimulus visuel débute par l'arrivée des photons, réfléchis par le stimulus, sur la rétine qui après avoir traversé la cornée, l'humeur aqueuse, le cristallin et l'humeur vitrée, viennent activer les cellules photoréceptrices de la rétine assurant ainsi la transduction de l'information lumineuse en messages nerveux. La rétine est composée de différents types cellulaires : des cellules photoréceptrices ou photorécepteurs appelées cônes et bâtonnets, des cellules de relais appelées cellules bipolaires, des cellules de projections appelées cellules ganglionnaires (divisées en deux classes P et M) et des interneurones appelés cellules amacrines et horizontales.

Les photorécepteurs ont chacun une sensibilité propre à la lumière. Les bâtonnets, les plus nombreux (environ 120 millions), sont très sensibles. Leur activité est souvent saturée en condition d'éclairage normal. Du fait de cette importante sensibilité, ils participent plutôt à la vision crépusculaire et nocturne. Au contraire, les cônes, moins nombreux (5 millions de cellules environ) ont une sensibilité plus faible et participent donc plutôt à la vision diurne, lorsque la luminosité est normale. Les cônes ont une sensibilité spectrale différente des bâtonnets, leur permettant de participer à la vision des couleurs tandis que les bâtonnets ne permettent qu'une vision achromatique, en noir et blanc. La répartition des photorécepteurs n'est pas uniforme. La concentration des cônes est maximale au centre de la rétine, appelée

fovéa, permettant ainsi une acuité visuelle importante au centre du champ visuel. Plus on s'éloigne de la fovéa, plus le nombre de cônes diminue, et, par conséquent, plus l'acuité visuelle sera faible.

Chaque photorécepteur va coder un point précis du champ visuel et l'information visuelle sera transmise aux cellules ganglionnaires. Les cellules ganglionnaires, quant à elles, recevant des inputs de plusieurs cellules nerveuses, vont intégrer l'information provenant des photorécepteurs. A ce stade, il est important d'introduire la notion de champ récepteur d'un neurone, qui constitue la région du champ visuel qui lorsqu'elle est stimulée va moduler la réponse du neurone. Ainsi le photorécepteur a un champ récepteur plus petit, tandis que les cellules ganglionnaires qui vont être activées par un grand nombre de photorécepteurs auront un champ récepteur un peu plus grand et ainsi de suite. Plus l'information progresse dans la hiérarchie visuelle, plus les champs récepteurs des cellules vont être de grande taille.

Les cellules ganglionnaires dont les axones constituent le nerf optique, vont ensuite générer des potentiels d'action se propageant vers le corps genouillé latéral (CGL, noyau du thalamus). Au sein du CGL, deux types de couches cellulaires ont été identifiées : quatre couches parvocellulaires qui reçoivent des inputs des cellules ganglionnaires de type P et deux couches magnocellulaires recevant des informations transmises par les cellules M. On retrouve aussi des cellules dites koniocellulaires moins nombreuses entre ces différentes couches. Le système parvocellulaire est impliqué dans la discrimination des formes et des couleurs et le système magnocellulaire est impliqué dans la détection du mouvement. Les informations traitées au niveau du CGL vont ensuite être transmises vers le cortex visuel primaire (V1). V1 (aire 17 selon l'architecture cérébrale décrite par Brodmann) est situé sur la face interne du lobe occipital de chaque hémisphère, à l'intérieur de la scissure calcarine. Il est appelé cortex strié en raison de son architecture cellulaire particulière. En effet, V1 est formé de 6 couches cellulaires. La couche granulaire externe (couche 4) est très développée et reçoit les inputs en provenance du CGL. En effet, les couches parvocellulaires projettent sur la couche 4C beta et les couches magnocellulaires sur les couches 4C alpha et 4B. Les cellules de la couche 4C, organisées en colonnes de dominance oculaire, d'orientation et de couleur, vont quant à elles, projeter sur les couches 2 et 3 dont l'organisation en strates fait

alterner des cellules sélectives à l'orientation (interblobs) et des cellules non-sélectives à l'orientation (blobs).

D'autre part, V1 est organisé de manière rétinotopique, c'est-à-dire que cette région reconstruit une « carte topographique », point par point, du champ visuel à partir des informations en provenance des cellules ganglionnaires. Une décussation partielle des fibres sensorielles au niveau du chiasma optique a lieu entre la rétine et le CGL. De ce fait, les informations provenant des cellules ganglionnaires de l'hémichamp droit, captées par les photorécepteurs de l'hémi-rétine temporale de l'œil gauche et l'hémi-rétine nasale de l'œil droit, seront envoyées vers le cortex visuel primaire gauche et vice versa.

1.2.1.2 Implication de la voie visuelle ventrale dans la reconnaissance des objets visuels

Deux voies visuelles corticales

Les informations visuelles présentes dans V1 vont ensuite être envoyées à différentes aires visuelles associatives interconnectées (travaux pionniers chez le singe : Felleman et Van Essen, 1991 ; Van Essen et al., 1992) qui vont extraire à chaque étape les différents indices de la scène visuelle permettant d'en analyser chacune des caractéristiques (Figure 15).

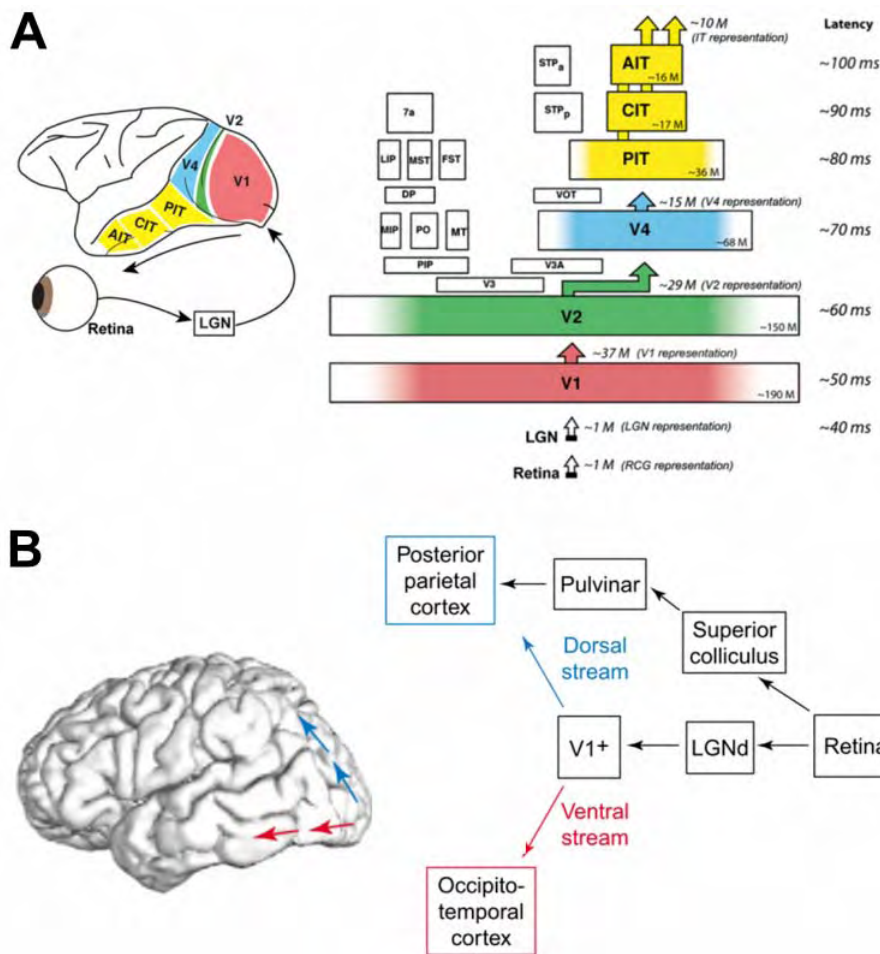


Figure 15 Représentation schématique de l'organisation de la voie visuelle ventrale chez le singe (en haut) et l'homme (en bas). (A) Modèle de la voie visuelle ventrale extrait de Dicarlo et al., 2012. Une représentation schématique des aires impliquées dans la voie visuelle ventrale chez le singe macaque rhesus depuis la rétine jusqu'au cortex inféro-temporal (IT) est représentée à gauche. Chaque aire est ensuite décrite plus précisément et représentée par des rectangles dont la taille est proportionnelle à la surface corticale (Felleman et Van Essen, 1991). Le nombre approximatif de neurones présents dans les deux hémisphères cérébraux est présenté dans le coin droit de chaque rectangle (en million) et le nombre de neurones de projection est présenté sous chaque rectangle (Collins et al., 2010). Enfin les latences médianes des réponses neuronales sont présentées à droite (Nowak et Bullier, 1997; Schmolesky et al., 1998). (B) Représentation schématique, des deux voies visuelles chez l'homme, proposée par Goodale et Westwood (2004). A gauche est représentée la localisation approximative des voies, basée sur une reconstruction 3D à partir d'IRM anatomiques. Les flèches indiquent l'implication d'interconnexions complexes. Les aires sont ensuite décrites plus précisément à droite : Depuis la rétine, l'information est envoyée vers la partie dorsale du noyau géniculé latéral du thalamus (LGNd) qui va ensuite envoyer des projections vers le cortex visuel primaire. La voie ventrale (ici en rouge) projette depuis les aires visuelles primaires de bas niveau (V1) vers le cortex occipito-temporal. La voie dorsale (représentée en bleu) projette vers le cortex pariétal postérieur. Le cortex pariétal reçoit également des informations en provenance du colliculus supérieur via le pulvinar.

Certaines informations seront transmises sur les aires V2 (Aire de Brodmann 18) puis vers des régions de plus en plus antérieures, selon deux voies principales mises en évidence dans le cerveau du singe par Mishkin et al. (1983).

La première est la voie visuelle dorsale (vers le lobe pariétal) ou voie du « où » qui code les informations relatives au mouvement, à la position dans l'espace permettant ainsi la localisation des objets dans notre environnement. Cette voie inclut l'aire MT (aire médiale temporale).

La seconde voie est la voie visuelle ventrale (vers le lobe temporal) ou voie du « quoi » qui depuis le lobe occipital, se prolonge à la surface ventrale du cerveau jusqu'au lobe temporal. Cette voie permet la perception des formes et des couleurs, la reconnaissance et l'identification des objets (DiCarlo et al., 2012). L'information entre les différentes aires circule de manière bien définie : Le signal ascendant (feedforward) entrant est reçu dans la couche 4 et transmis par les couches supérieures (3-1) aux aires plus antérieures. Le signal descendant (feedback) lui est produit par les couches inférieures et reçu dans les couches supérieures et inférieures des aires plus postérieures (Felleman et Van Essen, 1991). Cette organisation des aires et la circulation de l'information dans l'axe postéro-antérieur (occipital vers temporal) supportent l'idée d'une organisation hiérarchique de la voie visuelle ventrale. Plus l'on progresse dans la voie ventrale, plus les neurones présentent une forte invariance jusqu'à la représentation du concept dans les régions les plus antérieures, du cortex perirhinal au pôle temporal.

Cette dissociation voie ventrale/voie dorsale a également été décrite chez l'homme, grâce aux études lésionnelles ou à l'imagerie cérébrale chez les sujets sains (Haxby et al., 1991; Haxby et al., 1994). Cependant, de nombreuses interactions semblent entrer en jeu entre les régions de ces deux voies notamment depuis les aires pariétales de la voie visuelle dorsale vers les régions appartenant à la voie visuelle ventrale (Bullier, 2001 ; Goodale et Westwood, 2004).

Aires cérébrales impliquées dans la reconnaissance des visages

Les premières hypothèses concernant la localisation des régions cérébrales spécifiquement activées dans le traitement des visages émanent des études de cas de patients cérébro-lésés.

Par exemple, les données issues de la neuropsychologie ont montré que les patients prosopagnosiques (voir partie 1.1.1.1), incapables de reconnaître les visages au niveau individuel, présentent des lésions des régions postérieures (régions occipito-temporales ventrales comprenant notamment le gyrus lingual, le gyrus fusiforme et le cortex parahippocampique (De Renzi et al., 1994). Au contraire, l'agnosie des personnes (voir partie 1.1.2), perte du concept de la personne, serait plutôt due à des lésions plus antérieures notamment du pôle temporal et des structures temporales internes (Joubert et al., 2003, 2006 ; Gainotti et al., 2003).

Les techniques d'imagerie cérébrale telles que l'imagerie par résonance magnétique fonctionnelle (IRMf) ou la tomographie à émissions de positons (TEP) ont permis de mieux circonscrire ces différentes régions cérébrales rapportant un *large réseau* de régions cérébrales activé spécifiquement en réponse aux visages comparés à d'autres catégories d'objets (Sergent et al., 1992; Halgren et al., 1999; Haxby et al., 1999; Gauthier et al., 2000; Rossion et al., 2000).

L'hypothèse de l'existence d'un large réseau d'aires cérébrales a été confirmée par des études en IRMf menées chez le singe. En effet, Tsao et al. (2008) ont identifié un large réseau d'aires cérébrales spécialisées dans le traitement des visages divisé en 6 patchs situés dans le lobe temporal du macaque (Figure 16). Il semblerait même que ces régions soient largement interconnectés formant un réseau organisé de manière hiérarchique (Moeller et al., 2008).

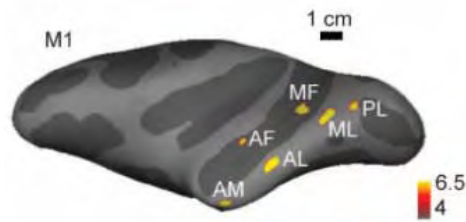


Figure 16 Figure extraite de Freiwald et Tsao (2010) montrant les six différents patches sélectifs aux visages identifiés dans le lobe temporal du macaque en IRMf.

Chez l'homme, trois régions principales ont été rapportées : (1) la Fusiform Face Area (FFA) située à la jonction du lobe occipital et du lobe temporal inférieur, au niveau du gyrus fusiforme, (2) la Occipital Face Area (OFA) (Kanwisher et al., 1997; Puce et al., 1995 ; Gauthier et al., 2000) localisée au niveau occipital inférieur droit et (3) la partie postérieure du sillon temporal supérieur (pSTS) (Haxby et al., 1999 ; Weiner et Grill-Spector, 2010) (Figure 17).

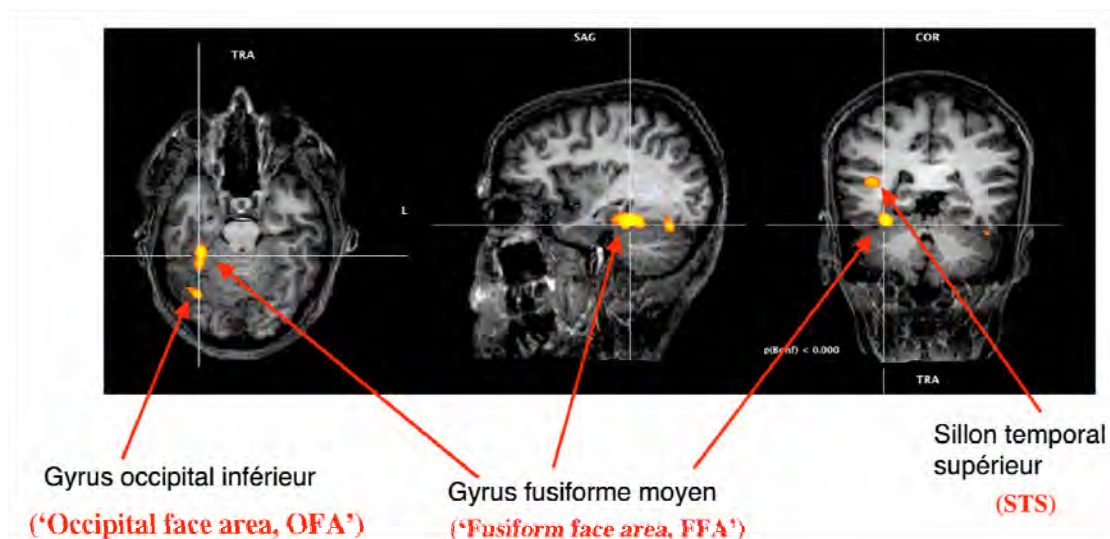


Figure 17 Figure extraite d'un chapitre de livre de Rossion (2009). Les aires spécifiques des visages identifiées en IRMf. Les visages activent de manière préférentielle et bilatérale (même si plus importante à droite) le gyrus occipital inférieur (OFA), le gyrus fusiforme moyen (FFA), et la partie postérieure du sillon temporal supérieur.

Certains auteurs se sont interrogés sur l'organisation de ces trois régions et se sont demandés si les mécanismes impliqués dans le traitement des visages était réalisé de manière purement feedforward comme le suggère l'organisation hiérarchique de la voie visuelle ventrale. Pour tester cette hypothèse, Rossion et al. (2003) rapportent le cas d'une

patiente prosopagnosique, PS, présentant une lésion de la OFA mais avec une FFA préservée. En réalisant une étude en IRMf, les auteurs montrent que la FFA de la patiente est activée lors de la présentation de stimuli visages et ce, en l'absence de toutes informations en provenance de la OFA. Cette étude suggère que l'intégrité de la OFA est nécessaire pour un traitement normal des visages et que des connections feedback depuis la FFA jusqu'à la OFA permettrait de traiter correctement les visages.

Dans ce sens, une récente méta-analyse réalisée par Bouvier et Engel en 2006 rapporte 92 cas de patients présentant une achromatopsie, qui se traduit par une absence de vision des couleurs, acquise suite à des lésions du cortex occipital ventral. Les auteurs utilisent comme contrôle les lésions de patients propopagnosiques rapportés dans la littérature. De manière intéressante et totalement indépendante de Rossion et al. (2003), les auteurs rapportent que les régions lésées communes chez tous les patients prosopagnosiques sont localisées au niveau de la OFA et non de la FFA (Figure 18).

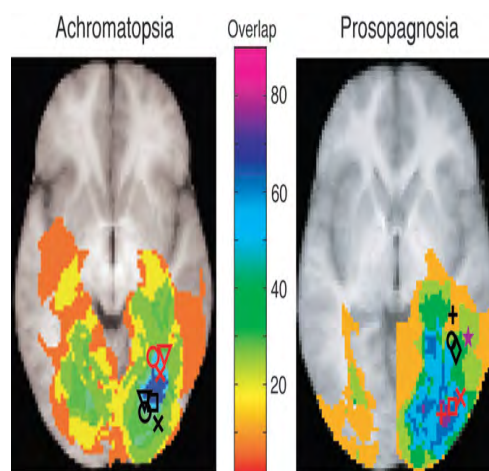


Figure 18 Figure extraite de Bouvier et Engel (2006) montrant la proportion de chevauchement des lésions. Les formes noires correspondent à la FFA, les formes rouges correspondent à la OFA et les violettes au STS. La région présentant le chevauchement le plus important à travers les patients prosopagnosiques est la OFA.

Au-delà de la comparaison de la catégorie des visages avec d'autres catégories d'objets, certains travaux en IRMf se sont intéressés à la représentation des différentes caractéristiques du visage et en particulier au *codage de l'identité*. Une étude de Rotshtein et al. (2005) montre en effet que certaines régions cérébrales sont impliquées dans le

traitement des caractéristiques physiques de l'image tandis que d'autres semblent sélectives à l'identité. En se basant sur des paires de photographies de personnes célèbres, ils ont créé plusieurs continua symétriques de 11 images depuis une identité (ex : Margate Thatcher) à une autre (ex : Marilyn Monroe) en augmentant graduellement de 10% les caractéristiques physiques à chaque étape. Les résultats sont ensuite analysés selon trois conditions : (1) une condition où des visages identiques (caractéristiques physiques et identités similaires) sont présentés, (2) une condition où les stimuli présentent une variation de 30% des caractéristiques physiques mais présentent la même identité et (3) une troisième condition où les deux images diffèrent de 30% également au niveau des caractéristiques physiques mais à une étape du continuum où les deux images ne présentent plus la même identité.

L'analyse du signal BOLD montre que le gyrus occipital inférieur est sélectif aux caractéristiques physiques de l'image (modulation de l'activité dans la condition où les caractéristiques physiques diffèrent quelle que soit l'identité, l'effet étant significativement plus faible lorsque les deux images présentent les mêmes caractéristiques). Au contraire, le gyrus fusiforme moyen est sélectif à l'identité, sa réponse hémodynamique étant modulée seulement lorsque l'identité est différente, indépendamment des caractéristiques physiques de l'image. Les auteurs observent des résultats similaires chez certains sujets lorsqu'il considère seulement la OFA et la FFA (Figure 19).

De plus, les sujets ayant au préalable réalisé un jugement de familiarité sur les différentes personnalités présentées, les auteurs investiguent la possible corrélation entre l'activité BOLD sur l'ensemble du cerveau et la force de la familiarité. Ils mettent ainsi en évidence une forte corrélation entre le niveau de familiarité et l'activité dans les régions temporales antérieures (hippocampe antérieur et pôle temporal).

Cette étude suggère donc un traitement hiérarchique des visages où que le gyrus occipital inférieur serait impliqué dans le traitement des propriétés physiques de l'image, la FFA dans le traitement de l'identité et enfin, les représentations situées dans les régions les plus antérieures du lobe temporal seraient impliquées dans le traitement de la familiarité à long terme.

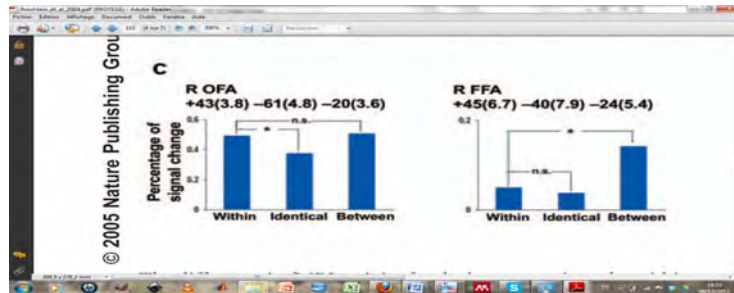


Figure 19 Figure extraite de Rotshtein et al. (2004) montrant la modulation du signal BOLD pour les trois conditions (within=même identité mais caractéristiques différentes, identical=même identité et même caractéristiques physiques, between=identité différentes et caractéristiques physiques différentes) dans la OFA à gauche et la FFA à droite. La OFA semble sélective aux caractéristiques physiques de l'image tandis que la FFA semble plutôt sélective à l'identité.

Cependant, ce résultat est sujet à controverse. Par exemple, une autre étude réalisée également en IRMf par Kriegeskorte et al. (2007) suggère plutôt que l'individualisation du visage serait réalisée à la fois au niveau de la FFA et dans la partie antérieure du cortex inféro-temporal (aIT, région homologue de patch antérieur temporel chez le macaque sélectif à l'identité, Rajimehr et al., 2008), cette dernière région produisant un pattern différent selon l'identité. Les auteurs suggèrent plutôt que la FFA permettrait la détection du changement de l'identité préparant ainsi l'identification dans aIT.

Enfin, une seconde étude en IRMf proposée par Ishai et al. (2002) adresse la question de l'imagerie mentale des visages. Les auteurs montrent qu'un large panel de régions cérébrales, incluant la FFA, la OFA, le STS et l'amygdale, est activé lorsque des visages célèbres sont présentés. De manière très intéressante, lorsqu'il est demandé au sujet d'imaginer les visages célèbres, d'autres régions, en plus des régions précitées, sont activées. En effet, les auteurs rapportent une activation de l'hippocampe, du précuneus, du sillon intra-pariétal et du gyrus frontal inférieur (IFG) pendant cette tâche d'imagerie.

Finalement, même si le rôle de ces différentes régions n'est pas encore très clair, il semblerait que le traitement des visages familiers soit sous-tendu par un large réseau d'aires cérébrales (Haxby et al., 2000 ; Ishai, 2008). En effet, des études utilisant des enregistrements intracrâniens sont venues confirmer le caractère distribué des régions

cérébrales impliquées permettant également d'apporter des éléments de réponse sur le rôle de ces régions dans le traitement des visages.

Les travaux d'Allison et al. (1994, 1999), réalisés à l'aide d'enregistrements à partir de grilles d'électrodes implantées dans le cortex occipito-temporal ventral et latéral de patients épileptiques (dans le cadre du diagnostic de leur épilepsie pharmacorésistante sévère) ont confirmé l'implication du cortex occipito-temporal dans le traitement des visages en rapportant des activités spécifiques aux visages comparés à d'autres catégories d'objets.

De même, dans une étude récente, Parvizi et al. (2012) ont adressé la question du rôle du gyrus fusiforme dans la perception des visages chez des patients épileptiques implantés avec des électrodes sous-durales dans le lobe temporal. Dans cette étude, les auteurs allient trois techniques de neuro-imagerie ; l'électrocorticographie, l'IRMf et la stimulation électrique cérébrale. A l'aide de l'IRMf, ils identifient deux patches sélectifs aux visages précédemment rapportés que sont la FFA et la OFA. L'enregistrement du signal électrophysiologique dans la FFA, montre en effet, une activité sélective aux visages. De plus, lorsqu'une stimulation des électrodes situées dans ces régions est appliquée, les patients rapportent que les visages des personnes présentes dans la pièce sont « métamorphosés ». Cette stimulation semble, au contraire, ne pas perturber la perception des objets ou la dénomination de personnes célèbres. Cette étude utilisant une approche multimodale apporte donc des données convergentes sur la spécificité du gyrus fusiforme dans la perception des visages.

D'autres études utilisant des électrodes cérébrales profondes ont montré une implication des régions hippocampique et sous-hippocampique dans la reconnaissance des visages familiers. Par exemple, les travaux de Dietl et al. (2005) et de Trautner et al. (2004) montrent une activation des cortex rhinaux (cortex périrhinal et entorhinal) et de l'hippocampe lors d'une tâche de reconnaissance de visages célèbres comparés à des visages inconnus. Ces travaux ont été confirmés par une étude de Barbeau et al. (2008) qui montre également une activation spécifique des régions hippocampiques et sous-hippocampiques dans une tâche de reconnaissance de visages célèbres. De plus, les auteurs montrent que la présentation des visages célèbres et inconnus activent de manière simultanée un large réseau d'aires

cérébrales incluant le la OFA, la FFA, l'IFG et le pôle temporal venant confirmer l'existence d'un large réseau d'aires cérébrales impliqué dans le traitement des visages.

Même si l'IRMf et les études en potentiels de champ intracérébraux ont permis d'identifier les différentes régions impliquées dans le traitement des visages, ces techniques ne permettent pas d'analyser la nature du codage de l'information visuelle et catégorielle des représentations. Dans ce sens, plusieurs études, à la fois chez l'homme et chez l'animal, ont eu l'opportunité d'enregistrer l'activité de neurones unitaires situés des les régions identifiées en IRMf.

L'équipe de Fried, dans plusieurs études chez l'homme (Fried et al. 1997; Kreiman et al., 2000) a rapporté que certains neurones situés dans le lobe temporal médian répondaient de manière sélective à la présentation d'images de visages, d'animaux, d'objets ou de scènes. Ces données corroborent les résultats obtenus à l'aide d'enregistrements de neurones unitaires chez le singe rapportant que les neurones situés en avant dans la voie visuelle ventrale (IT, STS) répondaient à des images complexes comme des visages ou des objets et présentaient un certain degré d'invariance à des propriétés métriques comme la taille du stimulus, la position ou l'angle de vue (Gross et al., 1972 ; Logothetis et Sheinberg, 1996).

Pour aller plus loin, l'équipe de Tsao a publié une étude dans laquelle les auteurs ont tenté de comprendre comment les différents patchs identifiés en IRMf (Figure 16) participaient à la représentation des visages (Freiwald et Tsao, 2010). Deux sets de stimuli étaient présentés aux singes : un premier set composé d'une centaine de stimuli appartenant à trois catégories d'items (visages, objets et corps), un second set de 200 visages provenant de 25 individus différents présentés selon huit orientations distinctes. L'activité électrophysiologique unitaire des neurones était enregistrée dans les deux patchs postérieurs médians et les deux patchs antérieurs de manière concomitante. Les résultats de cette étude montrent que les neurones situés dans la partie latérale médiane étaient spécifiques à l'angle de vue (réponse spécifique à une seule orientation) tandis que les neurones situés dans la partie antérieure latérale étaient spécifiques à la symétrie en miroir de chaque identité (deux orientations). Enfin, les neurones du patch le plus antérieur montraient une invariance presque complète aux différentes vues suggérant son implication dans le codage de l'identité.

D'autres études, chez l'homme, se sont également intéressées au codage de l'identité, plus particulièrement au codage du concept. De manière surprenante, l'équipe de Fried rapporte l'existence de neurones répondant spécifiquement à l'identité de la personne (Quiroga et al., 2005). Dans cette étude, des patients épileptiques, dans le cadre du diagnostic d'une épilepsie pharmaco-résistante grave, étaient implantés à l'aide de macro-électrodes profondes munis de micro-fils permettant d'enregistrer l'activité unitaire des neurones dans différentes régions temporales (hippocampe, amygdale, cortex entorhinal, gyrus parahippocampique). Le protocole se déroulait en deux phases. Lors d'une première phase de pré-test, l'expérimentateur présentait de manière aléatoire de nombreuses images aux patients, composées de différentes catégories d'objets telles que des célébrités, des animaux, des bâtiments, des paysages et des objets. Cette première phase permettait de sélectionner uniquement les images qui engendraient une réponse des neurones. Ainsi, environ 90 images correspondant à 14 célébrités ou objets différents ont été présentées au patient. Afin de maintenir l'attention du patient, une tâche de détection de visages humains lui était proposée, tâche s'avérant très facile puisque les performances étaient proches de 100%. Les résultats obtenus montrent une réponse sélective de certains neurones activés spécifiquement lors de la présentation d'images correspondant à certaines célébrités, paysages ou objets et parfois même à leur nom écrit. En effet, certains neurones répondent de manière spécifique à des photographies de Jennifer Anniston quels que soit le point de vue ou la taille du visage, suggérant qu'ils possèdent un certain degré d'invariance (Figure 20). De plus, ces neurones ne semblent pas répondre lorsque cette célébrité est accompagnée de Brad Pitt. D'autres neurones répondent plutôt à une autre célébrité, Halle Berry, même lorsqu'elle est déguisée en Catwoman. Enfin, ces neurones s'activent également lorsque le nom écrit est présenté. Le fait que ces neurones répondent également au nom suggère qu'ils ne répondent pas seulement aux caractéristiques visuelles communes aux différents exemplaires du visage d'Halle Berry mais supporte plutôt l'hypothèse d'un réel codage du concept de la personne. Cette hypothèse est confortée par une autre étude de la même équipe révélant qu'en plus de la modalité visuelle, ces neurones du lobe temporal médian répondraient également lorsque le nom de la personne est entendu (Quiroga et al., 2009). Cette activité plurimodale pourrait correspondre à l'activation du NIP proposé par Bruce et Young (voir partie 1.1.2).



Figure 20 Figure adaptée de Quiroga et al. (2005). Les réponses pour 60 images (30 en haut, 30 en bas) sont présentées. Pour chaque image, le « raster plot » des réponses classées par ordre de présentation du haut vers le bas et l'histogramme des réponses (en dessous) sont présentés. Les barres verticales en pointillés correspondent à la présentation et à la disparition du stimulus (1 s de présentation).

✓ A travers l'étude de cas de patients cérébro-lésés et de différentes techniques de neuro-imagerie, nous avons montré que les visages sont largement représentés dans le cerveau des primates. Même si leurs rôles restent discutés, trois régions principales semblent activées lors de la présentation d'un visage : la FFA, la OFA et la partie postérieure du STS. De plus, des régions situées plus en avant dans la voie ventrale telles l'hippocampe, les régions sous-hippocampiques et le pôle temporal semblent impliquées

dans le codage de l'identité voire même du concept de la personne. Cette organisation hiérarchique de la représentation des visages qui devient de plus en plus invariante lorsque l'on progresse dans cette hiérarchie a donné lieu à différents modèles théoriques que nous présentons dans la partie suivante.

1.2.2 Modèles de la reconnaissance des visages

Comme présenté dans la section 1.2.1.1, l'information neuronale peut être véhiculée de manière ascendante depuis les régions les plus postérieures vers les aires plus antérieures (information feedforward) ou au contraire de manière descendante (information feedback) qui, en théorie, permettrait aux régions plus antérieures de moduler l'activité des régions de plus bas niveau.

Même si la plupart des auteurs s'accordent sur le fait que des mécanismes à la fois feedforward et feedback interviennent dans la plupart des fonctions cognitives, la question de savoir si certaines fonctions peuvent être effectuées en se basant uniquement sur l'information feedforward donne lieu à de nombreux débats.

Ainsi, inspirés des travaux issus de la neuro-imagerie et de l'enregistrement de neurones, des modèles computationnels et anatomo-fonctionnels ont été proposés pour expliquer le fonctionnement de la voie visuelle ventrale dans la reconnaissance des objets visuels d'une part, et des modèles portant sur la reconnaissance des visages d'autre part.

1.2.2.1 Modèle de la voie visuelle ventrale

Modèles feedforward

Dans une revue récente, DiCarlo (2012) propose un modèle purement *ascendant* (feedforward), basé sur l'organisation hiérarchique du système visuel, dans lequel chaque aire visuelle, en progressant dans la hiérarchie, apporte successivement un pouvoir de calcul supérieur à la précédente permettant ainsi de résoudre des tâches de complexité croissante (Marr, 1982 ; Riesenhuber et Poggio, 1999 ; Dicarlo et al., 2012). Un exemple concret permettant d'illustrer cette hypothèse est le montage en chaîne d'une voiture. A travers les différentes stations de travail d'une usine automobile, en partant de la matière première, des pièces simples sont élaborées puis associées, pour former des structures de plus en plus complexes jusqu'à la construction finale de la voiture entière. Ce modèle est conforté par les latences d'activation des neurones dans ces différentes aires décalées d'environ 10 ms par

aire en progressant dans la hiérarchie (Nowak et Bullier, 1997) et un traitement des traits visuels de complexité croissante depuis V1 jusqu'à IT (Kobatake et Tanaka, 1994 ; Serre et al., 2007). Ainsi environ 100 ms après que l'image ait été projetée sur la rétine, une première vague de sélectivité neuronale semble observée dans IT (Desimone et al., 1984 ; Logothetis et Sheinberg, 1996, Tanaka., 1996) chez le singe tandis que chez l'homme cette activité différentielle liée à la reconnaissance d'objets est observée dès 100-150 ms après la présentation du stimulus (Liu et al., 2002 ; Thorpe et al., 1996 ; Halgren et al., 2000).

Modèle non-hiérarchique

Certains phénomènes visuels vont cependant à l'encontre d'une hypothèse purement hiérarchique. Par exemple, lorsque des scènes ou des objets sont présentés rapidement, les humains sont capables d'en extraire les caractéristiques globales (Oliva et Torralba, 2006) mais sont incapables d'en extraire rapidement les détails (Bar et al., 2006). Ce contraste entre un traitement global rapide et un traitement en détails plus tardif, s'oppose à un traitement purement « feedforward ».

Dans ce sens, l'hypothèse alternative d'un *codage par interaction ascendante et descendante* a été proposée. Elle suggère que la reconnaissance dépendrait de l'information véhiculée via des boucles cortico-corticales dites récurrentes permettant de préciser les représentations à la fois dans les aires visuelles de bas niveau et de haut niveau en parallèle (Lee et Mumford, 2003). Ces interactions entre les différentes couches seraient un algorithme clé de la construction du percept visuel (Lee et Mumford, 2003 ; Bullier, 2001). D'autres auteurs proposent plutôt une hiérarchie inverse comme Hochstein et Ahissar (2002 ; Figure 21). Dans leurs modèles, ces auteurs suggèrent qu'il existerait deux étapes de traitement de l'information (1) Une perception globale initiale basée sur une attention étendue (à large champ récepteur) permettant d'extraire les caractéristiques globales d'une scène (gist de la scène, plage, montagne). (2) un second traitement de l'information incorporant le traitement du détail. Le premier traitement de l'information serait sous-tendu par une première vague d'information purement feedforward tandis que la perception fine des détails serait sous-tendue par une seconde vague d'information feedback, des aires visuelles de haut niveau (IT ou cortex préfrontal (PFC) par exemple) aux aires d'intégration de plus en plus bas niveau si nécessaire. Une illustration concrète de cette hypothèse

pourrait être celle d'une armée où les fantassins (neurone de V1) transmettraient une information incertaine (« je vois peut-être un contour ») au sergent (V2), qui transmettrait à son tour l'ensemble des informations reçues au lieutenant (V4) et ainsi de suite. Les généraux à la tête de l'armée (IT, PFC) obtiendraient une idée grossière de la scène visuelle (« the gist ») qui permettrait par des processus descendants de préciser le travail local des fantassins (Bar et al., 2006 ; Hochstein et Ahissar, 2002).

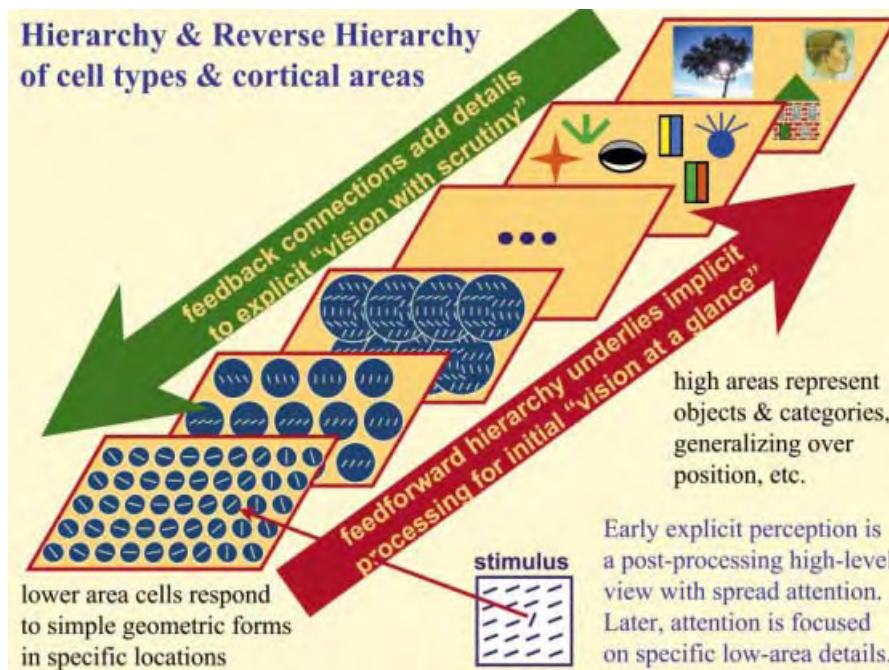


Figure 21 Représentation schématique du modèle de hiérarchie inverse extraite de Hochstein & Ahissar (2002). La flèche rouge représente le traitement hiérarchique classique basé sur l'information feedforward (DiCarlo, 2012). Cependant, considérant le fait qu'il existe anatomiquement des connexions feedback dont la fonction reste inconnue, la flèche verte représente l'hypothèse de hiérarchie inverse qui suggère que l'information feedforward sous-tendrait des processus implicites tandis la perception explicite commencerait dans les aires corticales de haut niveau d'intégration. Ainsi, le « gist » d'une scène serait sous-tendu par une première vague d'information feedforward suivi plus tardivement par un retour de l'information vers les aires de bas niveau via des connexions feedback permettant un traitement détaillé de l'information.

Cependant, en stoppant l'investigation au niveau des aires visuelles, ces modèles, n'expliquent pas la reconnaissance d'objets particuliers, c'est-à-dire des objets que l'on aurait mémorisés et qui nécessiteraient d'être reconnus au niveau individuel. En effet, selon

l'hypothèse d'une organisation hiérarchique de la voie ventrale, les représentations seraient de plus en plus invariantes tout au long de la voie jusqu'à l'élaboration d'une représentation du concept dans les structures temporales. Effectivement, pour reconnaître un objet, il faut tout d'abord le connaître c'est-à-dire l'avoir encodé et mémorisé.

1.2.2.2 Modèle de reconnaissance des visages

Les travaux issus de la neuro-imagerie et des cas de patients cérébro-lésés suggèrent que le gyrus fusiforme serait impliqué dans la perception des visages tandis que l'activation des informations biographiques relatives à l'identité de la personne seraient plutôt sous-tendue par le pôle temporal (Joubert et al., 2004). Elaborés à partir de ces observations, plusieurs modèles axés sur la reconnaissance des visages en particulier, ont été proposés. Ces modèles tentent d'expliquer comment un visage est reconnu par rapport à un autre visage, c'est-à-dire comment il est individualisé.

Un premier modèle proposé par Haxby et al. (2000) synthétise les hypothèses concernant le traitement visuel des visages (Figure 22). Ce modèle suggère que le traitement des visages seraient sous-tendu par un réseau distribué au sein de différentes aires cérébrales divisé en deux grands systèmes : (1) Un système principal (« core system »), constitué d'une part du gyrus fusiforme (FFA) et du cortex occipital inférieur (OFA) participant au traitement des aspects invariants du visage, et d'autre part de la partie postérieure du STS (pSTS) spécialisée dans l'analyse des traits variants. (2) Après traitement par le système principal, un autre système recevrait ces informations et mettrait en jeu un ensemble de mécanismes traitant chacune des caractéristiques du visage (informations liées à la personne, information émotionnelle du visage, etc). Ainsi, ce modèle propose un traitement hiérarchique des visages depuis les régions les plus postérieures vers les régions antérieures de la voie visuelle ventrale jusqu'au pôle temporal. Cependant, selon ce modèle, l'individualisation des visages serait liée à l'activation du gyrus fusiforme et non aux structures mnésiques, ces dernières étant uniquement associées à l'accès à l'identité et aux informations biographiques sur la personne.

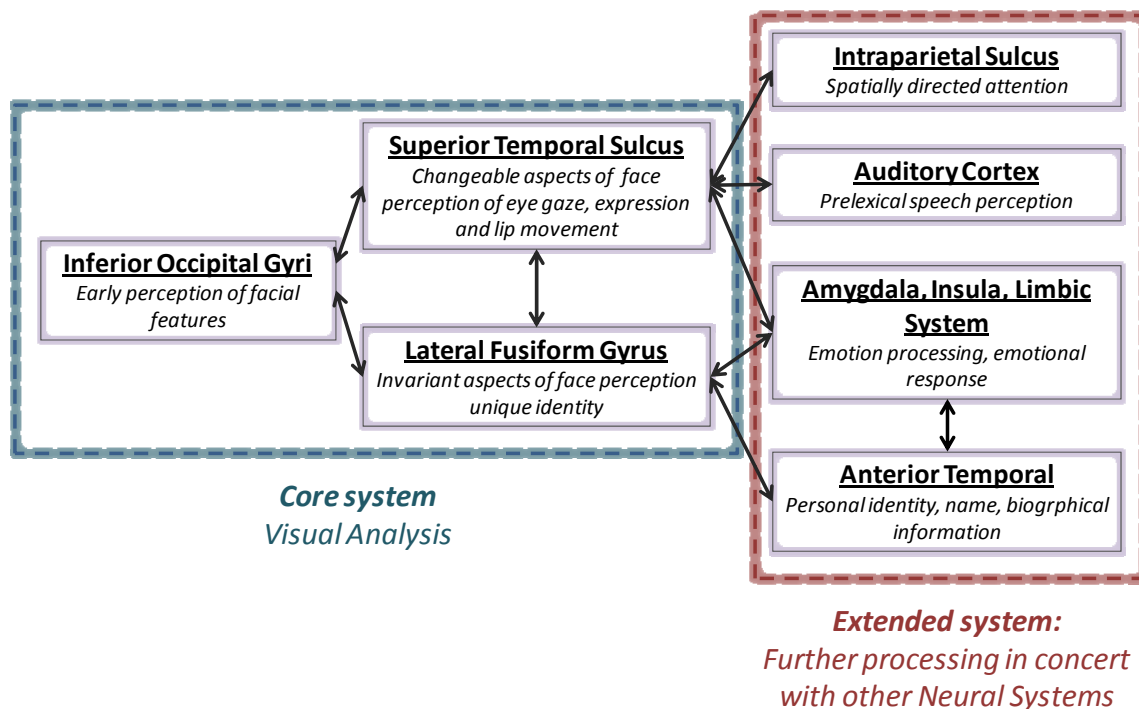


Figure 22 Adaptation du modèle anatomo-fonctionnel suggéré par Haxby (2000). Trois régions cérébrales (gyrus occipital inférieur ou OFA, gyrus fusiform latéral ou FFA et pSTS) formeraient le système principal de traitement perceptif des visages. Ce système serait ensuite connecté à un panel d'autres régions cérébrales participant plutôt à l'identification de la personne et de ses attributs. Ce modèle suit parfaitement l'hypothèse hiérarchique du traitement des visages.

Un second modèle, suggéré par Rossion (2008), propose une révision du modèle hiérarchique d'Haxby (2000). Ce qui le différencie majoritairement du modèle d'Haxby concerne l'organisation du « core system » (Figure 23). En effet dans son modèle, le système principal véhiculerait l'information visuelle selon 2 voies : (1) une voie directe véhiculant de l'information feedforward depuis les aires visuelles primaires directement vers la FFA, sans passer par la OFA et (2) une seconde boucle d'information feedback (réentrante) entre la FFA et la OFA. Ce traitement plus fin via des « allers-retours » entre la OFA et la FFA permettrait de reconstruire une représentation du visage au niveau individuel (Schiltz et Rossion, 2006 ; Rossion, 2009).

La voie directe permettrait de détecter et de catégoriser rapidement un stimulus visage sur la base d'une représentation holistique tandis que la seconde boucle d'information recrutant des régions de plus bas niveau qui présentent des champs récepteur plus petits,

comme la OFA, permettrait grâce à un traitement plus fin de réaliser une analyse détaillée du visage (trait par trait) qui contribuerait à l'individualiser.

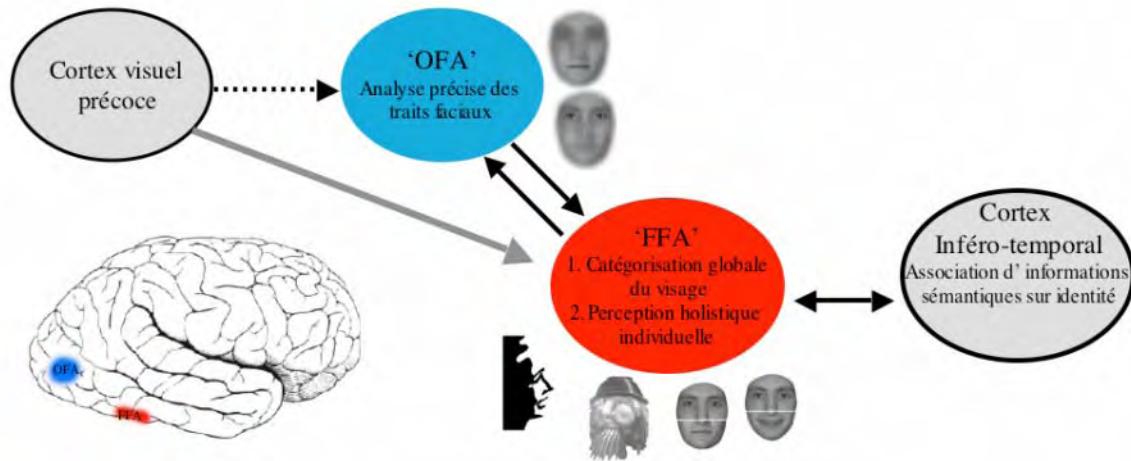


Figure 23 Modèle neuro-fonctionnel du traitement des visages proposé par Rossion (2008, 2009). Ce modèle propose une voie de traitement direct des aires visuelles de bas niveau vers la FFA , permettant une détection rapide du stimulus visage d'après diverses photographies (Mooney face, visages d'Arcimboldo). Cette information globale s'affinerait grâce à une boucle d'activité entre la FFA et la OFA permettant d'obtenir une représentation du visage au niveau individuel associée à des informations sémantiques sur la personne dans les régions temporales antérieures.

Ce modèle ne supporte donc pas une hypothèse strictement hiérarchique de traitement des objets visuels mais propose que des allers-retours entre la FFA et la OFA permettent d'accéder à la familiarité avec le visage. L'accès à l'identité serait ensuite possible grâce à des échanges entre la FFA et les régions temporales.

Un dernier modèle, suggéré par les travaux de Barbeau et al. (2008) réalisés chez des patients épileptiques, propose un traitement massivement parallèle des visages. En enregistrant les potentiels de champ intracrâniens de patients engagés dans une tâche de reconnaissance explicite de visages célèbres, les auteurs montrent une activation massive dans plusieurs régions de la voie ventrale de manière simultanée dans le gyrus fusiforme, le gyrus frontal et le cortex périrhinal lors de la présentation de visages (Figure 24).

Les auteurs proposent qu'au contraire d'un traitement purement sériel ou hiérarchique de l'information comme proposé par Haxby et al. (2000), l'information relative à la familiarité

avec le visage serait traitée en parallèle (activation simultanée de différentes régions cérébrales).

De plus, contrairement au modèle de Rossion (2008), des régions situées plus en avant dans la voie ventrale, telles que le cortex périrhinal, seraient également impliquées dans l'individualisation et la familiarité avec le visage.

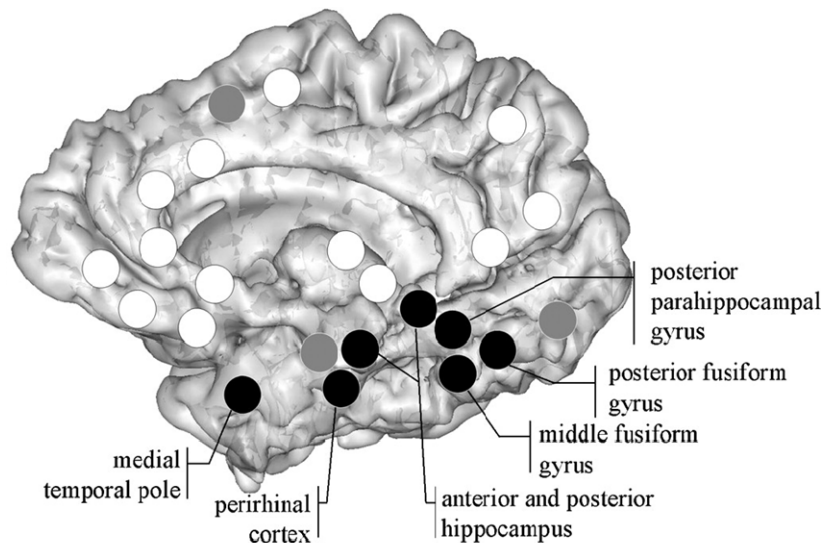


Figure 24 Représentation schématique des différentes régions activées dans l'étude de Barbeau et al. (2008). Plusieurs régions semblent activées simultanément supposant l'implication de l'ensemble des régions de la voie visuelle ventrale en parallèle pour reconnaître un visage.

✓ Dans cette partie, nous avons présenté deux modèles de la voie visuelle ventrale et trois modèles de la reconnaissance des visages familiers. Le modèle hiérarchique d'Haxby (2000) suggère l'implication spécifique du « core system » (FFA, OFA, pSTS) dans l'individualisation des visages. Ce modèle revisité par Rossion (2008) propose au contraire que l'information ne soit pas véhiculée de manière strictement hiérarchique mais grâce à des boucles récurrentes au sein du « core system » permettant d'affiner l'information. Un troisième modèle proposé par Barbeau et al. (2008) suggère plutôt l'implication parallèle de l'ensemble de la voie ventrale incluant le cortex périrhinal.

✓ Ces modèles sont en accord avec les données anatomiques et/ou fonctionnelles permettant d'apporter des informations importantes sur les mécanismes mis en jeu

lorsqu'on reconnaît un visage. Certains modèles proposent un traitement purement feedforward tandis que d'autres suggèrent au contraire l'implication de boucles réentrantes permettant d'affiner l'information ou l'activation de régions éloignées en parallèle. La question de savoir quels types d'informations (feedforward ou feedback) est strictement nécessaire pour reconnaître un visage reste non résolue, les études de cas ou l'IRMf ne permettant pas vraiment d'émettre d'hypothèse sur la dynamique des mécanismes mis en jeu. C'est pourquoi, la dynamique neurophysiologique de la reconnaissance des visages a été largement investiguée à l'aide de l'électroencéphalographie (EEG) et en particulier l'analyse en potentiels évoqués. En effet, un moyen de tester la validité de ces modèles serait de déterminer les latences cérébrales associées aux différentes étapes de traitement, depuis la détection jusqu'à la familiarité afin de confronter nos latences aux modèles théoriques.

1.2.3 Potentiels évoqués et dynamique de la reconnaissance des visages

De nombreuses études utilisant la technique des potentiels évoqués visuels ont permis d'identifier des composantes évoquées (variations du voltage obtenues après moyennage d'un grand nombre d'essais) liées au traitement des visages.

1.2.3.1 Les composantes liées à la détection et à la catégorisation des visages

La présentation de stimuli visages comparée à la présentation d'autres catégories d'objets, semble associée à une augmentation de l'amplitude, et un pic plus précoce de la P100 et de la N170 (Figure 25). La P100, onde positive, débiterait avant 100 ms sur les électrodes postérieures centrales et latérales. Cette composante semblerait trouver sa source dans les régions visuelles extrastriées ventrales et dorsales (Itier et Taylor, 2004 ; Liu et al., 2002). Cependant, sa spécificité pour les visages reste controversée.

Certaines études comme les études de Thierry et al. (2007a, 2007b) et de Liu et al. (2002) suggèrent que la catégorie des visages serait accessible dès 100 ms après la présentation du stimulus. De récentes études en IRMf montrent en effet, qu'il existerait une corrélation entre cette onde et l'activité des régions sensibles aux visages au sein du cortex occipital inférieur latéral droit (Sadeh et al., 2010). De plus, l'application précoce (avant 100 ms) d'une stimulation magnétique transcranienne (TMS) dans ces régions entraînerait un déficit

dans le traitement des visages (Pitcher et al., 2007). Des travaux issues d'enregistrements intracérébraux chez le patient épileptiques ont également montré que des informations relatives à la catégorie des visages comparée à la catégorie des objets étaient disponibles dans les aires appartenant à la voie visuelle ventrale dès 100 ms après la présentation du stimulus (Liu et al., 2009).

Cependant, la plupart des travaux suggèrent que la sensibilité de la P100 serait en fait due à des effets bas niveau (Rossion et al. 1999) en particulier à la différence de spectre d'amplitude entre les visages et les autres catégories d'objets (Tanskanen et al., 2005 ; Rossion et Caharel, 2011 ; Honey et al., 2008). En effet, une modulation spécifique par les stimuli visages serait retrouvée seulement lorsque les stimuli seraient des photographies et ne serait pas rapportée dans toutes les études (Rossion et Jacques, 2008).

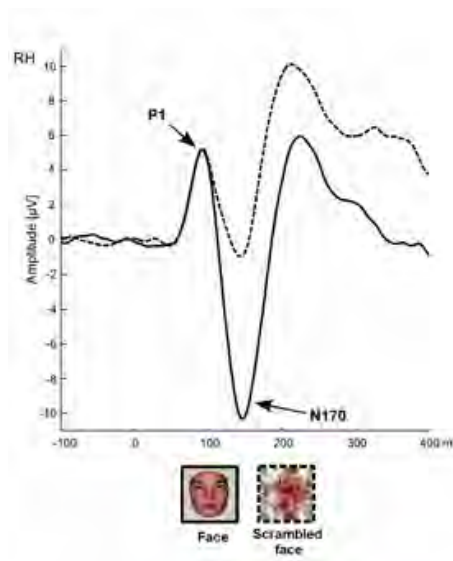


Figure 25 Exemple de potentiels évoqués montrant les composantes P100 et N170 (extraite de Rossion et Caharel, 2011).

La N170 est une onde négative, atteignant son amplitude maximale entre 150 et 180 ms (Bentin et al, 1996 ; George et al., 1996) sur les électrodes occipito-temporales ou occipitales latérales. Cette onde est associée à une onde positive, la VPP (Vertex Positive Potential), apparaissant en même temps sur les électrodes centrales (Jeffreys, 1989 ; Rossion et al., 1999 ; voir Figure 26). Contrairement à la composante P100, une augmentation d'amplitude et un pic plus précoce de la N170 seraient rapportés que l'on présente des visages dessinés,

des caricatures (George et al., 2005; Jeffreys, 1996 ; Latinus et Taylor, 2005), des visages d’Arcimboldo (Figure 26 à gauche) ou des Mooney faces (George et al., 2005 ; Latinus et Taylor, 2005), les Mooney faces étant des photographies de visages fortement contrastées (Figure 26). Ainsi, certains auteurs proposent que cette onde reflète l’accès au réel percept du visage (représentation haut niveau) activé lors de l’étape d’encodage structural du modèle de Bruce et Young (1986) (Eimer, 2000).

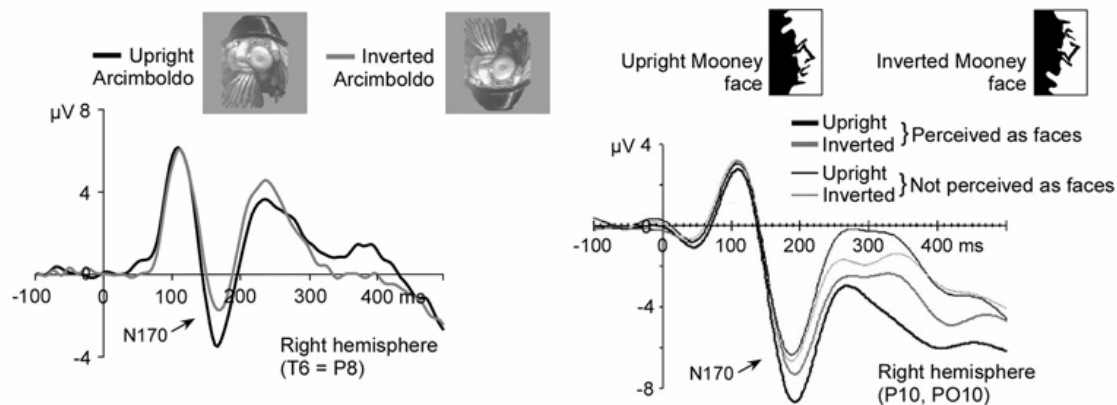


Figure 26 Extrait d’une étude en cours de Rossion et Jacques (page web, *The N170: understanding the time course of face perception*, <http://www.nefy.ucl.ac.be/facecatlab/resProject/N170.htm>) montrant un effet sur l’onde N170 dans la perception des visages en utilisant des visages d’Arcimboldo (à gauche) et des Mooney faces (à droite). Les visages d’Arcimboldo sont des visages suggérés par des fruits, des légumes ou des animaux. Les Mooney faces sont des photographies de visages en noir et blanc fortement contrastées.

En effet, une étude récente de Rossion et Caharel (2011) a développé une expérience permettant de clarifier le rôle de ces deux composantes dans le traitement des visages. Ces auteurs montrent une modulation de la P100 par le spectre d’amplitude (effet bas niveau) accompagnée d’une modulation de la N170 par la phase (représentation haut niveau du stimulus visage) (Figure 27). Dans leur expérience, 172 stimuli sont présentés à 15 sujets qui ont pour consigne d’appuyer sur une touche lorsqu’un objet (visage ou voiture) est présenté, et sur une autre touche lorsqu’il s’agit d’une texture, leur activité EEG étant enregistrée de manière concomitante. La subtilité de cette étude réside dans le contrôle des stimuli. En effet, le premier set de stimuli *objets* est constitué de visages et de voitures dits intacts égalisés en luminance et contraste mais non en spectre d’amplitude. Le second set de stimuli *textures* est constitué quant à lui des mêmes stimuli dont la phase a été remplacée par des valeurs aléatoires. Ainsi, les deux sets de stimuli divergent uniquement par leur phase,

permettant ainsi d'obtenir dans le premier cas, un objet auquel il est possible d'attribuer un sens et dans le second cas un objet dépourvu de sens (aucune catégorie ne pouvant lui être attribuée ; Figure 27).

Ainsi, les auteurs rapportent une augmentation de l'amplitude de la P100 et de la N170 lors de la présentation des visages comparés aux véhicules. Cependant, la modulation de la P100 persiste quelle que soit la phase de l'image (objet ou texture) suggérant que l'effet observé serait dû au spectre d'amplitude. Au contraire, cet effet bas niveau n'est pas retrouvé sur la N170 puisqu'une augmentation de son amplitude est observée seulement lorsque le visage est intact. Les auteurs concluent qu'il existe une différence fonctionnelle entre la P100 et la N170 : (1) la modulation de la P100 par les visages reflèterait des effets bas niveau (spectre d'amplitude) inhérents aux visages qui pourrait rendre compte des saccades très rapides rapportées dans la littérature (Crouzet et al., 2010), certains travaux suggérant que les saccades vers les visages seraient expliquées par le spectre d'amplitude de l'image (VanRullen, 2006; Honey et al., 2008 ; Crouzet et Thorpe, 2011). (2) La N170 indexerait plutôt l'accès à une réelle représentation haut niveau du visage indépendamment des caractéristiques bas niveaux de l'image.

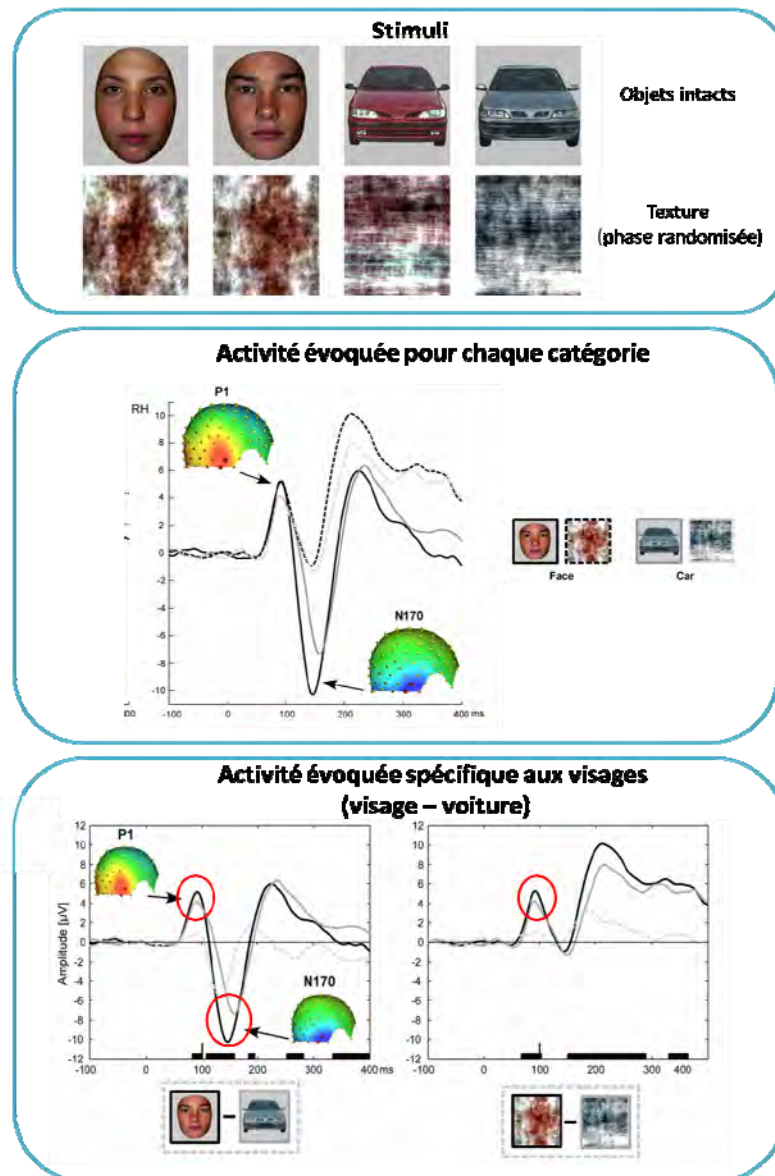


Figure 27 Principaux résultats extraits de Rossion et Caharel (2011) illustrant le rôle fonctionnel différent des composantes P100 et N170. Des exemples de stimuli sont présentés en haut (objets intacts vs objets dont la phase a été randomisée appelée texture). L'activité évoquée obtenue par moyennage est ensuite présentée (au centre) pour chaque catégorie de stimuli (visage intacte en trait plein noir, texture-visage en pointillés noirs, voiture intacte en trait plein gris et texture-voiture en pointillés gris) et montre une activité différentielle spécifique des visages dans les fenêtres temporelles de la P100 et de la N170. Le troisième graphique (en bas) présente l'activité évoquée spécifique aux visages intacts d'une part et du visage texture d'autre part en contrastant leur activité évoquée avec celle des véhicules. Cette analyse point par point (ANOVA), dont la significativité est signalée par la barre horizontale noire sur l'axe des x, révèle qu'une modulation de la N170 est observée uniquement lorsque le visage est intact contrairement à la P100, suggérant son rôle dans la perception consciente du stimulus visage.

1.2.3.2 Les composantes liées à la reconnaissance des visages familiers

L'activité différentielle la plus précoce liée au traitement des visages familiers comparés à des visages inconnus a été rapportée entre 140-200 ms après la présentation du stimulus soit, pendant la fenêtre de la N170 (Caharel et al., 2002, 2005, 2007; Wild-Wall et al., 2008, augmentation de l'amplitude en faveur des visages familiers). Cependant, cet effet ne semble pas retrouvé dans la plupart des études. Plusieurs travaux rapportent en effet une modulation de l'amplitude dans le sens inverse (diminution de l'amplitude pour les visages familiers ; Jemel et al., 2003; Todd et al., 2008). De plus, la plupart des études rapportant un effet à cette latence utilise un faible nombre de stimuli différents qui sont de fait, répétés.

En réalité, la majorité des études ne rapportent pas de modulation de la N170 par la familiarité en particulier lorsque des tâches de reconnaissance *explicite* sont utilisées (Anaki et al., 2007; Bentin et Deouell, 2000; Eimer, 2000; Schweinberger et al., 2004; Zheng et al., 2012; Eimer et al., 2012).

De manière générale, deux opinions divergentes sont retrouvées dans la littérature à propos du rôle de la N170 dans la familiarité. Certains auteurs proposent que la N170 reflète les mécanismes nécessaires à l'individualisation du visage (Jemel et al, 2009 ; Rossion, 2008), alors que d'autres suggèrent plutôt que la N170 serait uniquement liée à la construction de la représentation holistique du visage (Gosling et Eimer, 2011). Dans le premier cas, la familiarité serait accessible dès la fenêtre de la N170 ou proche de cette latence tandis que dans le second cas, la familiarité serait disponible à une étape ultérieure.

La N170 est en effet suivie d'une autre composante négative, la N250, qui semble de manière robuste, indexer la familiarité. Cette composante plus tardive est rapportée en réponse aux visages familiers comparés aux visages inconnus sur les électrodes occipito-temporales et est différente de la N250r qui est une autre composante modulée par la répétition de stimuli visages. Même si cette composante est appelée la N250, suggérant que son amplitude maximale est aux alentours de 250 ms, la plupart des études en potentiels évoqués, rapporte une latence de pic dans une fenêtre très variable après 250 ms (270-330 ms in Zheng et al., 2012; 230-400 ms in Gosling et Eimer, 2011; 240-300 ms in Jemel et al,

2009). Ainsi, ces études suggèrent que l'activité électrophysiologique liée à la familiarité avec les visages serait disponible 100 ms plus tardivement que la N170.

Cependant, la majorité des études s'est focalisée sur l'étude de la moyenne de cette composante (latence du pic) plutôt que sur le début de l'activité différentielle. En effet, une récente étude de Zheng et al. (2011), utilisant un continuum de visages (du même type que celui utilisé dans les études en IRMf) pour différents individus (à partir d'une identité jusqu'à un visage moyen), a proposé une tâche d'identification à des sujets tout en enregistrant leur activité EEG de surface. Les auteurs réalisent une analyse en potentiels évoqués et montrent que les premiers effets de familiarité sont retrouvés dès 200-250 ms après la présentation du stimulus et observent une augmentation de l'amplitude de la N250 en fonction de la force de l'identité. Au contraire, aucune modulation n'est observée pour les composantes précoces comme la P100 et la N170.

De même, une étude récente de Gosling et Eimer (2011), proposant une tâche explicite de reconnaissance de visages célèbres parmi des inconnus à des sujets dont l'activité EEG était enregistrée de manière concomitante, montre que la première différentielle observée en EEG est rapportée dès 230 ms et non plus précocement. De manière intéressante, ces auteurs rapportent une activité positive dès 200 ms largement distribuée sur les électrodes antérieures.

De plus, une étude récente de Jemel et al. (2010), s'est intéressée à la dynamique spatio-temporelle de la reconnaissance implicite et explicite des visages célèbres. Pour cela, les auteurs utilisent une tâche de reconnaissance explicite de stimuli manipulés comme dans certaines études en IRMf (Rotshtein et al., 2004 ; Rossion et al., 2001 en TEP). A partir de 50 photographies de personnes célèbres et 25 visages inconnus, les auteurs constituent pour chaque individu un continuum de 7 images contenant du bruit, depuis une des images très bruitées (niveau 7 ne contenant que du bruit) à des images contenant de moins en moins de bruit (niveau 6 à 1) (Figure 28). Il est demandé au sujet d'appuyer sur un bouton s'il considère qu'il s'agit d'un visage célèbre, sur un autre bouton lorsqu'il s'agit d'un visage inconnu tandis que son activité électrophysiologique est enregistrée.

En utilisant ce protocole, les auteurs cherchent à montrer si la reconnaissance des visages célèbres est un processus continu ou discret. En effet, l'utilisation de continua permet de comparer les essais où le sujet a explicitement reconnu le visage (réponse comportementale de type « oui ») avec les essais précédents la reconnaissance (trop bruités pour être explicitement reconnus).

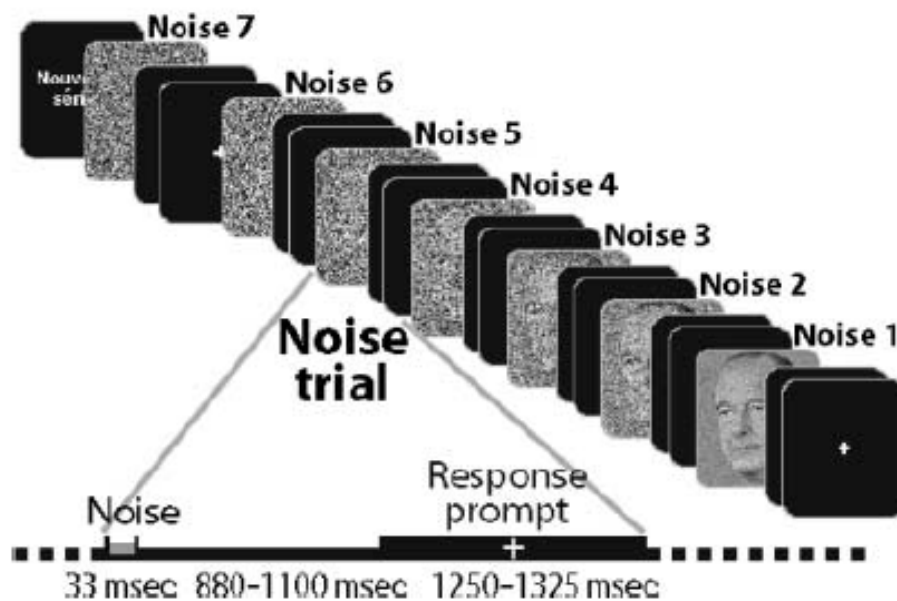


Figure 28 Design expérimental de l'étude de Jemel et al. (2010) montrant un exemple de continuum pour la personnalité de Jacques Chirac depuis le le niveau ne contenant que du bruit (Bruit 7) au niveau le moins bruité (bruit 1). Pour chaque essai le bruit est présenté pendant 33 ms suivi d'un écran noir pour un intervalle randomisé variant de 880 à 1100 ms. Cet écran est ensuite suivi d'une croix de fixation pendant une période aléatoire de 1250 à 1325 ms où le sujet doit donner sa réponse. Cette méthode permet d'étudier la reconnaissance explicite et implicite des visages célèbres.

Les auteurs réalisent ainsi une analyse en potentiels évoqués comparant l'amplitude du signal entre les reconnaissances correctes et les essais précédents la reconnaissance pour chaque catégorie d'images (célèbres et inconnus) d'une part et entre les visages célèbres et inconnus d'autre part, sur les différentes composantes évoquées rapportées dans la littérature que sont la P100, la N170, la N250. Ils étudient également des composantes plus tardives ; la N400 (Bentin et Deouell, 2000 ; Eimer, 2000 ; Itier et Taylor, 2004 ; Boehm et Paller, 2006 ; Schweinberger et al., 2002 ; Kutas et Federmeier, 2011 pour revue) et la P600 (Yovel et Paller, 2004 ; Sun et al., 2012 ; Bentin et Deouell, 2000 ; Eimer, 2000). En effet, la N400 semble être une composante associée à l'activation d'informations sémantiques

verbales ou non, liée aux traitements des visages (Jemel et al., 1999 ; Barret et Rugg, 1989) ou des objets (Kutas et Federmeier, 2011 pour revue) et la P600 semble indexer l'accès aux informations contextuelles sur l'objet (Yovel et Paller, 2004 ; Sun et al., 2012 ; Bentin et Deouell, 2000 ; Eimer, 2000).

Leurs résultats montrent que la reconnaissance des visages célèbres est réalisée à la fois de manière graduelle (modulation dès l'essai précédent la reconnaissance) et de manière tout-ou-rien (une modulation seulement sur les essais où le visage est reconnu). Ils montrent qu'alors qu'aucun effet n'est observé pour les visages inconnus, la N170 et la N250 présentent une amplitude plus importante lors de la reconnaissance explicite des visages célèbres comparés aux essais précédents cette reconnaissance. Au contraire les composantes plus tardives que sont la N400 et la P600 semblent être modulées dès l'essai précédent la reconnaissance (n-1) comparé à l'essai précédent (n-2) (c'est-à-dire deux essais avant la reconnaissance). Cette étude suggère que la dynamique de la reconnaissance des visages serait sous-tendue par des processus complexes, mettant en jeu des activités dès 170 ms (mais également à 250 ms) après la présentation de l'image, impliquées dans l'activation de la trace mnésique du visage célèbre permettant une reconnaissance explicite. Les activités plus tardives, présentant quant à elles un effet avant la reconnaissance explicite de l'image, pourraient être à l'origine de processus de type top-down permettant de guider et de faciliter l'accès aux traces mnésiques représentées en mémoire. Cette étude suggère donc que la familiarité serait accessible dès 170 ms mais dépendrait de mécanismes complexes mis en jeu de manière interactive et en parallèle.

Dans ce sens, les travaux de Barbeau et al. (2008) rapportent une activation massivement parallèle de l'information dans différentes régions de la voie visuelle ventrale (partie 1.2.2.2). L'analyse en potentiels évoqués du signal intracérébral montre que les différentes composantes semblent être retrouvées simultanément dans des régions cérébrales distinctes (à la fois dans le lobe temporal mais également dans les régions frontales inférieures). La synthèse des activités électrophysiologiques rapportées dans leur étude est présentée en Figure 29.

Plus particulièrement, il semblerait qu'un effet de familiarité soit rapporté dès 240 ms en parallèle dans le gyrus parahippocampique et le cortex périrhinal et plus tardivement dans l'hippocampe. Des activités plus tardives rapportées après 400 ms dans les cortex rhinaux et à 600 ms dans l'hippocampe sont également rapportées venant confirmer les précédents travaux de Dietl et al. (2005) et de Trautner et al. (2004).

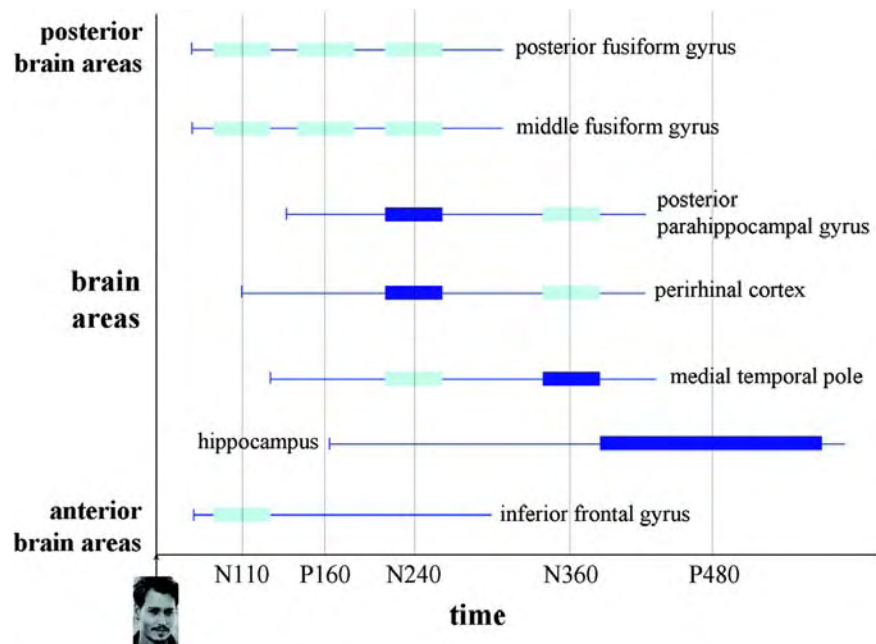


Figure 29 Représentation schématique des activités électrophysiologiques rapportées dans l'étude de Barbeau et al. (2008) dans différentes régions cérébrales. Les différentes composantes électrophysiologiques sont indiquées en abscisse. Les rectangles représentent les principaux pics des composantes dans chaque région avec en bleu foncé, les différentielles significativement différentes entre visages célèbres et visages inconnus.

En résumé, il n'existe pas de réel consensus sur les corrélats électrophysiologiques de la familiarité. En particulier, la latence minimale électrophysiologique liée à la familiarité pourrait débuter dès 150 ms (pendant la fenêtre de la N170) jusqu'à 300 ms (pendant la fenêtre de la N250) après la présentation du stimulus. De plus, la localisation de ces activités liées à la familiarité (électrodes d'intérêt) ne semble pas non plus très bien caractérisée, des activités différentielles étant rapportées à la fois sur les électrodes occipito-temporales mais également sur les électrodes plus antérieures (Gosling et Eimer, 2012).

✓ *La dynamique du traitement des visages a donc été largement étudiée via l'utilisation de l'analyse en potentiels évoqués. Ainsi, plusieurs composantes évoquées ont été rapportées : la P100 serait liée au traitement des informations bas niveau contenues dans le visage qui permettrait sa détection dans une scène pour par exemple effectuer des saccades rapides vers un visage. Le rôle de la N170 semble quant à lui plus controversé certains auteurs proposant un rôle dans l'accès à une représentation holistique haut niveau du visage uniquement, tandis que d'autres travaux suggèrent également qu'elle jouerait un rôle dans l'individualisation et la familiarité avec le visage. Les composantes plus tardives que sont la N250, la N400 et la P600 semblent toutes impliquées dans la reconnaissance des visages familiers. Ainsi, la latence minimale de la familiarité demeure toujours inconnue et pourrait être retrouvée dans une fenêtre variant, selon les études, de 150 ms (N170) à de 300 ms (N250) après la présentation du stimulus. De plus, les électrodes montrant un effet de la familiarité semblent également variables. Finalement, la dynamique de la reconnaissance des visages familiers reste une question non résolue à ce jour.*

✓

Problématique générale

Dans ce premier chapitre, nous avons présenté différentes études suggérant que la reconnaissance des visages serait sous-tendue par un système de reconnaissance très efficace nous permettant d'interagir *rapidement* avec nos congénères. Cependant les travaux portant sur la rapidité de la reconnaissance rapportent des latences comportementales et neurales très variables et par conséquent ne convergent pas vers un réel consensus qui permettrait de mieux caractériser les processus nécessaires à la reconnaissance d'un visage familier. Cette variabilité dans les temps de réaction pourrait s'expliquer par différents facteurs comme un mauvais contrôle de la stratégie des sujets (familiarité vs identification, la familiarité étant sensée être plus rapide que l'identification) qui pourrait peut-être être évitée si les sujets étaient contraints à répondre rapidement. En effet aucune étude n'a étudié la dynamique de la reconnaissance des visages familiers en appliquant des contraintes temporelles. *La reconnaissance des visages au niveau de la familiarité est-elle réellement si rapide ? Ne serait-ce pas simplement une idée reçue ? Rapide, peut-être, mais à quelle vitesse ? Que nous apprennent ces latences sur les processus mis en jeu ?*

Objectif 1 : *Déterminer les latences les plus rapides pour reconnaître un visage familier au niveau comportemental.*

L'état de l'art rend bien compte de la nécessité de mettre *un chiffre* sur le temps de traitement nécessaire à cette habileté. Pour cela, nous avons mis en place une stratégie méthodologique et conceptuelle particulière permettant d'apporter des éléments de réponse à cette problématique.

Sous-problématique 1 : *Déterminer le coût temporel entre la détection et la reconnaissance.*

Nous avons adapté et développé des paradigmes similaires à ceux utilisés pour étudier la rapidité du système visuel (catégorisation visuelle rapide) afin d'identifier et d'améliorer la compréhension des mécanismes strictement nécessaires à la reconnaissance d'un visage familier. Afin de pouvoir comparer nos résultats à une valeur de référence, nous utiliserons également une tâche de détection (visages humains vs visages animaux) et une tâche de

catégorisation rapide du genre (hommes vs femmes ou femmes vs hommes). Nous proposons de vérifier s'il existe une hiérarchie dans le traitement visuel des visages célèbres en comparant directement ces différentes étapes de traitement. Aucune étude n'a étudié cette question lorsqu'un sujet est forcé à répondre rapidement dans un protocole de catégorisation visuelle rapide. Nous avons pour objectif de déterminer le coût du passage de la détection à la reconnaissance d'un visage au niveau de la familiarité.

Sous-problématique 2 : Déterminer le coût d'une reconnaissance bottom-up comparé à une reconnaissance top-down. Comme défini dans la partie 1.1.3.2, il existe deux types de reconnaissance, la reconnaissance bottom-up et la reconnaissance top-down. Nous avons donc développé une tâche de reconnaissance top-down (ex : reconnaître plusieurs photographies de Johnny Hallyday parmi des inconnus) que nous comparerons à une reconnaissance bottom-up (plusieurs visages célèbres parmi des inconnus).

Objectif 2 : *Déterminer les latences électrophysiologiques minimales.*

Au vu des données de la littérature utilisant généralement la technique des potentiels évoqués pour étudier la dynamique des processus cérébraux, nous proposons ici d'utiliser une analyse essai par essai (multi-variate pattern analysis (MVPA)) qui nous permettra d'analyser le signal essai pas essai sans a priori sur les électrodes à considérer. En appliquant cette analyse pour décoder le signal EEG correspondant aux essais cibles des essais distracteurs, nous déterminerons la latence minimum de la reconnaissance des visages au niveau de la familiarité.

Objectif 3 : *Investiguer directement le lien entre comportement et activité électrophysiologique.*

Pour tenter d'identifier précisément les activités cérébrales strictement nécessaires à la reconnaissance des personnes, il semble important de mettre en relation les réponses comportementales (obtenues sous contraintes temporelles) avec les réponses électrophysiologiques. Nous tenterons de confronter les latences comportementales et électrophysiologiques en vérifiant s'il existe une corrélation directe entre les latences comportementales minimales et les latences électrophysiologiques minimales.

Objectif 4 : *Confronter nos latences aux modèles théoriques*

Nous faisons l'hypothèse que la détermination, en temps de réaction et en électrophysiologie, du temps de passage d'un niveau de catégorisation à un autre fournira des informations sur les processus et mécanismes supplémentaires mis en jeu entre ces deux étapes de traitement. Nous pourrions ainsi discuter des mécanismes (feedforward et/ou feedback) mis en jeu lors de la reconnaissance des visages familiers.

2 Chapitre 2. A la recherche du temps minimum

Étudier la « vitesse » des traitements cognitifs constitue la composante fondamentale de la méthodologie mise en place dans cette thèse. En effet, la composante temporelle est ici cruciale pour tenter de comprendre la dynamique du traitement des visages tant d'un point de vue comportemental que neural. C'est en contraignant le sujet à aller le plus vite possible que nous pourrions le forcer à utiliser la stratégie optimale, ou du moins la plus rapide, afin de caractériser les mécanismes strictement nécessaires pour effectuer cette tâche cognitive (Fabre-Thorpe, 2011). Déterminer la latence minimale permettra de proposer des hypothèses et de modéliser les mécanismes mis en jeu.

Déterminer la latence comportementale minimale pour reconnaître le visage d'une personne familière

Nous avons développé un ensemble de paradigmes expérimentaux contraignant le sujet à répondre le plus rapidement possible afin d'analyser les différentes latences nécessaires pour réaliser la tâche.

2.1.1 Tâches de catégorisation visuelle rapide : tâche de Go/no-go et de Speed accuracy boosting (SAB)

2.1.1.1 Tâche de go/no-go

Nous avons tout d'abord adapté un paradigme très utilisé au laboratoire (développé initialement pour étudier la rapidité du système visuel) qui est une tâche de catégorisation visuelle rapide appelée tâche de go/no-go (Figure 30). Elle consiste à répondre lorsqu'un item cible est présenté et d'inhiber sa réponse lorsqu'il s'agit d'un item distracteur. Des travaux de Bacon-Macé et al. (2007) ont montré que cette méthode permettait d'améliorer significativement les temps de réactions comportementaux comparés à une tâche de oui/non, cette dernière méthode consistant à produire une réponse spécifique aux items cibles d'une part et aux distracteurs d'autre part. Afin d'obtenir une estimation précise du

temps strictement nécessaire à la réalisation de la tâche, l'utilisation de cette méthode semblait donc être un choix judicieux.

De plus, nous avons utilisé différentes stratégies pour optimiser les performances des sujets :

- (1) Un entraînement avec des stimuli spécifiques afin que le sujet se familiarise avec la tâche.
- (2) un design expérimental organisé en série d'essais (ou bloc) permettant ainsi de montrer au sujet son temps de réaction ainsi que le nombre de fausses reconnaissances (fausses alarmes) à la fin de chaque bloc.
- (3) un encouragement oral qui consistait à demander systématiquement au sujet de « battre son score » au prochain bloc.
- (4) Des stimuli flashés pour une durée de seulement 100 ms favorisant ainsi des temps de réaction rapides et ne laissant pas assez de temps au sujet pour réaliser une saccade.

2.1.1.2 Speed and accuracy boosting procédure

Dans un second temps, nous avons développé une nouvelle méthodologie basée sur la tâche de go/no-go décrite précédemment. Cette méthode, appelée, procédure de « Speed and Accuracy Boosting » permet, en théorie, d'optimiser les performances des sujets en leur faisant utiliser d'emblée une stratégie plus rapide et efficace (Besson et al., sous presse). Pour cela, deux améliorations de la tâche de go/no-go sont mises en place :

- (1) Au lieu de disposer de 1000 ms pour répondre (temps de présentation initial de l'écran suivant le stimulus permettant au sujet de répondre), le temps de présentation de l'écran est variable permettant de contraindre le sujet à répondre de plus en plus rapidement et de ce fait de déterminer le temps de réponse minimal pour effectuer la tâche.
- (2) Un feedback auditif indique au sujet si sa réponse est correcte ou incorrecte (voir Figure 30) pour que le sujet ait un retour sur ses performances en temps réel, ce feedback ayant pour but d'augmenter la précision du sujet.

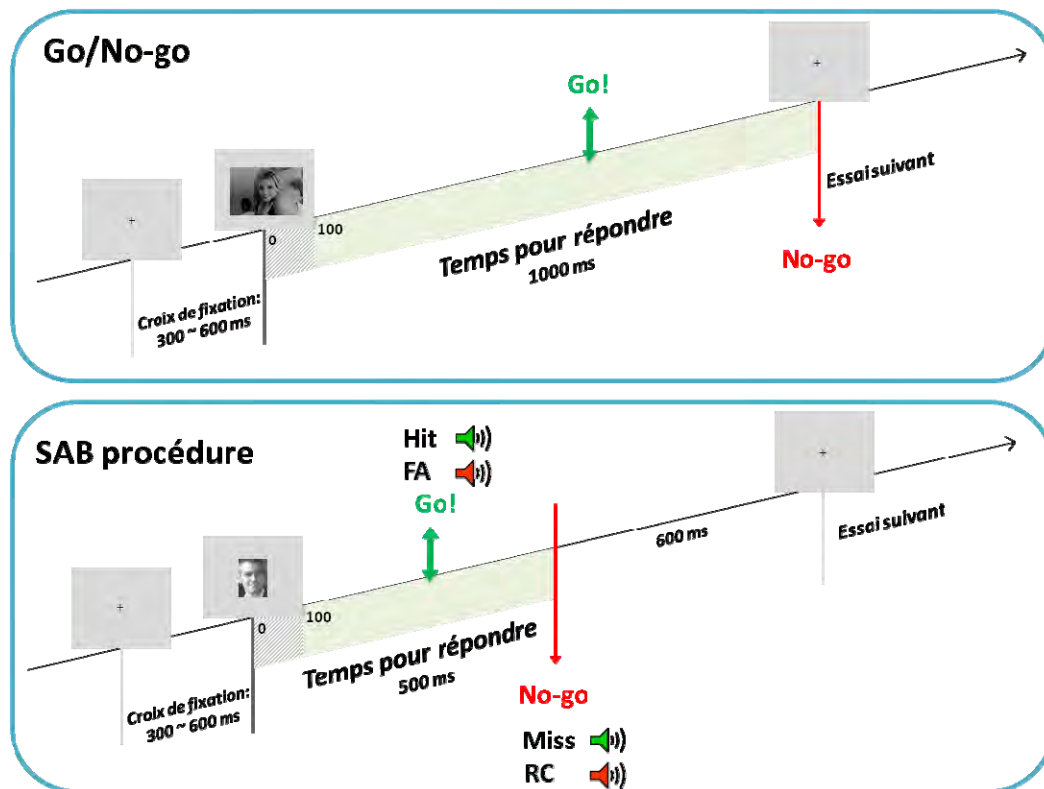


Figure 30 Schémas explicatifs des deux protocoles utilisés. Dans la tâche de go/no-go, la croix de fixation est présentée pendant un intervalle de temps aléatoire qui varie entre 300 et 600 ms suivie de l'apparition du stimulus flashé pendant 100 ms. Enfin, un écran noir permettant au sujet de répondre est présenté pendant 1000 ms avant le début de l'essai suivant. Dans la tâche de SAB, la croix de fixation est présentée pendant 300 à 600 ms avant la présentation du stimulus (100 ms). Un écran gris est ensuite présenté pendant seulement 500 ms après la présentation du stimulus. Si le sujet répond correctement (réponse go correcte (Hit) ou absence de réponse à un distracteur (rejet correct, CR), un feedback auditif positif est appliqué. A contrario, si la réponse du sujet est incorrecte (réponse go incorrecte (Fausse alarme, FA) ou réponse no-go incorrecte (omissions ou Miss en anglais) un feedback auditif négatif est appliqué.

Dans notre protocole de reconnaissance rapide, nous avons utilisé un temps de présentation de l'écran de réponse de 500 ms. Nous avons basé cette limite sur de précédents travaux réalisés par Besson et al. (sous presse) étudiant la mémoire de reconnaissance pour différentes catégories d'objets (objets concrets, objets abstraits et visages célèbres) montrant un effondrement des performances pour la reconnaissance des visages célèbres lorsque la limite de temps était en dessous de 500 ms.

2.1.2 Utilisation de stimuli naturels en contexte : optimisation des performances comportementales

Lorsque l'on cherche à étudier le traitement des objets visuels, deux alternatives sont possibles. Il est possible d'utiliser des stimuli bien contrôlés présentés de manière isolée. Cette méthodologie a l'avantage de permettre de faire varier de manière discrète une ou plusieurs propriétés de l'objet (taille, position,...) et de moduler idéalement de manière paramétrique les réponses neuronales face à ces variations (Hung et al., 2005). Une seconde possibilité est l'utilisation d'une large banque de photographies naturelles d'objets, qui présente de manière inhérente une combinaison continue de toutes les propriétés possibles (Thorpe et al., 1996).

La plupart des études portant sur le traitement des visages utilisent des stimuli artificiels (images composites ou détournées avec ou sans cheveux, Figure 31). Or, certaines études ont montré que trois facteurs jouent un rôle déterminant dans les classements effectués par les sujets dans des tâches d'identification: l'âge (« vieux » vs « jeune »), la forme du visage ("rond" vs "ovale") et la longueur des cheveux (« longs, frisés et sales » vs « courts et nets »). Il est intéressant de remarquer que les cheveux constituent également le trait facial le plus fréquemment cité (24 % des cas) lorsque l'on demande à des sujets de décrire un visage, présent ou absent, en vue de sa reconnaissance ultérieure par une tierce personne (Ellis et al., 1977, 1979).

De plus, des résultats neurophysiologiques montrent que l'activité unitaire des neurones est plus importante pour les stimuli naturels comparée à des stimuli artificiels témoignant d'un codage plus efficace (Vinje et Gallant, 2000). En d'autres termes, certains auteurs suggèrent que le codage de l'information visuelle serait « optimisée » pour le traitement des scènes naturelles (Simoncelli et Olshausen, 2001).

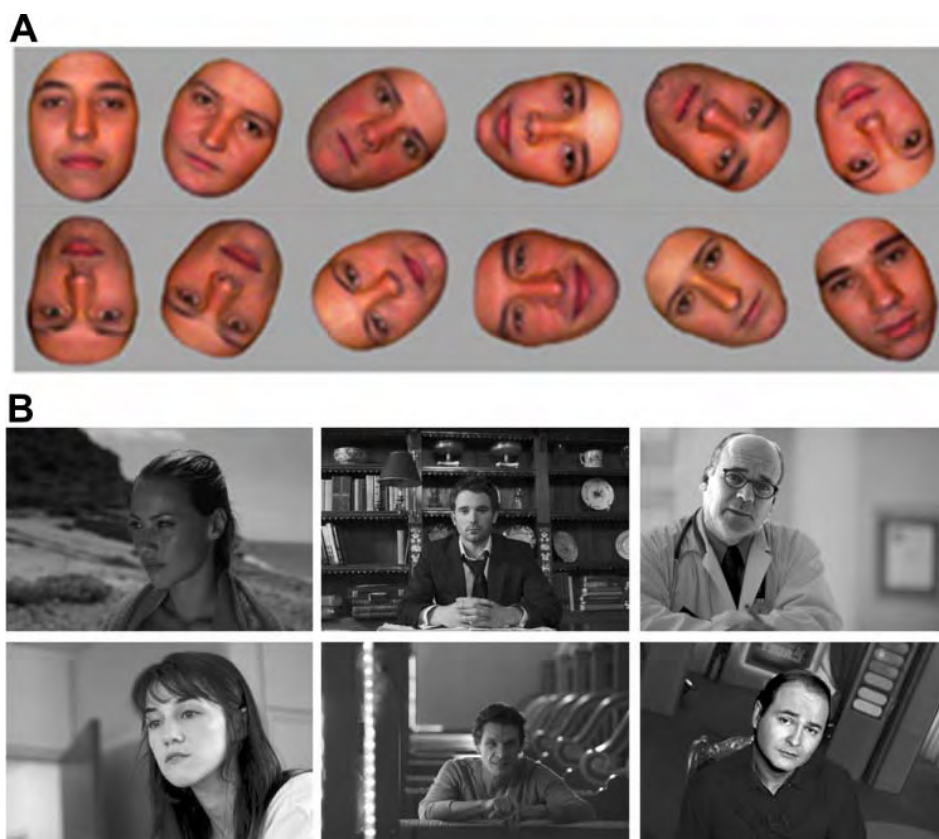


Figure 31 Exemple de stimuli présentant en (A) des stimuli extraits de Rossion et Jacques (2008) qui sont des visages détournés ne laissant pas apparaître les oreilles et les cheveux des individus et en (B) des stimuli en contexte naturels utilisés dans nos études montrant en haut des exemples de visages inconnus et en bas des exemples de visages célèbres.

Dans notre tâche de go/no-go, nous avons donc choisi d'utiliser des stimuli naturels. Une base de photographies, comportant 540 visages inconnus, 270 visages célèbres et 360 visages présentés dans leur contexte naturel, a été utilisée. Les visages pouvaient être en gros plan ou de loin, centrés ou non. Les photos ont été choisies pour être aussi variées que possibles. Les visages d'humains ou d'animaux pouvaient être de face, de trois quart, ou de profil. Les personnes expriment ou non des émotions. Certains font des gestes sociaux. Cette diversité d'angles de vue, de contextes et d'émotions semble écologiquement plausible.

Nous avons contrôlé les indices visuels de bas niveau pour toutes les photos des trois groupes de stimuli (animaux, inconnus, célébrités). Ces trois groupes ne diffèrent pas statistiquement par la luminance ($p > 0,05$), le contraste ($p > 0,1$). Nous avons aussi effectué ces contrôles pour le visage lui-même (luminance $p > 0,1$, contraste $p > 0,1$). Nous avons contrôlé que tous les visages des trois groupes de photos avaient une taille similaire (en

nombre de pixel) ($p > 0,1$). Nous avons également contrôlé d'autres facteurs comme la direction du regard, l'émotion, la présence d'objets sur le visage (lunettes de soleil par exemple). Le détail de ces facteurs est présenté en tableau 1 et en annexe 2.

Tableau 1 Comparaison des caractéristiques visuelles de l'image entre les visages célèbres et les visages inconnus comportant l'orientation de la tête (horizontale, verticale, inclinaison), l'ethnie (noir, asiatique, caucasien), la direction du regard (regard dans les yeux ou non), la présence des dents (visibles ou non), l'émotion (la présence d'un sourire ou non).

	Echelle/ Catégorie	Célebres (%)	Inconnus (%)	p
Horizontale	1	3.0	4.2	0.5
	2	15.3	15.7	0.9
	3	59.2	61.2	0.8
	4	20.7	14.5	0.02*
	5	1.8	4.4	0.08
Verticale	1	0.7	0.8	0.9
	2	3.4	7.4	0.04*
	3	91.5	88.9	0.3
	4	4.4	2.7	0.2
	5	0.0	0.2	0.5
Inclinaison	1	0.4	0.9	0.4
	2	3.3	2.3	0.3
	3	88.9	90.5	0.7
	4	7.4	5.9	0.3
	5	0.0	0.4	0.3
Ethnie	Caucasien	93.3	91.9	0.4
	Noir	6.3	4.1	0.1
	Asiatique	0.4	4.0	0.004*
Autres	Objets	35.2	34.4	0.5
	Regard	51.8	62.4	0.2
	Sourire	52.2	61.7	0.3
	Dents	41.5	53.7	0.06

*Un test de chi-2 a été appliqué pour comparaisons statistiques. Les valeurs rapportées sont exprimées en proportion et les symboles « * » indiquent une différence significative. Les p-values ne sont pas corrigées pour comparaisons multiples. Trois comparaisons sont significativement différentes mais liées à une faible proportion de photographies. Le détail de la valeur des échelles est précisé en annexe 2.*

Dans notre procédure SAB, nous avons utilisé les stimuli les mieux reconnus de la tâche de go/no-go mais découpés en rectangle autour du visage pour que ce dernier soit centré (Figure 32). Cette manipulation a été effectuée dans le but de faciliter la tâche, la SAB procédure étant plus difficile.



Figure 32 Exemples de stimuli utilisés dans la tâche de go/no-go en (A) et dans la procédure SAB en (B). Les stimuli cibles sont encadrés en vert et les distracteurs en rouge.

2.1.3 Calcul du temps de réaction minimum

La plupart des études en temps de réaction rapporte le temps de réaction moyen ou dans le meilleur des cas le temps de réaction médian. Cependant, certains détails fondamentaux sont capturés par l'étude de la distribution des temps de réaction et des temps de réaction minimaux (Rousselet et al., 2003 ; VanRullen et Thorpe, 2001).

En effet, le temps de réaction minimum constitue une information primordiale permettant de comprendre précisément à quelle vitesse le cerveau peut réaliser un traitement cognitif. Ainsi, une méthode très utilisée au laboratoire consiste à déterminer statistiquement le premier pas de temps pour lequel les bonnes réponses (Hit) sont significativement supérieures aux fausses alarmes (Figure 33). Ce calcul peut être réalisé pour chacun des

sujets mais également à travers les essais, cette seconde méthode permettant de modéliser un méta-sujet idéal.

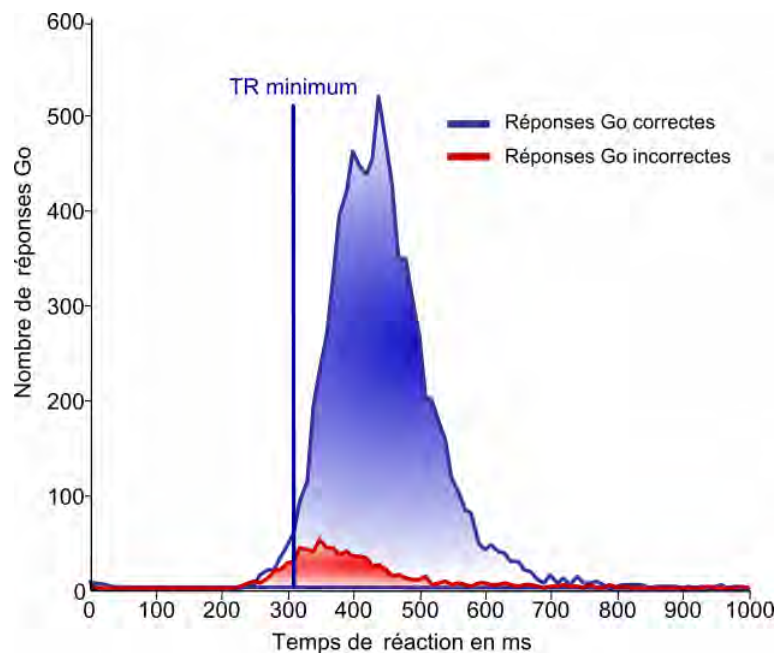


Figure 33 Exemple de calcul du TRmin à travers les essais dans une tâche de go no-go. La distribution des bonnes réponses Go (ou Hits) est représentée par la courbe bleue tandis que les mauvaises réponses go (ou fausses alarmes) sont représentées en rouge. La ligne bleue verticale représente le TR minimum calculé à l'aide d'un test de chi 2 où $p = 0.01$ avec des bins de 10 ms. Ce temps de réaction minimum correspond au premier pas de temps où les bonnes réponses sont significativement plus importantes que les fausses alarmes révélant ainsi le temps de réaction minimal pour faire un traitement cognitif donné.

Dans nos travaux, nous avons calculé le *temps de réaction minimum individuel* (pour chaque sujet) en effectuant un test de Fisher exact ($p < 0,05$) sur des bins de 30 ms. Pour corriger pour comparaisons multiples, la significativité devait durer 60 ms au minimum. Nous pouvons ainsi comparer les temps de réaction minimaux individuels entre nos différentes conditions (t test ou Wilcoxon selon le nombre de sujets considérés, Figure 34). Grâce à cela nous avons pu déterminer un *temps de réaction minimum moyen* à travers les sujets.

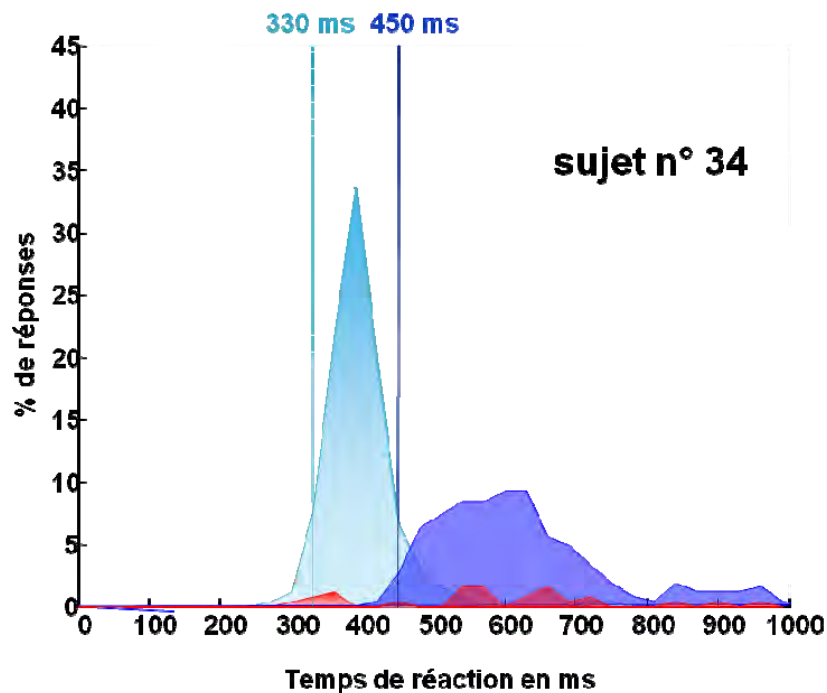


Figure 34 Exemple de deux distributions des temps de réaction obtenues pour un sujet dans la tâche de go/no-go dans deux conditions où les réponses go correctes apparaissent en bleu (visages humains vs visages animaux en bleu clair et visages célèbres vs visages inconnus en bleu foncé). Les fausses alarmes apparaissent en rouge. Les barres verticales indiquent le temps de réaction minimum pour chaque condition en utilisant un test de Fisher exact sur des bins de temps de 30 ms.

Nous avons également calculé un temps de réaction minimum à travers tous les essais. Pour cela nous avons réalisé un test de chi-2 ($p < 0,01$) comparant les réponses go correctes et incorrectes sur des bins de 10 ms. Afin de corriger pour comparaisons multiples, la significativité devaient durer 50 ms (soit 5 bins consécutifs). Nous avons ainsi déterminé le temps de réaction minimum du méta-sujet idéal, c'est-à-dire modéliser le comportement d'un sujet qui aurait effectué plus de 10000 essais. Cette valeur qui peut faire l'objet de critique au niveau statistique nous permet tout de même d'apporter un complément d'information sur la dynamique des mécanismes mis en jeu.

Déterminer la latence neurophysiologique minimale pour reconnaître le visage d'une personne familière

Identifier et comprendre les réseaux recrutés lors de différentes tâches cognitives suppose que l'on puisse mesurer les signaux reflétant l'activité de populations neuronales des différentes régions cérébrales impliquées dans cette tâche. L'enregistrement de l'activité d'une population peut se faire de différentes manières; de manière indirecte

(enregistrement de l'activité métabolique et/ ou métabolique consécutive à l'activité neuronale) ou de manière directe (accès aux signaux électriques ou magnétiques). Cette dernière méthode (l'enregistrement direct des signaux électriques), présente l'immense avantage d'avoir une résolution temporelle de l'ordre de la milliseconde contrairement aux méthodes indirectes.

Ici, la composante temporelle apparaît fondamentale lorsqu'il s'agit d'étudier la reconnaissance des visages familiers, l'être humain étant supposé être très rapide pour réaliser cette tâche. Ainsi, les techniques d'enregistrement de l'activité cérébrale utilisées doivent avoir une résolution temporelle suffisante. Parmi les différentes méthodes disponibles, l'électroencéphalographie (EEG) et la stéréo-EEG (sEEG) avec leur résolution temporelle à la milliseconde semblent être de bons candidats pour étudier la reconnaissance des visages.

2.2.1 Des techniques de haute précision temporelle

2.2.1.1 EEG de surface

L'EEG, est apparu au début du 20ème siècle avec Hans Berger, qui commença à l'étudier dans les années 1920 et enregistra le premier signal d'activité cérébrale en 1929. L'EEG permet de montrer les variations temporelles de l'activité électrique cérébrale avec une grande précision temporelle, milliseconde par milliseconde; elle est enregistrée au moyen d'électrodes placées sur le cuir chevelu du sujet, et amplifiée. En sciences cognitives, ces électrodes sont directement montées sur un casque EEG pouvant comporter de 32 à plus de 250 électrodes d'enregistrement.

Le signal obtenu montre les modifications de différence de potentiels électriques par rapport à une référence. Ces potentiels électriques enregistrés en surface semblent résulter de la sommation des potentiels d'action post-synaptiques synchrones issus d'un grand nombre de cellules pyramidales du cortex. Ces cellules, perpendiculaires à la surface du crâne, peuvent, lors du passage d'un influx nerveux, former un dipôle. Lorsque des zones entières réagissent de façon synchrone, une différence de potentiel suffisamment grande peut être captée. Par contre, le potentiel électrique mesuré en EEG est très faible (nécessité d'utiliser un amplificateur de signal) et très diffus, ce qui limite l'information spatiale fournie

par l'examen EEG. Il est donc difficile, sans traitements additionnels comme la localisation de source, de déterminer quelles sont les structures cérébrales d'où provient le signal EEG.

Lors de nos expériences, l'activité électrique était enregistrée à l'aide de deux systèmes (Figure 35). Dans les expériences présentées en chapitre 3 et 4, un bonnet de 32 électrodes est utilisé. Ce bonnet est en conformité avec le system 10-20 (Oxford Instruments). Des électrodes extra-occipitales ont été ajoutées. L'électroencéphalogramme est enregistré avec un amplificateur SynAmps (Neuroscan, Inc). Le signal a été numérisé à un taux de 1000 Hz. Un filtre Notch (50Hz) était appliqué lors de l'enregistrement ainsi qu'un filtre passe bas de 200 Hz et un filtre passe haut de type DC qui il permet de contrôler les dérives lentes du signal. L'électrode Cz a été utilisée comme référence pendant l'enregistrement. Les impédances ont été gardées en dessous de 5k Ω .

Neuroscan 32 électrodes



ANT 128 électrodes



Figure 35 Photographies des bonnets utilisés dans nos expériences avec le système d'acquisition Neuroscan en haut et le système ANT en bas. En haut, des photographies du bonnet EEG 32 électrodes permettant de montrer les électrodes extra-occipitales sont présentées. En bas sont présentées des photographies de l'amplificateur ANT, du bonnet EEG 128 électrodes ainsi que des électrodes permettant d'enregistrer les mouvements oculaires. Un système de localisation des électrodes appelé polhémus permettant d'enregistrer la position exacte des électrodes est également présenté à droite.

Dans les expériences préliminaires présentées en discussion, l'activité électrique a été enregistrée à l'aide d'un bonnet EEG comportant 128 électrodes conforme au système 10-20 (Oxford Instruments) avec une référence en M1 et une terre en AFFz. Toutes les impédances étaient maintenues inférieures à 5 k Ω . Les électrodes étaient reliées à un amplificateur (ANT) relié à un PC d'acquisition permettant l'enregistrement de l'EEG avec ASA version 4. Simultanément à l'enregistrement EEG, les mouvements oculaires du sujet étaient enregistrés à l'aide de 4 électrodes bipolaires ; 2 électrodes horizontales qui enregistrent les mouvements latéraux et 2 verticales qui enregistrent les clignements des yeux. L'acquisition du signal s'effectue en 1024 Hz. Un filtre Notch (50hz) est appliqué lors de l'enregistrement. Une photodiode, placée sur l'écran de stimulation, a été utilisée pour la synchronisation de la présentation des photographies et de l'enregistrement électrophysiologique.

2.2.1.2 Stéreo-electroencéphalographie (sEEG)

La sEEG est une technique opératoire de neurochirurgie, utilisée dans le cadre du diagnostic de l'épilepsie consistant à implanter chirurgicalement par voie stéréotaxique des électrodes profondes dans le cerveau du patient afin de localiser le foyer épileptogène (zone génératrice des crises). Ces électrodes profondes sont munies de multiples contacts électriques permettant de recueillir les signaux électriques d'activité encéphalique. Le patient est placé sous surveillance vidéo permanente et les signaux électriques recueillis sont enregistrés puis corrélés aux enregistrements vidéo et analysés afin de définir la (les) zone(s) à l'origine des crises d'épilepsie. Cette méthode n'est utilisée que dans le cas d'épilepsies sévères pharmacorésistantes.

Rappelons que l'épilepsie est une affection neurologique définie par la répétition de crises d'épilepsie spontanées, une crise épileptique étant l'ensemble des conséquences cliniques d'une décharge neuronale excessive et hypersynchrone d'une partie plus ou moins importante du cortex cérébral. Au sein des épilepsies partielles, l'épilepsie temporale est la forme la plus fréquente chez l'adulte (Bancaud, 1987). Ces épilepsies deviennent fréquemment pharmaco-résistantes. Dans ce cas, une intervention chirurgicale visant à retirer le foyer épileptogène est parfois proposé au patient. Lorsque le foyer épileptogène ne peut pas être déterminé par des méthodes non invasives (EEG de surface, tomographie par émission de positrons (TEP), IRM et bilan neuropsychologique), une investigation pré-

chirurgicale invasive consistant à implanter des électrodes intra-cérébrales (sEEG) peut être proposée au patient. Cette méthode permet d'analyser les foyers et réseaux épileptogènes et permet aussi d'établir une cartographie fonctionnelle des régions implantées avant l'acte chirurgical. Cette investigation se déroule dans une unité spécialisée, « l'unité d'épileptologie » (dirigée par le Dr Luc Valton) appartenant au service de Neurologie et d'explorations fonctionnelles du Centre hospitalier universitaire de Rangueil à Toulouse.

Les patients implantés conservent ces électrodes intracrâniennes pour une durée d'une à deux semaines, période pendant laquelle ils restent dans une chambre spécialement conçue pour cet examen. Les patients demeurent cependant relativement mobiles et peuvent vaquer à leurs occupations tandis que leur activité cérébrale est enregistrée en continue. En dehors des périodes d'investigations cliniques visant à caractériser leur épilepsie, les patients sont disponibles et nous pouvons leur proposer de réaliser certaines tâches cognitives pendant lesquelles leur activité électrique intracérébrale est enregistrée.

Les électrodes utilisées dans la sEEG sont des électrodes multi-plots (de 10 à 15 plots par électrode environ) (Figure 36).

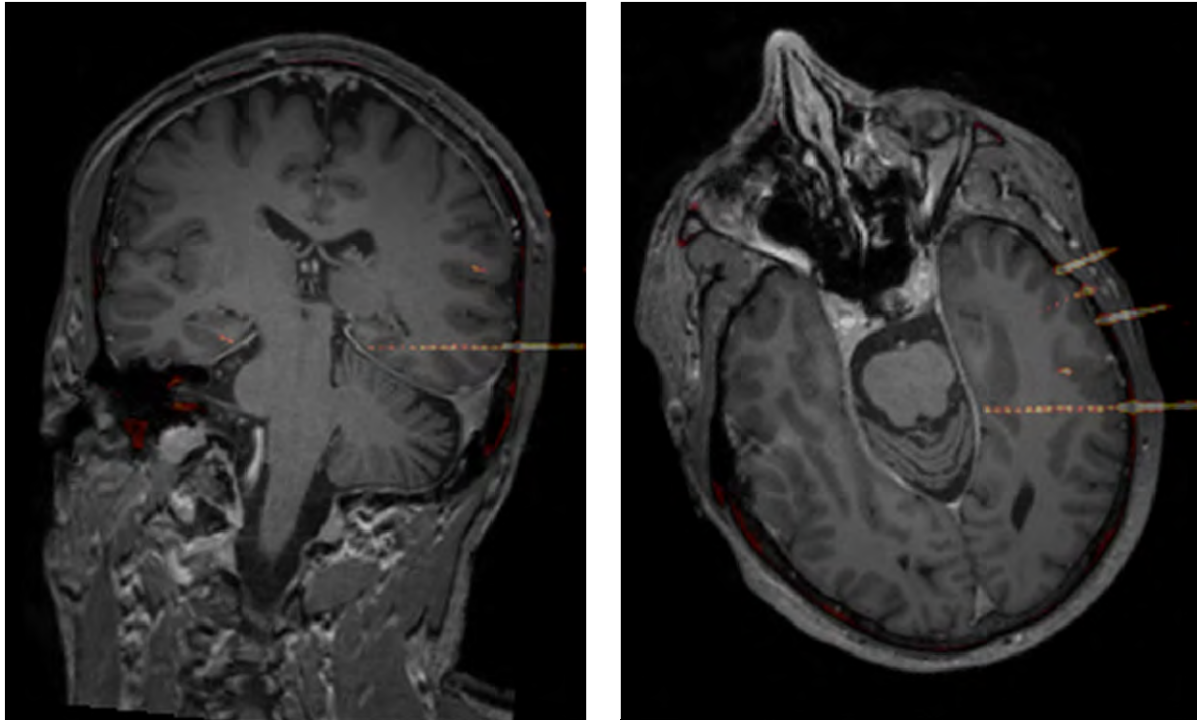


Figure 36 Photographie d'une macroélectrode en haut et du recalage d'une électrode sur l'IRM du patient implantées obtenues avec le logiciel Sysiphe. Il s'agit de l'électrode OT' (occipito-temporale gauche) laissant apparaître chacun des plots d'enregistrement.

Chaque plot constitue un site d'enregistrement. Les électrodes utilisées ont un diamètre de 0,8 mm, elles comportent cinq à quinze plots d'enregistrement de 2 mm de longueur, séparés par une distance de 1,5 mm. Cette technique permet donc d'enregistrer l'activité cérébrale avec une résolution de l'ordre du millimètre. Les électrodes sont implantées de manière orthogonale au plan sagittal inter-hémisphérique, traversant différentes régions cérébrales, cortex et substance blanche. Ce type d'implantation permet d'enregistrer différents champs électriques le long d'une seule et même électrode. La localisation précise des plots de chacune des électrodes peut être réalisée grâce à une méthode de recalage de la position des électrodes sur l'IRM du patient (logiciel Sysiphe développé par le Dr Jean-Albert Lotterie (Figure 36).

Le signal EEG enregistré par les plots des électrodes intracérébrales est probablement issu des fluctuations du voltage de larges populations de cellules pyramidales. Chaque neurone pyramidal a une arborisation dendritique dense dont l'activité des synapses excitatrices et inhibitrices engendre des flux électriques entre les dendrites et le soma. Chaque cellule pyramidale agit alors comme un dipôle de courant qui fluctue. Lorsque les cellules pyramidales sont toutes orientées dans le même axe, les champs électriques induits par chaque neurone se somment et peuvent être enregistrés localement ou à distance en fonction de la conductance des tissus environnants (Figure 37).

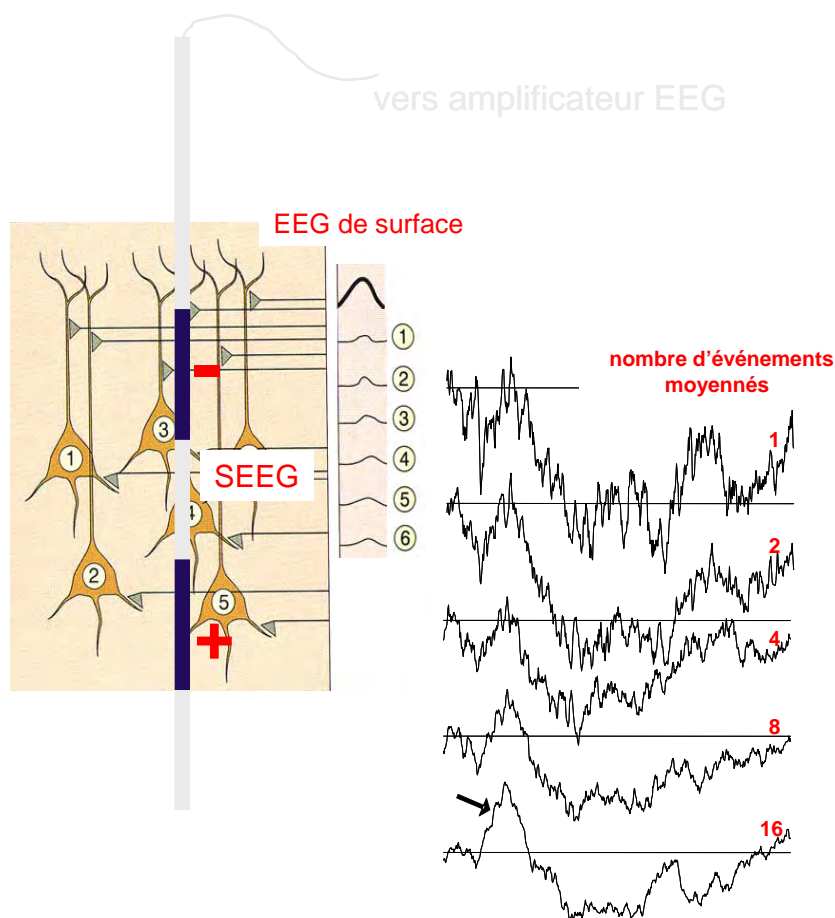


Figure 37 Représentation schématique de l'origine des signaux enregistrés en électrophysiologie intracérébrale à droite et en EEG de surface à gauche. À droite sont représentés les différents signaux électrophysiologiques enregistrés pour chaque événement. L'analyse en potentiel évoqués consiste à moyennner ces différents signaux.

2.2.2 A la recherche de l'activité électrophysiologique la plus précoce : Moyennage vs essai par essai

Pour étudier la reconnaissance des visages familiers au niveau neural, il faudrait idéalement une technique qui nous permettrait de quantifier l'information neurale avec précision pour chaque présentation de stimuli, d'évaluer la dynamique du traitement de cette information et de faire un lien avec le comportement c'est-à-dire entre traitement neuronal et réaction comportementale à chaque essai.

2.2.2.1 Les potentiels évoqués

Classiquement, l'analyse du signal en potentiels évoqués a été largement utilisé pour étudier la dynamique des différents traitements cognitifs. En effet, lorsque l'on présente un stimulus, comme une image par exemple, l'activité neuronale et donc l'activité électrique recueillie à la surface du scalp (EEG) ou au niveau des tissus cérébraux conduisant l'activité nerveuse locale (sEEG) va être modifiée. Comme le cerveau ne reste jamais inactif, cette modification est perdue dans l'activité cérébrale de fond. Cependant, si l'on répète de nombreuses fois la même stimulation, il est possible, par moyennage, d'extraire la séquence des événements électriques entraînés liée au traitement cognitif associé à cette tâche (Dawson, 1951 ; voir Figure 37).

Lorsque l'on fait la moyenne de l'activité électrique enregistrée après la présentation de l'image, seule l'activité qui s'est reproduit à chaque présentation de ce stimulus est visible. Ces petites variations électriques, appelées potentiels évoqués ou « Event Related Potentials » (ERPs), vont varier en polarité (positifs ou négatifs), en latences, amplitudes et topographie.

Cette approche consiste donc à étudier à quel temps ces événements électriques se produisent et sur quelle électrode. La majorité des travaux s'intéresse à la latence à laquelle l'amplitude maximale du potentiel est rapportée permettant ainsi de définir différentes composantes évoquées. Ensuite, une analyse classique appelée analyse de pic permet d'étudier si cette amplitude diffère d'une condition à une autre. D'autres travaux se limitent à donner une latence minimale de traitement de la fonction cognitive d'intérêt (Thorpe et al., 1996) qui correspond au début de la divergence de l'activité évoquée pour une condition

par rapport à une autre. Cette technique d'analyse a une grande précision temporelle et permet de suivre la progression du traitement d'un stimulus à la milliseconde près. Cependant, même si ces méthodes ont été extrêmement efficaces pour délimiter les temps de traitement associés aux différents mécanismes cognitifs, elles ne permettent pas d'accéder aux contenus mentaux, essai par essai. De plus, cette technique nécessite généralement de considérer des électrodes a priori pour mettre en évidence des effets. Dans notre cas, il n'existe pas de réel consensus à propos des électrodes présentant un effet de familiarité. C'est pourquoi, nous avons décidé d'utiliser dans cette thèse une méthode d'analyse alternative permettant également d'étudier le signal avec précision mais en utilisant l'ensemble du signal présent sur le scalp sans a priori sur les électrodes.

En effet, comme nous allons le voir, le champ émergent consistant à transférer les techniques d'apprentissage artificiel et de classification aux neurosciences semblent être l'outil idéal pour étudier les représentations neuronales.

2.2.2.2 L'analyse essai par essai

Depuis une dizaine d'années la combinaison des méthodes d'imagerie cérébrale et des algorithmes d'apprentissage par ordinateur a permis de développer des algorithmes d'apprentissage automatique. Leur rôle est d'extraire un sens à partir de données brutes pouvant posséder de nombreuses dimensions. On pose généralement deux types de problèmes à ces algorithmes (1) retrouver dans les données des classes connues au préalable, on parlera d'apprentissage supervisé ou (2) d'extraire une certaine organisation des données, sans classe préalablement déterminée, c'est l'apprentissage non supervisé.

Ici, nous utilisons donc l'approche la plus commune d'apprentissage supervisé: la classification. Les données brutes auxquelles on aura recours seront les potentiels de champs enregistrés sur l'ensemble des électrodes (donc sans a priori). Les classes que nous souhaitons « lire » dans cette activité sont celles des catégories d'objets visuels présentées au sujet (visages animal, visages humains, visages célèbres). On prédira quelle est la classe la plus probable de la représentation observée, parmi un nombre limité de classes connues qui ont servi à entraîner le classifieur au préalable.

Tous les algorithmes d'apprentissage supervisé fonctionnent de la même manière :

- (1) Dans une première phase, dite d'entraînement, l'algorithme va essayer de déterminer l'espace dans les données qui permet de séparer les différentes classes de manière optimale. Ici, cette phase consiste à donner au classifieur les extraits du signal EEG labélisé (par exemple visages humains et visages animaux) correspondant à la moitié des essais choisis de manière aléatoire parmi l'ensemble des essais. Le classifieur peut utiliser différentes caractéristiques du signal, dans notre cas, il utilise l'amplitude. Durant cette phase d'entraînement, le classifieur va rechercher des critères permettant de définir le meilleur hyperplan permettant de séparer les deux sets de données correspondant aux deux conditions dans notre espace multidimensionnel correspondant à nos 32 électrodes (ou à différents plots si l'on considère le signal sEEG).
- (2) La seconde phase est appelée test, puisqu'elle va permettre d'évaluer comment de nouveaux exemplaires vont être classifiés dans cet espace et ainsi de vérifier si la machine a bien appris à catégoriser. Cette seconde phase consiste à donner au classifieur la seconde moitié des essais non-labélisés qui sera projetée dans l'espace multidimensionnel de classification. On obtiendra ainsi un nombre d'images correctement classées qui correspondra à la précision du décodage.

La procédure est répétée 30 fois (cross validation) pour éviter un biais lié au set de stimuli d'entraînement et de test nous permettant d'obtenir un intervalle de confiance. La moyenne et l'intervalle de confiance à 95% de cette moyenne à travers ces 30 cross validations est ensuite rapportée. Ensuite, afin d'effectuer une analyse statistique sur ce décodage, une procédure similaire est réalisée sur les mêmes images mais lorsque le label de ces images (ex : cibles ou distracteurs) est randomisé lors de la phase d'entraînement afin d'obtenir un niveau de chance. Enfin, à l'aide d'un t test, le décodage réel et le décodage randomisé sont comparés. Cette analyse peut être réalisée à chaque milliseconde de signal nous permettant de déterminer à la milliseconde près et essai par essai à partir de quand le classifieur est capable de décoder l'activité électrophysiologique liée aux deux sets de stimuli.

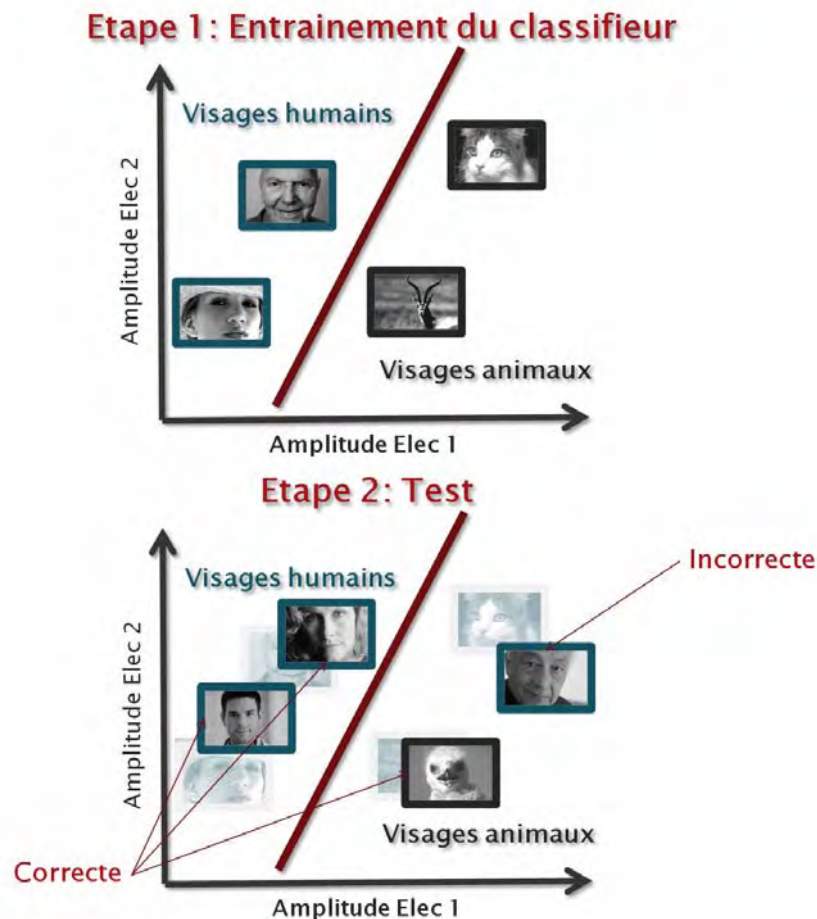


Figure 38 Illustration schématique du décodage. En haut est représentée la phase d'entraînement du classifieur. Pour simplifier, cet exemple considère un espace à seulement deux dimensions correspondant au signal enregistré sur deux électrodes. Ainsi, le classifieur détermine le meilleur hyperplan qui sépare les deux sets de stimuli dans cet espace à deux dimensions. Lors de la phase de test (en bas), de nouvelles images sont fournies au classifieur qui va les classer selon les critères préalablement établis lors de la phase d'entraînement. Dans cet exemple, trois images sur quatre ont été correctement classifiées. Ce résultat correspond à un décodage d'une précision de 75%. Dans nos travaux, il s'agit d'un espace à 32 dimensions correspondant à nos 32 électrodes, chaque set de stimuli contenant 270 images.

Il existe de nombreux algorithmes d'apprentissage automatique, différenciables par la fonction qu'ils appliquent pour séparer les différentes classes, et la méthode d'optimisation qu'ils utilisent pour y parvenir. La fonction de séparation, aussi appelée noyau, peut être extrêmement complexe et généralement, en s'appuyant sur des hypothèses fortes de la distribution des données, elle est adaptée par des spécialistes du domaine pour chaque problème à résoudre. Dans le cas des neurosciences et en particulier pour décoder des représentations mentales, on applique le principe de parcimonie, qui stipule que pour que ces représentations soient utilisables et interprétables, par exemple pour un neurone chargé

de prendre une décision, elles doivent être séparables le plus simplement possible, c'est-à-dire de manière linéaire. Dans ce cas, le but de l'algorithme est donc de trouver un ou des hyperplan(s) décrit(s) par la combinaison linéaire des mesures séparant au mieux les différentes classes. Même au sein de cette sous-classe d'algorithmes supervisés linéaires, ou classifieurs linéaires, il existe de nombreuses possibilités, néanmoins les études comparant l'efficacité de ces différents algorithmes en neurosciences, tendent à montrer qu'ils fournissent des résultats sensiblement similaires (Mehring et al. 2003; Lotte et al. 2007).

Pour des raisons d'accès libre sur internet et de rapidité de traitement, nous avons choisi d'utiliser des machines à vecteurs de support linéaire (« SVM for support vector machine » en anglais). Sans entrer dans les détails mathématiques de cet algorithme, la notion clé sur laquelle il repose est celle de marge maximale. La marge est la distance entre la frontière (hyperplan) de séparation et les essais les plus proches. Ces derniers sont appelés vecteurs supports. Dans les SVMs, la frontière de séparation est choisie comme celle qui maximalise la marge. Le problème est de trouver cette frontière séparatrice optimale, à partir de la série d'entraînement. Ceci est fait en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus (Fan et al. 2008).

On dira que l'on est capable de *décoder le signal catégoriel* si, à un instant t , on est capable de prédire l'état cognitif d'un sujet (ce qu'il voit, ce qu'il imagine...). Un SVM, s'il a été entraîné à classer des états cognitifs dans un espace d'activité neurale permettant une séparation des classes, sera capable, si on lui fournit le signal neural à un nouvel instant t , de prédire l'état cognitif du sujet. On parlera de « décodage » du signal.

Présentation des études

Une première étape a consisté à développer un protocole permettant de contraindre le sujet à effectuer la tâche de reconnaissance de visage célèbre le plus rapidement possible. Pour cela, nous avons adapté le protocole de catégorisation visuelle rapide (tâche de go nogo), très utilisé dans notre laboratoire dans des études portant sur le traitement des objets visuels en général (Thorpe et al., 1996 ; Rousselet et al., 2003). Nous avons développé une tâche de catégorisation de visages célèbres parmi des visages inconnus. Nous avons pris soin de caractériser si le visage nécessitait d'être identifié ou non pour effectuer la tâche. Ainsi, pour effectuer cette tâche, un simple sentiment de familiarité avec le visage semblait nécessaire. En effet, c'est l'étape de familiarité avec le visage que nous souhaitons étudier dans ce travail car selon la plupart des études, la familiarité serait plus rapide que l'accès à l'identité (Bruce & Young, 1986). D'autre part, nous avons utilisé un large set de personnes célèbres différentes (supérieurs à 200 célébrités) à propos desquelles le sujet n'avait aucune information au préalable, nous plaçant ainsi dans le cadre d'une reconnaissance de type « bottom-up ».

Partant de l'hypothèse que la reconnaissance des visages familiers serait un processus rapide, il nous semblait crucial de comparer cette tâche à des tâches déjà connues pour être rapides. Si l'on considère le modèle séquentiel de Bruce et Young (1986) ou les travaux de Macé et al. (2009), une tâche de catégorisation d'un visage au niveau superordonné (visages humains vs animaux) semblait être une bonne alternative. Cette première étude permettra de répondre à la première question « la reconnaissance bottom-up des visages familiers est-elle rapide ? », en le comparant à une tâche connue pour être rapide (catégorisation au niveau superordonnée, catégorisation du genre, Article 1).

Ensuite en utilisant notre protocole contraignant fortement les sujets à répondre rapidement (SAB), nous pourrions mettre un chiffre sur cette vitesse de traitement chez le sujet sain jeune dans des conditions bottom-up (Article 2) ou top-down (Article 3).

Une seconde étape a consisté à déterminer la dynamique du signal électrophysiologique dans ces mêmes tâches de détection et de reconnaissance des visages familiers. Nous pourrions ainsi confronter nos latences électrophysiologiques obtenues lorsque le sujet est contraint de répondre rapidement avec les latences rapportées dans la littérature. En utilisant une technique originale de traitement du signal (MVPA), nous pourrions identifier le temps de réaction minimal permettant de reconnaître le plus rapidement possible des visages familiers parmi des inconnus. Nous pourrions ainsi mettre en relation les latences comportementales, obtenues lors de la détection et de la reconnaissance des visages familiers, avec les latences électrophysiologiques minimales et ainsi apporter des éléments de réponses quant à la dynamique et la relation des activités électrophysiologiques rapportées avec le comportement du sujet.

Enfin, nous discuterons nos résultats et confronterons nos latences aux modèles théoriques afin de proposer des hypothèses sur les mécanismes mis en jeu lors de la reconnaissance des visages au niveau de la familiarité.

3 **Chapitre 3. A quelle vitesse reconnaît-on un visage célèbre?**

L'idée selon laquelle la reconnaissance des visages familiers est un processus rapide semble évidente dans la littérature scientifique et dans l'opinion publique. Cependant les données rapportées dans la littérature ne semblent pas converger vers un consensus sur cette vitesse de traitement. Connaître précisément cette vitesse semble pourtant crucial pour contraindre les modèles actuels dans le domaine de la vision et de la mémoire. En effet, c'est à l'aide de protocoles utilisant des contraintes temporelles que certaines fonctions cognitives, comme la reconnaissance des objets, ont été plutôt bien caractérisées.

Nous avons donc développé différentes expériences permettant d'étudier la reconnaissance rapide des visages familiers. Nous comparerons ces latences avec des tâches dites rapides comme la catégorisation superordonnée des visages (détection) ou la catégorisation du genre (Article 1 et Article 2). Nous nous intéresserons ensuite à un facteur conceptuel important, souvent oublié dans la littérature qui est le type de reconnaissance : la reconnaissance « bottom-up » et la reconnaissance « top-down ». La reconnaissance bottom-up correspond à la recherche d'un visage déterminé (ex : rechercher Nicolas Sarkozy) et la reconnaissance top-down correspond à la recherche de plusieurs personnes familières différentes (sans savoir à l'avance de qui il s'agit) parmi des inconnus. De tels paradigmes peuvent être comparés à des situations de la vie courante, la recherche "bottom-up" pouvant être comparée à la situation écologique de "tomber" par hasard et de manière inattendue sur une personne familière dans la rue et la recherche top-down pouvant correspondre à l'activité de recherche d'une personne connue dans un groupe ou dans une foule. Nous comparerons ces deux types de reconnaissance dans l'Article 3 et déterminerons ainsi le coût temporel d'une reconnaissance bottom-up.

ARTICLE 1 : How fast is famous face recognition?

Gladys Barragan-Jason, Fanny Lachat, Emmanuel Barbeau

(2012), Frontiers in Psychology, 3, 454

Article 1: How fast is famous face recognition? (publié)

Introduction : Reconnaître rapidement un visage familier est crucial dans nos interactions sociales. Cependant la vitesse de la reconnaissance des visages familiers reste inconnue. Plusieurs paradigmes et définitions ont mené à des conclusions contradictoires, la vitesse de reconnaissance n'ayant jamais été comparée à des tâches dites rapides. Dans cette première étude, nous tenterons de déterminer le coût temporel entre la reconnaissance de visages célèbres au niveau de la familiarité et deux autres tâches supposées rapides (catégorisation superordonnée et catégorisation du genre).

Méthode : 18 sujets sains doivent catégoriser le plus rapidement possible 240 visages célèbres parmi des visages inconnus en se basant sur leur familiarité avec le stimulus présenté. Nous comparerons les performances et les temps de réaction obtenus avec une tâche contrôle dite de catégorisation superordonnée (détection) de 240 visages inconnus parmi des visages animaux. Les tâches de détection et de reconnaissance impliquant l'utilisation de stimuli d'homogénéité différente (l'une ne met en jeu que des visages humains contrairement à l'autre), les sujets effectuent également une tâche de catégorisation du genre (Visages masculins vs visages féminins et vice versa) utilisant exactement les mêmes stimuli que dans la tâche de reconnaissance (visages célèbres et inconnus) permettant d'évaluer l'impact de ce facteur sur les temps de réaction. Nous avons ensuite réalisé une expérience complémentaire au cours de laquelle 7 sujets devaient apprendre les stimuli cibles puis réaliser de nouveau la tâche de reconnaissance à 4 reprises. Cette tâche permettait d'une part d'augmenter la précision des sujets afin d'obtenir des expériences de difficulté comparable et de vérifier si les sujets pouvaient diminuer leur temps de réaction au cours des sessions d'autre part.

Résultats : Reconnaître un visage célèbre nécessite au minimum 467 ms à l'échelle du groupe, soit environ 150-180 ms de plus que la détection ou la catégorisation du genre. Deux sujets semblent cependant utiliser une stratégie plus efficace et arrivent à effectuer la tâche de reconnaissance en 390 ms. Sept sujets dont les deux sujets rapides ont ensuite effectué la tâche d'apprentissage. Les résultats montrent que :

(1) les sujets lents peuvent atteindre un temps de réaction minimum proche de 400 ms comme les sujets rapides

(2) les sujets rapides ne peuvent pas améliorer leur temps de réaction mais atteignent un temps limite à 360-390 ms incompressible

(3) un coût temporel de plus de 150 ms est observé entre reconnaissance et catégorisation

(4) ce coût temporel ne peut pas être seulement expliqué ni par la similarité des stimuli (un délai similaire étant observé entre la catégorisation du genre et reconnaissance), ni par la difficulté de la tâche (l'apprentissage permet aux sujets d'atteindre une précision en reconnaissance comparable aux tâches de détection et de catégorisation du genre).

Discussion : Ces résultats suggèrent la reconnaissance d'un visage familier implique des mécanismes additionnels à la catégorisation superordonnée et du genre qui prennent du temps. Nos résultats sont en effet compatibles avec un traitement hiérarchique depuis le niveau superordonné jusqu'au niveau de la familiarité.

✓ *Même si les temps de réaction les plus rapides étaient aux alentours de 360-390 ms, ceci concernait seulement deux sujets les autres participants répondant majoritairement après 450 ms. Ainsi, nous avons fait l'hypothèse que notre protocole de reconnaissance rapide en go/no-go ne permettait peut-être pas de forcer les sujets à utiliser leur stratégie la plus rapide. Nous avons donc développé dans une seconde étude un nouveau protocole, la Speed and Accuracy Boosting procédure qui est une méthode de haute contrainte temporelle. Nous espérons ainsi contraindre un maximum de sujets à répondre rapidement afin de déterminer une valeur minimale robuste pour reconnaître un visage célèbre.*

**ARTICLE 2 : Fast and Famous: Looking at the fastest
speed at which a face can be recognized**

Gladys Barragan-Jason, Gabriel Besson, Mathieu Ceccaldi, Emmanuel Barbeau

(2013), Frontiers in Psychology, 4, 100

Article 2: Fast and Famous: Looking for the fastest speed at which a face can be recognized (publié)

Introduction : Suite à une première étude montrant que certains sujets (n=2) pouvaient reconnaître un visage célèbre en 390 ms, nous avons utilisé un nouveau protocole permettant, en théorie, de contraindre les sujets à utiliser d'emblée leur stratégie la plus rapide. Nous avons voulu tester un grand nombre de sujets afin de déterminer un temps minimum robuste pour reconnaître un visage familier.

Méthode : Nous avons développé deux expériences. Dans l'*Expérience 1*, 31 sujets ont réalisé une tâche de go/no-go similaire à celle présentée dans l'Article 1 (270 célèbres vs 270 inconnus). Dans l'*Expérience 2*, 101 sujets ont catégorisé 140 visages célèbres (les mieux reconnus de l'expérience 1) parmi 140 inconnus en utilisant le nouveau paradigme (SAB) permettant de forcer les sujets à utiliser une stratégie rapide.

Résultats : Comme nous le supposions, l'expérience 1 révèle que même sur 31 nouveaux sujets, seulement 2 sujets utilisent la stratégie rapide (390 ms), le temps médian des temps de réaction minimums des sujets étant de 555 ms. Ces résultats sont tout à fait comparables aux résultats rapportés dans l'Article 1. Au contraire, à l'aide de la SAB, près de 40 % des sujets semblent obtenir un temps de réaction minimum entre 360 et 390 ms, la médiane des minimums étant de 440 ms. Ainsi, de nouveau et de manière cohérente avec l'Article 1, il semble qu'un temps incompressible de 360-390 ms soit nécessaire pour reconnaître un visage familier dans une situation bottom-up.

Discussion : La reconnaissance des visages familiers ne semble donc pas si rapide. Ce temps incompressible de 360-390 ms confirme nos hypothèses sur l'implication de mécanismes additionnels à la catégorisation superordonnée. Les processus mis en jeu méritent d'être investigués en détails.

✓ *Ce temps de reconnaissance minimum est obtenu dans une reconnaissance de type bottom-up, c'est-à-dire lorsque le sujet n'a aucun a priori sur les visages qui vont être*

présentés. Il nous semblait donc intéressant de comparer cette condition de reconnaissance bottom-up à une reconnaissance top-down et de déterminer ainsi le coût temporel entre ces deux types de reconnaissance. Nous espérons ainsi que ce résultat nous donnera des informations supplémentaires sur les mécanismes impliqués dans la reconnaissance des visages familiers.

ARTICLE 3 : Top-down vs bottom-up face recognition

Gabriel Besson, Gladys Barragan-Jason, Simon Thorpe, Sébastien Puma,

Mathieu Ceccaldi, Emmanuel Barbeau

Article en préparation

Article 3 : Top-down vs bottom-up face recognition (en préparation)

Introduction : Dans nos deux premières études, nous nous sommes focalisés sur la reconnaissance de type bottom-up, c'est-à-dire la reconnaissance de plusieurs visages familiers parmi des inconnus. Cependant, il existe un autre type de reconnaissance, la reconnaissance d'une personne en particulier parmi des inconnus que l'on appellera la reconnaissance top-down. Ces deux types de reconnaissance peuvent être comparés aux situations écologiques de rencontrer par hasard une personne dans la rue (reconnaissance bottom-up), ou de rechercher activement une personne en particulier au milieu d'une foule (reconnaissance top-down). Dans le cas d'une reconnaissance bottom-up, on peut supposer que le sujet n'a pas construit d'attente précise sur le visage à reconnaître contrairement à la reconnaissance top-down qui devrait mettre en jeu des mécanismes de pré-activation de caractéristiques bas niveau de la cible à reconnaître disponibles déjà très tôt dans la voie visuelle ventrale, permettant peut-être de répondre plus rapidement. Nous voulons donc ici déterminer le coût temporel comportemental de la reconnaissance bottom-up par rapport à la reconnaissance top-down afin d'améliorer la compréhension des mécanismes sous-jacents.

Méthode : Afin de déterminer le temps de réaction minimum pour reconnaître un visage 15 sujets réalisent une procédure SAB à travers 6 conditions. En effet, les sujets doivent réaliser une reconnaissance de type bottom-up (visages célèbres parmi des inconnus), une reconnaissance de type top-down (70 photographies différentes d'une même célébrités parmi des visages inconnus appariés un à un) et une catégorisation superordonnée (visages humains vs visages inconnus). De plus, afin de tester l'effet d'inversion, les sujets réalisent également ces trois conditions mais lorsque les stimuli sont présentés à l'envers.

Résultats : Reconnaître un visage en condition bottom-up nécessite au minimum 470 ms contre 260 ms en condition top-down et 220 ms en catégorisation superordonnée. D'autres part, l'inversion des stimuli entraîne une dégradation des performances en reconnaissance bottom-up (les sujets n'arrivent pas à réaliser la tâche) alors que les tâches de catégorisation

et de reconnaissance top-down peuvent toujours être réalisées. Cependant, alors que l'inversion des stimuli n'a pas d'effet sur la catégorisation superordonnée, une diminution des performances (baisse du d' de 4 à 2,3 et augmentation des temps de réaction minimaux de 60 ms) est rapportée pour la reconnaissance top-down.

Discussion : Cette étude permet de déterminer le temps de réaction minimum de la reconnaissance top-down à 260 ms et nous permet de confirmer nos résultats obtenus dans les deux premiers articles sur la reconnaissance bottom-up à 380 ms et la catégorisation à 240 ms. La similarité des temps de réaction obtenus dans des tâches de catégorisation superordonnée et de reconnaissance top-down suggère que la reconnaissance top-down serait une forme de catégorisation. Cependant, l'effet d'inversion semble avoir une influence sur la reconnaissance top-down uniquement suggérant une perturbation de l'accès à une représentation peut-être de premier ordre (Maurer et al., 2002) tandis que la reconnaissance bottom-up, complètement perturbée par l'inversion pourrait impliquer des mécanismes additionnels à la catégorisation et à la reconnaissance top-down, notamment l'accès à une représentation holistique du visage. Pour tester ces hypothèses, il serait intéressant d'étudier la catégorisation superordonnée et la reconnaissance bottom-up au niveau cérébral.

✓ *Ainsi, dans ce chapitre, nous avons montré que la reconnaissance bottom-up d'un visage au niveau de la familiarité était de 360-390 ms au minimum, soit environ 150 ms de plus que dans des tâches de détection, catégorisation du genre ou reconnaissance top-down. Ceci suggère l'implication de mécanismes additionnels. Afin de mieux caractériser ces mécanismes, le chapitre suivant s'intéressera aux latences électrophysiologiques rapportées pendant des tâches de détection et de reconnaissance et tentera de lier comportement et électrophysiologie.*

Top-down versus bottom-up face recognition

Running title : Top-down versus bottom-up recognition

Besson G^{1,2}, Barragan-Jason G¹, Thorpe SJ¹, Puma S¹, Ceccaldi M^{3,4}, Barbeau EJ^{1,2}

¹Université de Toulouse; UPS; Centre de Recherche Cerveau et Cognition; France

²CNRS; CerCo; Toulouse, France

³APHM, Service de Neurologie et Neuropsychologie, CHU Timone, Marseille, France

⁴Institut de neurosciences des systèmes, Aix-Marseille Université, INSERM U 751, Marseille, France

Correspondence:

Gabriel Besson

Centre de recherche Cerveau & Cognition (CerCo)

CNRS CERCO UMR 5549

Pavillon Baudot, CHU Purpan, BP 25202

31052 Toulouse Cedex, France.

besson@cerco.ups-tlse.fr

Abstract

In daily life, two types of face recognition can be distinguished: unexpected encountering, for example meeting an acquaintance in the street without any previous clue that he/she would be there, or looking for someone in the crowd. We call these two types Bottom-Up and Top-Down Face Recognition respectively (BUFR and TDFR). It is expected that, unlike BUFR, TDFR relies on the preactivation of diagnostic facial clues that will speed up and enhance recognition. We surmise that identifying the speed gain between TDFR and BUFR will reveal whether these diagnostic facial clues are available early or late during visual processing. We also compared TDFR to face categorization (categorizing human among animal faces, CAT), as it has been repeatedly documented that this could be achieved very fast. All three tasks were performed with upright or inverted faces to assess the need to access face configuration.

To identify the minimum reaction time needed for TDFR, BUFR and CAT, we used a novel Speed and Accuracy Boosting procedure (SAB, Besson et al, in press) based on a response deadline at 600 ms. In the BUFR condition, subjects had to recognize faces of different famous faces among unknown ones. In the TDFR condition, subjects had to recognize various exemplars of faces of the same famous person (either Nicolas Sarkozy or Johnny Hallyday, a famous French rock singer). 18 subjects underwent all conditions. TDFR yielded a minimum reaction time at ~260 ms, much faster than BUFR (~380 ms) and close to CAT (~240 ms). Inverted faces did not affect CAT at all, severely impaired BUFR and in contrast partly impaired TDFR.

These results underlie the importance of distinguishing top-down from bottom-up face recognition, supply values for the speed of TDFR and BUFR and indicate that TDFR has to rely on the activation of diagnostic facial clues early in the visual ventral stream. The nature of these early facial clues and the reasons why BUFR can't rely on these early clues are discussed.

Keywords : Recognition memory; Minimal Reaction Time; Go/no-go; Faces; Speed

INTRODUCTION

Recognizing familiar faces rapidly and accurately is crucial for social interactions, for example to distinguish a friend from a foe. Traditionally, face recognition is understood as the recognition which occurs when someone unexpectedly meet an acquaintance (Figure 1A). In such case, no clue is available to help recognition, a type of recognition we call “Bottom-Up Face Recognition” (BUFR). Finding someone one is looking for (e.g. in a crowd) paradoxically would also be referred to as “face recognition” (Figure 1B). However in such case facial features diagnostic at the individual level can be preactivated by means of top-down mechanisms and probably speed up and enhance recognition. We call such recognition type “Top-Down Face Recognition” (TDFR). Face recognition thus appears to be an ambiguous expression, used indifferently for both TDFR and BUFR. To our knowledge, these two types of face recognition haven’t been contrasted explicitly, although many studies have relied on one or the other aspect of recognition. As noted, TDFR should be faster and more accurate than BUFR. Understanding how much faster could help revealing how early facial features diagnostic at the individual level are available during visual processing. Also, many people tend to think that face recognition is fast or should be fast. However is it BUFR or TDFR which is fast? And fast compared to what?

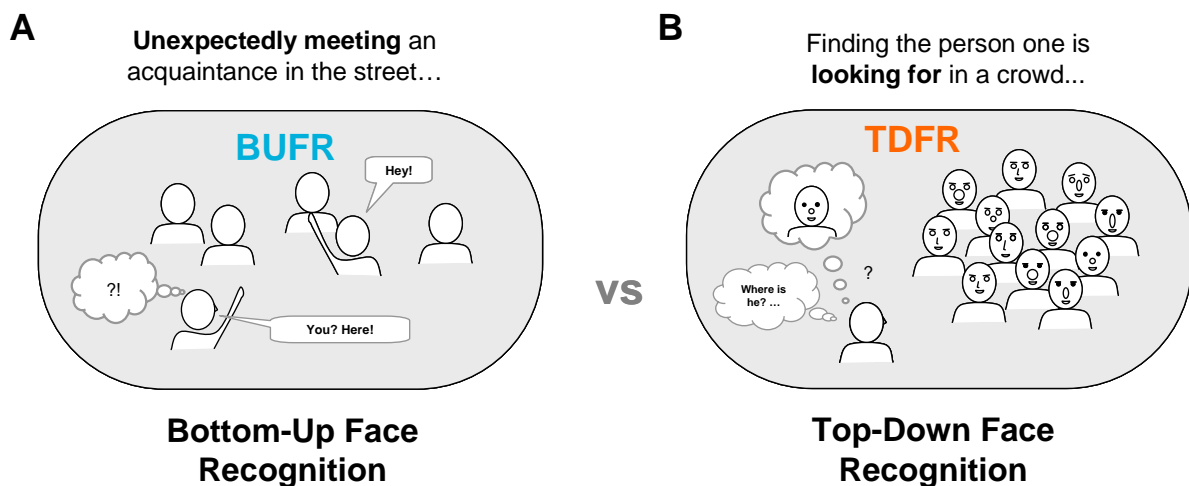


Figure 1.A) Bottom-up face recognition (BUFR) and B) top-down face recognition (TDFR) are two types of face recognition. Typically, BUFR occurs when one unexpectedly meet an acquaintance. In contrast, TDFR typically occurs when finding someone one is looking for (e.g. in a crowd).

BUFR corresponds to the cognitive processes evaluated when investigating recognition memory. Classically studied using old/new paradigms, recognition of a known face can either be based on familiarity only or on identification; familiarity being faster than identification following classic hierarchical models of face recognition (Bruce and Valentine, 1986). Moreover, it has been hypothesized that subjects may have difficulties preventing identification (Bruce & Young, 1986; Valentine, 2001), suggesting that generally, subjects

may be rather slow to recognize a face in the BUFR condition. This hypothesis was confirmed recently in a study where BUFR reaction times were compared to the time needed to categorize faces at the superordinate (human face vs animal faces) or gender level (Barragan-Jason et al., 2012). Participants needed a huge 200 ms more to recognize faces than to categorize them.

Contrary to BUFR, TDFR involves processes such as goal directed attention, for example to establish a feature-based attentional set, in order to guide face recognition. Attending selectively for relevant features or stimuli has been shown to be supported by top-down modulation that affects both the magnitude of neural activity and the speed of neural processing (Gazzaley et al., 2005).

TDFR has been largely used in the face literature, using 'category-verification task' (Tanaka, 2001), 'individual-level' verification task (Anaki et al., 2009) or simply 'face-identification task' (e.g. Reddy et al., 2006). Basically, a verbal label is proposed to the subject (e.g. "Brad Pitt") and subjects have to verify whether a face that is presented afterwards matches or not the label. Object can be identified at different level of abstraction (e.g. a picture of a dog can be categorized at the superordinate level as an "animal", at the basic level as a "dog" or at the subordinate-level as a "collie"). Categorizing an object at the basic level is faster than at the subordinate level (Rosch et al., 1976). Experts usually process subordinate levels for their domain of expertise faster than basic level (Gauthier & Tarr, 1997; Johnson & Mervis, 1997; Scott et al., 2006; Tanaka & Taylor, 1991). It has been suggested that all adults may be "face experts" (Carey 1992; Carey & Diamond 1977; Tanaka & Gauthier 1997), and it has been shown that categorizing a face at the subordinate level (e.g. as "Bill Clinton") is faster than at the basic level (e.g. as "human") (Tanaka, 2001; Anaki et al., 2009).

Overall, TDFR should therefore be faster than BUFR. However several problems prevent reaching such conclusion. Categorizing faster does not mean being processed first (Mack et al., 2009). Most of TDFR tasks have been run without any speed constraints. It is unclear if results showing that subordinate processing may be faster than basic level hold when speed constraints are used to study only the processes necessary to perform the task. For example, it has been shown that, contrary to studies which did not used speed constraints and reported only mean RTs (e.g. Rosch et al., 1976), categorization of objects at the superordinate level is in fact faster than at the basic level (Macé et al., 2009). This would support the idea of a hierarchy from fast RTs at the superordinate level to slower RTs at subordinate levels. Interestingly, instead of reporting mean RTs, this study reported minimum RTs, the first latency at which correct responses significantly outnumber incorrect ones (Rousselet et al., 2003, Macé et al., 2009). This problem of speed-constraint is even more critical in BUFR tasks since, as noted, subjects tend to rely on identification rather than on familiarity, thus be rather long overall. Again, most studies have been run without any speed constraints to our knowledge. A recent study by our group showed that using stringent speed constraints, subjects could perform a BUFR task in 360-390 ms (Barragan-Jason et al, submitted). Would TDFR be much faster than this or in reality close to such speed?

In sum, TDFR and BUFR have not been compared directly to our knowledge. Estimating the difference between these (under speed constraints) is critical to verify if they really differ. In particular, much could be inferred about the visual processes involved in both, depending on whether they are “relatively” close or “relatively” different. Our aim in this study was therefore to assess this difference.

We therefore set out to compare TDFR and BUFR. In order to constrain subjects to go fast, we used the Speed and Boosting Accuracy procedure (SAB), a novel paradigm constraining subjects to use a fast strategy (Besson et al in press, Barragan-Jason et al, submitted). It is important in such approach to be able to compare the speed of the BUFR and TDFR tasks with a reference speed in the same subjects. We chose here a face/non-face categorisation task (henceforth called CAT) since such tasks are known to be performed remarkably fast with high accuracy (reviewed in Fabre-Thorpe, 2011). In order to investigate whether these different tasks require access to face configuration (Maurer et al, 2002) or could be performed without, all were also run with inverted, single-trial, stimuli (Rousselet et al, 2003).

MATERIEL AND METHODS

Participants

18 participants (11 women) were included in this study (mean age: 24 [21–27], 2 left-handed). All participants had normal or corrected-to-normal vision.

Experiment

The experiment included 3 tasks: 1) a “bottom-up face recognition” (BUFR) task, 2) a “top-down face recognition” (TDFR) task and 3) a categorization (CAT) task. Each task was made of two blocks: one using upright stimuli, the other inverted stimuli. No stimulus was repeated. The order of the three tasks was pseudo-randomized and counter-balanced across participants, as well as was the order of upright and inverted conditions. Each task was run using the Speed and Accuracy Boosting procedure (SAB) with a response deadline (RDL) set at 600 ms from stimulus onset, a deadline inferred from a previous study (Besson et al., in press). Briefly, the SAB procedure, based on a classical Go/No-Go task, constrains participants to answer before a RDL (to boost speed), and provides an audio-feedback (positive or negative) at the response (to boost accuracy). If no Go-response is made before the RDL, the response is considered as a No-Go-response and, at the RDL, the audio-feedback is played. Before each item presentation, a fixation cross is displayed for a pseudo-random time between 300 and 600 ms. Items are presented, one by one, in the centre of a grey screen, for 100 ms (comprised in the RDL). The SAB is highly demanding and training is necessary. Each experiment was preceded by a training block (20 targets to be recognized among 20 distractors), which could be repeated if needed following participants’ request. A self-paced pause was proposed each 20 items.

Tasks

In the BUFR task, targets were famous faces of different persons (persons were never repeated) and distractors unknown faces. The task was made of one block of 140 targets mixed among 140 distractors. Participants were instructed to respond as quickly as they could to famous faces.

In the TDFR task, targets were different picture of the same famous person and had to be recognized among unknown faces (distractors). All unknown faces were paired one by one with each of the famous person (Figure 1), although all faces were presented randomly during the block. The task was made of two blocks of 70 targets mixed among 70 distractors. One block consisted in recognizing all faces of Nicolas Sarkozy (NS), the other all faces of Johnny Hallyday (JH), a famous French rock singer. Therefore, all subjects underwent two blocks of TDFR with a different target person each. This was carried out to appreciate whether similar RTs would be observed for these two famous faces. The order of the two blocks was pseudo-randomized and balanced across participants. Whereas one block was performed with upright faces, the other was performed with inverted faces.

In the CAT task, targets were human faces and distractors animals faces. The task was made of one block of 140 targets mixed among 140 distractors and participants were instructed to respond as quickly as they could to human faces.

Stimuli

Sample stimuli for each task are presented in Figure 2. Famous faces in the BUFR task were selected as the best recognized famous faces of a large database used in previous experiments with participants of the same age as in this study (Barragan-Jason et al., submitted). Distractors for famous faces were unknown faces randomly selected in the same previous experiment. Unknown faces were chosen so that they “looked like” they could be famous.

Pictures for the TDFR task were selected from the web. All original pictures were of correct quality, at least 200x200 pixels for the face area. We avoided as much as possible pictures with paraphernalia and with an objects hiding the face (e.g. hand, microphone, other person, etc.). In the target sets, all pictures were chosen relatively close to one prototype face (e.g. in the same period of life, colour of hair, etc.), but we avoided to choose pictures too similar (e.g. from a same photo shooting, etc.). The distractor sets were made by matching each target picture with an unknown face picture. Picture were matched on the types of clothes, type and colour of hair (and beard), colour of eyes, age, paraphernalia or hiding object if needed, head orientation, face expression (e.g. smiling, laughing, neutral, talking, etc.) as much as possible.

Unknown human faces and animals faces of the CAT task were randomly chosen from a previous similar experiment (Barragan-Jason et al, 2012).

All pictures for the three experiments were greyscale. Each picture was framed manually around the faces following the same procedure (i.e. a rectangle delimited at the bottom by the chin, on the side by the point between the face itself and the ear or, when the ear was

masked in three-quarter profiles, the farthest point between the eyebrow arch and the cheekbone, and at the top by the midpoint of the natural front hairline). Pictures could then be resized so that each face had the same size and could be cropped all identically around the face, using a home-made script on Matlab. Thus, all pictures were similar close-up greyscale pictures of centred faces of the same size (208x279 pixels, visual angle: $\sim 4.7 \times 6.3$ degrees). Lastly, all stimuli were equalized to the same luminance and contrast (computed as the standard error of pixels luminance) across conditions.

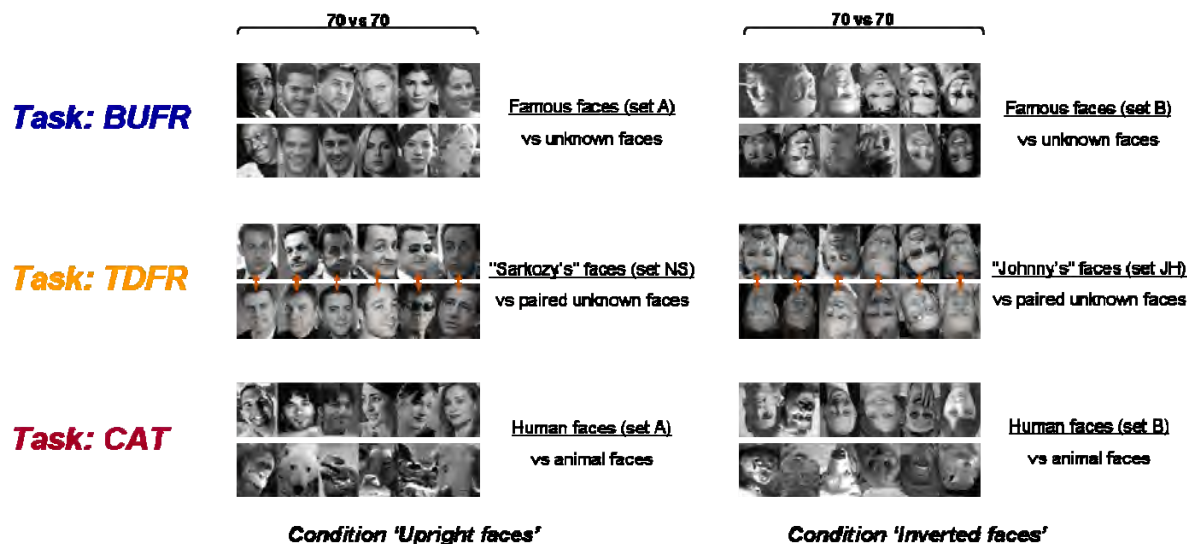


Figure 2. Figure 2. Samples of the stimuli used in each of the three tasks. Each task consists of recognizing 70 upright target faces among 70 upright distractor faces in a block ('Upright' condition) and 70 inverted target faces among 70 inverted distractor faces in another block ('Inverted' condition). In the BUFR task, targets are famous faces, distractors unknown faces; in the TDFR task, target are pictures of a famous person (either the ex-president of France, Nicolas Sarkozy (NS), or the famous French singer Johnny Hallyday (JH)), distractors paired unknown faces; in the CAT task, targets are human faces, distractors animal faces.

Set-up

Participants sat in a dimly lit room, at 90 cm from a computer screen piloted by a PC. Image presentation and behavioural responses recordings were carried out using the E-prime v2. Participants responded to the stimuli by raising their fingers from an infrared response pad.

Minimal reaction times

To obtain an estimation of the minimal processing time required to recognize targets, the minimal behavioral reaction time (minRT) was computed by determining the latency at which correct go-responses (hits) started to significantly outnumber incorrect go-responses (false-alarms) (Rousselet et al., 2003). For each task, analyses were performed either across

trials (by pooling together all trials from all participants for a given condition) and across participants. Across trials analyses have been used in previous studies (Rousselet et al., 2003; Barragan-Jason et al., 2012; Besson et al., in press) and are like building a “meta-participant”, reflecting the performance over all the population. MinRTs across trials were computed using 10 ms time bins and determined as the middle of the first bin that was significant, χ^2 -test, $p < .05$, followed by at least three significant consecutive bins. Across participants, in order to accommodate for the lower statistical power than across trials data since there were fewer trials, we used 30 ms time bins and a Fisher’s exact test ($p < .05$), followed by at least two significant consecutive bins. A minRT can’t be computed if the distribution of hits and false alarms do not allow satisfying the statistical conditions described. Thus some participants don’t have a minRT, in particular when d' are low.

Statistical analysis

Performance (accuracy) and bias were computed using d' and C based on the SDT (corrected according to Snodgrass and Corwin, 1988). As some participants failed to do the BUFR task in the upright condition (χ^2 -test between hits and false alarms (FA), $p < .05$), their data were discarded from further analyses (see Table 2 for details). Statistical comparison were computed using ANOVA for group test and paired t-test for post-hoc analysis when parametrical conditions applied (normality checked with Lilliefors test, $p < .05$; variance equality tested with Brown & Forsythe test, $p < .05$). Friedman and Wilcoxon non-parametrical tests were used when these conditions did not apply. In this study, accuracy (d') and bias (C) showed a Gaussian distribution, but minRTs did.

RESULTS

TDFR task

Results of TDFR tasks are presented in Table 1. There wasn’t any difference between the two targets (Nicolas Sarkozy, TDFR-NS or Johnny Hallyday, TDFR-JH) on any variable assessed (accuracy, bias or minRT). Performance was very good as assessed by the d' .

Table 1- performance on the two TDFR task. *N* = number of participants. *SD* = standard deviation.

		<i>Upright TDFR-NS</i>	<i>Inverted TDFR-NS</i>	<i>Upright TDFR-JH</i>	<i>Inverted TDFR-JH</i>
<i>Succeeded to do the task/underwent the task</i>	<i>N</i>	5/5	10/10	10/10	5/5
<i>Kept for analysis</i>	<i>N</i>	5	10	10	5
Accuracy (d')	<i>Mean</i>	4.04	2.34	3.96	2.11
	<i>SD</i>	0.96	0.23	0.26	0.50
	<i>range</i>	[2.64; 4.91]	[2.04; 2.75]	[3.56; 4.49]	[1.75; 2.98]
	<i>Across trials</i>	3.82	2.34	4.13	2.05
Bias (C)	<i>Median</i>	-0.00	-0.12	-0.23	-0.22
	<i>1st and 3rd quartiles</i>	[-0.26; -0.00]	[-0.35; -0.05]	[-0.40; -0.11]	[-0.28; 0.11]
	<i>range</i>	[-0.40; -0.00]	[-0.42; 0.10]	[-0.46; 0.11]	[-0.41; 0.16]
	<i>Across trials</i>	-0.17	-0.18	-0.29	-0.16
<i>Obtained a minRT</i>	<i>N</i>	5/5	10/10	10/10	5/5
Minimal RTs	<i>Median</i>	300	345	300	360
	<i>1st and 3rd quartiles</i>	[270; 330]	[330; 360]	[270; 300]	[330; 360]
	<i>range</i>	[270; 330]	[300; 390]	[270; 330]	[330; 360]
	<i>Across trials</i>	310	300	260	330

Comparison of the task: BUFR vs TDFR vs CAT for upright faces

As no effect of the target was observed in TDFR tasks, we pooled trials of TDFR-NS and TDFR-JH tasks for each participant in the following analyses. Detailed results of BUFR, TDFR and CAT tasks are reported in Table 2. Figure 2 shows results across trials and Figure 3 across subjects. Post-hoc statistical analyses are reported in Figure 3.

Table 2: performance on the three tasks for upright and inverted faces.

		<i>Upright BUFR task</i>	<i>Inverted BUFR task</i>	<i>Upright TDFR task</i>	<i>Inverted TDFR task</i>	<i>Upright CAT task</i>	<i>Inverted CAT task</i>
<i>Succeeded / underwent the task</i>	<i>N</i>	15/18 (83.33%)	3/18 (16.67%)	18/18	18/18	18/18	18/18
<i>Kept for analysis</i>	<i>N</i>	15	15	15	15	15	15
Accuracy (d')	<i>Mean</i>	1.13	0.17	3.99	2.26	4.18	3.90
	<i>SD</i>	0.48	0.27	0.55	0.34	0.52	0.47
	<i>range</i>	[0.50; 2.16]	[-0.21; 0.61]	[2.64; 4.91]	[1.75; 2.98]	[2.78; 4.91]	[3.05; 4.91]
	<i>Across trials</i>	1.08	0.16	4.03	2.24	4.20	3.99
Bias (C)	<i>Median</i>	0.21	0.27	-0.21	-0.16	-0.19	-0.11
	<i>1st and 3rd quartiles</i>	[0.02; 0.42]	[-0.18; 0.69]	[-0.40; -0.00]	[-0.33; -0.01]	[-0.30; -0.00]	[-0.28; -0.00]
	<i>range</i>	[-0.10; 1.13]	[-0.62; 2.46]	[-0.46; 0.11]	[-0.42; 0.16]	[-0.52; 0.32]	[-0.40; 0.00]
	<i>Across trials</i>	0.25	0.27	-0.25	-0.17	-0.13	-0.16
<i>Obtained a minRT</i>	<i>N</i>	11/15 (73.33%)	4/15 (26.66%)	15/15	15/15	15/15	15/15
Minimal RTs	<i>Median</i>	420	510	300	360	270	270
	<i>1st and 3rd quartiles</i>	[390; 442.5]	[465; 555]	[270; 322.5]	[330; 360]	[270; 270]	[270; 300]
	<i>range</i>	[330; 540]	[450; 570]	[270; 330]	[300; 390]	[240; 300]	[240; 330]
	<i>Across trials</i>	380	520	260	300	240	250

BUFR was more difficult than both TDFR and CAT and was the only task not succeeded by every participant. In contrast, TDFR and CAT were performed very well, with no difference between them. A conservative bias in BUFR was observed, while both TDFR and CAT showed an identical liberal bias. A minRT could not be calculated for some participants in the BUFR task only. Again, a clear difference between BUFR and TDFR or CAT on minRTs was observed, with participants being 120 ms slower in BUFR than in TDFR. Whereas, as noted, participants showed similar accuracy and bias in TDFR and CAT tasks, they were however somewhat slower in TDFR than in CAT, with a mean significant difference of 30 ms across subjects and 20 ms across trials.

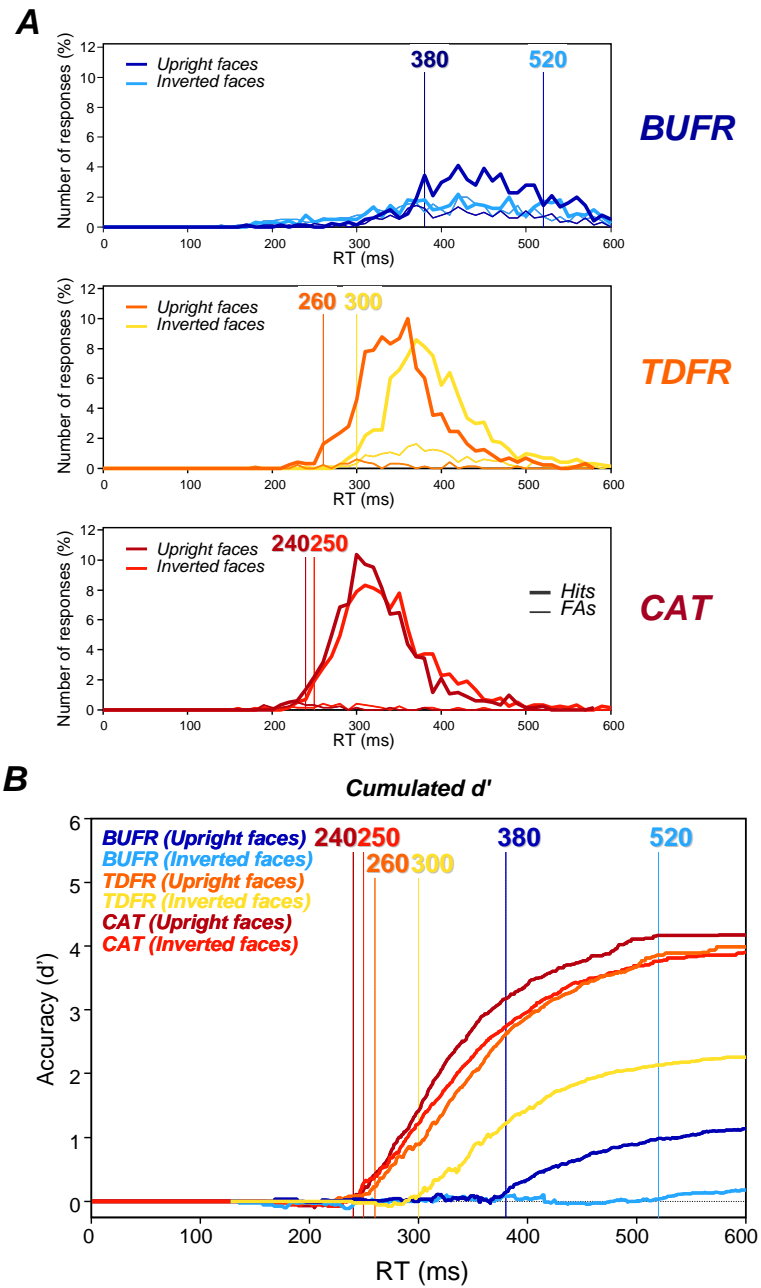


Figure 2. A) Performances obtained on each of the three task. Across trials distribution of RTs obtained in each task in the upright or inverted condition. Vertical bars and numbers correspond to across trials minRTs obtained in each task. B) Cumulated d' for the three tasks.

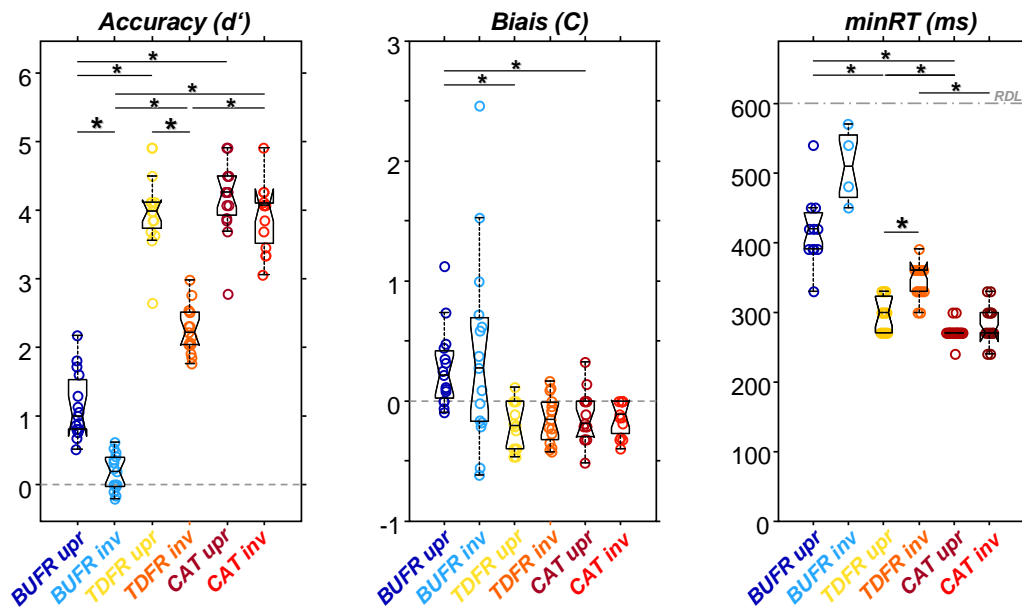


Figure 3: Comparison of performances, bias to respond and minRTs computed 'across participant' in each task. *: statistical comparison reached significance ($p < .05$, corrected with Bonferroni for multiple comparisons)

Inverted vs upright faces

Inverted faces were processed strikingly differently according to task (Table 2, Figures 2 and 3). Inverted faces could not be recognized in BUFR ($d' = 0.17$). In strong contrast, inverted faces did not change results in the CAT condition (d' , bias and minRT did not show any statistical difference). In TDFR however, performance remained correct but decreased sharply (from 3.99 to 2.26) and minRT slowed down by 60 ms.

Effect of TDR preparation on the first target

RTs of each participant were ordered by the rank of presentation of targets, with no RT if no go-response was made on a particular target by a given participant (omission). Here, we focus on the RTs of the first target (targets were different faces across subjects as targets were presented randomly for each subject) (Figure 4). Importantly, the large difference observed between BUFR and TDFR was apparent right from the beginning, i.e. for the first target (81 ms). Similarly, TDFR was close to CAT even for the first target (22 ms).

Effect of priming

We then focused on the RTs of the first target compared to the rest of the RTs in the same

task to investigate a possible role of priming. Statistical differences between each pair of rank of targets presented were computed. No priming effect was observed (Figure 4).

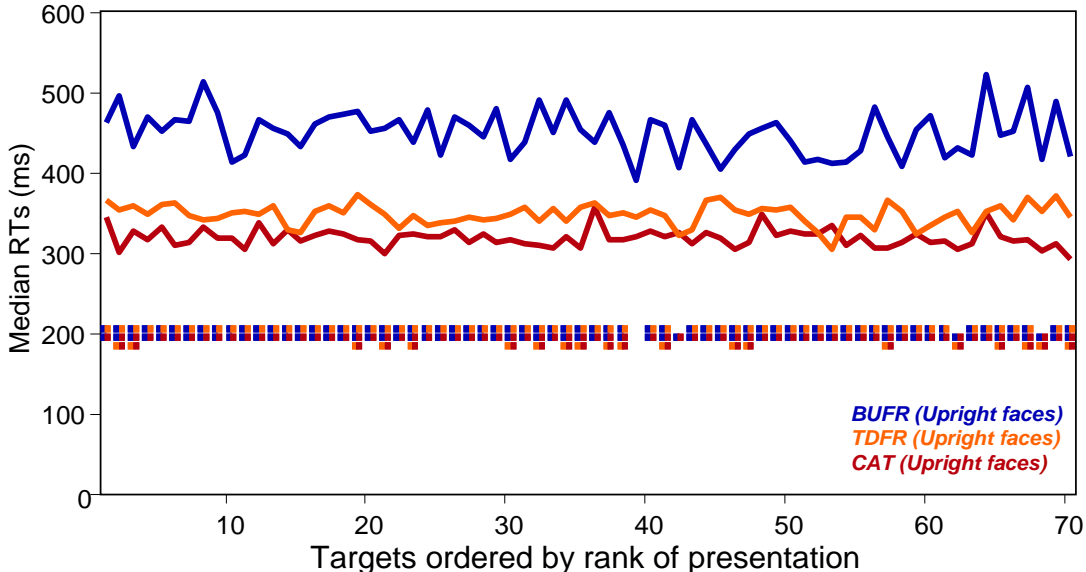


Figure 4. Comparison of RTs obtained on targets ordered by rank of presentation. Coloured rectangles represent significance between conditions. BUFR and TDFR tasks. Coloured rectangles under median RTs represent significance reached between CAT and TDFR tasks.

In a complementary analysis, we focused on the first go-responses (rather than on the first target as in the previous analysis). Each go-response, given either on a target or a distractor, was thus correct (hit) or incorrect (false-alarm). Therefore, this allowed us to compute minRT (Figure 5). Results were largely convergent with the previous analysis. No error was made on either the TDFR or CAT, which was not the case for the BUFR or the inverted conditions.

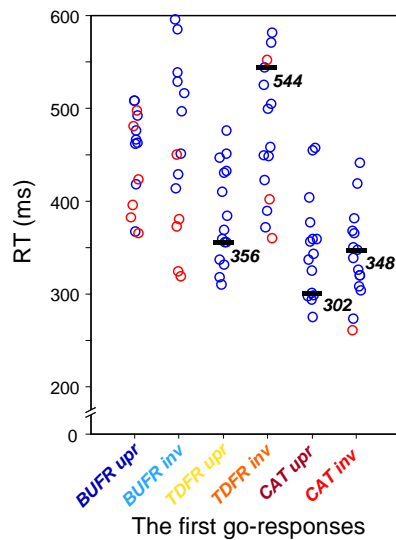


Figure 5. RTs obtained for the first go-responses, in each of task. Each dot represents an RT obtained by a participant either on a target (i.e. hit, in blue) or on a distractor (i.e. false alarm, in red). Horizontal lines indicate the minRT. No minRT could be computed for BUFR because of too many false-alarms.

DISCUSSION

The main goal of this study was to investigate the speed of face recognition, with an emphasis on the difference between two types of face recognition, namely bottom-up and top-down recognition. Accuracy was very high in the TDFR but low in the BUFR. The bias was different between tasks suggesting that subjects used different strategies. Importantly, BUFR was delayed compared to the TDFR by 120 ms. Such delay or enhanced accuracy appears to hold even for the very first answer (Figures 3 and 4). Overall, this pattern of results strongly suggests that TDFR relies on different mechanisms than BUFR.

TDFR is thus faster than BUFR. How fast is it however? A comparison with a face categorization task (CAT) widely regarded as a “fast” task (Fabre-Thorpe et al, 2010) revealed that there was only a 30 ms difference between CAT and TDFR. This result also holds for the first answer (22 or 54 ms difference depending on the analysis). Accuracy for the two tasks was very high and did not differ. The bias was similar. This indicates that TDFR is roughly as fast as other “fast” tasks.

This would seem to suggest that TDFR and CAT rely on very similar mechanisms (same accuracy, same bias, small minRT delay). However, the investigation of inverted face processing using each of the three tasks revealed striking differences between tasks. BUFR

could not be performed if inverted faces were used. In contrast, performance or minRT of the CAT was not modified by face inversion, in agreement with a previous study (Rousselet et al, 2003). Inverted faces processed in the TDFR revealed yet another pattern, as accuracy decreased sharply (yet remained at a fair level) and minRTs were delayed by about 60 ms. Such pattern indicates that TDFR is optimal for upright faces and that the identification of facial features diagnostic at the individual level is partly disrupted by inversion.. Therefore, TDFR and CAT do not rely on exactly the same mechanisms.

Several levels of face configuration have been distinguished, such as first-order configuration (eyes above nose above mouth), holistic configuration (gestalt processing making it harder to process individual features) and second-order configuration (metric distance between features) (Maurer et al, 2002). CAT does not require access to face configuration as indicated by preserved performance and similar RTs in the inverted condition. In contrast, TDFR does. Given that TDFR is temporally close to CAT, access to a simple first order configuration could be required for TDFR. This would make sense if TDFR relies on the identification of individual features since those have to be looked for somewhere in the face. Inversion would partly prevent subjects to look efficiently for these features. With this respect our data suggest that accessing such first-order face configuration would take no more than 20-30 ms compared to CAT. BUFR on the hand, as it is much delayed compared to TDFR (and even more compared to CAT) and because face inversion impaired completely performance in contrast to CAT, probably requires accessing a more complex form of configuration, i.e. holistic representation at the individual level. It is this sort of representation that has been proposed to be selectively impaired in prosopagnosic patients (Busigny et al, 2010).

Our results are in agreement with the idea that TDFR and BUFR rely on widely different brain mechanisms. Given the long RTs observed in BUFR, it probably relies on the activation of the whole visual ventral pathway as developed elsewhere (Barbeau et al, 2008; Barragan-Jason et al, 2012). In contrast, TDFR could rely on the activation of posterior visual areas through top-down mechanisms. These areas could be close to those involved in fast categorization tasks such as CAT. What could be the local mechanisms enabling such high accuracy and efficiency? There could be “multi-purpose” face neurons activated in this area through top-down mechanisms. These neurons could be tuned for the target (e.g. to some features characteristic of the face one is looking for). Ohayon et al (2012) recorded face-selective neurons in IT in the monkey and found that half of these neurons showed a preference for

contrast polarity as seen in faces (e.g. between eye and nose for example). Interestingly, these authors found out that these cells were also tuned to the “geometry of facial features, suggesting cells encode information useful both for detection and recognition”. Hence, these cells could be involved both in face perception and face recognition and probably be tuned following top-down mechanisms.

It is interesting to note that Reddy et al, (2006) investigated a face identification task – consisting in recognizing a particular famous face, e.g. Brad Pitt – in a double-task paradigm. Subjects had to pay attention to an attentional demanding central task, detecting a T among Ls, while trying to detect whether a face presented in the periphery (outside of the focus of attention) was the target or not. The finding was that, even in the near-absence of attention, such a face identification task was still possible. Previous work on such paradigm showed that superordinate-level task (i.e. animal vs non-animal task) is possible in the near absence of attention, while other tasks, that appear rather simple, are not (e.g. orientation detection of bi-coloured disks; VanRullen et al., 2004). It has therefore been proposed that the peripheral task can be achieved only if neuronal populations selective to the target category of stimulus exist (VanRullen et al., 2004). Hence, there may exist neuronal populations for individual faces that could be preactivated rather low in the visual ventral stream. A feed-forward stream of activity would suffice to activate these preactivated neurons rapidly, hence could account for the fact that RTs for TDFR are close to those found in CAT. Tasks similar to CAT are indeed thought to rely on such feed-forward sweep. An alternative could be that there exists neurons specific to single faces low in visual areas, rather than face neurons that are activated following top-down control. However, this seems unlikely since in this case it would be difficult to understand why BUFR is so long.

It would interesting to clarify what are the diagnostic features at the individual level. Although it is tempting to think that it could be a simple explicit feature such as a mole on the face, it is likely that these features are more abstract (e.g. broad shape, texture, color, of skin, etc.) and that it is actually a combination of these that are implicitly preactivated.

A possibility could be that the differences we observe between TDFR and BUFR is related to our stimuli, which were highly different in BUFR compared to TDFR. Accuracy was rather low for BUFR and one could make the hypothesis that with well-known stimuli, accuracy would increase while RTs would decrease. This appears unlikely however. Low accuracy is expected since BUFR for famous faces assesses knowledge about famous people. From our

experience, not two people, even of the same age and from the same background has the same knowledge. Furthermore, previous analyses have shown that there was no trade-off between minRTs and accuracy (Barragan-Jason et al, 2012). In addition, extensive analyses have shown that the fastest RTs in BUFR never was below 360 ms (across 132 subjects, Barragan-Jason et al, submitted).

In conclusion, these results underlie the importance of distinguishing top-down from bottom-up face recognition and supply values for the speed of TDFR and BUFR. Our results also indicate that TDFR relies on the activation of diagnostic facial clues early in visual areas.

REFERENCES

- Anaki, D., & Bentin, S. (2009). Familiarity effects on categorization levels of faces and objects. *Cognition* 111, 144–149.
- Barragan-Jason, G., Lachat, F., and Barbeau, E. J. (2012). How fast is famous face recognition? *Front. Psychology* 3, 454.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology* 77, 305–327.
- Carey, S. (1992). Becoming a face expert. *Philos Trans R Soc Lond B Biol Sci* 335(1273):95-102; discussion 102-3.
- Carey, S., Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science* 195(4275):312-4.
- Chanceaux, M., Vitu, F., Bendahman, L., Thorpe, S., Grainger, J. (2012). Word processing speed in peripheral vision measured with a saccadic choice task. *Vision Research*, 56:10-9.
- Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision* 10, 1–17.
- Dicarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434.
- Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J Cogn Neurosci* 17(3):507-17.
- Gauthier, I., Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res* 37(12):1673-82.
- Johnson, K.E., Mervis, C.B. (1997). Effects of varying levels of expertise on the basic level of categorization. *J Exp Psychol Gen* 126(3):248-77
- Kalaska, J., and Crammond, D. (1992). Cerebral cortical mechanisms of reaching movements. *Science* 255, 1517 –1523.
- Li, F. F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 9596–9601.
- Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., and Fabre-Thorpe, M. (2009). The time-course of visual categorizations: you spot the animal faster than the bird. *PLoS one* 4, e5927.

Mack, M. L., Wong, A. C., Gauthier, I., Tanaka, J. W., Palmeri, T. J. (2009). Time course of visual object categorization: fastest does not necessarily mean first. *Vision Res* 49(15):1961-8.

Maurer, D., Le Grand, R.L., Mondloch, C.J. The many faces of configural processing. (2002). *Trends Cogn Sci* 6(6):255-60.

Ohayon, S., Freiwald, W.A., Tsao, D.Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron* 74(3):567-81.

Poncet, M., Reddy, L., and Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision* 12(1):15, 1–16.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.

Reddy, L., Reddy, L., Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Res* 46(15):2336-43.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39(4):1959-79.

Rousselet, G. A., Macé, M. J.-M., and Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of vision* 3, 440–55.

Scott, L.S., Tanaka J.W., Sheinberg D.L., and Curran T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *J Cogn Neurosci* 18(9):1453-65.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J Exp Psychol Gen* 117, 34–50.

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of experimental psychology General* 130, 534–543.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive psychology* 23(3), 457-482.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. *Psychology of learning and motivation* 36, 83-125.

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

Valentine, T. (2001). Face-space models of face recognition. In: Wenger M.J. and Townsend J.T. (eds.) Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Mahwah: LEA, 83–113.

VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci* 13, 454–461.

VanRullen, R., Reddy, L., and Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *J Cogn Neurosci* 16, 4–14.

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.

4 Chapitre 4. Quelles sont les latences électrophysiologiques de la catégorisation rapide des visages célèbres ?

Une seconde étape a consisté à déterminer la dynamique du signal électrophysiologique dans ces mêmes tâches de détection et de reconnaissance des visages familiers.

Le traitement des visages a été largement investigué via l'utilisation de la méthode des potentiels évoqués. Cependant, les données de la littérature sur la dynamique ne semblent pas converger vers un réel consensus sur les latences de la reconnaissance des visages familiers laissant supposer que les latences électrophysiologiques les plus rapides se situent entre 140 (N170) et 300 ms (N250). De plus, les électrodes d'intérêt pour étudier la familiarité ne semblent pas être très bien caractérisées.

D'autre part, l'utilisation d'images naturelles pour étudier le traitement des visages semblent poser quelques problèmes méthodologiques, certains effets classiquement rapportés ne pouvant pas être répliqués (Rousselet et al., 2004).

Nous proposons donc aborder la question du traitement des visages en appliquant la technique de MVPA (multivariate pattern analysis) sur nos données électrophysiologiques. Nous pourrions ainsi déterminer la latence minimale électrophysiologique en utilisant le signal enregistré sur l'ensemble du scalp (sans a priori sur les électrodes). Cette méthode nous permettra de confronter nos latences électrophysiologiques obtenues lorsque le sujet est contraint à répondre rapidement avec les latences rapportées dans la littérature.

Pour cela, nous avons développé deux études, une première étude s'intéressant à la dynamique électrophysiologique de la catégorisation rapide de visages inconnus parmi des animaux, la seconde s'intéressant plus particulièrement à la reconnaissance des visages célèbres.

**ARTICLE 4 : The neural dynamics of human face detection
in the wild**

*Maxime Cauchoix *, Gladys Barragan-Jason*, Thomas Serre, Emmanuel Barbeau*

Article en préparation

Article 4: The neural dynamics of human face detection in the wild (en préparation)

Introduction : Comment détectons-nous un visage humain dans la nature ? Quelle est la dynamique de l'information permettant de réaliser cette tâche cognitive ? La dynamique neurale du traitement des visages a été largement étudiée en analyse en potentiels évoqués du signal EEG de surface en utilisant des stimuli bien contrôlés et homogènes. Cependant, certains effets spécifiques aux visages ne semblent pas être retrouvés lorsque les visages sont présentés dans des scènes naturelles (Rousselet et al., 2003) ou lorsque le set de stimuli présente une forte variabilité (Thierry et al., 2007). Ce type de résultats soulève la question de la validité de ces observations d'un point de vue écologique, c'est-à-dire si les résultats que l'on obtient en laboratoire dans des conditions très contrôlées peuvent être transférés à la perception des visages dans la nature. Nous proposons donc d'étudier la question de la détection des visages dans des scènes naturelles à l'aide d'une nouvelle méthode d'analyse permettant d'analyser et de quantifier l'information disponible essai par essai sans a priori sur les électrodes à considérer.

Méthode : Nous avons réalisé une analyse en MVPA sur l'ensemble du signal EEG (32 électrodes) chez 28 sujets qui réalisaient une tâche de go/no-go de visages humains parmi des visages animaux présentés dans des scènes naturelles de forte variabilité.

Résultats : Nos résultats montrent que l'information relative à la catégorie visage est décodée dès 80 ms après la présentation du stimulus. Plus particulièrement, nos résultats révèlent l'existence de trois phases dans le décodage de l'information : Une première phase débutant dès 80 ms, une phase de plateau entre 140-200 ms et une dernière phase au cours de laquelle la précision de l'information du classifieur augmente de nouveau. De plus, une analyse par quartile étudiant l'évolution du décodage pour les essais appartenant au premier, second, troisième et dernier quartile montre une modulation significative du signal après 160 ms environ. Ainsi, un lien avec le comportement est observé 80 ms plus tardivement que le début du décodage.

Discussion : Les visages présentés dans des scènes naturelles seraient donc traités selon deux phases successives, l'une traitant les informations bas niveau en accord avec la latence de la composante évoquée P100 et l'autre liée à la décision de catégorisation en accord avec la latence de la N170. Le lien avec le comportement rapporté seulement après 160 ms suggère la nécessité d'accéder à une représentation peut-être holistique du visage indexée par la N170. Ces travaux sont également en accord avec les travaux portant sur le traitement visuel des objets suggérant qu'une représentation haut niveau de la catégorie serait disponible après 150 ms. Ainsi, notre étude propose une nouvelle méthode permettant d'étudier le traitement des visages présentant une très forte variabilité (stimuli naturels), réconciliant ainsi les données de la littérature.

✓ *Nous avons donc étudié la dynamique de la détection des visages présentés dans un contexte naturel en revisitant la question à l'aide d'une analyse en MVPA. Nous avons montré qu'il était possible de décomposer le traitement des visages en deux étapes successives qui correspondent aux composantes évoquées rapportées dans la littérature que sont la P100 et la N170. De la même manière, afin d'investiguer l'évolution du traitement de l'information après la détection du visage, il nous a semblé intéressant d'appliquer la même analyse lors de la tâche de reconnaissance de visages familiers. La comparaison des latences obtenues dans la tâche de détection et de reconnaissance nous apportera des informations précieuses sur les mécanismes impliqués dans ces deux tâches.*

The neural dynamics of face detection in the wild

Running title : Face detection in the wild

Cauchoix M.*¹, Barragan-Jason G.*¹, Serre T.^{#2}, Barbeau E.J.^{#1}

¹ Centre de recherche Cerveau et Cognition, Université de Toulouse, CNRS-UMR 5549, Toulouse, France

² Cognitive Linguistic & Psychological Sciences Department, Institute for Brain Sciences, Brown University, Providence, RI, USA

*,# equal contributions

Correspondence:

Maxime Cauchoix

Centre de recherche Cerveau & Cognition (CerCo)

CNRS CERCO UMR 5549

Pavillon Baudot, CHU Purpan, BP 25202

31052 Toulouse Cedex, France.

mcauchoixx@gmail.com

Abstract

Face detection is thought to be fast. Given that most studies of face processing are carried out using cropped faces with low variability, the aim of the present study was to assess how early neural activity could be selective to faces presented “in the wild”, that is, in natural scenes. Here, we recorded EEG activity while participants performed a rapid go/no-go categorization task in which they had to detect the presence or absence of human vs. animal faces in complex and highly variable natural scenes. We performed a multivariate pattern analysis (MPVA) to assess the amount of task-related information across all electrode channels over time. This analysis, revealed that face category information could be readout very early, around ~80 ms post-stimulus onset. Such latency is remarkably fast, despite the fact that our stimuli were highly complex and variable. This leaves the visual cortex only 50 ms to detect a face in a scene given that information flows in V1 around 30 ms at the earliest. However, a relation with behavior (reaction times) could only be observed during a later time window starting ~160 ms post stimulus onset. This activity could reflect high-level category processes independent of low-level visual properties of the sort necessary to drive behavioral go/no-go responses. Overall thus, face detection relies on two consecutive stages, each with its own properties.

Introduction

Let's imagine you are hiking in the mountain. How much time would you need to detect a face in the scene you have before your eyes? It is likely that you would need very little time if you consider that faces have probably been for our ancestors crucial visual cues that signaled either a danger or an opportunity. With this respect, it has been shown that human observers can recognize faces presented in complex natural scenes very rapidly (~260-290 ms post-stimulus) and with very high accuracy (~95%) despite large changes in appearance due to scale, position, viewpoint and clutter (Fabre-Thorpe, 2011 for a recent review). Furthermore, human participants have been shown to produce reliable saccades towards faces presented in natural scenes as early as 100-110 ms post-stimulus onset (Crouzet et al., 2010), faster than towards other object categories. Such results highlight the formidable robustness and efficacy of the primate visual system to detect faces in natural scenes.

Given that the earliest saccades are made towards faces around 100 ms and that about 20 ms are needed to initiate a saccade (Crouzet et al., 2010), the hypothesis can be made that neural activity specific to faces should start around 80 ms post-stimulus. In agreement, event related potential (ERP) studies have sometimes reported an ERP response over occipito-temporal sites that is selectively modulated by faces 80-120 ms post stimulus onset (Liu et al., 2002; Halgren et al., 2000; Eimer & Holmes, 2002; Itier et al, 2004; Rossion & Caharel, 2011). However, a subsequent N170 component occurring later around 140-200 ms post stimulus has been identified as a more reliable correlate of face perception (Bentin et al. 1996, Jeffreys, 1989; Rossion et al., 1999). The problem is that these studies have relied on the use of cropped face with limited stimulus variability. The visual system doesn't have in this case to "detect" a face since it is the only stimulus presented. Hence, the early face selective activity found around 80-120 ms could be artificial. Whether largely variable faces presented in complex natural scenes would elicit a distinct pattern of neural activity at such early latencies remains to be investigated (Rousselet et al., 2004).

One major limitation inherent to classical (univariate) ERP analysis is that electrodes need to be selected before averaging. This has ignited a vigorous debate concerning the link between face-related components and stimulus variability, and in this context about which electrode should be chosen for analysis (Thierry et al, 2007a; Bentin et al. 2007; Rossion and Jacques, 2008).

The aim of the present study was to assess how early neural activity could be selective to faces "in the wild", that is, in natural scenes. To be in the most difficult condition, distractors were faces of animals in natural scenes (Rousselet et al., 2004). To overcome the limitation of electrode(s) choice, here all the more critical that the simultaneous presentation of scenes could modify the classic topography of the neural activity evoked by faces, we used whole brain multivariate pattern analysis technique to investigate the dynamics of face processing. This decoding method provides the double advantage of getting rid of most controversial issue associated with ERP analyses (Rousselet and Pernet, 2011) and at the same time, to obtain a single temporal quantification of face processing resulting from all electrodes at the millisecond precision. An early face specific activity should be correlated to

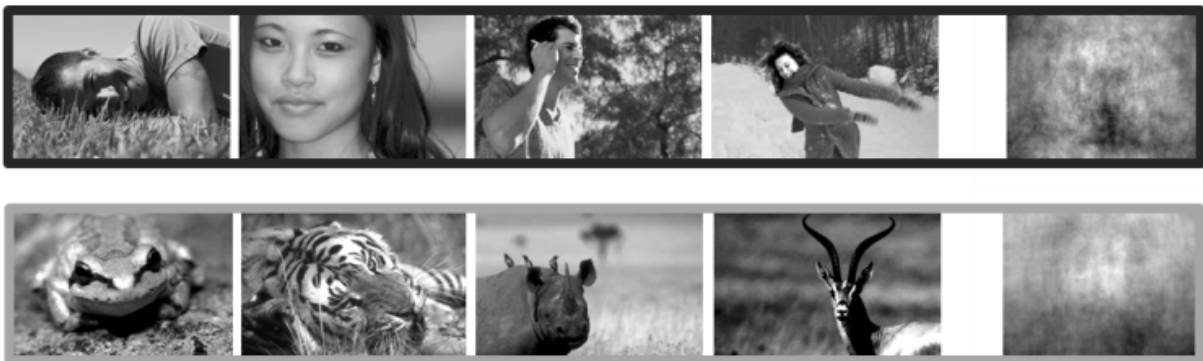
behavior if it is to be useful (e.g. to escape from a detected foe for example). We therefore investigated correlation with reaction times (RTs). We made the hypothesis, in accordance with the literature reviewed above, that the earliest task-related face information should be readout as early as 80 ms post stimulus onset.

Methods

Participants

Fifteen females and thirteen males (n=28, mean age: 25.4 yrs, std=3.8 yrs, 25 right-handed) with normal or corrected-to-normal vision signed informed consent to participate in the experiment.

A



B

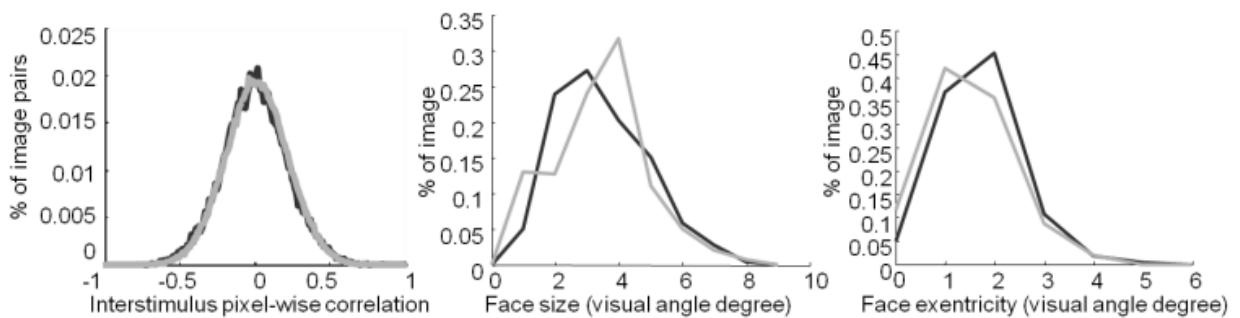


Figure 1: Stimulus statistics. A) Representative stimuli used in the study and pixel-based sample mean average over the entire set of stimuli. B) Distribution of inter-stimulus perceptual variance (ISPV) as defined by Thierry et al. (2007a), size and eccentricity for targets (black) and distractors (gray).

Stimulus set

Target images consisted in grayscale photographs of human faces (270 images) presented in their natural contexts (i.e., images included some background clutter and no face was pasted). We selected face exemplars that exhibited significant intra-class variations (such as view-point, gender and race, Figure 1 and Table S1 as well as eccentricity and size, see Figure 1A top row). Distractor stimuli consisted in (non-human) animal faces (270 images), which included different species (mammals, birds, reptiles, etc). The stimulus set was previously

used in Rousselet et al., (2004), see Figure 1A bottom row for examples. Images were normalized to be 320 x 480 pixels in size. The global luminance and RMS contrast was similar between the two groups.

To characterize this intra-class variability, we computed a histogram of inter-stimulus perceptual variance (ISPV) as defined by Thierry et al. (2007a) as well as size and eccentricity for targets (black) vs. distractors (gray), see Figure 1B. This ISPV measure is estimated using pixel-wise correlation and provides a quantitative measure of similarity within the set of targets and distractor images. A low pixel-wise correlation (close to 0) between a pair of stimuli corresponds to a high ISVP while a high pixel-wise correlation (close to 1) corresponds to a low ISVP. According to this measure, both target and distractor images present a comparable high degree of inter-stimulus variability (target: mean= 0.004, std= 0.212; distractor: mean=0.013, std=0.207). To further verify that the set of target and distractor images did not differ in low level visual differences, we used a classification approach similar to Torralba et al. (2008) based on low-resolution (32x48 pixels) grayscale images. A linear SVM classifier was fed directly with these grayscale pixel values using a classification procedure identical to the one used for neural decoding (see below). Classification was not significantly different from chance level (mean: 52%, $p = 0.24$). These results suggest that the low-level visual processing based on local brightness are insufficient for the task.

Experimental Setup

Participants sat in a dimly lit room at 90 cm from a CRT computer screen. Photographs were displayed on a black background and subtended a visual angle of $\sim 7^\circ \times 11^\circ$. The experiment consisted in a go/no-go paradigm which was divided into three blocks of 180 photographs each (90 targets and 90 distractors). Participants were familiarized with the experiment using a small set of stimuli not used for the actual experiment.

Participants were instructed to respond as quickly and accurately as possible by raising their finger from an infrared response pad when a target stimulus was presented (human face target / go response). They were asked to keep their finger on the response pad if a distractor stimulus (an animal face) was presented (no-go response). At the beginning of each trial, a fixation cross appeared for a random time interval to prevent anticipatory responses (300-600 ms) as much as possible. This was followed by the presentation of the stimulus (100 ms) and a blank screen (1000 ms, Figure 2). The order of the stimuli was randomized across blocks and participants.

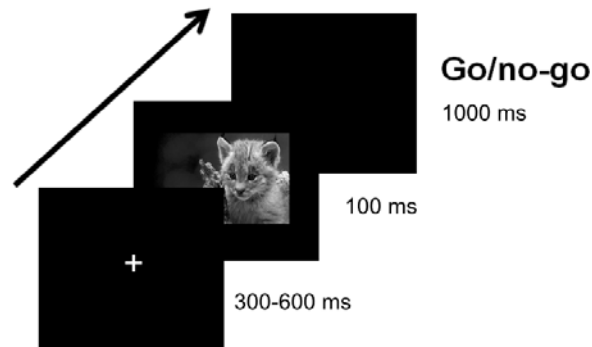


Figure 2: Experimental design. After the presentation of a fixation cross for a random time interval (300–600 ms), the stimulus was flashed (100 ms) followed by a blank screen. Participants had 1 sec to respond using an infrared pad if they perceived a human face in the picture (go response). Otherwise, they had to withhold their response (no go response).

EEG recording

EEG activity was recorded from 32 electrodes mounted on an elastic cap in accordance with the 10-20 system (Oxford Instruments) with the addition of extra occipital electrodes, using a SynAmps amplifier system (Neuroscan). The ground electrode was placed along the midline, in front of Fz, and impedances were kept below 5 k Ω . Signals were digitized at a sampling rate of 1,000 Hz and low-pass filtered at 100 Hz. Potentials were referenced online to the Cz electrode and average-referenced off-line. EEG data analysis was performed using EEGLAB (Delorme, 2004), a freely available open source toolbox (<http://www.sccn.ucsd.edu/eeglab>) running under MATLAB (The Mathworks).

First, data were down-sampled to 128 Hz and then digitally filtered using a bidirectional linear filter that preserves the phase information (pass-band 0.1–40 Hz). Independent Component Analysis (ICA) was performed to remove eye blinks for each subject individually. Data epochs were extracted (-100 to 700 ms) and baseline corrected (-100 to 0 ms). A total of 540 epochs for each individual participant was then visually inspected. Epochs containing artifacts (e.g., due to swallowing, electrode cable movements, etc.) were excluded from further analysis. Following this procedure, the mean percentage of rejected epochs across all participants was 19.9 % (across subject SD=7%, across subject range: 8%–37%). For two of the participants, one of the channels also had to be excluded from further analysis because of the presence of significant permanent artifacts. Correct trials only were considered for further EEG analysis, which accounted for 94.8% % (across subject SD=3%, across subject range: 89.1% – 99.6%) of the overall trials.

Behavioral performance analysis

To estimate the minimal processing time required to detect target images, we computed the shortest reaction time (RT) at which correct go responses started to significantly outnumber incorrect go responses (Rousselet et al., 2003) using an exact Fisher test and 30 ms time bins. The minimal RT was estimated by considering the onset of the first significant bin followed by at least five consecutive significant bins (Thorpe et al. 1996).

Multi-Variate Pattern Analysis (MVPA)

MVPA was conducted directly on the EEG signal. A (temporal) decoding analysis was conducted (target vs. distractor) and an accuracy measure was derived by feeding individual time bins of the EEG signal across all electrodes to a classifier (see below). Such decoding analysis characterizes the temporal evolution of the category signal. Each input feature (electrode potential) was normalized (Z-score) across trials and a linear Support Vector Machine (SVM) was used as classifier using the libsvm implementation (Fan et al., 2005). The classification procedure ran as follows: (1) The stimulus set was split equally in a training and a test set that contained an equal proportion of target and distractor images; (2) An optimal cost parameter C was determined through line search optimization using 8-fold cross-validation on the training set; (3) An SVM classifier was trained and tested on each set.

For each subject, this procedure was repeated over 300 cross-validations whereby different training and test sets were selected each time at random. A measure of chance level was obtained by performing the same analysis on permuted labels. This allowed us to estimate the latency of task-related information across all participants via a paired t-test for every time bin individually (accuracy on permuted vs. non-permuted labels; $p < 0.01$). To correct for multiple comparisons, a bin was considered significant when followed by at least 5 consecutive significant bins (Liu et al., 2009). Reported decoding latencies correspond to the earliest significant bin.

Results

Participants performed the categorization task (human vs. animal face) with a very high level of accuracy (mean = 94.8%, std = 3.3%, range: 89.1% – 99.6%) and fast reaction times (mean RT = 445 ms, std = 44 ms). The mean minimum RT across subjects was 354 ms (std=38 ms, range: 270 ms – 450 ms).

The plot in Figure 3A shows the ERPs from all electrodes between -100 ms and 700 ms post stimulus onset, for both human target and animal distractor conditions. The classical P1-N1-P2 complex can be readily observed. At the same time, the N170 component appears very small compared to components obtained using cropped homogenous stimuli (e.g. Rossion & Caharel, 2011). The maximal N170 amplitude (mean=-3.6 μ V, std= 2.4 μ V) on human face stimuli was recorded on the right temporal electrodes (T6' or TP8) at 179 ms post stimuli onset. ERPs averaged over classically reported electrodes from the right hemisphere T6 and CB2 (equivalent to P8 and PO8; Rossion & Jacques, 2008) showed an even smaller N170 amplitude (T6: mean=-2.7 μ V, std= 3.4 μ V; CB2: mean=-0.6 μ V, std=2.4 μ V, respectively) both peaking at 164 ms post stimuli onset (Fig S1). Most importantly, consistent with the results by Rousselet et al. (2004), none of these electrodes showed any significant amplitude modulation on the N170 for target human faces vs. distractor animal faces using a classical pic analysis (paired t-test; CB2: tstat=-0.86, pval=0.40, df=27; T6: tstat=-0.98, pval=0.33, df=27; T6': tstat=-0.88, pval=0.38, df=27).

To finely quantify the amount of face information at each time point we conducted a decoding (MVPA) analysis. Face category information could be readout significantly above chance as early as 82 ms post-stimulus onset (Figure 3B). By looking at the zero crossings of the temporal derivative of the accuracy curve, we identified three distinct phases in the temporal evolution of face information. During a first phase (82-140 ms), the amount of face information increased monotonously until a plateau was reached (~60% accuracy around 140 ms after stimuli onset). In a second phase (140-195 ms), the decoding accuracy remained constant (~60% until 195 ms post stimuli onset). During a third phase (195-350 ms), a monotonic increase was observed again (reaching 78% of decoding accuracy around 350 ms).

To link the face information contained in the EEG signal to behavioral responses, we conducted an MVPA for each participant independently. For this analysis, trials were sorted in quartiles according to reaction times (Fig. 3C). A significant modulation of the decoding accuracy by reaction times was observed starting at 160 ms (1-way ANOVA, $p < 0.01$, 5 consecutive time bins).

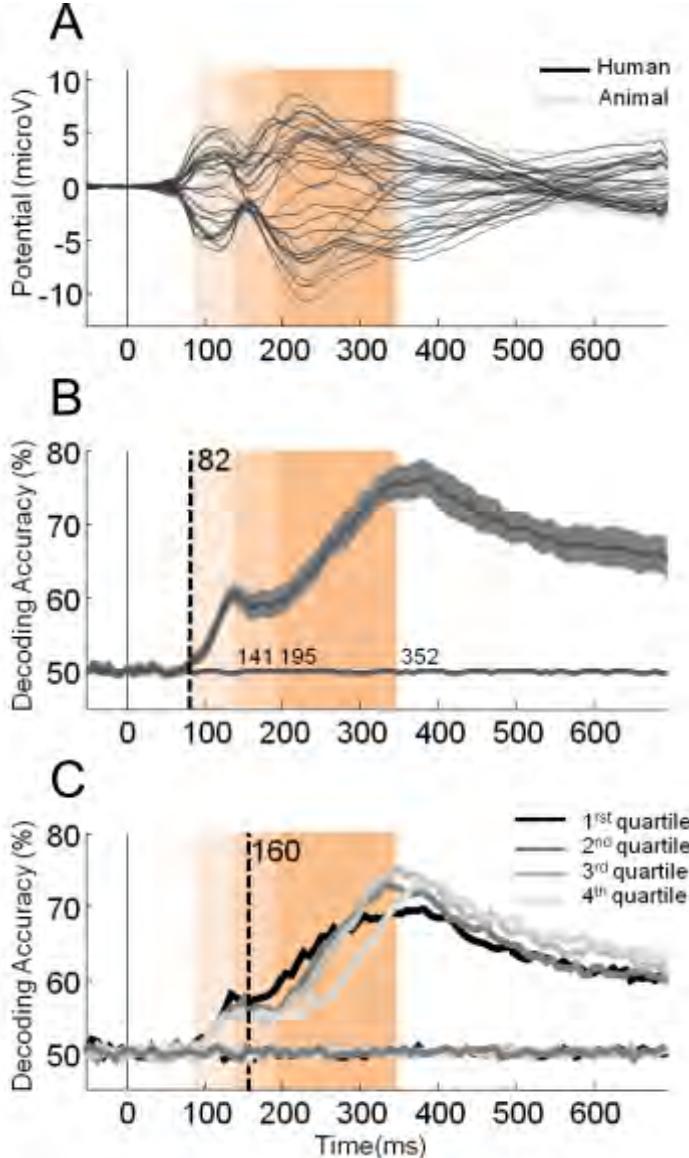


Figure 3: Neural timing of face detection in the wild. A) Grand average ERP (n=28) for each of the 32 electrodes for human target (black) and animal distractor (grey) conditions. B) Average temporal decoding accuracy (+/- 95% CI) across all subjects. The decoding accuracy estimated from permuted labels was used to assess chance level and is shown in gray. The vertical dotted line indicates the latency of the first significant decoded bin. Based on the latency of this first bin and changes in the sign of the derivative of the temporal decoding accuracy curve, we observed three distinct temporal windows (shown in shades of orange: 82-141 ms; 141-195 ms; 195-352ms). These three windows are also shown on A and C for improved readability. C) Temporal decoding by RT quartile. The vertical dot line indicates the latency of the first significant difference between the four quartiles.

Discussion

The present study investigated the neural dynamics of face detection in natural scenes and focused on the earliest latency at which face-specific neural activity emerged. Using multivariate pattern analysis on EEG activity recorded during a rapid go/no-go face categorization task, we found that face category information could be readout very early, starting around 80 ms after stimulus presentation. This result was actually predicted from the literature and implies that faces presented in natural scenes elicit specific neural activity at similar early latencies than cropped faces with low variability. Such latency is remarkably fast, despite the fact that our stimuli were highly complex and variable, and underlies the incredible efficiency of the visual system given that information flows in V1 around 30 ms at the earliest (Kirchner et al, 2009). This leaves the visual cortex only 50 ms to detect a face in a scene. Such result emphasizes the fact that the visual system is highly optimized to process complex natural stimuli (Vinje and Gallant, 2000; Simoncelli and Olshausen, 2001).

The temporal evolution of the decoding accuracy revealed three different phases at ~80-140 ms, ~140-195 ms and ~195-350. Decoding accuracy started to decrease after 350 ms, supporting the idea that face detection is carried out fast. Interestingly, the 140-195 ms time-period showed a plateau (decoding accuracy remained constant) but it is at this time, and not during the first phase starting at 80 ms, that a correlation with behavior was found.

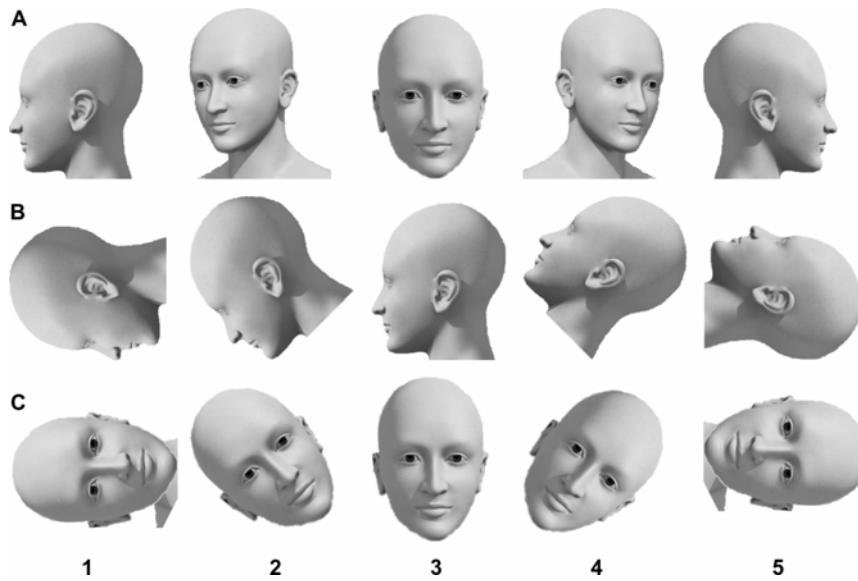
The latencies of the first (~80-140 ms post stimulus onset) and second phase (~140-195 ms post stimulus onset) were highly consistent with the latencies of the two P100 and N170 ERP components that have been traditionally associated with face processing. However, a direct link between neural activity and behavioral responses was only observed during the second phase. Trials corresponding to the fastest RTs (sorted according to quartiles) were decoded with higher accuracy than the other starting at 160 ms. This result is consistent with a two-stage model of face categorization proposed by Rossion & Caharel (2011). In this model, the P100 would reflect the processing of low-level image properties (such as an image frequency content as estimated via the Fourier power spectrum), while the later N170 would more directly reflect high-level face processes (see however, Liu et al., 2002). Numerous studies have reported a similar category-selective activity between 140-180 ms (e.g. Bentin et al., 1996; Thorpe et al., 1996) leading to the hypothesis that this activity could reflect high-level

category processes independent of low-level visual properties (VanRullen and Thorpe, 2001; Rossion & Caharel, 2011). We speculate that the first phase may be sufficient to drive saccades in a rather automatic fashion as observed during rapid categorization tasks using the saccadic choice paradigm (Kirchner & Thorpe, 2005; Crouzet et al., 2010), while the second phase would be necessary to drive behavioral go/no-go responses (VanRullen & Thorpe, 2001) and decision making (Philiastides & Sajda, 2006) as directly tested in the present experiment. In this context, it would be interesting to apply a decoding technique to the EEG recorded during a saccadic choice paradigm towards faces in order to assess whether in this case decoding around 80 ms could be related to saccade RTs.

Interestingly, the amplitude modulation of the N170 component obtained in the present study, with face stimuli embedded in natural background scenes, seems lower than is usually reported with artificial cropped face stimuli. This seems consistent with previous studies that have shown weak or no significant face selective modulation of the P100 and N170 using more natural faces embedded in natural scenes (Rousselet et al., 2004). One explanation for this discrepancy is that background clutter in natural scenes disrupts the selectivity of the N170 component. This would seem consistent with monkey electrophysiology (Duncan & Desimone, 1999; Zhang et al., 2011) as well as imaging studies (Reddy & Kanwisher, 2009) that have shown that the patterns of brain activity associated with object categories are disrupted by the presentation of multiple stimuli on screen. Another likely explanation for the decreased modulation of the N170 is the high variability of the face stimuli used (in terms of eccentricity, size and shape) compared to previous studies, which might decrease the time-locking to stimulus onset and thus decrease the N170 average amplitude.

To the best of our knowledge, the results presented here are the first to demonstrate at the single-trial level the availability of behaviorally-relevant information from EEG signal during a difficult detection task. While MVPA has been extensively used in the context of fMRI studies (see Tong & Pratte, 2012 for a recent review), the use of decoding techniques for M/EEG analysis has been mainly limited to the field of brain computer interface (P300 speller, Farwell and Donchin, 1988). Very few studies have investigated the possibility to read visual category information from non invasive human electrophysiological signals. Among them, a MEG study has demonstrated the possible readout of the position of objects with relatively low accuracy and late latencies (incompatible with behavioral results from Crouzet et al., 2010) using artificial cropped stimuli of faces vs. textures and houses (Carlson et al., 2011). Another EEG study reached higher decoding accuracy to decode line drawing of animals versus tools (Simanova et al., 2010). However, the very few number of exemplars used by category, the evident low level visual differences between the two category and the high number of presentation for each stimuli used for decoding, was inconclusive regarding the real nature of the decoded information. On the contrary here we use a large database of stimuli presented only once to each subject.

In conclusion, detection of faces in complex natural stimuli relies on two early consecutive stages, each with their own properties and starting as early as 80 ms.



Supplementary figure 1. Description of horizontal (A), vertical (B) and slope (C) orientations rating from 1 to 5 with number 1 used when the face was orientated between the first and second position, 2 between position 2 and 3, 3 when the full-face was presented etc... Details of the results are provided in supplementary Table 1.

Supplementary Table 1. Visual characteristics of human face stimuli showing the proportion of the different head orientations (horizontal, vertical, slope, see Supplementary Figure 1); male, race; other (paraphernalia, teeth being visible or not, presence of an expression (here limited to the presence of a smile or not since it was the only expression shown), and gaze direction (if the face is looking in the eyes or not).

	Rate/Category	Proportion (%)
Horizontal	1	4.2
	2	15.7
	3	61.2
	4	14.5
	5	4.4
Vertical	1	0.8
	2	7.4
	3	88.9
	4	2.7
	5	0.2
Slope	1	0.9
	2	2.3
	3	90.5
	4	5.9
	5	0.4
Sex	Male	47.1
Race	Caucasian	91.9
	Black	4.1
	Asian	4.0
Other	Paraphernalia	34.4
	Gaze	62.4
	Smile	61.7
	Teeth	53.7

References

- Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological Studies of Face Perception in Humans. *Journal of cognitive neuroscience* 8:551–565.
- Bentin S, Taylor MJ, Rousselet GA, Itier RJ, Caldara R, Schyns PG, Jacques C, Rossion B (2007) Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nature Neuroscience* 10:801–802.
- Bötzel K, Schulze S, Stodieck SR (1995) Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res* 104:135–143.
- Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. *J Vis* 11 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21920851> [Accessed September 26, 2012].
- Crouzet SM, Kirchner H, Thorpe SJ (2010) Fast saccades toward faces: face detection in just 100 ms. *J Vis* 10:16.1–17.
- Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. *Annual review of neuroscience* 18, 193–222.
- DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434.
- Eimer M, Holmes A (2002) An ERP study on the time course of emotional face processing. *Neuroreport* 13:427–431.
- Fabre-Thorpe M (2011) The characteristics and limits of rapid visual categorization. *Front Psychol* 2:243.
- Fan R-E, Chen P-H, Lin C-J (2005) Working Set Selection Using Second Order Information for Training Support Vector Machines. *J Mach Learn Res* 6:1889–1918.
- Farwell LA, Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol* 70:510–523.
- Halgren E, Raji T, Marinkovic K, Jousmäki V, Hari R (2000) Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb Cortex* 10:69–81.
- Itier RJ, Taylor MJ (2004) N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb Cortex* 14:132–142.
- Jeffreys DA (1989) A face-responsive potential recorded from the human scalp. *Exp Brain Res* 78:193–202.
- Kirchner, H., Barbeau, E.J., Thorpe, S.J., Régis, J., Liégeois-Chauvel, C., 2009. Ultra-rapid sensory responses in the human frontal eye field region. *J. Neurosci.* 29, 7599–7606.
- Kirchner H, Thorpe SJ (2006) Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res* 46:1762–1776.
- Liu H, Agam Y, Madsen JR, Kreiman G (2009) Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62:281–290.

- Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. *Nat Neurosci* 5:910–916.
- 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.
- Reddy, L., Kanwisher, N., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr. Biol.* 17, 2067–2072.
- Rossion B, Caharel S (2011) ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Res* 51:1297–1311.
- Rossion B, Delvenne JF, Debatisse D, Goffaux V, Bruyer R, Crommelinck M, Guérit JM (1999) Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol Psychol* 50:173–189.
- Rossion B, Jacques C (2008) Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39:1959–1979.
- Rousselet GA, Macé MJ-M, Fabre-Thorpe M (2003) Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J Vis* 3:440–455.
- Rousselet GA, Macé MJ-M, Fabre-Thorpe M (2004) Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *J Vis* 4:13–21.
- Rousselet GA, Pernet CR (2011) Quantifying the Time Course of Visual Object Processing Using ERPs: It's Time to Up the Game. *Front Psychol* 2:107.
- Simanova I, van Gerven M, Oostenveld R, Hagoort P (2010) Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS ONE* 5:e14465.
- Simoncelli, E.P., Olshausen, B.A., 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.
- Spiridon M, Kanwisher N (2002) How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35:1157–1165.
- Thierry G, Martin CD, Downing P, Pegna AJ (2007a) Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci* 10:505–511.
- Thierry G, Martin CD, Downing PE, Pegna AJ (2007b) Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nature Neuroscience* 10:802–803.
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
- Tong F, Pratte MS (2012) Decoding patterns of human brain activity. *Annu Rev Psychol* 63:483–509.
- Torralba, A., 2009. How many pixels make an image. *Visual neuroscience* 26, 123–131.
- VanRullen R, Thorpe SJ (2001) The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci* 13:454–461

- Vinje, W.E., Gallant, J.L., 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276.
- Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., Desimone, R., 2011. Object decoding with attention in inferior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8850–8855.

**ARTICLE 5 : The speed of face recognition: relating EEG to
behaviour using MVPA**

Gladys Barragan-Jason, Maxime Cauchoix *, Emmanuel Barbeau*

Soumis dans NeuroImage

Article 5: The speed of face recognition (soumis)

Introduction : La dynamique électrophysiologique de la reconnaissance explicite des visages reste toujours inconnue à ce jour, les données de la littérature rapportant des latences variant de 150 à plus de 400 ms. De plus, aucun consensus concernant les électrodes à considérer ne semble réellement émerger. Afin de déterminer la latence électrophysiologique minimale pour reconnaître un visage familier dans des conditions bottom-up, nous avons décidé d'appliquer notre méthode de MVPA permettant de s'abstraire de la sélection des électrodes d'une part et d'étudier la reconnaissance des visages dans des conditions écologiques d'autre part (c'est-à-dire présentés dans un contexte naturel, voir Article 4). De plus, afin d'identifier les latences minimales, les sujets sont contraints à répondre le plus rapidement possible.

Méthode : Une analyse MVPA a été réalisée sur le signal EEG 32 de 28 sujets engagés dans une tâche de reconnaissance rapide de visages célèbres parmi des inconnus.

Résultats : Les visages célèbres peuvent être décodés des visages inconnus dès 230 ms après la présentation du stimulus. Cette latence en décodage est très prédictive de la vitesse de reconnaissance du sujet. En effet, les essais les plus rapides (inférieurs au temps de réaction médians) sont décodés 90 ms plus rapidement (à 203 ms) que les essais lents (supérieurs au temps de réaction médian décodés à 289 ms) d'une part et les latences individuelles en décodage sont fortement corrélées au temps de réaction minimum de chacun des sujets. Cependant, même si elle n'est pas significative, une petite variation de la précision du décodage est rapportée dans la fenêtre de la N170.

Discussion : La reconnaissance explicite des visages célèbres semble être possible dès 200-300 ms après la présentation du stimulus. Cette latence est en accord avec la composante N250, supposée indexée sur la familiarité mais ne supporte pas l'hypothèse d'une familiarité liée à la N170. Cependant, il semble évident que les processus ayant lieu pendant la fenêtre de la N170 participent à cette reconnaissance, l'accès à une représentation holistique du visage étant indispensable à la reconnaissance. De plus, cette latence de décodage à 200 ms

semble d'une part (1) beaucoup plus tardive que celle observée pour la détection (80 ms) confirmant l'implication de processus additionnels et d'autre part (2) directement corrélée au comportement des sujets contrairement à la détection suggérant l'implication de mécanismes qualitativement différents.

✓ *Dans ce chapitre nous avons montré que la reconnaissance des visages au niveau de la familiarité est possible après 200 ms, révélant qu'elle nécessite des processus additionnels à ceux impliqués dans la tâche de détection. Le dernier chapitre discutera des hypothèses sur les mécanismes sous-tendant cette reconnaissance.*

The speed of face recognition: relating EEG to behavior using MVPA

Running title: The speed of face recognition

Barragan-Jason G.*¹, Cauchoix M.*¹, Barbeau E.J.¹

1 Centre de recherche Cerveau et Cognition, Université de Toulouse, CNRS-UMR 5549, Toulouse, France

Corresponding author:

Gladys Barragan-Jason
Centre de recherche Cerveau & Cognition (CerCo)
CNRS CERCO UMR 5549
Pavillon Baudot, CHU Purpan, BP 25202
31052 Toulouse Cedex, France.
gladou86@gmail.com

Abstract

Because of its critical role in our everyday life, recognizing familiar people from their faces is supposed to be fast. However, the actual neural dynamics associated with the explicit recognition of familiar faces remains unknown. In particular, the respective implication of EEG components peaking around 170 and 250 ms (N170 and N250) is unclear with respect to this issue. We recorded scalp EEG activity in 28 subjects performing a go/no-go, famous/non-famous face recognition task under speed constraints. Using temporal multivariate EEG analyses, we report that famous faces could be robustly distinguished from unknown faces as soon as 230 ms after stimulus onset. This familiarity-selective signal was directly linked to the subject's recognition speed. Photos corresponding to fast responses were decoded roughly 90 ms faster (~203 ms) than photos corresponding to slow answers (~289 ms). Furthermore, EEG decoding latencies in individual subjects were highly correlated with their minimum reaction times. Thus, a familiar face selective EEG activity related to behaviour is reported, starting at the earliest around 200 ms but spreading over longer times depending on the subject. Implications regarding the N170 and N250 and the mechanisms involved in face recognition are discussed.

Keywords: famous faces, explicit recognition; MVPA; EEG; familiarity

INTRODUCTION

Many people think that face recognition is fast (e.g. Sagan, 1995). Part of this belief is probably related to the idea that face recognition *should* be fast so as to differentiate friends from foes or because face processing is so important in daily life. The problem, however, is that little is known about the actual speed at which a given face can be recognized. Whereas many studies have reported neural latencies related to the presentation of faces, few have focused on the fastest differentiation between known and unknown faces.

Here, we focus on the speed at which a face can be found to be familiar or not (explicit familiarity decision). This would correspond to the ecological situation of meeting someone unexpectedly in the street and assessing whether the person is known or not from his/her face. Recognizing faces among unknown ones can be performed in ~400-900 ms (mean or median reaction times, Anaki et al., 2007, 2009; Baird and Burton, 2008; Barragan-Jason et al., 2012; Caharel et al., 2005; Herzmann et al., 2004; Kampf, 2002; Ramon et al., 2011). The high variability of these behavioural latencies can be explained by differences in experimental procedure (e.g. go/no-go, category-verification, repetition) or the nature of stimuli used (e.g. famous faces, personally familiar faces, novel familiar faces, cropped stimuli, natural stimuli...). Another source of variability is related to the fact that most studies have allowed a long time for subjects to respond. Thus, subjects could use different strategies, either familiarity-based ("I know this face") or identity-based ("This is Brad Pitt, a famous actor"). Following classic hierarchical models of face recognition, familiarity-based responses are thought to be faster than identity-based responses (Bruce and Young, 1986). However, it is likely that subjects cannot help trying to recognize a face at the level of identity (Bruce and Young, 1986; Valentine, 2001) and hence giving rather slow answers overall.

Using EEG, the earliest neural differential activity between known and unknown faces has been found in the 140-200 ms time window after stimulus onset, i.e. during the N170 (e.g. Caharel et al., 2002, 2005, 2007; Wild-Wall et al., 2008, known faces showing higher amplitude than unknown ones). However, these results are inconsistent with other studies that have found an amplitude modulation in the opposite direction (Jemel et al., 2003; Todd et al., 2008). Moreover, most of these studies used a small number of stimuli that were repeated or primed. Actually, most studies did not find any difference during the N170, especially when an explicit face recognition task was used (e.g. Anaki et al., 2007; Bentin and Deouell, 2000; Eimer, 2000; Schweinberger et al., 2004; Zheng et al., 2012; Eimer et al., 2012). Overall, there are two contrasted positions regarding the N170 at present. Some authors argue that it reflects mechanisms necessary for face individuation (e.g. Jemel et al., 2009), while others link the N170 to configural/holistic face representation only (e.g. Gosling & Eimer, 2011). In the first case, familiarity could occur concomitantly or close to the N170. In the second case, familiarity would occur in a subsequent stage.

Following the N170, the N250 component has been more reliably identified as a likely neural correlate of familiarity. This slow component is more negative in response to known than to unknown faces over lateral occipito-temporal electrodes and differs from the N250r in that

the latter component is elicited by repeated presentation of faces. Although this component has been called N250, suggesting it should peak around 250 ms, most authors average data for their analyses over a long, and variable, period after 250 ms (e.g. 270-330 ms in Zheng et al., 2012; 230-400 ms in Gosling & Eimer, 2011; 240-300 ms in Jemel et al, 2009). This would suggest that familiarity with faces occurs rather late compared to the N170: at least 100 ms later. However, to date, authors have focused on a component averaged over a long period rather than on the onset of this component. Actually, Zheng et al, 2012 suggest that this component could start around 200 ms, thus right after the N170. Gosling & Eimer (2011) report similar findings in their Figure 4. Most importantly, these authors report that they have identified a large positive component related to familiarity, starting “much earlier” [than the usual N250, which is around 200 ms] but broadly distributed over anterior electrodes.

All in all, several issues are still unresolved. The onset of the neural correlate of familiarity is unknown, possibly ranging from 150 (in the N170 time-window) to 300 ms (in the N250 time-window) post stimulus onset. Although some studies suggest that the onset should be at around 200 ms, this has never been assessed. As far as we know, no speed constraints have been used in any studies, and this allows subjects to use different strategies. It could be that using speed constraints enhances neuronal synchronization, thus evoking a clearer, and possibly earlier, neural correlate of familiarity. Furthermore, the study by Gosling & Eimer (2011) suggests that the topography of this familiarity correlate is actually not very well known. This underlines how little we know about the neural correlate of familiarity. Lastly, if some neural activity is related to familiarity, it should then also be related to behaviour (e.g. RTs). To our knowledge, no such relation has been studied yet (see, however, Bentin and Deouell, 2000).

Our aims in this study are to (1) identify the minimal neural latencies sufficient to explicitly recognize faces and (2) to analyse how these latencies relate to behaviour. We expect to find these neural latencies in the 160-300 ms time-window.

The timing of the neural mechanisms underlying familiar face recognition was investigated using an averaging method. Traditionally, studies have focused on peak latencies or on a period of activity rather than on the onset of the differential activity. Furthermore, such analyses rely strongly on the identification of a given set of electrodes where a familiarity effect can be found. The size of the effect varies greatly depending on the electrodes that are included in the analyses (Rossion and Jacques, 2008) and there is no agreement to date on which electrodes should be included to analyse a neural correlate of familiarity.

To address our aims and overcome these issues, 28 subjects were asked to recognize famous faces among unknown ones using speed constraints. Multivariate pattern analysis (MVPA) of EEG recorded concurrently with the task allowed the information involved in the recognition of famous faces to be analysed at the single trial level and at millisecond resolution, without any *a priori* about the ERP components or scalp location of the familiarity effect.

Furthermore, we investigated the relation between neural activity and behavioural responses.

Methods

Participants

Twenty-eight healthy subjects, (15 females, mean age: 25.4, SD = 3.8, 25 right-handers) with normal or corrected-to-normal vision gave their written informed consent to participate in the experiment.

Stimuli

A set of 270 greyscale photographs of famous faces (e.g. Brad Pitt, George Clooney, Bill Clinton, Madonna...), selected as being recognized by adults between 20 and 40 years old, and 270 photographs of unknown faces were presented in their natural context (i.e. they were not cropped and some background could be seen) (Rousselet et al., 2003). Unknown faces were chosen to be comparable in terms of quality (professional photographs), attractiveness (models) and emotional expression. Examples of photographs are presented in Figure 1.

Each image was 320 x 480 pixels and luminance and RMS contrast were comparable between famous and unknown images. We verified that images were not different in terms of low level features: we tried to directly decode the image category from different image features such as (1) raw grey level of pixels, which can be seen as a model of the retina, (2) weibull parameters, which can be seen as a model of the LGN (Scholte et al., 2009) and (3) the gist; a model proposed by Torralba and Oliva (2003) that extracts orientation at different scales in the picture and performs well in categorizing natural images (this model can be interpreted as a V1-like model). None of these models allowed image category to be significantly decoded from picture features (retina-like: decoding accuracy = 52%, $p = 0.28$; LGN-like: decoding accuracy = 52%, $p = 0.26$; V1-like: decoding accuracy = 54%, $p = 0.15$).

We also controlled other variables such as head orientation, paraphernalia, race and facial expressions between famous and unknown faces. Details are provided in supplementary Table 1.

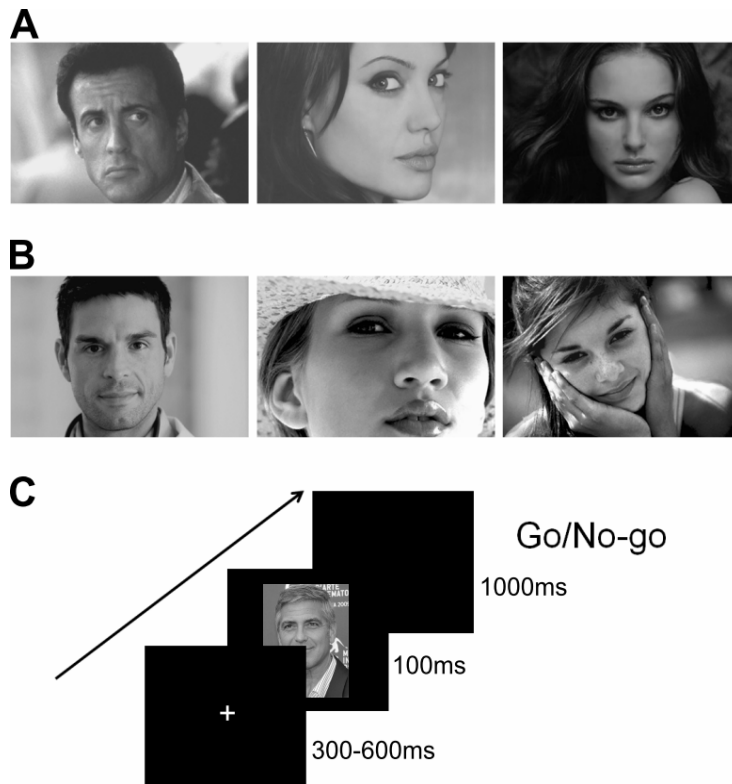


Figure 1. Experimental design. Example of famous faces in A and unknown faces in B. Example of one trial in C. After the presentation of a fixation cross for a random interval of 300-600 ms, the stimulus was flashed on the screen for 100 ms followed by a black screen.

Experimental setting

Subjects sat in a quiet, dimly lit room at ~90 cm from a 19" CRT computer screen (resolution: 1024 x 768; vertical refresh rate: 100 Hz). Responses were provided through a custom-made infrared response pad. Stimuli were presented using Eprime 2.0.

Experimental design

The go-no go task was divided into three blocks of 180 photographs (90 targets and 90 distractors). Participants had to raise their fingers from the response pad as quickly as possible when a famous face (go-response) was presented among unknown faces (no-go response). Each trial started with the presentation of a fixation cross for a random interval to prevent anticipatory responses (300-600 ms). The fixation cross was followed by a photograph flashed for 100 ms and a black screen for 1000 ms. Stimuli were randomly displayed across blocks for each subject and each photograph was seen only once by each subject.

At the end of the experiment, each individual famous face was randomly displayed and subjects had to rate their familiarity with the face using a scale from 0 (unfamiliar face) to 6 (highly familiar face). Famous faces rated as unknown (0 on the familiarity scale) were discarded from analyses, as were reaction times below 200 ms or above 1000 ms (Barragan-Jason et al., 2012).

Speed constraints

We developed various strategies to ensure that subjects answered as fast as they could. Stimuli were flashed quickly (100 ms) and subjects had to answer before 1000 ms post-stimulus. Subjects performed a training session with specific stimuli and could repeat the training session if they wanted. After each block, including the training session, mean reaction times and false-alarm rates were displayed so that subjects could monitor their performance. Lastly, they received strong encouragement before and between blocks to answer as fast as possible. In particular, after each block, they were asked to “beat” their RT score.

EEG recording

We recorded neural activity from 32 electrodes mounted in an elastic cap in accordance with the 10-20 system (Oxford Instruments) with the addition of extra occipital electrodes, using a SynAmps amplifier system (Neuroscan). The ground electrode was placed along the midline, ahead of Fz, and impedances were systematically kept below 5 k Ω . Signals were digitized at a sampling rate of 1000 Hz and low-pass filtered at 100 Hz. Potentials were referenced on-line to the Cz electrode and averaged-referenced off-line. EEG data analysis was performed using EEGLAB (Delorme and Makeig, 2004), an open source toolbox running under Matlab (Mathworks).

Data were first down-sampled to 128 Hz and were then digitally filtered using a bi-directional linear filter that preserved the phase information (pass band 0.1 to 40 Hz) to minimize drifts and line noise. ICA decomposition was applied to the signal for each subject. Data epochs were extracted (-100 to 700 ms) and baseline corrected (-100 to 0 ms). All epochs, corresponding to correct trials only (69.5%, mean: 376 epochs, range: 323-459), were then inspected visually, in each individual dataset in each condition. Epochs containing single non-stereotyped artefacts (e.g. electrode cable movements) were rejected from further analysis, whereas epochs containing eye blinks were corrected (subtraction) using ICA. Following this procedure, the mean percentage of rejected epochs was 17.8% (mean: 68 epochs, range: 10–126). For two subjects, one bad channel was also excluded from further analysis.

Behavioural performance analyses

To obtain an estimation of the minimal processing time required to recognize targets, the latency at which correct go-responses started to significantly outnumber incorrect go-responses was determined for each subject (e.g. Rousselet et al., 2003). We applied an exact Fisher test with 30 ms time bins on RTs for each individual participant (Barragan-Jason et al., 2012). To correct for multiple comparisons, significance had to last 60 ms at least and the beginning of the first significant bin was considered as the minimum RT. We also calculated the d' (performance accuracy) for each participant as a function of processing time with 10-ms time bins using cumulative numbers.

Multivariate Pattern Analysis (MVPA)

The temporal evolution of famous vs unknown information in the EEG signal was assessed, at the single trial and single subject level, using temporal multivariate pattern analysis (MVPA).

For every time point independently, a linear Support Vector Machine (SVM) was used to distinguish correct famous from correct unknown trials based on the potential recorded over the 32 electrodes. The classification procedure was as follows: (1) the trial set was equally split into a training set and a test set, each containing an equal proportion of famous and unknown images; (2) an optimal cost parameter C was determined through line search optimization using 8-fold cross-validation on the training set of images; (3) an SVM classifier was trained and tested on each set.

For each subject, this procedure was repeated 50 times, different training and test sets being selected each time at random (Cross-Validation, CV). A measure of chance level was obtained by performing the same analysis on permuted labels. This single subject decoding was considered above chance when a t-test between randomized and real decoding was significant ($p < 0.01$) across the 50 CVs.

Across-subject decoding was considered significant when real decoding (averaged over the 50 CVs) was significantly higher than the permuted one over all subjects (paired t test, $p < 0.01$).

Single subject and across subject decoding latencies correspond to the earliest significant decoding bin. To correct for multiple comparisons, significance had to last for at least 5 consecutive bins and the first significant bin was considered as the decoding onset value (Liu, 2009).

Slow vs fast responses

To test for an intra-subject link between the speed of face recognition and the speed of EEG decoding, trials were split into slow ($RT > \text{median } RT$) and fast ($RT < \text{median } RT$) trials for each subject. We then performed the temporal decoding analysis as explained above for the two categories of trials independently.

Results

Mean d' was 1.32, $SD=0.44$, range: 0.46-2.12. The task was performed above chance starting at 546 ms (minimum RT averaged across participants, $SD=79$; range: 390-720 ms).

The neural dynamics was estimated using a single trial temporal decoding accuracy measure for each subject averaged across participants (Figure 2A). Significant decoding values could not be calculated for one subject. The onset of the neural decoding accuracy was significant 227 ms after stimulus onset. A small, but not significant, bump of decoding power was observed between 130-180 ms. In order to assess whether decoding onset and reaction times were related, we contrasted fast ($< \text{median } RT$) and slow ($> \text{median } RT$) trials from each individual subject (Figure 2B). Fast trials were decoded 90 ms faster than slow trials, the

fastest responses being decoded as soon as 203 ms after stimulus presentation and the slowest responses 289 ms after stimulus onset.

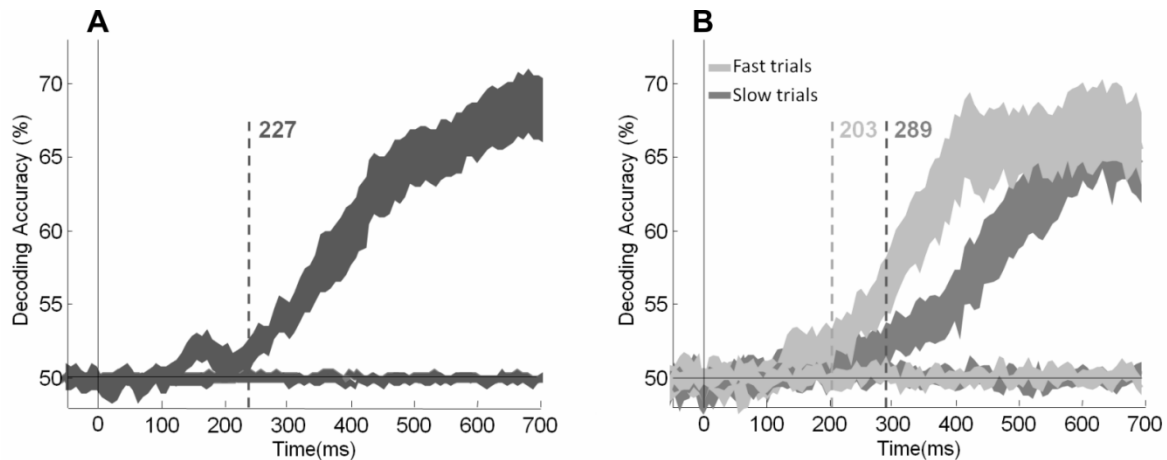


Figure 2. A) 95% confidence interval decoding accuracy across subjects on true and randomized labels (plotted on the x axis). The earliest significant decoding onset is indicated by the vertical dotted line. B) Confidence intervals (95%) of the decoding accuracy across subjects on fast (light grey) and slow (dark grey) trials. An estimation of the chance level is made using the same measure computed after label permutation (plotted on the x axis).

We investigated the correlation between minimum RTs and accuracy (d') to investigate a possible trade-off. No significant correlation was found ($R^2 = 0.00002$, $p = 0.9$). We then investigated the correlation between the decoding onset values (mean = 299 ms, SD = 75, range: 200-461) and the corresponding accuracies (Figure 3A). Again, no correlation was found ($R^2 = 0.004$, $p = 0.7$). The correlation between the decoding onset values and minimum reaction time (Figure 3B) was significant however ($R^2 = 0.46$, $p < .001$). The mean intra-subject difference between decoding onset and minimum RT was 250 ms (SD = 62, range: 153-342). Figure 3A also shows that 17/27 (63%) subjects' decoding onset values were in the 200-300 ms time window, some decoding onset values being around 200 ms.

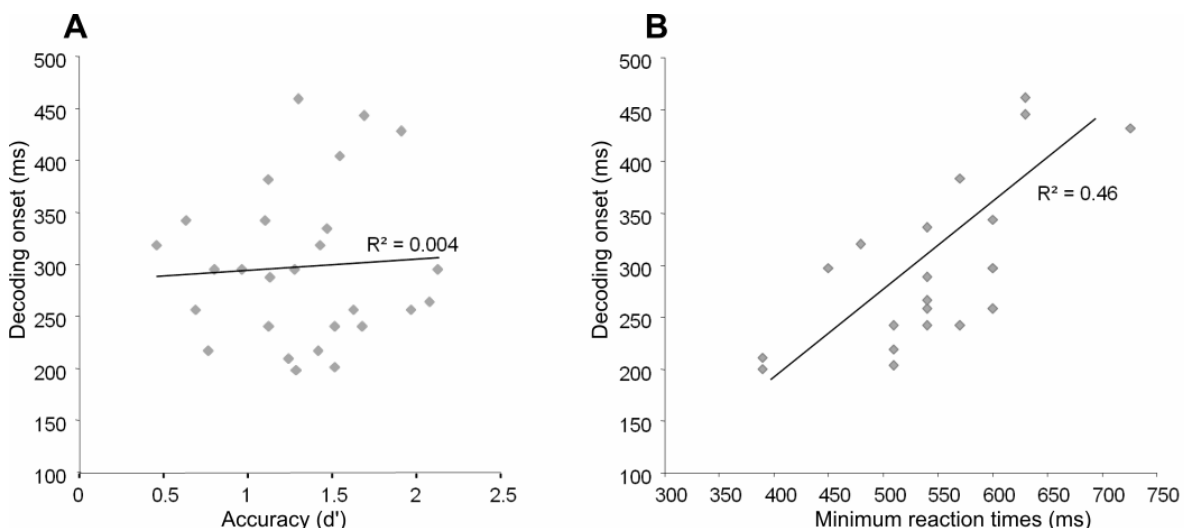


Figure 3. A) Pearson correlation between accuracy and decoding onset. Each point represents a subject. Decoding onset could not be calculated for one subject. B) Pearson correlation between minimum reaction time and decoding onset. Each point represents a subject. Minimum RTs could not be calculated for 6 subjects.

Discussion

The aim of the current study was to track down the minimum neural latencies underlying explicit face recognition. Using single-trial temporal-decoding analyses, we found that famous faces could be accurately distinguished from unknown faces as early as ~230 ms after stimulus onset at the group level. 63% of the subjects had individual decoding onset in the 200-300 ms time-window. Individual decoding onsets were around 200 ms for some subjects. The decoding onset was related to behaviour. Fast (<median RT) trials could be decoded at 203 ms, while slow trials (>median RT) were decoded at 289 ms. Furthermore, a significant correlation ($R^2 = 0.46$) was found between decoding onset values and minimum reaction times. Overall, these results show that the neural activity underlying explicit face recognition is available early, starting around 200 ms and that this activity is related to behaviour.

A minimum neural boundary incompatible with the N170

This study provides the first minimal neural boundary for rapid famous face recognition of around 200 ms after stimulus onset for the fastest responses, suggesting that information available at this time is sufficient for behavioural responses related to explicit face recognition to be performed. Such a boundary at 200 ms was predicted by two previous studies (Gosling & Eimer, 2011; Zheng et al, 2012). However, most decoding onsets were not at 200 ms but spread over the 200-300 ms time range. Importantly, no correlation was found between decoding onset values and accuracy, suggesting that the fastest decoding onset values were not related to poor accuracy. These decoding onsets are consistent with the familiarity-sensitive components, which are more or less pronounced and peak more negatively around 250 ms in scalp EEG, called the N250 (e.g. Bentin and Deouell, 2000; Eimer et al., 2012; Kaufmann et al., 2009; Tanaka et al., 2006).

Such latencies are not consistent with the modulation of the N170 component by familiarity. As already mentioned, most studies reporting that the N170 could be related to face individuation, familiarity or identification did not actually find such modulation under conditions of explicit, unprimed recognition. Interestingly, we report a little bump of decoding power during the 130-180 ms time window (Figure 2A), where the N170 is traditionally observed but the decoding accuracy was weak and it was not significantly different from chance. This result points out the importance of quantifying effects in EEG research (Rousselet & Pernet, 2011). Importantly, as far as we know, there is no direct evidence in the literature that this early N170 differential activity can be related to behaviour.

Overall, this strongly suggests that explicit face familiarity relies on dedicated processes different from those reflected in the N170 and taking place *after* this component. There have been many studies investigating the specificity of the N170 to faces (e.g. George et al., 2005; Itier & Taylor, 2004) and this component is usually related to face perception (“in the sense of an interpretation of the stimulus as a face”, Rossion and Caharel, 2011). An important related question for the future will be to assess whether the processes taking place during the period starting at 200 ms also show some specificity for faces or whether they underlie familiarity for any kind of stimulus.

In this respect, the N170 is a large component easily identified over a known set of electrodes (Rossion and Jacques, 2008). It is related to increased EEG power phase-locked and time-locked to stimulus onset (Rousselet et al, 2007). In contrast, and as already noted, no clear ERP component signalling familiarity for faces has been identified, although converging evidence points to an effect around 250 ms. This suggests that there may be far fewer neurons involved in face familiarity (coding would be performed by fewer neurons) and that time-locking to the stimulus could be partially lost following the N170.

There are not one but several processes taking place in the period during which familiarity with a given face emerges. First, familiarity requires face processing at the exemplar level, which is beyond the faceness representation presumably reflected in the N170. Second, familiarity requires access to memory, i.e. some internal representation signalling that the face has already been seen. Third, familiarity is more than a cognitive process, it is also a more or less strong “sense” of familiarity (e.g. the butcher-on-the-bus phenomenon, Yovel and Paller, 2004). Last but not least, explicit familiarity requires that the subject be aware of this familiarity. In other words, explicit familiarity probably requires access to a “global neuronal workspace” (Dehaene and Changeux, 2011 for a review). This type of awareness is a characteristic of declarative memory and is usually referred to as “noetic consciousness” (Tulving, 1985): “I know that I know this face”. All these processes, and possibly others, presumably take “some” time beyond the processes reflected in the N170.

The facets of face recognition

Although we used time constraints, some minimum reaction times were short (around 400 ms) while others were long (600-700 ms). It could be argued that such variability was the result of an accuracy trade-off. However, no correlation was found between accuracy and minimum reaction times. Another explanation is that subjects may have used different strategies despite the speed constraints imposed. A sense of familiarity with the face was sufficient to perform our task but complementary processing related to activation of person knowledge or identification may also have been performed, leading to a variety of reaction times. Some authors have indeed posited that trying to retrieve the identity of a person is a mandatory process, i.e. we are “generally not satisfied that we know a face until more than a sense of familiarity is achieved” (Bruce & Young, 1986).

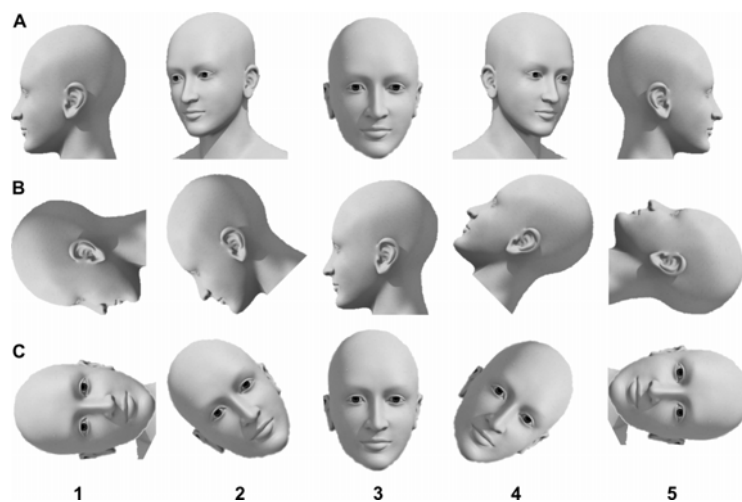
These different strategies may have been used both within subjects (a given subject may have based some of his answers on familiarity and the others on identification, Figure 2B)

and across subjects (some subjects may have had an overall bias toward a familiarity strategy or towards an identification strategy, Figure 3B). In agreement with these hypotheses, we showed that fast RTs could be associated with decoding as soon as 200 ms after stimulus onset, whereas slow RTs involved later decoding, around 290 ms. The strong correlation we found between decoding onset values and minimum reaction times ($R^2 = 0.46$) also implies that fast decoding at the individual level was associated with short minimum RTs (presumably familiarity-based, e.g. the two subjects at bottom left of Figure 3B, whose decoding onset and minimum RT were fast) and that slow decoding was associated with long minimum RTs.

The fastest answers could be related to the activation of the ventral visual pathway up to the most anterior areas (e.g. the perirhinal cortex). Hochstein and Ahissar (2002) have suggested that “vision-with-scrutiny” of the type required to identify a face could be based on processes beginning at the top of the visual hierarchy. Many studies (e.g. Aggleton and Brown, 1999) furthermore suggest that the anterior visual pathway is necessary to compute a familiarity signal. Intracranial EEG recordings have consistently found prominent activity evoked by face recognition tasks in the anterior visual pathway (Dietl et al, 2005; Trautner et al, 2004). A similar area has been identified in the monkey (Rajimehr et al, 2009). This is also in keeping with the many fMRI studies in both animals and humans that have identified more patches of activity that respond to individual faces or identity in anterior than in posterior temporal lobe areas (Rothshtein et al., 2005; Kriegeskorte et al, 2007; Freiwald & Tsao, 2010). Furthermore, the activity evoked by faces in this area shows a large N240, a negative component peaking at 240 ms (Barbeau et al, 2008; Trautner et al, 2004). The decoding associated with the fastest RTs could thus be related to the rapid activation of the whole ventral stream in parallel as already posited (Barbeau et al, 2008).

In contrast, later decoding associated with longer reaction times could be a hallmark of semantic activation about the person, which would require recurrent activity between the ventral stream and medial temporal lobe structures such as the hippocampus and the medial temporal pole. The differential activity between famous and unknown faces starts around 320 ms in the temporal pole (Barbeau et al, 2008), which is compatible with the ~290 ms decoding activity identified in this study for long RTs. A large negative peak at 360 (N360) is also recorded in the anterior visual pathway and medial temporal pole, the N240 and N360 together making up the large AMTL-N400 (Anterior Medial Temporal Lobe) reported in some studies (Trautner et al, 2004). The temporal pole is critical for semantic memory and for person knowledge in particular (Brambati et al, 2010; Joubert et al, 2004; 2006). Within this context, the fact that the scalp-N250 is a large component could be related to an overlap of the familiarity and identification processes across trials. Some trials could lead to a fast familiarity decision followed by fast identity retrieval, others to a slow familiarity decision but fast identity retrieval, etc. Furthermore, the different processes associated with familiarity identified above (processing at the exemplar level, access to memory and to awareness, etc.) probably add even more variability into the system. Overall, such a pattern of non-time-locked evoked potentials would lead to a slow averaged potential.

Famous faces activate the hippocampus strongly and reliably (Barbeau et al, 2008; Dietl et al, 2005; Quiroga et al., 2005; Trautner et al, 2004). However, the differential activity recorded in the hippocampus between famous and unknown faces does not start before 320 ms post-stimulus (Barbeau et al, 2008; Dietl et al, 2005; Trautner et al, 2004). Moreover, hippocampal single-neuron activity occurs later than neocortical activity for the same stimuli (Mormann et al, 2008). This suggests, in accordance with dual models of familiarity (Aggleton & Brown, 1999), that the hippocampus is not necessary for fast, familiarity-based, answers. Hence, our data support the idea that familiarity-based responses can be given on the basis of a neocortical route only, a possibility that has been under debate for some time (Barbeau et al, 2005, 2011; Holdstock et al, 2008; Manns et al, 2004; Mayes et al, 2002). In conclusion, our study provides useful time limits that may be applied to constrain models of face recognition and illustrates how MVPA can be very efficient for reading out evoked potentials without any a priori on the components or electrodes of interest. Results suggest that explicit face recognition takes place after the well-known N170 component.



Supplementary figure 1. Description of horizontal (A), vertical (B) and slope (C) orientations rating from 1 to 5 with number 1 used when the face was orientated between the first and second position, 2 between position 2 and 3, 3 when the full-face was presented etc... Details of the results are provided in supplementary Table 1.

Supplementary Table 1. Comparison of visual characteristics across groups of photographs: head orientation (horizontal, vertical, slope, see Supplementary Figure 1); race; other (paraphernalia, teeth being visible or not, presence of an expression (here limited to the presence of a smile or not since it was the only expression shown), and gaze direction (if the face is looking in the eyes or not). A chi-square test was applied (between famous vs non-famous face) for statistical comparisons. Values are reported in proportion (%) and the * sign corresponds to a significant p value. The ps here were uncorrected for multiple comparisons. Three comparisons were found to be significant. Please note however that they are related to small percentage of photographs.

	Rate/ Category	Famous (%)	Non-Famous (%)	<i>p</i>
Horizontal	1	3.0	4.2	0.5
	2	15.3	15.7	0.9
	3	59.2	61.2	0.8
	4	20.7	14.5	0.02*
	5	1.8	4.4	0.08
Vertical	1	0.7	0.8	0.9
	2	3.4	7.4	0.04*
	3	91.5	88.9	0.3
	4	4.4	2.7	0.2
	5	0.0	0.2	0.5
Slope	1	0.4	0.9	0.4
	2	3.3	2.3	0.3
	3	88.9	90.5	0.7
	4	7.4	5.9	0.3
	5	0.0	0.4	0.3
Race	Caucasian	93.3	91.9	0.4
	Black	6.3	4.1	0.1
	Asian	0.4	4.0	0.004*
Other	Paraphernalia	35.2	34.4	0.5
	Gaze	51.8	62.4	0.2
	Smile	52.2	61.7	0.3
	Teeth	41.5	53.7	0.06

References

- Aggleton, J.P., Brown, M.W., 1999. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci* 22, 425–444; discussion 444–489.
- Anaki, D., Bentin, S., 2009. Familiarity effects on categorization levels of faces and objects. *Cognition* 111, 144–149.
- Anaki, D., Zion-Golombic, E., Bentin, S., 2007. Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *Neuroimage* 37, 1407–1416.
- Baird, L.M., Burton, A.M., 2008. The bilateral advantage for famous faces: interhemispheric communication or competition? *Neuropsychologia* 46, 1581–1587.
- Barbeau, E.J., Felician, O., Joubert, S., Sontheimer, A., Ceccaldi, M., Poncet, M., 2005. Preserved visual recognition memory in an amnesic patient with hippocampal lesions. *Hippocampus* 15, 587–596.
- Barbeau, E.J., Pariente, J., Felician, O., Puel, M., 2011. Visual recognition memory: a double anatomo-functional dissociation. *Hippocampus* 21, 929–934.
- Barbeau, E.J., Taylor, M.J., Regis, J., Marquis, P., Chauvel, P., Liégeois-Chauvel, C., 2008. Spatio temporal dynamics of face recognition. *Cereb. Cortex* 18, 997–1009.
- Barragan-Jason, G., Lachat, F., Barbeau, E.J., 2012. How fast is famous face recognition? *Front. Psychology* 3, 454.
- Bentin, S., Deouell, L.Y., 2000. Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cogn Neuropsychol* 17, 35–55.
- Boehm, S.G., Paller, K.A., 2006. Do I know you? Insights into memory for faces from brain potentials. *Clin EEG Neurosci* 37, 322–329.
- Brambati, S.M., Benoit, S., Monetta, L., Belleville, S., Joubert, S., 2010. The role of the left anterior temporal lobe in the semantic processing of famous faces. *Neuroimage* 53, 674–681.
- Bruce, V., Young, A., 1986. Understanding face recognition. *Br J Psychol* 77 (Pt 3), 305–327.
- Caharel, S., Bernard, C., Thibaut, F., Haouzir, S., Di Maggio-Clozel, C., Allio, G., Fouldrin, G., Petit, M., Lalonde, R., Rebaï, M., 2007. The effects of familiarity and emotional expression on

face processing examined by ERPs in patients with schizophrenia. *Schizophr. Res.* 95, 186–196.

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 Cogn* 59, 96–100.

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.

Dehaene, S., Changeux, J.-P., 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.

DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? *Neuron* 73, 415–434.

Dietl, T., Trautner, P., Staedtgen, M., Vannuchi, M., Mecklinger, A., Grunwald, T., Clusmann, H., Elger, C.E., Kurthen, M., 2005. Processing of famous faces and medial temporal lobe event-related potentials: a depth electrode study. *Neuroimage* 25, 401–407.

Eimer, M., 2000. Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin Neurophysiol* 111, 694–705.

Eimer, M., Gosling, A., Duchaine, B., 2012. Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain* 135, 542–554.

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. *Brain Res Cogn Brain Res* 24, 663–673.

Herzmann, G., Schweinberger, S.R., Sommer, W., Jentsch, I., 2004. What's special about personally familiar faces? A multimodal approach. *Psychophysiology* 41, 688–701.

Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.

Holdstock, J.S., Parslow, D.M., Morris, R.G., Fleming, S., Abrahams, S., Denby, C., Montaldi, D., Mayes, A.R., 2008. Two case studies illustrating how relatively selective hippocampal lesions in humans can have quite different effects on memory. *Hippocampus* 18, 679–691.

Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb. Cortex* 14, 132–142.

Jacques, C., Rossion, B., 2006. The speed of individual face categorization. *Psychol Sci* 17, 485–492.

Jemel, B., Pisani, M., Calabria, M., Crommelinck, M., Bruyer, R., 2003. 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, 431–446.

Joubert, S., Felician, O., Barbeau, E., Ranjeva, J.-P., Christophe, M., Didic, M., Poncet, M., Ceccaldi, M., 2006. The right temporal lobe variant of frontotemporal dementia: cognitive and neuroanatomical profile of three patients. *J. Neurol.* 253, 1447–1458.

Joubert, S., Felician, O., Barbeau, E., Sontheimer, A., Guedj, E., Ceccaldi, M., Poncet, M., 2004. Progressive prosopagnosia: clinical and neuroimaging results. *Neurology* 63, 1962–1965.

Kampf, M., Nachson, I., Babkoff, H., 2002. A serial test of the laterality of familiar face recognition. *Brain Cogn* 50, 35–50.

Kaufmann, J.M., Schweinberger, S.R., Burton, A.M., 2009. N250 ERP correlates of the acquisition of face representations across different images. *J Cogn Neurosci* 21, 625–641.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62, 621–647.

Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.

Maillard, L., Barbeau, E.J., Baumann, C., Koessler, L., Bénar, C., Chauvel, P., Liégeois-Chauvel, C., 2011. From perception to recognition memory: time course and lateralization of neural substrates of word and abstract picture processing. *J Cogn Neurosci* 23, 782–800.

Manns, J.R., Hopkins, R.O., Reed, J.M., Kitchener, E.G., Squire, L.R., 2003. Recognition memory and the human hippocampus. *Neuron* 37, 171–180.

Mayes, A.R., Holdstock, J.S., Isaac, C.L., Hunkin, N.M., Roberts, N., 2002. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus* 12, 325–340.

Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., Koch, C., 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* 28, 8865–8872.

Pfütze, 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, 140–160.

Pickering, E.C., Schweinberger, S.R., 2003. N200, N250r, and N400 event-related brain potentials reveal three loci of repetition priming for familiar names. *J Exp Psychol Learn Mem Cogn* 29, 1298–1311.

Puce, A., Allison, T., McCarthy, G., 1999. Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458.

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., Caharel, S., Rossion, B., 2011. The speed of recognition of personally familiar faces. *Perception* 40, 437–449.

Rossion, B., Caharel, S., 2011. ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Res.* 51, 1297–1311.

Rossion, B., Jacques, C., 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39, 1959–1979.

Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2007. Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 36, 843–862.

Rousselet, G.A., Macé, M.J.-M., Fabre-Thorpe, M., 2003. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J Vis* 3, 440–455.

Rousselet, G.A., Pernet, C.R., 2011. Quantifying the Time Course of Visual Object Processing Using ERPs: It's Time to Up the Game. *Front Psychol* 2, 107.

Sagan, C., 1995. *The Demon-Haunted World - Science as a Candle in the Dark*. New York: Random House. ISBN 0-394-53512-X.

Scholte, H.S., Ghebreab, S., Waldorp, L., Smeulders, A.W.M., Lamme, V.A.F., 2009. Brain responses strongly correlate with Weibull image statistics when processing natural images. *J Vis* 9, 29.1–15.

Schweinberger, S.R., Huddy, V., Burton, A.M., 2004. N250r: a face-selective brain response to stimulus repetitions. *Neuroreport* 15, 1501–1505.

Schweinberger, S.R., Pickering, E.C., Burton, A.M., Kaufmann, J.M., 2002a. Human brain potential correlates of repetition priming in face and name recognition. *Neuropsychologia* 40, 2057–2073.

Schweinberger, S.R., Pickering, E.C., Jentsch, I., Burton, A.M., Kaufmann, J.M., 2002b. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Res Cogn Brain Res* 14, 398–409.

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 Cogn Neurosci* 18, 1488–1497.

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

Todd, R.M., Lewis, M.D., Meusel, L.-A., Zelazo, P.D., 2008. The time course of social-emotional processing in early childhood: ERP responses to facial affect and personal familiarity in a Go-Nogo task. *Neuropsychologia* 46, 595–613.

Torralba, A., Oliva, A., 2003. Statistics of natural image categories. *Network* 14, 391–412.

Trautner, P., Dietl, T., Staedtgen, M., Mecklinger, A., Grunwald, T., Elger, C.E., Kurthen, M., 2004. Recognition of famous faces in the medial temporal lobe: an invasive ERP study. *Neurology* 63, 1203–1208.

Tulving, E., 1985. Memory and consciousness. *Canadian Psychology/Psychologie Canadienne* 26, 1.

VanRullen, R., Thorpe, S.J., 2001. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception* 30, 655–668.

Wild-Wall, N., Dimigen, O., Sommer, W., 2008. Interaction of facial expressions and familiarity: ERP evidence. *Biol Psychol* 77, 138–149.

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.

Zheng, X., Mondloch, C.J., Segalowitz, S.J., 2012. The timing of individual face recognition in the brain. *Neuropsychologia* 50, 1451–1461.

5 Discussion générale et perspectives

La reconnaissance des visages familiers, un traitement rapide ?

A l'âge adulte, chacun de nous connaît des centaines, si ce n'est des milliers de personnes différentes, ces chiffres ne tenant même pas compte de l'avènement des réseaux sociaux virtuels tels que Facebook ou MSN. Le visage d'une personne joue un rôle clé dans cette reconnaissance.

Malgré son apparente complexité, la reconnaissance des visages est supposée être un processus rapide et efficace. Cependant, la vitesse de reconnaissance d'un visage familier restait inconnue, les données de la littérature étant très variables.

L'objectif principal de cette thèse était de déterminer cette vitesse au niveau comportemental et électrophysiologique afin de s'en servir pour contraindre les modèles de reconnaissance des visages. Un second objectif était de comparer cette vitesse de reconnaissance aux latences obtenues dans des tâches de catégorisation, tâches supposées rapides, nous permettant ainsi d'affiner nos hypothèses sur les mécanismes cérébraux sous-jacents.

5.1.1 A quelle vitesse reconnaît-on un visage familier ?

5.1.1.1 Un temps de réaction incompressible de 360 ms

L'objectif principal de nos travaux était de déterminer la vitesse de la reconnaissance de type bottom-up des visages familiers (visages célèbres vs visages inconnus). La vitesse minimale de la reconnaissance des visages familiers est de 360-390 ms.

Comment avons-nous déterminé cette valeur ? Les résultats de la tâche de go/no-go montrent un temps minimum d'environ 470 ms à travers les sujets (440 ms à travers les essais) dans l'Article 1 et de 555 ms (460 ms à travers les essais) dans l'Article 2. Cependant, l'analyse individuelle des temps de réaction minimaux révèle que quatre sujets sur 59 (deux sujets dans l'Article 1 et deux dans l'Article 2) sont capables de réaliser la tâche dès 360-390

ms. Ainsi, afin de vérifier si ces temps de réaction plus rapides sont anecdotiques ou réellement représentatifs, nous avons développé une tâche d'apprentissage (Article 1) et un nouveau paradigme (Speed and Accuracy Boosting procedure, SAB , voir 2.1.1.2) dans le but de pousser les sujets à répondre le plus rapidement possible.

La tâche d'apprentissage consistait à faire apprendre les stimuli cibles (visages célèbres) à sept sujets (cinq sujets dits normaux choisis aléatoirement et les deux sujets rapides, Article 1) qui réalisaient ensuite quatre sessions additionnelles de la tâche de go/no-go. Les résultats de cette première expérience complémentaire montrent que tous les sujets sont capables de répondre en 390 ms (Figure 39). De manière intéressante, les deux sujets qui étaient spontanément rapides semblent quand à eux incapables d'améliorer leur temps de réaction, une diminution de seulement 30 ms étant observée entre la première et la dernière séance (soit un TR minimum de 360 ms). Ainsi, il semblerait que la reconnaissance ne puisse pas être effectuée avant 360-390 ms.

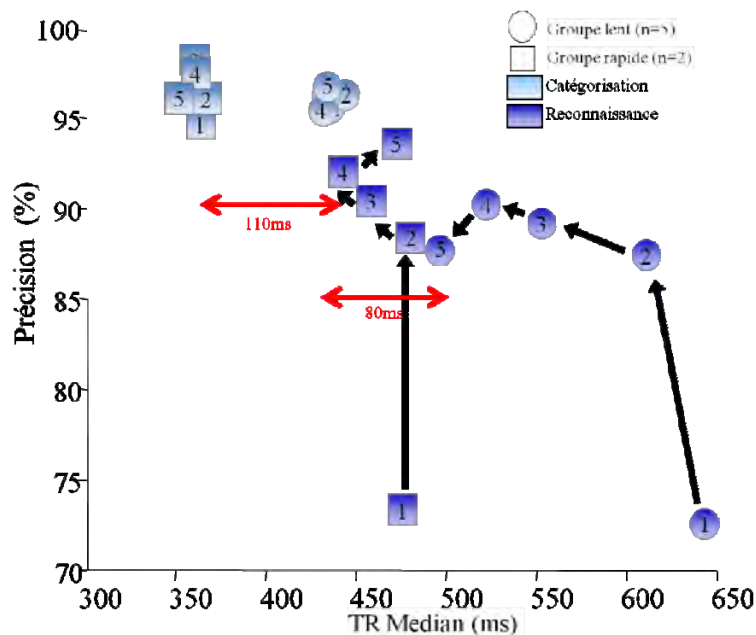


Figure 39 Résultats de la tâche d'apprentissage réalisée dans l'Article 1 représentant la précision en fonction du temps de réaction médian obtenus dans la tâche de reconnaissance en bleu foncé et dans la tâche de catégorisation superordonnée en bleu clair. Les cercles représentent les performances des sujets lents (dits normaux), les carrés, celles des sujets rapides. Chaque chiffre réfère à une session (la session 1 correspondant à la première session de la tâche de go/no-go de l'expérience 1 de l'Article 1). Dans la tâche de reconnaissance, une amélioration significative de la précision est observée pour les deux groupes au cours des sessions. Cependant, seuls les sujets lents améliorent significativement leurs temps de réaction, les sujets rapides n'améliorant leur temps que de 30 ms. Au contraire, aucune amélioration de la précision ou

des temps de réaction n'est observée pour la tâche de catégorisation superordonnée. Ainsi, un coût temporel de 80 ms au minimum est observé entre les deux conditions même après un apprentissage intensif. Cependant, puisque la répétition des stimuli (4 fois lors de l'apprentissage et 5 fois lors de la tâche de go no-go) remettait en question notre situation bottom-up (pré-activation d'indices contenus dans la photographie qui pourraient aider à la reconnaissance), nous avons développé un second protocole permettant de contraindre fortement les sujets à répondre rapidement d'emblée.

La SAB a donc été utilisée. Afin de déterminer une limite temporelle robuste à la reconnaissance des visages célèbres, un large nombre de sujets (n=101) a été testé. Encore une fois, malgré les hautes contraintes temporelle et le large nombre de sujets, les résultats révèlent que les sujets ne peuvent pas aller plus vite que 360-390 ms pour reconnaître un visage familier. De plus, la procédure de SAB semble très efficace pour contraindre les sujets à répondre rapidement car cette fois-ci, environ 37 % des sujets (32 sujets sur 87) répondent avant 420 ms contre 7 % (4 sujets sur 59) dans la tâche de go/no-go.

L'ensemble de ces travaux comportementaux permet donc de conclure de manière robuste qu'un temps incompressible de 360-390 ms est nécessaire pour reconnaître un visage familier dans une situation de reconnaissance bottom-up, ceci dans le cadre des conditions que nous avons utilisées bien entendu.

5.1.1.2 Observation de deux stratégies comportementales: familiarité vs identification

Une stratégie rapide, l'autre lente

Un deuxième point important suggéré par nos travaux est l'utilisation de deux stratégies comportementales différentes, l'une rapide, l'autre lente. En effet, les deux premières études réalisées en go/no-go et en SAB ([Article 1 et 2](#)) montrent que les sujets sont capables de reconnaître les visages familiers en 360-390 ms au minimum après la présentation du stimulus. Cependant, si l'on considère les résultats obtenus en go/no-go, la majorité des sujets (plus de 90%) utilise une stratégie beaucoup plus lente (100 ms de plus en moyenne). Il est important de noter qu'il ne s'agit en aucun cas d'un speed accuracy trade-off puisque les sujets rapides ont une précision comparable aux autres sujets. Cette stratégie lente,

quoique d'efficacité similaire, semble adoptée spontanément par nos sujets lorsqu'on leur demande de reconnaître des visages familiers.

Il semble cependant que notre procédure SAB permette de forcer nos sujets à utiliser la stratégie dite rapide. De plus, il est également intéressant de constater que l'analyse des résultats individuels obtenus dans la tâche de reconnaissance en SAB montre une corrélation inverse au speed accuracy trade-off, c'est-à-dire une précision plus importante pour les sujets qui répondent rapidement contre une précision moindre pour les sujets plus lents. Ce dernier résultat suggère que les sujets étant capables d'utiliser la stratégie la plus rapide réussissent également mieux la tâche. Comment peut-on interpréter ces deux stratégies ? Que nous apprennent-elles sur les mécanismes sous-jacents ?

Familiarité vs identification

Nous observons que nos sujets peuvent utiliser une stratégie rapide et une stratégie plus lente. Comme décrit dans la partie 1.1.3.2, la mémoire de reconnaissance pour les objets visuels serait supportée par deux processus : la familiarité et la recollection (Mandler, 1980; Yonelinas et Levy, 2002; Yonelinas et al., 2010), les réponses basées sur la familiarité supposées être automatiques et rapides, celles basées sur la recollection difficiles et lentes (Juola et al., 1971; Brown et Aggleton, 2001; Rugg et Curran, 2007; Staresina et al., 2012; Besson et al., sous presse). Ainsi, nous pouvons suggérer, bien que nous ne l'ayons pas démontré formellement, que la stratégie rapide serait basée sur l'accès à la familiarité avec le visage tandis que la seconde stratégie serait le résultat de processus de vérification ou d'identification permettant d'accéder aux informations relatives à la célébrité à reconnaître (métier, nom...). En effet, les travaux de Bruce et Young (1986) suggèrent que les sujets ne seraient pas satisfaits d'une reconnaissance accompagnée d'un simple sentiment de familiarité pour la personne mais auraient besoin d'accéder à son identité. Ainsi, notre latence minimale de 360 ms serait le reflet de réponses basées sur la familiarité tandis que nos réponses plus tardives seraient plutôt le résultats de processus liés à l'identification et/ou à l'activation d'informations relatives à la personne.

Dans une certaine mesure, on peut supposer que pour réussir à reconnaître des visages familiers sous les contraintes imposées par la procédure SAB, les sujets sont obligés de baser

leurs réponses sur la familiarité plutôt que l'identité. Les sujets échouant à cette tâche (car trop lents) ne parviendraient sans doute pas à répondre sans accéder à l'identité. Il serait intéressant d'investiguer cette hypothèse plus en détails dans de futures recherches en demandant par exemple au sujet d'indiquer après chacune de ses réponses s'il accédait à l'identité de la personne ou non. Cette donnée, bien que subjective, nous permettrait peut-être de clarifier ce point.

✓ *Dans ce premier point de discussion, nous avons déterminé que reconnaître un visage familier dans des conditions bottom-up nécessitait 360-390 ms au minimum, temps incompressible malgré un apprentissage intensif ou de très fortes contraintes temporelles. Cependant, il semble que deux stratégies différentes puissent être adoptées par nos sujets, une stratégie rapide probablement basée sur la familiarité, une plus lente impliquant probablement l'accès à l'identité et aux informations sur la personne. Qu'en est-il au niveau cérébral ? A quelle latence notre cerveau est-il capable de reconnaître des visages célèbres parmi des visages inconnus ? Le second point de cette discussion portera sur nos latences de décodage du signal EEG enregistré pendant la tâche de reconnaissance rapide et sur leurs liens avec la décision. Nous confronterons également nos latences avec la littérature sur les visages en électrophysiologie.*

5.1.2 A quelle latence notre cerveau reconnaît-il un visage familier ?

5.1.2.1 Une latence électrophysiologique minimale de la reconnaissance des visages familiers dès 200 ms

Nos analyses du signal EEG en MVPA dans la tâche de reconnaissance de visages familiers de type bottom-up en go/no-go montrent que les visages célèbres peuvent être décodés à 230 ms en moyenne, voire 200 ms si l'on considère les réponses les plus rapides ([Article 5](#)).

Ces latences de décodage concordent parfaitement avec nos latences comportementales. En effet, si l'on considère que 100-130 ms sont nécessaires pour produire une réponse motrice (Kalaska et Crammond, 1992 ; VanRullen et Thorpe, 2001), on peut faire l'hypothèse que ce comportement à 360 ms ([Article 1, 2 et 3](#)) soit sous-tendu par une activité neurophysiologique qui se déroulerait à 230-260 ms après la présentation du stimulus, soit exactement aux latences de 230 ms rapportées dans l'[Article 5](#).

5.1.2.2 Des latences en décodage prédictives du comportement

Encore plus convaincant, au-delà d'une simple concordance, nous avons montré dans l'Article 5 une réelle corrélation entre comportement et activité neurale : une forte corrélation est observée entre les temps de réaction minimaux et le début du décodage. Ces observations suggèrent d'une part que l'activité cérébrale décodée à ces latences sont à l'origine du comportement des sujets et d'autre part que les réponses rapides des sujets sont décodées 90 ms plus tôt (dès 200 ms environ) par rapport aux réponses lentes (290 ms environ). Ceci ajoute un argument fort en faveur de l'existence de deux stratégies, une rapide et une lente, dont l'activité cérébrale associée est décodée à des latences différentes.

5.1.2.3 Que nous apprennent ces latences électrophysiologiques ?

Plusieurs hypothèses peuvent être formulées à propos de ces latences électrophysiologiques, notamment d'après les études EEG utilisant la méthode des potentiels évoqués.

Premièrement, cette latence électrophysiologique minimale à 200-230 ms n'est pas compatible avec l'hypothèse d'un traitement précoce de la familiarité associé à la N170 ou à des composantes encore plus précoces (voir partie 1.2.3). En effet, cette discordance peut être expliquée par le fait que nos stimuli ne sont jamais répétés et que nous n'avons pas utilisé de tâche passive mais une tâche nécessitant de reconnaître explicitement le visage présenté. En effet, il semblerait qu'aucune modulation précédent 200 ms ne soit observée lors d'une tâche explicite de reconnaissance (Gosling et Eimer, 2011).

Au contraire, nos latences électrophysiologiques les plus précoces rapportées ici sont en accord avec différentes activités évoquées rapportées à 200-250 ms. Une première hypothèse serait que cette latence correspondrait au début de la composante N250 retrouvée dans des tâches de reconnaissance explicite de visages célèbres et généralement associée au processus de familiarité avec le visage (N250, voir 1.2.3).

En effet, comme mentionné précédemment, nous avons fait l'hypothèse que la stratégie rapide soit en fait basée sur la familiarité, alors que la stratégie lente serait le résultat d'un accès à l'identité. Il semblerait bien que les réponses les plus rapides de chaque sujet soient

également décodées plus tôt que celles des sujets lents. Le décodage plus tardif des réponses à 290 ms pourrait quant à lui être le résultat d'activités liées à l'activation d'informations sémantiques sur la personne (comme la N400 par exemple, Bentin et Deouell (2000)). Il serait très intéressant de tester cette hypothèse en utilisant de manière concomitante SAB et EEG. Ainsi, comme cette tâche est supposée favoriser les réponses basées sur la familiarité, le décodage du signal EEG réalisé à travers les sujets pourrait peut-être atteindre les 200 ms à travers les sujets (contre 230 ms dans les présents travaux).

Une seconde interprétation serait que cette activité décodée à 200 ms soit le reflet de processus de discrimination nécessaires pour réaliser une tâche nécessitant de traiter des exemplaires très similaires entre eux. En effet, d'autres études ont montré que lorsque des objets visuels devaient être reconnus au niveau subordonné (ex : colley au lieu de chien), une modulation de la N250 était observée (Scott et al., 2006; Tanaka et al., 2006). Cette modulation a été interprétée par le fait que ce type de tâche nécessitait une augmentation de la discrimination qui agirait directement sur la N250. Ainsi, notre activité à 200 ms pourrait également refléter une discrimination plus fine des visages, à un niveau individuel.

Enfin, une dernière hypothèse serait que notre tâche de reconnaissance explicite nécessiterait un accès à des processus de méta-mémoire (« je sais que je connais ce visage », Tulving, 1985) qui nécessiterait un accès à la conscience. En effet, une composante négative appelée la VAN (Visual Awareness Negativity) est rapportée à environ 200 ms lorsque l'on compare entre eux des stimuli visuels perçus consciemment ou non (revue de Railo et al., 2011). Cette hypothèse n'est pas en contradiction avec les autres et serait également intéressante à tester dans de futures recherches.

✓ *Dans cette partie, nous avons rappelé que les visages célèbres pouvaient être décodés des visages inconnus dès 200 ms après la présentation de l'image dans une tâche de go/no-go. Cette activité électrophysiologique à 200 ms pourrait être le résultat de plusieurs processus (familiarité, discrimination, méta-conscience) permettant l'émergence d'une représentation consciente du visage familier tandis que l'activité plus tardive rapportée à 290 ms serait le reflet de l'accès aux informations sémantiques relatives à la personne. Quelles hypothèses peut-on formuler à propos des mécanismes impliqués ? La*

partie suivante comparera nos latences comportementales et électrophysiologiques en reconnaissance à des tâches dites rapides plutôt bien caractérisées. Cette comparaison permettra de discuter des différentes alternatives en confrontant nos latences aux modèles théoriques. Nous pourrions ainsi présenter nos hypothèses sur les mécanismes impliqués dans la reconnaissance des visages familiers.

Reconnaissance vs catégorisation : un coût temporel révélateur

5.2.1 La reconnaissance d'un visage familier nécessite 100-200 ms de plus que sa catégorisation

A la question, la reconnaissance des visages familiers est-elle rapide ? La réponse est non, pas vraiment ! L'ensemble des résultats comportementaux ([Article 1, 2 et 3](#)) montrent en effet que reconnaître spontanément des visages célèbres parmi des inconnus en situation bottom-up nécessite en moyenne 180 ms (range : 117-270 à travers les études) de plus que des tâches dites rapides de catégorisation superordonnée (moyenne des minimums : 277, 250 ms à travers les essais) ou de catégorisation du genre (moyenne des minimums : 308 ms, 280 ms à travers les essais) ([Figure 40](#)). De plus, la comparaison des temps de réaction minimaux entre une reconnaissance top-down (plusieurs photographies d'une seule célébrité parmi des inconnus) et une reconnaissance bottom-up (plusieurs célébrités différentes parmi des inconnus) montre une différence du même ordre (150 ms environ, [Article 3](#)).

Il semble important de souligner que nos tâches contrôles de reconnaissance top-down et de catégorisation du genre ne diffèrent pas en latences (300 contre 308 ms), soit seulement 30 ms de plus que la catégorisation superordonnée (visages humains vs visages animaux). Ce dernier point suggère fortement que la reconnaissance top-down pourrait être une forme de catégorisation basée sur des mécanismes de pré-activation similaires à ceux impliqués dans la catégorisation (activation d'un template, Tsao et Livingstone, 2009). En d'autres termes, la reconnaissance top-down serait un type de catégorisation contrairement à la reconnaissance bottom-up.

		Détection	Reconnaissance	Coût temporel	
Comportement	Etude 1: Exp.1	285	467	182	
	Etude 1: Exp.2		402	117	
	Etude 2: Exp.1		555	270	
	Etude 2: Exp.2		420	135	
	TR minimum	Etude 3	270	450	180
		<i>Moyenne</i>	277	459	177
		<i>Min</i>	270	402	117
		<i>Max</i>	285	555	270
		<i>SD</i>	11	59	59
	EEG surface	Etude 4	80		
Décodage minimal	Etude 5		230	150	

Figure 40 Tableau récapitulatif des latences minimales comportementales (médiane des TR minimum) et électrophysiologiques (début du décodage moyen) rapportées dans les différentes études dans les tâches de détection et de reconnaissance. Le coût temporel correspond à la différence entre les latences en reconnaissance et les latences en détection. Ce coût temporel comportemental varie de 117 à 270 ms et présente une moyenne de 177 ms. De plus, il est important de noter que la latence obtenue en reconnaissance top-down est de 300 ms tandis que la latence minimum en catégorisation du genre est de 308 ms, latences comparables aux latences obtenues en détection.

Cette première analyse comportementale suggère que la reconnaissance des visages familiers impliquerait des mécanismes additionnels à la catégorisation (genre ou superordonnée) ou à la reconnaissance top-down qui nécessite environ 100-150 ms de plus. Qu'en est-il au niveau cérébral ?

5.2.2 Une différence électrophysiologique de 120-150 ms entre la catégorisation et la reconnaissance

L'analyse essai par essai en MVPA montre que le signal électrophysiologique des essais correspondant à la catégorie des visages humains peut être correctement différencié des essais correspondant à la catégorie des visages animaux dès 80 ms après la présentation du stimulus dans une tâche de catégorisation superordonnée (Article 4). Au contraire, lors d'une tâche de reconnaissance de visages familiers de type bottom-up, le signal électrophysiologique des essais correspondant aux visages célèbres est correctement différencié des essais correspondant aux visages inconnus seulement après 230 ms (Article 5), soit 150 ms plus tardivement que la catégorisation superordonnée. Cette différence entre les latences de décodage concorde parfaitement avec notre différence rapportée au niveau comportemental (Figure 40).

Au-delà d'une simple concordance, nous avons également investigué s'il existait une corrélation entre les résultats comportementaux des sujets et les latences comportementales. Dans la tâche de catégorisation superordonnée, les temps de réaction minimaux et le début du décodage ne sont pas corrélés. Par contre, une analyse par quartiles permettant de décoder le signal électrophysiologique à chaque milliseconde pour les réponses dont le temps de réaction appartient au premier, second, troisième ou dernier quartile, montre une relation avec le comportement après 160 ms (Article 4). Ce résultat suggère que l'activité décodée à 80 ms n'est pas prédictive du comportement du sujet et que seule l'activité électrophysiologique disponible après 160 ms est liée avec le comportement de catégorisation rapide.

En ce qui concerne la tâche de reconnaissance des visages familiers, une corrélation est observée entre comportement et activité neuronale. Comme mentionné précédemment dans la discussion (partie 5.1.2), les latences en décodage sont prédictives de la vitesse de la reconnaissance de chacun des sujets et modulées par la stratégie utilisée.

✓ *Nous avons donc montré qu'un coût temporel de 100-150 ms était observé entre la reconnaissance et la catégorisation au niveau comportemental et électrophysiologique. Ces résultats suggèrent l'implication de mécanismes additionnels. D'autre part, l'activité décodée dans la tâche de reconnaissance est directement liée au comportement, contrairement à la tâche de détection suggérant une différence qualitative des processus sous-jacents. Dans la prochaine partie, nous allons développer nos arguments sur l'implication de mécanismes distincts et proposerons une interprétation possible aux résultats rapportés dans cette thèse à travers l'élaboration d'un modèle théorique.*

Vers l'hypothèse d'un traitement massivement parallèle

Au-delà d'un temps additionnel important, l'ensemble de nos résultats suggèrent fortement que la reconnaissance et la détection des visages sont basées sur des mécanismes cérébraux distincts.

5.3.1 Reconnaissance vs catégorisation : des mécanismes distincts ?

Plusieurs arguments soutiennent l'hypothèse selon laquelle la reconnaissance et la catégorisation seraient sous-tendues par des mécanismes différents.

Comme le suggère les données de la littérature (pour revue, voir Fabre-Thorpe, 2011), nos latences en catégorisation sont très reproductibles à travers les sujets et à travers nos études (entre 240-290 ms, [Article 1, 2 et 3](#)). De plus, ces latences ne semblent pas pouvoir être améliorées par l'apprentissage. En effet, même si ces résultats ne sont pas présentés dans l'[Article 1](#), lors de la tâche d'apprentissage présentée précédemment, les sujets apprenaient également les images cibles de la condition superordonnée (visages inconnus) puis réalisaient quatre sessions supplémentaires de la tâche de go/no-go (Figure 39). Les résultats montrent que les sujets n'améliorent ni leur précision, ni leur temps de réaction. Ces temps de réaction minimaux ne semblent pas non plus améliorés avec une forte contrainte temporelle (SAB) confirmant ainsi l'hypothèse d'un mécanisme ultra-rapide et efficace d'emblée (Thorpe et al., 2006).

De plus, comme mentionné précédemment, la latence de décodage à 80 ms n'est pas liée au comportement du sujet, une relation étant observée seulement après 160 ms. Cette latence est en accord de nombreux travaux qui suggèrent qu'une étape cruciale de traitement des visages c'est-à-dire l'accès à une représentation haut niveau de la catégorie serait possible à cette latence (N170, Rossion et Caharel, 2011 ; voir VanRullen et Thorpe, 2001 pour la catégorie des objets, ou plus récemment Johnson et Olshausen, 2003).

Au contraire, les latences en reconnaissance semblent variables à travers les sujets. En effet certains sujets utilisent une stratégie lente alors que d'autres sont plus rapides ([Article 1 et 2](#)). Les temps de réaction peuvent être améliorés de manière significative par

l'apprentissage (Figure 39) ou de fortes contraintes temporelles (SAB, [Article 2](#)). Enfin, l'activité décodée en MVPA est directement corrélée au comportement du sujet.

De plus, les résultats du calcul du biais (Snoggrass et Corwin, 1988 ; [Article 1](#)) montrent que les sujets sont plus conservateurs lorsqu'il s'agit de reconnaître un visage, comparé aux tâches de catégorisation où ils semblent au contraire plus libéraux.

En plus d'être beaucoup plus lente, la reconnaissance bottom-up serait donc qualitativement différente de la catégorisation.

5.3.2 Qu'est-ce qui différencie la catégorisation de la reconnaissance bottom-up ?

5.3.2.1 Les mécanismes impliqués dans la catégorisation visuelle.

Catégoriser un objet visuel au niveau superordonné est possible très rapidement, dès 240-290 ms après la présentation du stimulus en réponse manuelle (Fabre-Thorpe et al., 2001 ; Fabre-Thorpe, 2011 ; Macé et al., 2009 ; Joubert et al., 2007 ; Rousselet et al., 2003 ; VanRullen et Thorpe, 2001) et même dès 100 ms lorsqu'un protocole de choix saccadique est utilisé (Crouzet et al., 2010). Cette rapidité de traitement alliée à la robustesse des temps de réaction rapportée à travers les études a mené à l'hypothèse que la catégorisation pourrait être effectuée en utilisant un traitement purement feedforward. En effet, sachant qu'environ 100-130 ms sont nécessaires pour produire une réponse motrice (20 ms s'il s'agit de saccades) (Kalaska et Crammond, 1992 ; VanRullen et Thorpe, 2001) , l'activité neuronale à l'origine des réponses devraient donc être enregistrées à environ 80-130 ms. Dans ce sens, le modèle de DiCarlo et al. (2012) propose un modèle purement feedforward dans lequel environ 10 aires (10 ms entre chaque aire) seraient recrutées afin d'atteindre une représentation haut niveau de l'objet dans IT dès 100 ms. Ce modèle concorde parfaitement avec nos latences à 80 ms obtenues en décodage dans la tâche de catégorisation superordonnée ([Article 4](#)).

Au contraire, même après un apprentissage intensif ou l'utilisation de fortes contraintes temporelles, le temps de réaction minimum en reconnaissance est toujours de 360-390 ms soit au moins 80-110 ms de plus que la catégorisation ([Article 1](#)). Nos données

électrophysiologiques obtenues dans les [Articles 4 et 5](#) révèlent en effet que le début du décodage dans la tâche de détection est significatif dès 80 ms, soit 120-150 ms plus tôt que dans la tâche de reconnaissance dont le décodage est rapporté à 230 ms. Que se passe-t-il pendant ces 100 ms additionnelles ?

5.3.2.2 Des mécanismes qui nécessitent 100 ms de plus que la catégorisation

Nous proposons de confronter nos latences aux trois grands modèles décrits en introduction (1.2.2) et de proposer une interprétation de nos résultats.

Plusieurs hypothèses peuvent être formulées. Une première hypothèse serait que la reconnaissance des visages serait basée sur les mêmes mécanismes feedforward que la détection mais nécessiteraient le recrutement d'aires de plus haut niveau dans la voie visuelle ventrale. En effet, les représentations visuelles étant de plus en plus invariantes au fur et à mesure que l'information progresse dans la voie ventrale, il n'est pas exclu que le temps additionnel rapporté ici soit dû à la nécessité de recruter certaines aires supplémentaires appartenant au lobe temporal. En effet, le cortex perirhinal, le cortex enthorhinal et l'hippocampe (Diana et al., 2007 ; Bowles et al., 2007) semblent nécessaires dans des tâches de mémoire de reconnaissance. Plus particulièrement, le cortex perirhinal semble être une structure impliquée dans les processus de familiarité (Aggleton et Brown, 2006; Barbeau et al., 2005, 2011). Enfin, le pôle temporal qui entraîne une agnosie des personnes lorsqu'il est lésé pourrait également participer à la reconnaissance (Joubert, 2004, 2006) (Figure 41).

Une seconde hypothèse est basée sur le modèle de Rossion et al. (2008) découlant du modèle d'Haxby (2000) (présentés en détails en section 1.2.2.2). L'auteur propose l'implication de boucles récurrentes entre la FFA et la OFA permettant d'individualiser le visage. Dans le cadre de son modèle, le temps additionnel que l'on rapporte ici pourrait être dû à ces allers retours entre la FFA et la OFA permettant d'affiner la représentation et d'individualiser le visage.

Une dernière alternative, qui se base sur le modèle de Barbeau et al. (2008) suggère que le temps additionnel soit dû à l'activation massivement parallèle de l'ensemble des structures appartenant à la voie visuelle ventrale. En effet, des activités précoces ont été rapportées

dès 130 ms dans le gyrus frontal (Barbeau et al., 2008, Bar et al., 2006) et une activation massive parallèle a été observée dans toute la voie ventrale à 240 ms pendant une tâche de reconnaissance de visages célèbres chez des patients épileptiques implantés (Barbeau et al., 2008). De même, les travaux récents en potentiels évoqués de Jemel et al. (2010) (partie 1.2.3.2) suggèrent l'existence de mécanismes complexes qui permettraient d'optimiser la reconnaissance des visages à travers différents processus mis en jeu de manière interactive et en parallèle plutôt qu'un traitement purement hiérarchique. Ces trois modèles sont présentés en Figure 41.

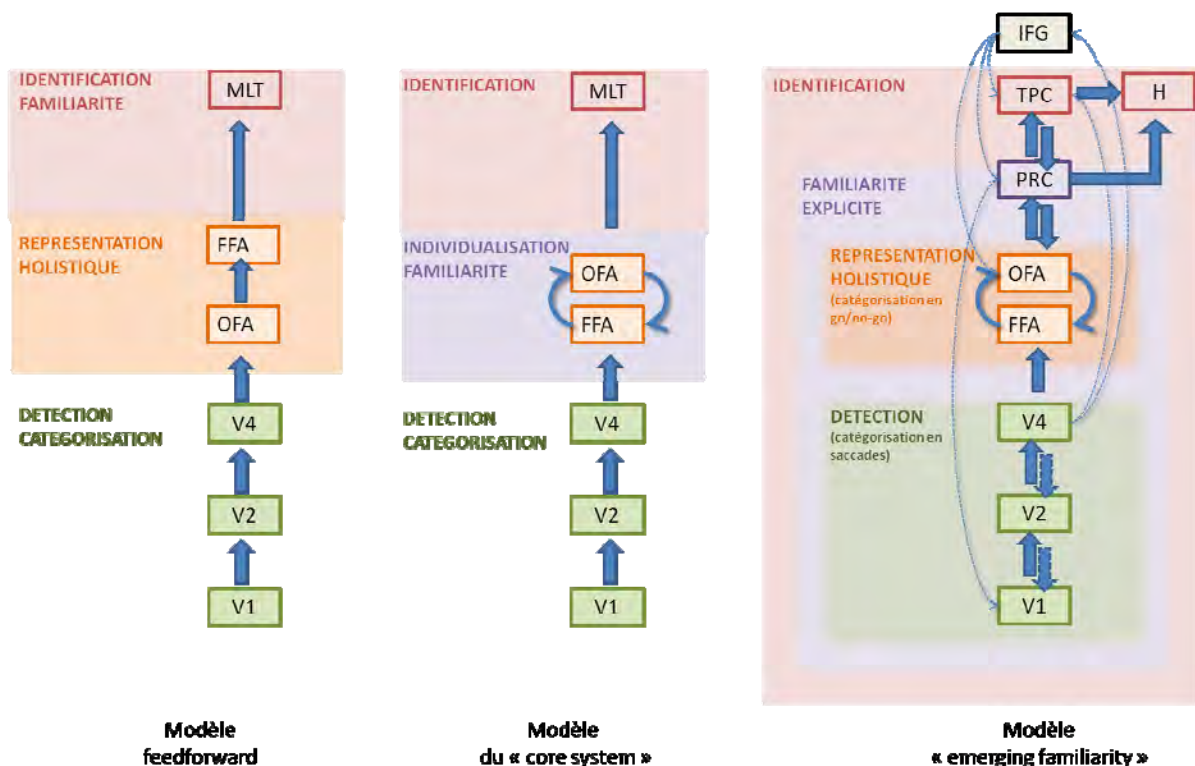


Figure 41 Présentation des modèles de la familiarité avec les visages dans une reconnaissance de type bottom-up. Le premier modèle, inspiré du modèle de Dicarlo et al. (2012), propose une reconnaissance basée sur des mécanismes purement feedforward où la familiarité (cadre rose) serait possible après l'activation successive de plusieurs aires incluant les régions appartenant au lobe temporal médian (MLT). Un second modèle, suggéré par Rossion (2008) propose que la familiarité soit le résultat de boucles récurrentes au sein du core network (cadre violet). Enfin, le troisième modèle que nous proposons est appelé le modèle de la familiarité émergente (« emerging familiarity ») qui serait le résultat de boucles récurrentes permettant d'élaborer et de maintenir une représentation à chaque niveau d'intégration. PRC : cortex périrhinal, TPC : Pôle temporal, IFG : gyrus frontal inférieur, H : hippocampe, FFA : face fusiform area, OFA : occipital face area.

5.3.3 Le modèle d' « emerging familiarity »

Notre opinion concernant ces différentes hypothèses favorise le troisième modèle.

En effet, si l'on considère le modèle de Dicarlo, 100 ms additionnelles correspondraient à l'activation de 10 relais supplémentaires. Pourtant, des connexions directes entre IT, le cortex perirhinal et le pôle temporal ont été rapportées (Suzuki et Amaral, 1994a, 1994b) n'ajoutant de ce fait qu'un temps de seulement 30-40 ms à l'information pour atteindre ces régions d'intérêt. Ainsi, un traitement purement feedforward semble inadapté.

Considérons maintenant le modèle de Rossion (2008). L'idée selon laquelle une représentation haut niveau du visage soit nécessaire pour reconnaître un visage familier n'est pas nouvelle, l'inversion du stimulus entraînant une dégradation importante des performances en reconnaissance de visage célèbre (Valentine, 1988). Nos résultats viennent confirmer cette idée, les sujets étant incapables de réaliser la tâche de reconnaissance en condition bottom-up lorsque les visages sont inversés ([Article 3](#)). Cependant, ce système de représentation holistique individuelle serait associé à des activités différentielles rapportées entre 140-180 ms (N170), activités 60 à 90 ms plus précoces (140 ms vs 230 ms) que les latences en décodage rapportées dans nos travaux. Une modulation non-significative de la précision du décodage des visages célèbres vs des visages inconnus est cependant observée pendant la fenêtre de la N170 dans notre tâche de reconnaissance bottom-up reflétant peut-être l'accès à cette représentation holistique individuelle du visage. Cette activité liée à des boucles entre la FFA et la OFA serait donc nécessaire mais pas suffisante pour reconnaître un visage familier dans des conditions bottom-up. De plus ce modèle n'explique pas pourquoi de fortes activations du cortex périrhinal ont été rapportées dans des tâches de reconnaissance de visages célèbres (Trautner et al., 2004 ; Dietl et al., 2005 ; Barbeau et al., 2008).

Finalement, nous proposons une amélioration du modèle suggéré par Barbeau et al. (2008) La reconnaissance des visages impliquerait l'ensemble de la voie visuelle ventrale activant des processus parallèles nécessitant des communications de type feedback locales mais également entre les aires visuelles et temporelles.

5.3.3.1 L'émergence d'une familiarité explicite

Plus particulièrement, l'information liée à la présentation d'une image contenant un visage serait tout d'abord véhiculée de manière feedforward. Dès que cette première vague d'activité atteint les régions de V4 et IFG (100-130 ms), le visage peut être détecté et localisé permettant l'élaboration de saccades rapides vers les visages (Crouzet et al., 2010 ; Kirchner et Thorpe, 2006). La détection ne nécessiterait pas l'accès à une représentation holistique du visage (représentée par un cadre vert).

Il serait ensuite nécessaire d'attendre l'activation de la FFA puis de la OFA (170 ms) pour obtenir une représentation haut niveau du visage, le réel percept qui permettrait d'accéder et de manipuler cette représentation afin, par exemple, de la mémoriser (encodage) ou de la discriminer d'un second visage. Cette étape apparaît comme une étape de transition dans notre [Article 4](#) (plateau dans le decoding apparaissant pendant la fenêtre de la N170), faisant partie intégrante du processus. Cette étape pourrait permettre de créer une représentation holistique complète du visage (comme suggéré par Rossion, 2008) qui pourrait ensuite être reconnu, mémorisé ou dans notre cas catégorisé rapidement (go/no-go) (cadre orange sur le modèle).

En parallèle, l'information suivrait son cours jusqu'au cortex perirhinal. Ce modèle n'est pas séquentiel. Autrement dit, l'information ne ferait pas que transiter mais déclencherait l'activation de plusieurs boucles d'activités permettant de garder actives les différentes représentations tout au long de la voie ventrale comme proposé par Bullier (2001) ou Lee et Mumford (2003) (notion de black-board).

L'implication d'allers-retours entre la FFA et la OFA (Rossion, 2008) est tout à fait compatible avec ce modèle permettant de traiter les traits individuels du visage. Cependant dans notre modèle, nous proposons que la familiarité explicite ne soit possible qu'après l'activation de toute la voie ventrale y compris du cortex périrhinal.

L'émergence d'une représentation méta-consciente du visage familier (« je connais ce visage ») permettant la reconnaissance explicite du visage comme familier, ne serait possible que via l'activation de boucles de type feedback locales et à distances dans les différentes aires appartenant à la voie visuelle ventrale (Figure 41).

Un dernier point reste cependant discutée. Dans ce modèle, nous suggérons que la catégorisation superordonnée nécessiterait d'accéder à une représentation holistique. Cependant, les résultats de l'Article 3 montre que l'inversion des stimuli dans la tâche de catégorisation superordonnée ne modifie la précision ou le temps de réponse des sujets comparé à la présentation des visages à l'endroit. Ces résultats viennent confirmer les résultats de Rousselet et al. (2003). Ce premier point suggère que la catégorisation superordonnée ne nécessiterait pas d'accéder à une représentation holistique du visage. Cependant, le lien entre l'activité électrophysiologique décodée et le comportement des sujets rapporté seulement après 160 ms (Article 4) nous amène à nous questionner sur la nature de la représentation activée dans la catégorisation superordonnée. Les résultats obtenus dans la tâche de reconnaissance de type top-down viennent apporter quelques pistes sur ce questionnement. En effet, la reconnaissance de type top-down est seulement 40 ms plus lente que la catégorisation superordonnée et de ce fait beaucoup plus rapide que la reconnaissance bottom-up. De plus, contrairement à la reconnaissance bottom-up, les sujets sont capables de réaliser la reconnaissance top-down lorsque les images sont présentées à l'envers. Cependant, l'inversion des images provoquent une baisse significative de la performance des sujets (augmentation des temps de réaction de 60 ms et diminution de la précision) dans la tâche de reconnaissance top-down. Ainsi, il semblerait que la reconnaissance de type top-down nécessite au moins l'accès à une représentation de premier ordre (deux yeux au dessus d'un nez et d'une bouche, Maurer et al., 2002), l'inversion du stimulus perturbant cette configuration rendrait plus difficile la recherche des caractéristiques du visage pré-activés telles que « un grand nez ou un gros grain de beauté ». Ces hypothèses demeurent purement spéculatives et mériterait grandement d'être adressées dans de futures travaux par exemple en enregistrant l'activité EEG des sujets pendant la tâche de reconnaissance top-down.

5.3.3.2 Qu'en est-il de l'identification ?

Selon les données issues d'enregistrements de l'activité électrophysiologique de surface et intracérébrale, des activités différentielles plus tardives observées après 300-350 ms (N400, Bentin et Deouell (2000)) et même à 600 ms (P600) seraient associées respectivement à la récupération d'informations sémantiques sur la personne, et à la récupération

d'informations contextuelles. Si l'on s'intéresse aux régions cérébrales impliquées, selon les données issues des cas de patients présentant une agnosie des personnes par exemple (cf 1.2.2.2), la récupération des informations relatives à la personne dépendrait du pôle temporal. De plus, l'activation des informations contextuelles, plus précisément liée aux processus de recollection, suggère une implication de l'hippocampe qui ne serait pas nécessaire aux processus de familiarité (Dietl et al., 2005 ; Trautner et al., 2004). Ces différentes notions, notamment de recollection n'ont pas été clairement définies pour le traitement des visages mais on peut supposer que ces deux structures, que sont le pôle temporal et l'hippocampe, ne seraient pas nécessaires pour reconnaître un visage au niveau de la familiarité mais seraient activées seulement pour identifier le visage et accéder aux différentes informations sémantiques et contextuelles relatives à cette personne. Dans ce sens, les travaux de Barbeau et al. (2008, Figure 42) mais également ceux de Trautner et al. (2004) montrent une activation tardive de l'hippocampe (après 400 ms) dans ce genre de tâche (soit plus de 100 ms après le cortex perirhinal).

✓ *Ainsi, nous proposons que la familiarité explicite pour les visages soit sous-tendue par l'activation de la voie ventrale de manière purement néocorticale contrairement à l'identification qui nécessiterait l'activation d'autres structures telles que l'hippocampe. Cette hypothèse mérite grandement d'être investiguée.*

✓ *Dans cette thèse, nous avons montré que la reconnaissance des visages familiers dans des conditions bottom-up était possible seulement après 360 ms au niveau comportemental et dès 200 ms au niveau électrophysiologique, temps beaucoup plus long comparé à des tâches de catégorisation réalisées environ 100-150 ms plus rapidement. La vitesse de la reconnaissance est directement corrélée au comportement contrairement à la vitesse de détection. Nous suggérons donc que contrairement à la catégorisation, la reconnaissance bottom-up ne peut pas être basée sur des mécanismes purement feedforward mais nécessite l'activation en parallèle de l'ensemble de la voie visuelle ventrale. De plus, nos données suggèrent que la familiarité explicite pourrait être le résultat d'une activation purement néocorticale, tandis que l'identification nécessiterait le recrutement d'autres structures appartenant au lobe temporal, comme l'hippocampe ou bien le pôle temporal.*

- ✓ *Un moyen de tester cette hypothèse serait d'analyser avec nos méthodes de MVPA l'activité des structures profondes, via l'implantation d'électrodes intra-cérébrales, pendant nos tâches de reconnaissance et de catégorisation.*

Latences électrophysiologiques intracrâniennes de la reconnaissance des visages célèbres : Etude préliminaire

Un dernier point important qui n'a pas encore été concrètement abordé est la question de montrer d'où proviennent les activités enregistrées en surface à 230 ms ou plus concrètement quelles sont les régions de la voie visuelle ventrale qui sont activées à ces latences lorsqu'un visage célèbre est présenté. Cette information pourrait permettre de valider ou non notre modèle d'« emerging familiarity ».

Au-delà des différentes techniques de localisation de source, une manière directe de répondre à cette question est l'utilisation d'enregistrements de potentiels de champs intracérébraux à l'aide de la méthode de sEEG. Les travaux de Trautner et al. (2004) et de Barbeau et al. (2008) par exemple ont investigué cette question en utilisant une analyse en potentiels évoqués sur du signal électrophysiologique obtenu à l'aide d'électrodes profondes. Ils ont montré une activité différentielle dont l'amplitude était maximale entre les visages célèbres et les visages inconnus dès 240 ms dans le cortex perirhinal. Cette activité semblait débiter plus tôt dans le cortex périrhinal que dans l'hippocampe activé plus tardivement. De plus, l'étude de Barbeau et al. (2008) montre que cette activation est massivement parallèle dans la voie visuelle ventrale (Figure 42).

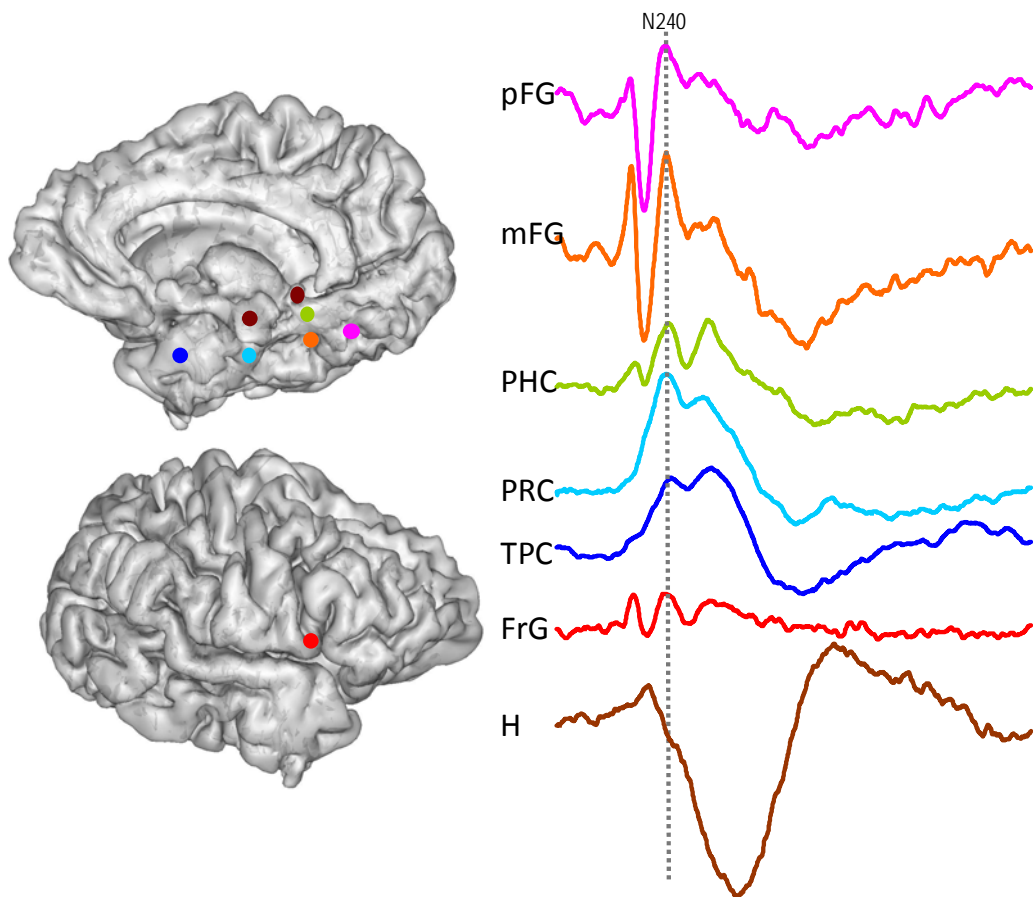


Figure 42 Figure adaptée de Barbeau et al. (2008) montrant une activation massivement parallèle à 240 ms dans toute la voie visuelle ventrale, y compris dans les régions frontales inférieures, en réponse à la présentation de visages célèbres. L'hippocampe quant à lui semble activé plus tardivement.

Ces deux études utilisent la méthode des potentiels évoqués. Notre méthode de MVPA a été largement utilisée pour analyser des données obtenues en IRMf (Haxby et al., 2001 ; Kay et al., 2008) parfois également sur du signal électrophysiologique de surface (Carlson et al., 2011) ou appliquée à du signal obtenu à l'aide de grilles d'électrodes (Electrocorticogramme) implantées à la surface du cortex (Liu et al., 2009), n'autorisant pas l'accès aux structures profondes. Nous avons eu l'opportunité de proposer nos tâches de reconnaissance et de détection rapide de visages célèbres en go/no-go à des patients épileptiques implantés. Ainsi, nous avons pu réaliser le même type d'analyse sur l'activité intracérébrale des patients lors de cette tâche.

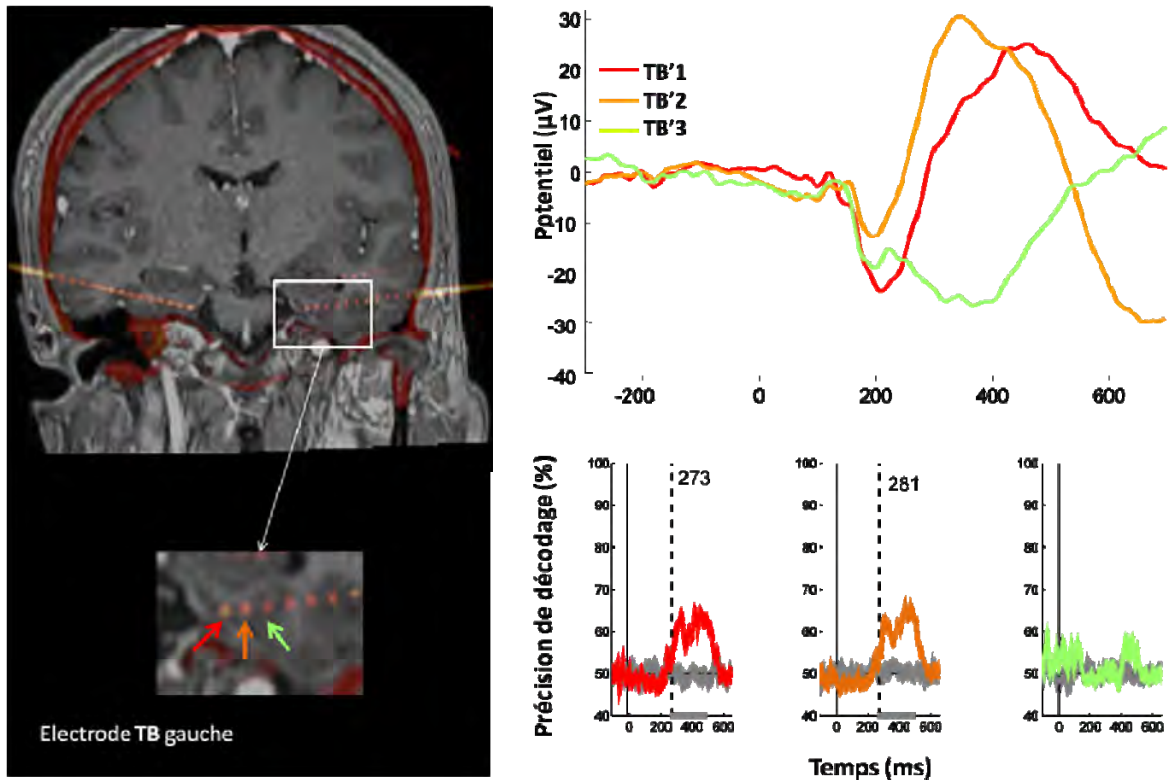


Figure 43 Exemples de réponses neuronales enregistrées sur TB gauche pendant la tâche de reconnaissance de visages célèbres. A gauche, une coupe du recalage des électrodes réalisé à l'aide du logiciel Sysiphe. A droite sont représentés les potentiels évoqués sur l'ensemble des essais (cibles et distracteurs) pour le plot 1 en rouge, le plot 2 en orange et le plot 3 en vert. En bas à droite sont présentées les analyses en décodage sur ces mêmes plots avec en gris le décodage permuté (précision du decodeur en y et le temps en x). Les barres verticales en pointillées désignent la latence minimale significative du décodage : 273 ms sur le plot 1, 281 ms sur le plot 2 et pas de décodage sur le plot 3.

Même si l'analyse des données est encore en cours, les résultats préliminaires révèlent que les latences de décodage au niveau de certains plots d'enregistrement sont en accord avec les latences rapportées en surface. L'activité enregistrée sur ces plots semblent être focale comme en témoigne l'analyse en potentiels évoqués réalisée sur l'ensemble des essais (cibles et distracteurs, voir Figure 43 et Figure 44). On observe que les visages célèbres peuvent être décodés des visages inconnus dès 273-281 ms après la présentation de l'image sur les deux plots les plus profonds de l'électrode TB gauche situés dans le cortex entorhinal. Cette latence semble un peu tardive mais tout de même comparable aux latences électrophysiologiques rapportées en surface. Ces données mériteraient d'être investiguées de plus près en analysant par exemple le lien avec le comportement du patient.

L'analyse du signal enregistré sur un autre patient dans une tâche de détection sur des plots très postérieurs situés dans la région du gyrus fusiforme, montre une activité beaucoup plus précoce (environ 100 ms plus rapide) lorsque l'on contraste les visages humains aux visages animaux. Cette latence électrophysiologique, même si elle est plus tardive que celle rapportée en surface, est tout à fait compatible avec la N200 rapportée en enregistrement intracrânien par Allison et al. (1999) par exemple. Ces analyses méritent d'être approfondies en investigant plusieurs plots situés dans la voie ventrale chez plusieurs patients et de lier cette activité aux performances comportementales.

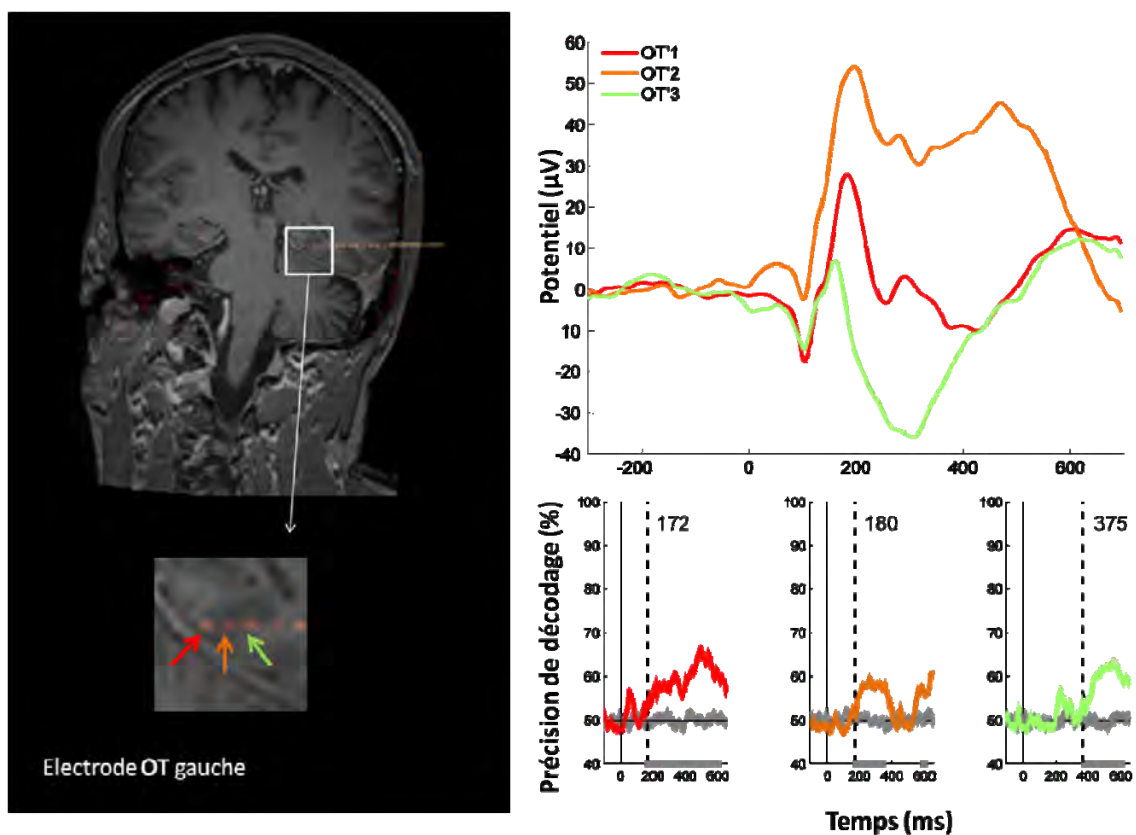


Figure 44 Exemples de réponses électrophysiologiques enregistrées sur OT gauche pendant la tâche de détection. A gauche, une coupe du recalage des électrodes réalisée à l'aide du logiciel Sysiphe. A droite sont représentés les potentiels évoqués sur l'ensemble des essais (cibles et distracteurs) pour le plot 1 en rouge, le plot 2 en orange et le plot 3 en vert. En bas à droite sont présentées les analyses en décodage sur ces mêmes plots avec en gris le décodage permuté (précision du décodeur en y et le temps en x). Les barres verticales en pointillées désignent la latence minimale significative du décodage : 172 ms sur le plot 1, 180 ms sur le plot 2 et un décodage très tardif probablement non lié à la tâche sur le plot 3.

Une autre analyse qui semblerait très intéressante serait d'utiliser le classifieur pour étudier la réelle nature des stimuli. Par exemple, certains des patients ont réalisé une tâche de

détection d'humains parmi des animaux puis une tâche de détection des animaux parmi les humains. En entraînant le classifieur à décoder le signal correspondant aux visages cibles vs animaux distracteurs, puis en utilisant les visages distracteurs vs animaux cibles pendant la phase de test, on pourrait décoder la catégorie du stimulus indépendamment du statut cible ou distracteur du stimulus. Au contraire, en entraînant le classifieur sur les visages cibles vs visages distracteurs et en testant ensuite le classifieur sur les animaux cibles vs animaux distracteurs, on décoderait plutôt le statut cible ou distracteur du stimulus.

En effet, une autre analyse très préliminaire utilisant le signal électrophysiologique de l'ensemble des électrodes implantées chez deux patients, l'un présentant une implantation frontale, l'autre une implantation temporale suggère que la catégorie serait codée dans les régions temporales dès 156 ms tandis que le statut cible ou distracteur semble être codé dans les régions frontales dès 133 ms environ (Figure 45). Ces résultats fournissent des pistes très intéressantes pour comprendre le codage des représentations dans le cerveau humain.

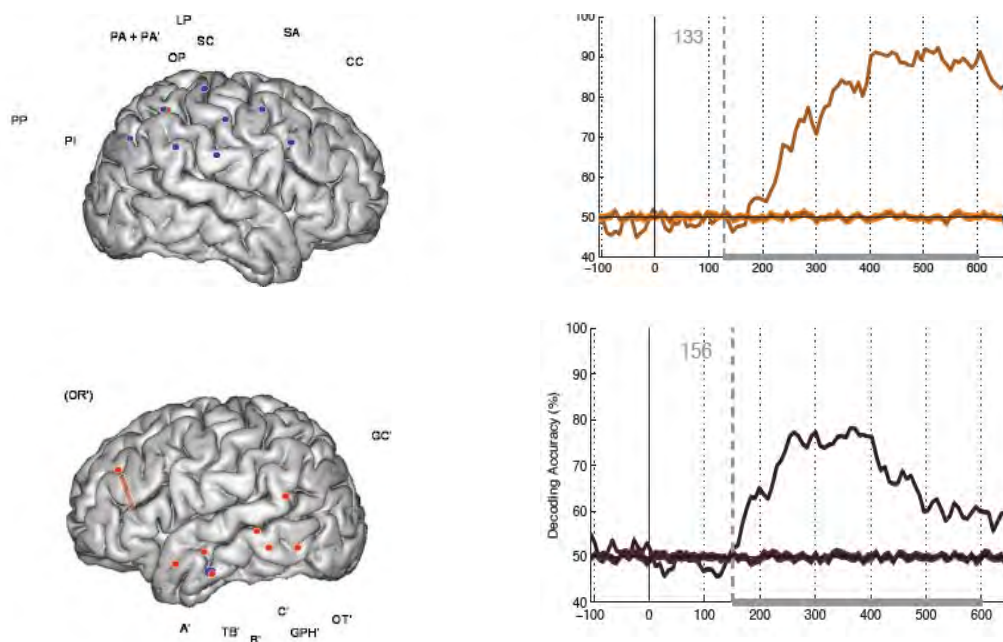


Figure 45 Résultats préliminaires de l'inversion de tâches permettant de décoder le statut cible ou distracteur (en haut) de la catégorie (en bas). A gauche sont présentés les schémas d'implantation approximatif des électrodes profondes avec leurs labels. A droite est présenté le résultat obtenu après la phase de test du décodage.

Pour aller plus loin... Perspectives et conclusion générale

Au cours de cette thèse, nous avons débuté certains travaux permettant d'adresser la question de la reconnaissance des visages d'une toute autre manière. En effet, au-delà de la familiarité avec le visage, l'accès à l'identité et au concept semble être un sujet d'étude tout aussi passionnant. En effet, ces aspects du traitement des personnes nous permettent d'ouvrir une porte vers l'étude de la mémoire en générale.

5.5.1 Le mot sur le bout de la langue : un accès direct à la méta-cognition

Nous avons évoqué l'idée d'un accès à la méta-conscience lorsqu'il s'agit de reconnaître explicitement un visage (« je sais que je connais ce visage »). Cette notion de méta-mémoire nous a intéressé tout particulièrement. Selon Smith (2009), le moyen de prédilection pour étudier la méta-cognition est l'étude d'un phénomène appelé le mot sur le bout de la langue (« tip of the tongue », TOT). Nous avons donc développé une autre étude portant sur la dénomination des visages célèbres, et plus particulièrement sur le manque du nom propre, comme modèle de l'accès à cette méta-mémoire.

Qu'est ce que le manque du nom propre ? Lorsque nous parlons, il arrive parfois que nous nous trouvions dans l'incapacité de produire un mot précis. Ce phénomène, appelé manque du mot, a été étudié dans différentes études en linguistique, en neuropsychologie et également en neuroimagerie (Evrard, 1998 ; Burke et al, 2004). Ces études ont permis de caractériser différents manques du mot (Maril et al, 2005). En effet, nous nous sommes tous déjà retrouvés dans la situation agaçante de chercher en vain un mot que l'on est pourtant sûr de connaître: « Quel est déjà le nom du présentateur de Question pour un Champion, son nom contient Pirse ou peut-être Perse... je suis bloqué... Je ne retrouve pas son nom, je l'ai presque pourtant...sur le bout de la langue ! » L'expérience du phénomène de mot sur le bout de la langue, est définie comme l'incapacité à produire un mot précis que l'on est pourtant certain de connaître et qui semble prêt à ressurgir d'un instant à l'autre (Brown et McNeill, 1966). Le mot sur le bout de la langue est le résultat d'une activation incomplète de la forme phonologique du mot (blocage phonologique). Le sujet peut alors posséder un panel important de connaissances phonologiques (la longueur de l'item, le nombre de syllabes, des fragments du mot ou encore le phonème initial) ou sémantiques à propos du mot recherché sans parvenir à le retrouver (Schwartz, 1999). Au niveau comportemental, le

sujet présente une obnubilation de la conscience et un sentiment de blocage intense, d'imminence et de frustration. Le sujet peut tout à coup retrouver le mot en question, on parle alors de mot sur le bout de la langue résolutif. Dans le cas contraire, un simple indiçage permettra de retrouver le mot recherché.

Une autre forme de manque du mot, appelé sentiment de familiarité pour le nom (FOK pour Feeling Of Knowing en anglais), se caractérise par l'incapacité à rappeler un mot précis tout en s'estimant cependant capable de le reconnaître s'il est présenté (Hart, 1965). Le sujet a tout simplement oublié le mot qu'il a pourtant connu à une époque. Il s'agit d'un défaut d'activation lexicale : l'information est présente mais non accessible et n'est accompagnée que d'une faible activation des informations sémantiques. Un questionnaire à choix multiples ou une ébauche orale permettra le rappel du mot recherché (Maril et al, 2005). Le manque du mot touche préférentiellement la catégorie des noms propres (Cohen, 1990 ; Burke et al, 1991 ; Cohen et Faulkner, 1986). Les noms de personnes dont font partie les personnes célèbres sont une catégorie de noms propres particulièrement touchée (Burke et al., 1988 ; Cohen et Faulkner, 1986).

Comme un manque du mot est par définition caractérisé par des latences anormalement longues en dénomination, nous avons donc développé un protocole permettant une « capture » efficace des différents manques du mot en alliant une étude comportementale (clé vocale enregistrant les temps de réaction) et électrophysiologique (EEG 128). Les sujets devaient dénommer le plus rapidement possible 110 photographies de personnes et 110 photographies d'objets de faibles fréquences dans le langage parlé (Figure 46). Nous voulions ainsi caractériser les latences comportementales des différents manques du mot d'une part (TOT vs FOK). Pour comparer l'état cognitif dans lequel le sujet se trouve lors d'un manque du mot à des tâches de production de mots et de lecture, nous avons aussi utilisé différentes tâches contrôles (tâche de fluence verbale, génération mentale de mot et lecture silencieuse).

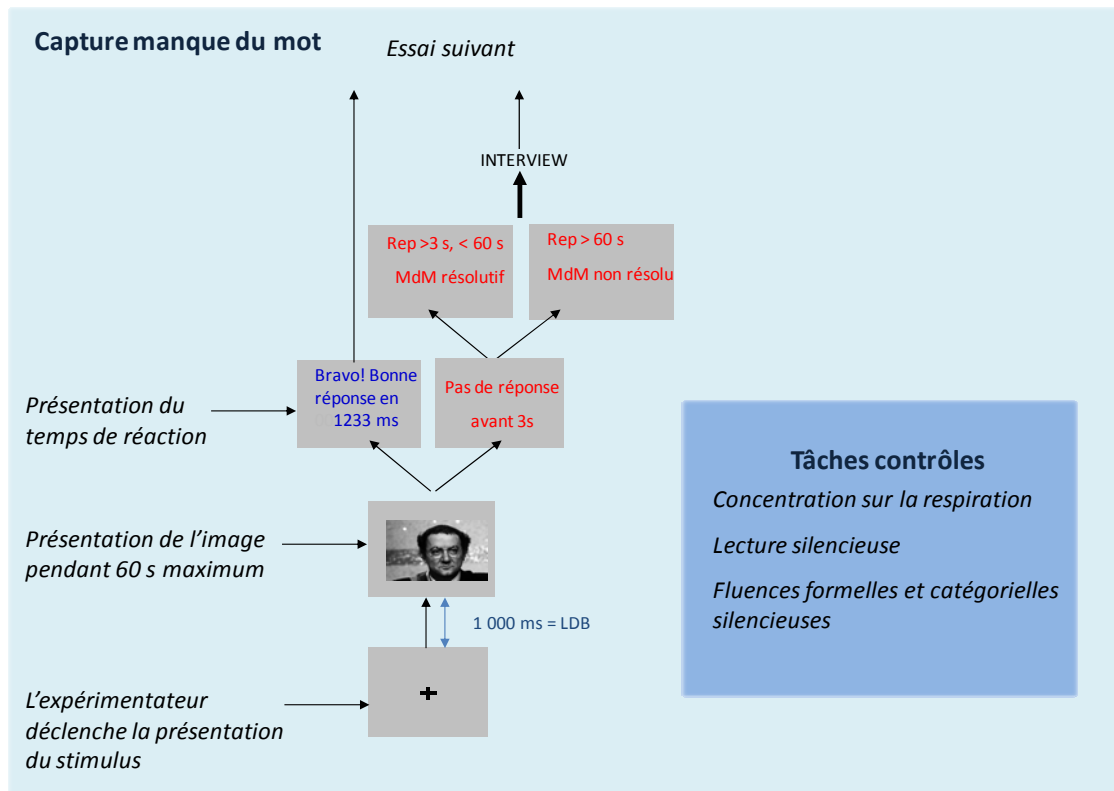


Figure 46 Protocole expérimental de « capture » du manque du mot. La présentation de la croix de fixation est déclenchée par l'expérimentateur. Cette croix de fixation reste à l'écran pendant 1 s suivie de la présentation de l'image. L'image reste à l'écran jusqu'à que sujet déclenche le microphone ou pendant 60 s maximum si le sujet ne répond pas. Un feedback présentant le temps de réaction du sujet est ensuite présenté. Si la réponse est correcte, l'expérimentateur déclenche l'essai suivant. Si ce n'est pas le cas, une interview destinée à déterminer le type de manque du mot est proposée. C'est le sujet lui-même qui détermine s'il est en condition de mot sur le bout de la langue ou de sentiment de familiarité. Avant et après cette expérience, des tâches contrôles de concentration sur la respiration, de lecture silencieuse et de fluence formelle (mots commençant par r, s et p) et catégorielle (animal, fleur) sont proposées au sujet.

Quelques résultats préliminaires sont présentés en Figure 47 et Figure 48. Malgré le faible nombre de participants ($n=10$), les résultats comportementaux montrent que les sujets sont plus rapides pour dénommer que lors d'un manque du mot résolu. De plus, un TOT est plus rapide qu'un FOK ce qui confirme l'hypothèse de Burke et al. (1991) proposant qu'un mot sur le bout de la langue est un blocage de l'accès à la phonologie tandis qu'un sentiment de familiarité est plutôt dû à un défaut d'activation lexical en amont dans le modèle.

Sujets	Nb. de manques du mot	Nb. de réponses correctes	Erreurs
1	26	109	85
2	5	131	84
3	32	130	58
4	25	124	71
5	39	169	12
6	34	176	10
7	13	182	25
8	34	163	23
9	27	175	18
10	29	162	29

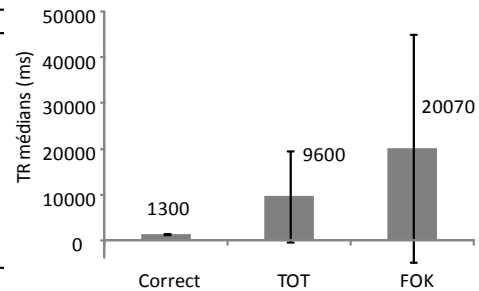


Figure 47 Résultats comportementaux préliminaires de l'expérience sur le manque du mot. A gauche, le nombre de manque du mot, de réponses correctes et d'erreurs (erreurs visuelles, paraphasies) pour chaque sujet est présenté. A droite, les temps de réaction médians pour les bonnes réponses, les mots sur le bout de la langue (TOT) et les sentiments de familiarité (FOK) pour le mot.

Les résultats électrophysiologiques ne montrent pas de différence significative en potentiel évoqué (analyse de pics) entre les manques du mot et les réponses correctes répliquant ainsi les travaux d'une équipe espagnol (Diaz et al., 2007 ; Bujan et al., 2009; Galdo-Alvarez et al, 2009).

Par contre, une analyse très préliminaire en temps-fréquence (décomposition du signal électrophysiologique dans les différentes bandes de fréquences en fonction du temps) montre une activité dans les bandes de fréquences gamma contrairement aux tâches contrôles. Le signal semble cependant très bruité car effectué sur un très faible nombre d'essais (n=10). En effet, une limite majeure de notre protocole est le faible nombre de manques du mot collecté par sujet (20 mots sur le bout de la langue en moyenne contre 6 sentiments de familiarité par sujet sur les 220 images, la majorité étant des manques du mot pour les personnalités).

Ainsi, il serait intéressant d'appliquer notre méthode de décodage essai par essai, poolés à travers les sujets pour tenter d'investiguer la dynamique de cette activité qui semble spécifique aux mots sur le bout de la langue.

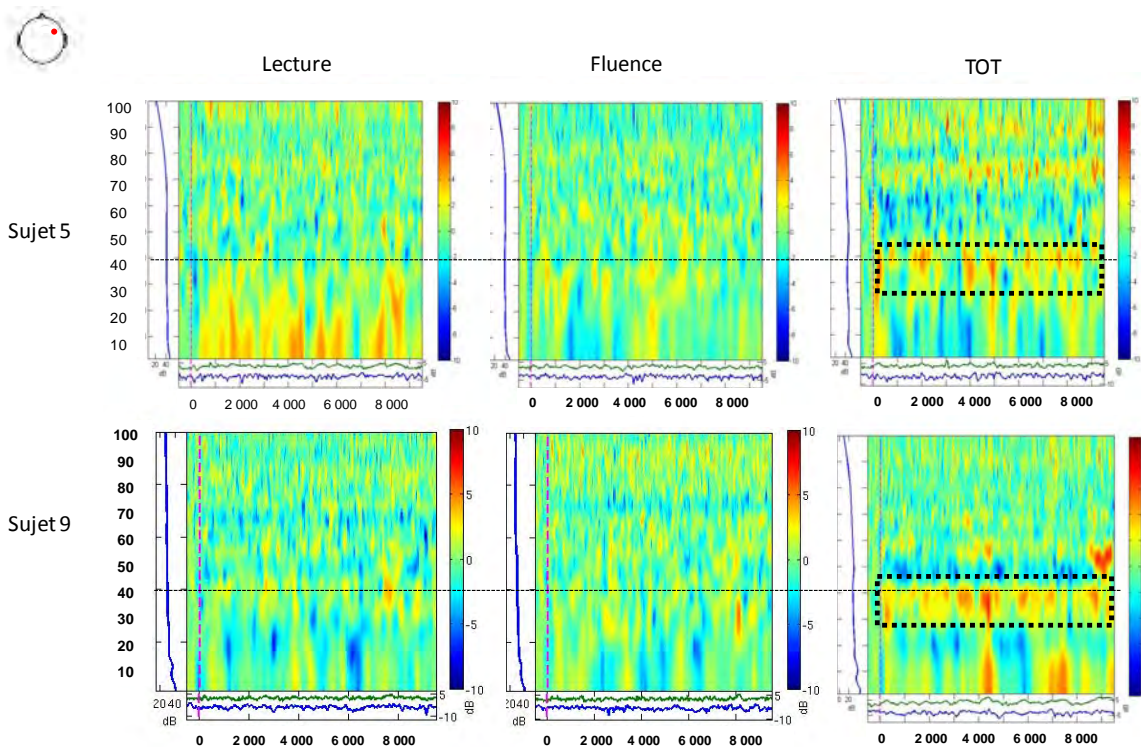


Figure 48 Résultats électrophysiologiques préliminaires de l'expérience sur le manque du mot. L'analyse en temps-fréquence obtenue en moyennant 10 essais de chaque condition pour deux sujets sur une électrode frontale droite (F4) montre de l'activité dans les bandes de fréquences gamma (cadre noir en pointillées) lors d'un mot sur le bout de la langue qui n'est pas observé pendant les tâches contrôles.

De plus, il serait intéressant d'essayer d'améliorer ce protocole en provoquant d'avantage de manque du mot chez les participants. D'après la théorie de Burke (1991, 2004), il semblerait que le manque du mot puisse être provoqué par l'activation accidentelle d'un mot proche sémantiquement (Marceau pour Adajani) ou phonologiquement (Pernaud pour Perrault). Ainsi, il serait intéressant de tester cet effet en réalisant un masquage par des photographies d'autres célébrités entraînant ainsi davantage de manque du mot.

5.5.2 Étudier la personnes sous tous ces aspects : décodage du concept

Dans les différentes études présentées ici, nous avons toujours utilisé des photographies pour étudier la reconnaissance. Cependant, comme le rapporte les patients prosopagnosiques, un grand nombre d'informations autres que le visage peuvent faciliter la reconnaissance. Par exemple, la démarche, les vêtements, la voix, ou encore le nom de la personne, sont autant d'entrées plurimodales permettant d'accéder au concept de la personne.

C'est pourquoi, l'idée d'un dernier protocole que nous avons développé en collaboration avec Maxime Cauchoix est une expérience visant à décoder le concept. Des stimuli particuliers constitués d'extraits de discours, de visages et de noms écrits de différentes célébrités ou d'autres catégories d'objets (véhicules, animal) ont été préparés.

Notre méthode de décodage pourrait être utilisée à travers une seule modalité d'une part mais également à travers les modalités. Par exemple, on pourrait entraîner le classifieur sur les voix des personnes et lors de la phase de test lui demander de classifier les images ou les noms écrits. Encore plus intéressant, il serait passionnant d'également demander au sujet de penser à une catégorie ou à une personne en particulier et d'essayer de décoder cette représentation (imagerie mentale).

5.5.3 Conclusion

Nos travaux suggèrent que la reconnaissance explicite des visages célèbres est un processus lent qui nécessiterait « le travail collaboratif » de l'ensemble de la voie ventrale. Cependant, il semble important de souligner que nos résultats ont été obtenus en utilisant des visages célèbres. La question se pose quant à l'extrapolation à l'ensemble des visages familiers tels que les visages personnellement connus dont la représentation est supposée être plus robuste. Cette question mériterait grandement d'être explorée.

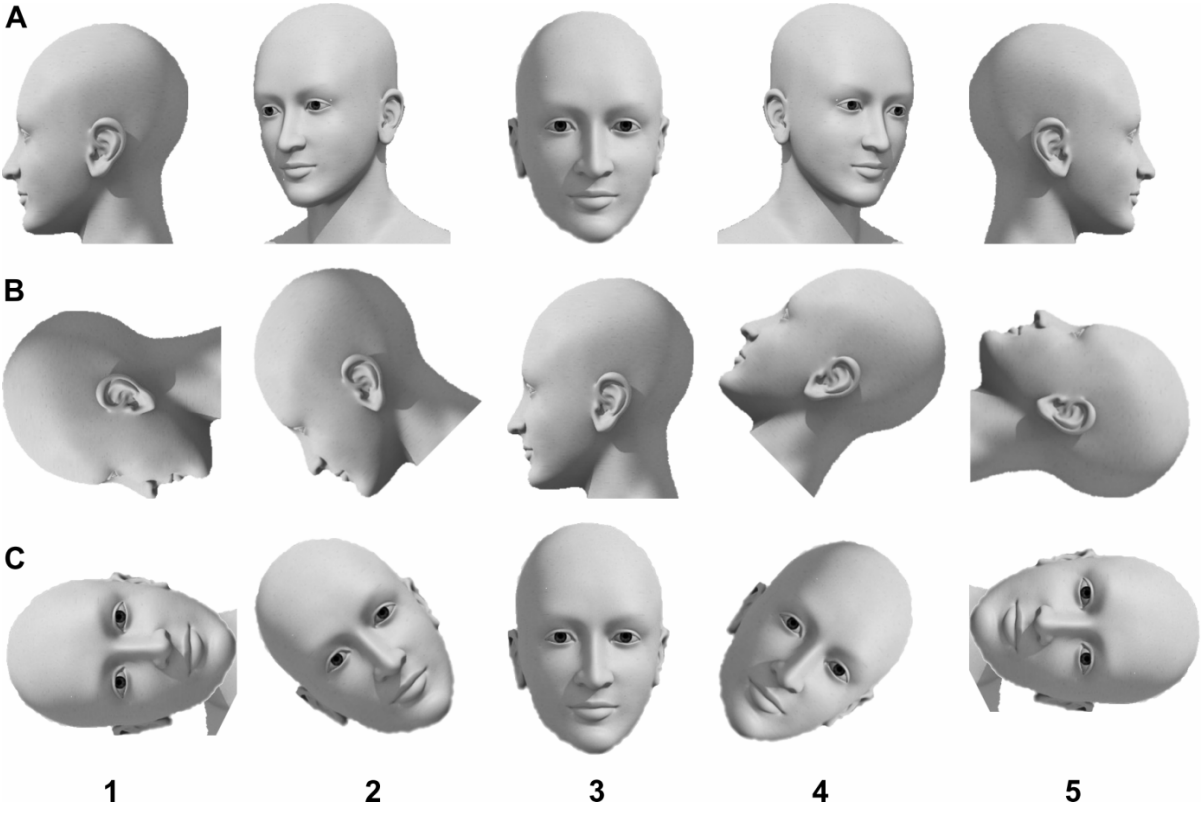
Un second point est que nos résultats ont été obtenus sous contrainte temporelle. Il s'agit d'un avantage si l'on veut déterminer les processus strictement nécessaires, mais les mécanismes que nous avons étudiés ici sont sans doute un peu minimalistes, omettant de prendre en compte différents paramètres relatifs à l'affect, le contexte ou encore plus important, la nécessité d'accéder à l'identité qui semble être une habileté sociale de haute importance dans notre société. Ces différents aspects nécessiteraient d'être adressés dans de futures recherches en comparant par exemple plusieurs catégories de personnes (amie, ennemi, star préférée...) et en contrôlant en direct (peut-être à l'aide de questionnaires) les déclencheurs de cette reconnaissance. Enfin, comme présenté dans deux études préliminaires, la méthode originale de MVPA que nous avons utilisé laisse présager une application beaucoup plus large qui nous permettrait d'étudier différentes fonctions cognitives comme la méta-mémoire, la conscience ou l'imagerie mentale.

Annexes

Annexe 1. Tableau récapitulatif des travaux portant sur la détection et la reconnaissance décrites en section 1.1.3.1. détaillant plusieurs études portant sur la détection et la reconnaissance des visages. Les références, les temps de réaction minimaux (RT min), les temps de réaction moyen ou médian (RT moy ou med), les paradigmes utilisés, le nombre de stimuli, la présence d'un contexte et le nombre de sujets sont rapportés pour chacune des études référencées.

	Références	RTmin.	RTmoy.ou med.	Paradigmes	Nb. de stimuli	Contexte	Sujets
DETECTION	Eimer, 2000		489	Choix forcé	16 humains/16 maisons	sans contexte	26
	Rousselet et al., 2003 (exp1)	260	329	Go nogo	384 humains / 384 (animaux+objets)	en contexte	24
	Rousselet et al., 2003 (exp2)	260	327	Go nogo	192 hum / 192 (animaux+objets)	en contexte	24
	Latinus et Taylor, 2006		522	Choix forcé	45 humains/45 images brouillées	en contexte	28
	Crouzet et al., 2011 (exp1)	100	147	Saccades oculaires	125 humains/125 (animaux, véhicules, objets)	en contexte	8
	Crouzet et al., 2011 (exp2)	100	140	Saccades oculaires	100 humains and 100 véhicules	en contexte	8
RECONNAISSANCE	Anaki et al., 2007		411	Choix forcé	180 célèbres/180 inconnus	sans contexte	24
	Anaki et al., 2009		800	Vérification de catégorie	8 célèbres	sans contexte	28
	Kampf et al., 2002		431	Choix forcé	144 célèbres/ 288 inconnus	sans contexte	96
	Baird et Burton, 2008		775	Saccades oculaires	16 célèbres vs 16 inconnus	sans contexte	26
	Lewis et Ellis, 2000		425	Choix forcé	320 visages de Diana /16 visages de Fergie	sans contexte	4
	Caharel et al., 2005		491	Choix forcé	20 célèbres/20 inconnus (x5)	sans contexte	28
	Caharel et al., 2005		537	Choix forcé	2 visages personnellement connu(x100)/20 visages inconnus (x5)	sans contexte	28
	Herzmann., 2004		875	Choix forcé	15 célèbres/ 30 inconnus	non précisé	24
	Herzmann., 2004		855	Choix forcé	15 personnellement connus/30 inconnus	non précisé	24
	Ramon et al., 2011370		494	Go nogo	26 personnellement connus/26inconnus	sans contexte	17

Annexe 2. Description des orientations horizontales (A), verticales (B) et des inclinaisons (C) utilisées pour les échelles du tableau 1. Les valeurs varient de 1 to 5 où 1 used est utilisé quand le visage est orienté entre la première et seconde position, 2 entre la seconde et mla troisième, etc.



Publications

Articles publiés

1. **Barragan-Jason, G.**, Lachat, F., & Barbeau, E. J. (2012). How fast is famous face recognition? *Frontiers in Psychology*, Perception science serie, 3, 454.
2. **Barragan-Jason, G.**, Besson, G., Barbeau, E.J. Fast and Famous : Looking for the fastest speed at which a face can be recognized, *Frontiers in Psychology*, Cognitive Science serie, 4, 100.

Articles soumis

3. **Barragan-Jason, G.***, Cauchoix *, Barbeau, E.J. The speed of famous face recognition: relating EEG to behavior using MVPA, *soumis*.

Articles en préparation

4. Cauchoix *,M., **Barragan-Jason*, G.**, Serre, T., Barbeau E.J. The neural dynamics of human face detection in the wild, *en préparation*.
5. Besson, G., **Barragan-Jason, G.**, Thorpe, S.J., Puma, S., Barbeau, E.J. Top-down vs bottom-up face recognition, *en préparation*.
6. Condret, V., **Barragan-Jason, G.**, Marlats, F., Denuelle, M., Nespoulous, J.-L., Valton, L., Barbeau, E.J. What is your name again? Proper name retrieval failures characterization in temporal lobe epilepsies, *en préparation*.

Présentations orales à des conférences internationales

Barragan-Jason, G., Valton, L., Serre, T., Barbeau E.J. The dynamics of face processing: from detection to recognition, *European society of cognitive and affective neuroscience (ESCAN)*, Marseille, Mai 2012.

Posters présentés à des conférences internationales

1. **Barragan-Jason, G.**, Lachat, F., Barbeau, E.J. (2011). Rapid recognition of famous faces: Isn't it in fact relatively slow? *Perception 40* European Conference of Visual Perception (ECPV) Abstract Supplement, page 119
2. **Barragan-Jason, G.**, Fabié, A., Barbeau, E.J. (2011). Categorizing faces, recognizing famous faces: Any relationship? *Perception 40* ECPV Abstract Supplement, page 158

3. Besson, G., **Barragan-Jason, G.**, Puma, S., Barbeau, E. J. (2011). Top-down vs bottom-up face identification. *Perception* 40 ECVF Abstract Supplement, page 158.
4. **Barragan-Jason, G.**, Lachat, F., Barbeau, E.J. The time course of individual face processing, *ICOM5*, York, Août 2011.
5. Condret, V., **Barragan-Jason, G.**, Marlats, F., Denuelle, M., Nespoulous, J.-L., Valton, L., Barbeau, E.J. What is your name again? Proper name retrieval failures characterization in temporal lobe epilepsies. *ICOM5*, Août 2011.
6. **Barragan-Jason, G.**, Lachat, F., Barbeau, E.J. (2011). The time course of individual face processing, *Journal of Vision* 11 (11), 593-593.
7. **Barragan-Jason, G.**, Lachat, F., Barbeau, E.J. Is face recognition fast, very fast or relatively slow? *Réunion d'Hiver de la Société de Neuropsychologie de Langue Française (SNLF)*, Toulouse, Décembre 2010.
8. **Barragan-Jason, G.**, Lachat, F., Barbeau E.J. Is face recognition fast, very fast or relatively slow? *Journées IFR jeunes*, Toulouse, Octobre 2010.
9. **Barragan-Jason, G.**, Lachat, F., Barbeau, E.J. The time course of individual face processing, *Second Meeting of the Federation of the European Societies of Neuropsychology*, Amsterdam, Septembre 2010.
10. Condret, V., **Barragan-Jason, G.**, Marlats, F., Denuelle, M., Nespoulous, J.-L., Valton, L., Barbeau, E.J. Proper name retrieval failures characterization in temporal lobe epilepsies, *second Meeting of the Federation of the European Societies of Neuropsychology*, Amsterdam, Septembre 2010.
11. Condret, V., Olivier, R., **Barragan-Jason, G.**, Marlats, F., Nespoulous, J.-L., Barbeau, E.J., Valton, L. Proper name retrieval failures characterization in temporal lobe epilepsies, *12èmes journées françaises de l'épilepsie*, Marseille, Novembre 2009.

Responsabilités et autres activités réalisées au cours de la thèse

- Représentante des étudiants et des post-doctorants au conseil du laboratoire.
- Membre active de l'association des jeunes chercheurs en sciences cognitives InCOGnu (vulgarisation scientifique, interventions lycée, semaine du cerveau).
- Stage d'un mois à l'étranger au « Centre for Speech, Language and the Brain » à Cambridge.

Bibliographie

- Adachi, I., Chou, D.P., Hampton, R.R., 2009. Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Curr. Biol.* 19, 1270–1273.
- Adolphs, R., 2003. Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178.
- Aggleton, J.P., Brown, M.W., 1999. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci* 22, 425–444; discussion 444–489.
- Aggleton, J.P., Brown, M.W., 2006. Interleaving brain systems for episodic and recognition memory. *Trends Cogn. Sci. (Regul. Ed.)* 10, 455–463.
- 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. *Cerebral Cortex* 9, 415–430.
- Anaki, D., Bentin, S., 2009. Familiarity effects on categorization levels of faces and objects. *Cognition* 111, 144.
- Anaki, D., Zion-Golumbic, E., Bentin, S., 2007. Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *Neuroimage* 37, 1407–1416.
- Anderson, B., Mruczek, R.E.B., Kawasaki, K., Sheinberg, D., 2008. Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral cortex (New York, N.Y. : 1991)* 18, 2540–52.
- Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., Thorpe, S.J., 2007. Effects of task requirements on rapid natural scene processing: from common sensory encoding to distinct decisional mechanisms. *J Exp Psychol Hum Percept Perform* 33, 1013–1026.
- Baird, L.M., Burton, A.M., 2008. The bilateral advantage for famous faces: interhemispheric communication or competition? *Neuropsychologia* 46, 1581–1587.
- Balas, B., Cox, D., Conwell, E., 2007. The effect of real-world personal familiarity on the speed of face information processing. *PloS one* 2, e1223.
- Bancaud, J., 1987. [Clinical symptomatology of epileptic seizures of temporal origin]. *Rev. Neurol. (Paris)* 143, 392–400.

- Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Schmidt, A.M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E., 2006a. Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454.
- Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Schmidt, A.M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E., 2006b. Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454.
- Barbarotto, R., Capitani, E., Spinnler, H., Trivelli, C., 1995. Slowly progressive semantic impairment with category specificity. *Neurocase* 1, 107–119.
- Barbeau, E.J., Felician, O., Joubert, S., Sontheimer, A., Ceccaldi, M., Poncet, M., 2005. Preserved visual recognition memory in an amnesic patient with hippocampal lesions. *Hippocampus* 15, 587–596.
- Barbeau, E.J., Joubert, S., Felician, O., 2008a. *Traitement et reconnaissance des visages : du percept à la personne*. Editions Solal.
- Barbeau, E.J., Pariente, J., Felician, O., Puel, M., 2011. Visual recognition memory: a double anatomo-functional dissociation. *Hippocampus* 21, 929–934.
- Barbeau, E.J., Taylor, M.J., Regis, J., Marquis, P., Chauvel, P., Liégeois-Chauvel, C., 2008b. Spatio temporal dynamics of face recognition. *Cereb. Cortex* 18, 997–1009.
- Barragan-Jason, G., Lachat, F., Barbeau, E.J., 2012. How fast is famous face recognition? *Front. Psychology* 3, 454.
- Barrett, S.E., Rugg, M.D., 1989. Event-related potentials and the semantic matching of faces. *Neuropsychologia* 27, 913–922.
- Besson, G., Ceccaldi, M., Didic, M., Barbeau, E.J., 2012. The speed of visual recognition memory. *Visual Cognition* 20, 1131–1152.
- Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological Studies of Face Perception in Humans. *Journal of cognitive neuroscience* 8, 551–565.
- Bentin, S., Deouell, L.Y., 2000. Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cogn Neuropsychol* 17, 35–55.
- Bentin, S., Taylor, M.J., Rousselet, G.A., Itier, R.J., Caldara, R., Schyns, P.G., Jacques, C., Rossion, B., 2007. Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nature Neuroscience* 10, 801–802.
- Bodamer, J., 1947. Die Prosop-Agnosie. *Arch. f. Psychiatr. u. Z. Neur.* 179, 6–53.

- Boehm, S.G., Paller, K.A., 2006. Do I know you? Insights into memory for faces from brain potentials. *Clin EEG Neurosci* 37, 322–329.
- Bötzel, K., Schulze, S., Stodieck, S.R., 1995. Scalp topography and analysis of intracranial sources of face-evoked potentials. *Exp Brain Res* 104, 135–143.
- Bouvier, S.E., Engel, S.A., 2006. Behavioral Deficits and Cortical Damage Loci in Cerebral Achromatopsia. *Cereb. Cortex* 16, 183–191.
- Bowles, B., Crupi, C., Mirsattari, S.M., Pigott, S.E., Parrent, A.G., Pruessner, J.C., Yonelinas, A.P., Köhler, S., 2007. Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16382–16387.
- Brambati, S.M., Benoit, S., Monetta, L., Belleville, S., Joubert, S., 2010. The role of the left anterior temporal lobe in the semantic processing of famous faces. *Neuroimage* 53, 674–681.
- Brédart, S., Bruyer, R., 1994. The cognitive approach to familiar face processing in human subjects. *Behavioural Processes* 33, 213–232.
- Brown, M.W., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* 2, 51–61.
- Brown, R., McNeill, D., 1966. The “tip of the tongue” phenomenon. *Journal of verbal learning and verbal behavior* 5, 325–337.
- Brown, S.D., Dooling, R.J., 1993. Perception of conspecific faces by budgerigars (*Melopsittacus undulatus*): II. Synthetic models. *J Comp Psychol* 107, 48–60.
- Bruce, V., Young, A., 1986. Understanding face recognition. *Br J Psychol* 77 (Pt 3), 305–327.
- Buján, A., Lindín, M., Díaz, F., 2009. Movement related cortical potentials in a face naming task: Influence of the tip-of-the-tongue state. *International Journal of Psychophysiology* 72, 235–245.
- Bullier, J., 2001. Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107.
- Bullier, J., Hupé, J., James, A., Girard, P., 1996. Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology-Paris* 90, 217–220.
- Bullier, J., Hupé, 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.

- Burke, D.M., MacKay, D.G., Worthley, J.S., Wade, E., 1991. On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of memory and language* 30, 542–579.
- Burke, D.M., Shafto, M.A., 2004. Aging and Language Production. *Curr Dir Psychol Sci* 13, 21–24.
- Burke, D.M., Worthley, J., Martin, J., 1988. I'll never forget what's-her-name: Aging and tip of the tongue experiences in everyday life. *Practical aspects of memory: Current research and issues* 2, 113–118.
- Bushneil, I.W.R., Sai, F., Mullin, J.T., 1989. Neonatal recognition of the mother's face. *British Journal of Developmental Psychology* 7, 3–15.
- Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., Rossion, B., 2010. Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia* 48, 4057–4092.
- Buttle, H., Raymond, J.E., 2003. High familiarity enhances visual change detection for face stimuli. *Perception & psychophysics* 65, 1296–306.
- Caharel, S., Bernard, C., Thibaut, F., Haouzir, S., Di Maggio-Clozel, C., Allio, G., Fouldrin, G., Petit, M., Lalonde, R., Rebai, M., 2007. The effects of familiarity and emotional expression on face processing examined by ERPs in patients with schizophrenia. *Schizophr. Res.* 95, 186–196.
- 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 and cognition* 59, 96–100.
- Caharel, S., Jacques, C., D' Arripe, O., Ramon, M., Rossion, B., 2011. Early electrophysiological correlates of adaptation to personally familiar and unfamiliar faces across viewpoint changes. *Brain Res.* 1387, 85–98.
- 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.
- Caldara, R., Schyns, P., Mayer, E., Smith, M.L., Gosselin, F., Rossion, B., 2005. Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *J Cogn Neurosci* 17, 1652–1666.

- Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011a. High temporal resolution decoding of object position and category. *J Vis* 11.
- Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011b. High temporal resolution decoding of object position and category. *J Vis* 11.
- Cohen, G., 1990. Recognition and retrieval of proper names: Age differences in the fan effect. *European Journal of Cognitive Psychology* 2, 193–204.
- Cohen, G., Burke, D.M., 1993. Memory for proper names: a review. *Memory* 1, 249–263.
- Cohen, G., Faulkner, D., 2011. Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology* 4, 187–197.
- Collins, C.E., Airey, D.C., Young, N.A., Leitch, D.B., Kaas, J.H., 2010. Neuron densities vary across and within cortical areas in primates. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15927–15932.
- Connor, C.E., 2005. Friends and grandmothers. *Nature* 435, 1036–1037.
- Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19, 261–270.
- Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision* 10, 1–17.
- Crouzet, S.M., Thorpe, S.J., 2011. Low-level cues and ultra-fast face detection. *Front Psychol* 2, 342.
- Damasio, A.R., Damasio, H., Hoesen, G.W.V., 1982. Prosopagnosia Anatomic basis and behavioral mechanisms. *Neurology* 32, 331–331.
- Dawson, G.D., 1951. A summation technique for detecting small signals in a large irregular background. *J. Physiol. (Lond.)* 115, 2p–3p.
- De Haan, E.H., Young, A.W., Newcombe, F., 1991. Covert and overt recognition in prosopagnosia. *Brain* 114 (Pt 6), 2575–2591.
- De Renzi, E., Perani, D., Carlesimo, G.A., Silveri, M.C., Fazio, F., 1994. Prosopagnosia can be associated with damage confined to the right hemisphere--an MRI and PET study and a review of the literature. *Neuropsychologia* 32, 893–902.
- Dehaene, S., Changeux, J.-P., 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227.

- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Delorme, A., Rousset, G.A., Macé, M.J.-M., Fabre-Thorpe, M., 2004. Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res* 19, 103–113.
- Delorme, A., Thorpe, S.J., 2001. Face identification using one spike per neuron: resistance to image degradations. *Neural Networks* 14, 795–803.
- Desimone, R., Albright, T.D., Gross, C.G., Bruce, C., 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4, 2051–2062.
- Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. *J Exp Psychol Gen* 115, 107–117.
- Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci. (Regul. Ed.)* 11, 379–386.
- Diaz, F., Lindin, M., Galdo-Alvarez, S., Facal, D., Juncos-Rabadán, O., 2007. An event-related potentials study of face identification and naming: The tip-of-the-tongue state. *Psychophysiology* 44, 50–68.
- Dicarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? *Neuron* 73, 415–434.
- Dietl, T., Trautner, P., Staedtgen, M., Vannuchi, M., Mecklinger, A., Grunwald, T., Clusmann, H., Elger, C.E., Kurthen, M., 2005. Processing of famous faces and medial temporal lobe event-related potentials: a depth electrode study. *Neuroimage* 25, 401–407.
- Donders, F.C. (1869). On the speed of mental processes. In W. G. Koster (Ed.), *Attention and Performance II*. *Acta Psychologica*, 30, 412-431. (Original work published in 1868.)
- Dubois, S., Rossion, B., Schiltz, C., Bodart, J.M., Michel, C., Bruyer, R., Crommelinck, M., 1999. Effect of familiarity on the processing of human faces. *NeuroImage* 9, 278–89.
- Eger, E., Schweinberger, S.R., Dolan, R.J., Henson, R.N., 2005. Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. *NeuroImage* 26, 1128–39.
- Eimer, M., 2000a. Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin Neurophysiol* 111, 694–705.

- Eimer, M., 2000b. The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport* 11, 2319–2324.
- Eimer, M., Gosling, A., Duchaine, B., 2012. Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain* 135, 542–554.
- Eimer, M., Holmes, A., 2002. An ERP study on the time course of emotional face processing. *Neuroreport* 13, 427–431.
- Ellis, A.W., Young, A.W., Flude, B.M., 1990. Repetition priming and face processing: priming occurs within the system that responds to the identity of a face. *Q J Exp Psychol A* 42, 495–512.
- Ellis, H.D., Davies, G.M., Shepherd, J.W., 1977. Experimental studies of face identification. *Journal of Criminal Defense* 3, 219–234.
- Ellis, H.D., Shepherd, J.W., Davies, G.M., 1979. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception* 8, 431–439.
- Eriksen, B.A., Eriksen, C.W., Hoffman, J.E., 1986. Recognition memory and attentional selection: serial scanning is not enough. *J Exp Psychol Hum Percept Perform* 12, 476–483.
- Evans, J.J., Higgs, A.J., Antoun, N., Hodges, J.R., 1995. Progressive prosopagnosia associated with selective right temporal lobe atrophy A new syndrome? *Brain* 118, 1–13.
- Evrard, M. (1998). L'accès aux noms de personnes. *Informations in Cognito*, 10, 19-28.
- Fabre-Thorpe, M., 2011. The characteristics and limits of rapid visual categorization. *Front Psychol* 2, 243.
- Fabre-Thorpe, M., Delorme, a, Marlot, C., Thorpe, S., 2001a. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience* 13, 171–80.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., Thorpe, S., 2001b. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci* 13, 171–180.
- Fabre-Thorpe, M., Richard, G., Thorpe, S.J., 1998. Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9, 303–308.
- Fan, R.-E., Chen, P.-H., Lin, C.-J., 2005a. Working Set Selection Using Second Order Information for Training Support Vector Machines. *J. Mach. Learn. Res.* 6, 1889–1918.

- Fan, Y., Shen, D., Davatzikos, C., 2005b. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Med Image Comput Comput Assist Interv* 8, 1–8.
- Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J., 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, 9, 1871–1874.
- Farah, M.J., 1996. Is face recognition “special”? Evidence from neuropsychology. *Behav. Brain Res.* 76, 181–189.
- Farah, M.J., Levinson, K.L., Klein, K.L., 1995. Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* 33, 661–674.
- Farwell, L.A., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol* 70, 510–523.
- Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* 1, 1–47.
- Field, T.M., Cohen, D., Garcia, R., Greenberg, R., 1984. Mother-stranger face discrimination by the newborn. *Infant Behavior and Development* 7, 19–25.
- Freiwald, W.A., Tsao, D.Y., 2010a. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.
- Freiwald, W.A., Tsao, D.Y., 2010b. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.
- Freiwald, W.A., Tsao, D.Y., 2010c. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.
- Fried, I., MacDonald, K.A., Wilson, C.L., 1997. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18, 753–765.
- Gainotti, G., Barbier, A., Marra, C., 2003. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain* 126, 792–803.
- Gainotti, G., Ferraccioli, M., Quaranta, D., Marra, C., 2008. Cross-modal recognition disorders for persons and other unique entities in a patient with right fronto-temporal degeneration. *Cortex* 44, 238–248.
- Galdo-Alvarez, S., Lindín, M., Díaz, F., 2009. The effect of age on event-related potentials (ERP) associated with face naming and with the tip-of-the-tongue (TOT) state. *Biological psychology* 81, 14–23.

- Galton, S.F., 1883. *Inquiries into human faculty and its development*. Macmillan.
- 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.
- George, N., Evans, J., Fiori, N., Davidoff, J., Renault, B., 1996. Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research* 4, 65–76.
- 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. *Brain Res Cogn Brain Res* 24, 663–673.
- George, N., Jemel, B., Fiori, N., Renault, B., 1997. Face and shape repetition effects in humans: A spatio-temporal ERP study. *NeuroReport* 8, 1417–1422.
- Gerson, A.D., Parra, L.C., Sajda, P., 2005. Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage* 28, 342–353.
- Girard, P., Salin, P.A., Bullier, J., 1991. Visual activity in areas V3a and V3 during reversible inactivation of area V1 in the macaque monkey. *J Neurophysiol* 66, 1493–1503.
- Gobbini, M.I., Haxby, J.V., 2006. Neural response to the visual familiarity of faces. *Brain research bulletin* 71, 76–82.
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., Goebel, R., 2011. From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cereb. Cortex* 21, 467–476.
- Goodale, M.A., Westwood, D.A., 2004. An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. *Curr. Opin. Neurobiol.* 14, 203–211.
- Gosling, A., Eimer, M., 2011. An event-related brain potential study of explicit face recognition. *Neuropsychologia* 49, 2736–2745.
- 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.
- 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.
- Halgren, E., Raji, T., Marinkovic, K., Jousmäki, V., Hari, R., 2000. Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb. Cortex* 10, 69–81.
- Hart, J.T., 1965. Memory and the feeling-of-knowing experience. *J Educ Psychol* 56, 208–216.

- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.
- Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., Rapoport, S.I., 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 88, 1621–1625.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends in Cognitive Sciences* 4, 223–233.
- Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* 14, 6336–6353.
- 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, 189–199.
- Hay, D.C., Young, A.W., 1982. The human face. Normality and pathology in cognitive functions 173–202.
- Hegd , J., 2008. Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439.
- Herzmann, G., Schweinberger, S.R., Sommer, W., Jentsch, I., 2004. What’s special about personally familiar faces? A multimodal approach. *Psychophysiology* 41, 688–701.
- Hintzman, D.L., Caulton, D. a, Levitin, D.J., 1998. Retrieval dynamics in recognition and list discrimination: further evidence of separate processes of familiarity and recall. *Memory & cognition* 26, 449–62.
- Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.
- Holdstock, J.S., Parslow, D.M., Morris, R.G., Fleming, S., Abrahams, S., Denby, C., Montaldi, D., Mayes, A.R., 2008. Two case studies illustrating how relatively selective hippocampal lesions in humans can have quite different effects on memory. *Hippocampus* 18, 679–691.
- Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, 8, 9.1–13.

- 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.
- Ishai, A., 2008. Let's face it: it's a cortical network. *Neuroimage* 40, 415–419.
- Ishai, A., Haxby, J.V., Ungerleider, L.G., 2002. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage* 17, 1729–1741.
- Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb. Cortex* 14, 132–142.
- Jacques, C., Rossion, B., 2006. The speed of individual face categorization. *Psychol Sci* 17, 485–492.
- Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., Taylor, E., 2011. Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance* 37, 1824.
- Jeffreys, D.A., 1989a. A face-responsive potential recorded from the human scalp. *Exp Brain Res* 78, 193–202.
- Jeffreys, D.A., 1989b. A face-responsive potential recorded from the human scalp. *Exp Brain Res* 78, 193–202.
- Jeffreys, D.A., 1996. Visual evoked potential evidence for parallel processing of depth- and form-related information in human visual cortex. *Exp Brain Res* 111, 79–99.
- Jemel, B., George, N., Olivares, E., Fiori, N., Renault, B., 1999. Event-related potentials to structural familiar face incongruity processing. *Psychophysiology* 36, 437–452.
- Jemel, B., Pisani, M., Calabria, M., Crommelinck, M., Bruyer, R., 2003. 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, 431–446.
- Jemel, B., Schuller, A.-M., Goffaux, V., 2010. Characterizing the spatio-temporal dynamics of the neural events occurring prior to and up to overt recognition of famous faces. *J Cogn Neurosci* 22, 2289–2305.
- Johnson, J.S., Olshausen, B.A., 2003. Time course of neural signatures of object recognition. *J Vis* 3, 499–512.
- Johnson, M.H., Dziurawiec, S., Ellis, H., Morton, J., 1991. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19.

- Jolicoeur, P., Gluck, M.A., Kosslyn, S.M., 1984. Pictures and names: making the connection. *Cogn Psychol* 16, 243–275.
- Joubert, O.R., Rousselet, G.A., Fize, D., Fabre-Thorpe, M., 2007. Processing scene context: fast categorization and object interference. *Vision Res.* 47, 3286–3297.
- Joubert, S., Felician, O., Barbeau, E., Ranjeva, J.-P., Christophe, M., Didic, M., Poncet, M., Ceccaldi, M., 2006. The right temporal lobe variant of frontotemporal dementia: cognitive and neuroanatomical profile of three patients. *J. Neurol.* 253, 1447–1458.
- Joubert, S., Felician, O., Barbeau, E., Sontheimer, A., Barton, J.J., Ceccaldi, M., Poncet, M., 2003. Impaired configurational processing in a case of progressive prosopagnosia associated with predominant right temporal lobe atrophy. *Brain* 126, 2537–2550.
- Joubert, S., Felician, O., Barbeau, E., Sontheimer, A., Guedj, E., Ceccaldi, M., Poncet, M., 2004a. Progressive prosopagnosia: clinical and neuroimaging results. *Neurology* 63, 1962–1965.
- Joubert, S., Mauries, S., Barbeau, E., Ceccaldi, M., Poncet, M., 2004b. The role of context in remembering familiar persons: Insights from semantic dementia. *Brain and Cognition* 55, 254–261.
- Juola, J.F., Fischler, I., Wood, C.T., Atkinson, R.C., 1971. Recognition time for information stored in long-term memory. *Perception & Psychophysics* 10, 8–14.
- Kalaska, J.F., Crammond, D.J., 1992. Cerebral cortical mechanisms of reaching movements. *Science* 255, 1517–1523.
- Kampf, M., Nachson, I., Babkoff, H., 2002. A serial test of the laterality of familiar face recognition. *Brain Cogn* 50, 35–50.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- Kaufmann, J.M., Schweinberger, S.R., Burton, A.M., 2009. N250 ERP correlates of the acquisition of face representations across different images. *J Cogn Neurosci* 21, 625–641.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452, 352–355.
- Kirchner, H., Thorpe, S.J., 2006. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* 46, 1762–1776.

- Kloth, N., Dobel, C., Schweinberger, S.R., Zwitserlood, P., Bölte, J., Junghöfer, M., 2006. Effects of personal familiarity on early neuromagnetic correlates of face perception. *Eur. J. Neurosci.* 24, 3317–3321.
- Kobatake, E., Tanaka, K., 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71, 856–867.
- 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.
- Kriegeskorte, N., Formisano, E., Sorger, B., Goebel, R., 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20600–20605.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62, 621–647.
- Latinus, M., Taylor, M.J., 2005. Holistic processing of faces: Learning effects with Mooney faces. *Journal of Cognitive Neuroscience* 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.
- Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20, 1434–1448.
- Leopold, D.A., Rhodes, G., 2010. A comparative view of face perception. *J Comp Psychol* 124, 233–251.
- Leveroni, C.L., Seidenberg, M., Mayer, A.R., Mead, L.A., Binder, J.R., Rao, S.M., 2000. Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience* 20, 878.
- Lewis, M.B., Ellis, H.D., 2000. The effects of massive repetition on speeded recognition of faces. *Q J Exp Psychol A* 53, 1117–1142.
- Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.
- Liu, J., Harris, A., Kanwisher, N., 2002. Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916.

- Logothetis, N.K., Sheinberg, D.L., 1996. Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.
- Lotte, F., Congedo, M., Lécuyer, a, Lamarche, F., & Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 4(2), R1–R13.
- Luo, A., Sajda, P., 2009. Comparing neural correlates of visual target detection in serial visual presentations having different temporal correlations. *Front Hum Neurosci* 3, 5.
- Macé, M.J.-M., Joubert, O.R., Nespoulous, J.-L., Fabre-Thorpe, M., 2009. The time-course of visual categorizations: you spot the animal faster than the bird. *PLoS ONE* 4, e5927.
- Mack, M.L., Palmeri, T.J., 2011. The timing of visual object categorization. *Front Psychol* 2, 165.
- Maillard, L., Barbeau, E.J., Baumann, C., Koessler, L., Bénar, C., Chauvel, P., Liégeois-Chauvel, C., 2011. From perception to recognition memory: time course and lateralization of neural substrates of word and abstract picture processing. *J Cogn Neurosci* 23, 782–800.
- Mandler, G., 1980. Recognizing: The judgment of previous occurrence. *Psychological Review* 87, 252–271.
- Manns, J.R., Hopkins, R.O., Reed, J.M., Kitchener, E.G., Squire, L.R., 2003. Recognition memory and the human hippocampus. *Neuron* 37, 171–180.
- Maril, A., Wagner, A.D., Schacter, D.L., 2001. On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron* 31, 653–660.
- Marino, L., Connor, R.C., Fordyce, R.E., Herman, L.M., Hof, P.R., Lefebvre, L., Lusseau, D., McCowan, B., Nimchinsky, E.A., Pack, A.A., Rendell, L., Reidenberg, J.S., Reiss, D., Uhen, M.D., Van der Gucht, E., Whitehead, H., 2007. Cetaceans Have Complex Brains for Complex Cognition. *PLoS Biol* 5.
- Marr, D., 1982. *Vision: A computational approach*. Freeman & Co., San Francisco.
- Martin, E., 2000. Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology* 111, 694–705.
- Mayes, A.R., Holdstock, J.S., Isaac, C.L., Hunkin, N.M., Roberts, N., 2002. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus* 12, 325–340.

- Maurer, D., Grand, R.L., Mondloch, C.J., 2002. The many faces of configural processing. *Trends in cognitive sciences* 6, 255–260.
- Mayes, A.R., Montaldi, D., 1999. The neuroimaging of long-term memory encoding processes. *Memory* 7, 613–659.
- Mehring, C., Rickert, J., Vaadia, E., Cardoso de Oliveira, S., Aertsen, A., & Rotter, S., 2003. Inference of hand movements from local field potentials in monkey motor cortex. *Nature neuroscience*, 6(12), 1253–4.
- Mervis, C.B., Rosch, E., 1981. Categorization of Natural Objects. *Annual Review of Psychology* 32, 89–115.
- Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., Poggio, T., 2008. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* 100, 1407–1419.
- Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y.T., Fryer, T.D., Williams, G.B., Hodges, J.R., Nestor, P.J., 2010. What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* 133, 3256–3268.
- Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983a. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences* 6, 414–417.
- Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983b. Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences* 6, 414–417.
- Moeller, S., Freiwald, W.A., Tsao, D.Y., 2008. Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320, 1355–1359.
- Montaldi, D., Mayes, A.R., 2010a. The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus* 20, 1291–1314.
- Montaldi, D., Mayes, A.R., 2010b. The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus* 20, 1291–1314.
- Montaldi, D., Mayes, A.R., 2011. Familiarity, recollection and medial temporal lobe function: an unresolved issue. *Trends Cogn. Sci. (Regul. Ed.)* 15, 339–340.
- Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., Koch, C., 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* 28, 8865–8872.
- Morton, J., Johnson, M.H., 1991. CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol Rev* 98, 164–181.

- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci. (Regul. Ed.)* 10, 424–430.
- Nowak, L.G., Bullier, J., 1998. The timing of information transfer in the visual system. *Cerebral cortex* 12, 205–241.
- Oliva, A., Torralba, A., 2006. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* 155, 23–36.
- Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731.
- Palermo, R., Rhodes, G., 2007. Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia* 45, 75–92.
- Paller, K. a, Voss, J.L., Boehm, S.G., 2007. Validating neural correlates of familiarity. *Trends in cognitive sciences* 11, 243–50.
- Parr, L.A., De Waal, F.B., 1999. Visual kin recognition in chimpanzees. *Nature* 399, 647–648.
- Pascalis, O., De Schonen, S., Morton, J., Deruelle, C., Fabre-Grenet, M., 1995. Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development* 18, 79–85.
- Parvizi, J., Jacques, C., Foster, B.L., Withoft, N., Rangarajan, V., Weiner, K.S., Grill-Spector, K., 2012. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920.
- Peelen, M.V., Glaser, B., Vuilleumier, P., Eliez, S., 2009. Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev Sci* 12, F16–25.
- Pernet, C., Basan, S., Doyon, B., Cardebat, D., Démonet, J.-F., Celsis, P., 2003. Neural timing of visual implicit categorization. *Brain Res Cogn Brain Res* 17, 327–338.
- Pfütze, 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, 140–160.
- Pickering, E.C., Schweinberger, S.R., 2003. N200, N250r, and N400 event-related brain potentials reveal three loci of repetition priming for familiar names. *J Exp Psychol Learn Mem Cogn* 29, 1298–1311.
- Pitcher, D., Walsh, V., Yovel, G., Duchaine, B., 2007. TMS evidence for the involvement of the right occipital face area in early face processing. *Curr. Biol.* 17, 1568–1573.
- Plotnik, J.M., Waal, F.B.M. de, Reiss, D., 2006. Self-recognition in an Asian elephant. *PNAS* 103, 17053–17057.

- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum, 1978.
- Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience* 16, 5205–5215.
- Puce, A., Allison, T., Gore, J.C., McCarthy, G., 1995. Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74, 1192–1199.
- Puce, A., Allison, T., McCarthy, G., 1999. Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458.
- Quian Quiroga, R., Kraskov, A., Koch, C., Fried, I., 2009. Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology* 19, 1308–1313.
- 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.
- Railo, H., Koivisto, M., Revonsuo, A., 2011. Tracking the processes behind conscious perception: a review of event-related potential correlates of visual consciousness. *Conscious Cogn* 20, 972–983.
- Rajimehr, R., Young, J.C., Tootell, R.B.H., 2009. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1995–2000.
- Ramon, M., Caharel, S., Rossion, B., 2011. The speed of recognition of personally familiar faces. *Perception* 40, 437–449.
- Ratcliff, R., Philiastides, M.G., Sajda, P., 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6539–6544.
- Reddy, L., Kanwisher, N., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr. Biol.* 17, 2067–2072.
- Reddy, L., Reddy, L., Koch, C., 2006. Face identification in the near-absence of focal attention. *Vision research* 46, 2336–43.
- Rhodes, G., Brake, S., Atkinson, A.P., 1993. What's lost in inverted faces? *Cognition* 47, 25–57.
- Rhodes, G., Jeffery, L., 2006. Adaptive norm-based coding of facial identity. *Vision research* 46, 2977–2987.

- Rhodes, G. (1996). *Superportraits: Caricatures and recognition*. Hove: The Psychological Press.
- Rhodes, G., Brennan, S., Carey, S., 1987. Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, 19, 473–497
- Riddoch, M.J., Johnston, R.A., Bracewell, R.M., Boutsen, L., Humphreys, G.W., 2008. Are faces special? A case of pure prosopagnosia. *Cognitive Neuropsychology* 25, 3–26.
- Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. *Nature neuroscience* 2, 1019–1025.
- Rosch, E., 1975. Cognitive representations of semantic categories. *Journal of experimental psychology: General* 104, 192.
- Rossion, B. (2009). Clarifying the functional neuro-anatomy of face perception by single case neuroimaging studies of acquired prosopagnosia. In: Jenkin M. and Harris L.R. (eds.) *Cortical Mechanisms of Vision*, 179–215.
- Rossion, B., 2002. Is sex categorization from faces really parallel to face recognition? *Visual Cognition* 9, 1003–1020.
- Rossion, B., 2008. Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *Neuroimage* 40, 423–426.
- Rossion, B., Caharel, S., 2011. ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Res.* 51, 1297–1311.
- Rossion, B., Caharel, S., Jacques, C., Ramon, M., 2010. The speed of familiar face recognition. *Journal of Vision* 10, 617.
- Rossion, B., Caldara, R., Seghier, M., Schuller, A.-M., Lazeyras, F., Mayer, E., 2003. 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., Curran, T., Gauthier, I., 2002. A defense of the subordinate-level expertise account for the N170 component. *Cognition* 85, 189–196.
- Rossion, B., Delvenne, J.F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., Guérit, J.M., 1999. Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol Psychol* 50, 173–189.

- Rossion, B., Dricot, L., Devolder, A., Bodart, J.M., Crommelinck, M., De Gelder, B., Zoontjes, R., 2000. Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J Cogn Neurosci* 12, 793–802.
- Rossion, B., Jacques, C., 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39, 1959–1979.
- Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., Crommelinck, M., 2001. How does the brain discriminate familiar and unfamiliar faces?: a PET study of face categorical perception. *Journal of cognitive neuroscience* 13, 1019–34.
- Rotshtein, P., Henson, R.N.A., Treves, A., Driver, J., Dolan, R.J., 2005. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* 8, 107–113.
- Rousselet, G. a, Macé, M.J.-M., Fabre-Thorpe, M., 2003. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of vision* 3, 440–55.
- Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2007. Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 36, 843–862.
- Rousselet, G.A., Macé, M.J.-M., Fabre-Thorpe, M., 2004. Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *J Vis* 4, 13–21.
- Rousselet, G.A., Pernet, C.R., 2011. Quantifying the Time Course of Visual Object Processing Using ERPs: It's Time to Up the Game. *Front Psychol* 2, 107.
- Rugg, M.D., Curran, T., 2007. Event-related potentials and recognition memory. *Trends Cogn. Sci. (Regul. Ed.)* 11, 251–257.
- Sadeh, B., Podlipsky, I., Zhdanov, A., Yovel, G., 2010. Event-related potential and functional MRI measures of face-selectivity are highly correlated: a simultaneous ERP-fMRI investigation. *Hum Brain Mapp* 31, 1490–1501.
- Schiltz, C., Rossion, B., 2006a. Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage* 32, 1385–1394.
- Schmolesky, M.T., Wang, Y., Hanes, D.P., Thompson, K.G., Leutgeb, S., Schall, J.D., Leventhal, A.G., 1998. Signal timing across the macaque visual system. *Journal of Neurophysiology* 79, 3272–3278.

- Scholte, H.S., Ghebreab, S., Waldorp, L., Smeulders, A.W.M., Lamme, V.A.F., 2009. Brain responses strongly correlate with Weibull image statistics when processing natural images. *J Vis* 9, 29.1–15.
- Schwartz, B.L., 1999. Sparkling at the end of the tongue: the etiology of tip-of-the-tongue phenomenology. *Psychon Bull Rev* 6, 379–393.
- Schweinberger, S.R., Huddy, V., Burton, A.M., 2004. N250r: a face-selective brain response to stimulus repetitions. *Neuroreport* 15, 1501–1505.
- Schweinberger, S.R., Pfütze, E.-M., Sommer, W., 1995. Repetition Priming and Associative Priming of Face Recognition: Evidence From Event-Related Potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 722–736.
- Schweinberger, S.R., Pickering, E.C., Burton, A.M., Kaufmann, J.M., 2002a. Human brain potential correlates of repetition priming in face and name recognition. *Neuropsychologia* 40, 2057–2073.
- Schweinberger, S.R., Pickering, E.C., Jentsch, I., Burton, A.M., Kaufmann, J.M., 2002b. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Res Cogn Brain Res* 14, 398–409.
- Schyns, P.G., 1998. Diagnostic recognition: task constraints, object information, and their interactions. *Cognition* 67, 147–179.
- Scott, L.S., Tanaka, J.W., Sheinberg, D.L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. *Journal of cognitive neuroscience* 18, 1453–1465.
- Seeck, M., Michel, C.M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., Schomer, D.L., 1997. Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport* 8, 2749–2754.
- Semenza, C., Zettin, M., 1989. Evidence from aphasia for the role of proper names as pure referring expressions.
- 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., Poncet, M., 1990. From covert to overt recognition of faces in a prosopagnosic patient. *Brain* 113 (Pt 4), 989–1004.
- Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429.

- Simanova, I., Van Gerven, M., Oostenveld, R., Hagoort, P., 2010. Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS ONE* 5, e14465.
- Simoncelli, E.P., Olshausen, B.A., 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.
- Smith, J.D., 2009. The study of animal metacognition. *Trends in cognitive sciences* 13, 389–396.
- Snodgrass, J.G., Corwin, J., 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J Exp Psychol Gen* 117, 34–50.
- Spiridon, M., Kanwisher, N., 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35, 1157–1165.
- Squier, R.W., Mew, J.R., 1981. The relationship between facial structure and personality characteristics. *Br J Soc Psychol* 20, 151–160.
- Squire, L.R., Zola-Morgan, J., 1991. Recognition memory and the medial temporal lobe: a new perspective. *Nat. Rev. Neurosci.* 8, 872–883.
- Staresina, B.P., Fell, J., Do Lam, A.T.A., Axmacher, N., Henson, R.N., 2012. Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nat. Neurosci.* 15, 1167–1173.
- Sternberg, S. (1966). "High speed scanning in human memory". *Science* 153 (3736): 652–654
- Sugase, Y., Yamane, S., Ueno, S., Kawano, K., 1999. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873.
- Sun, D., Chan, C.C.H., Lee, T.M.C., 2012. Identification and Classification of Facial Familiarity in Directed Lying: An ERP Study. *PLoS ONE* 7, e31250.
- Suzuki, W.A., Amaral, D.G., 1994a. Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J. Neurosci.* 14, 1856–1877.
- Suzuki, W.A., Amaral, D.G., 1994b. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J. Comp. Neurol.* 350, 497–533.
- Tanaka, J.W., 2001. The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen* 130, 534–543.
- Tanaka, J.W., Curran, T., 2001. A neural basis for expert object recognition. *Psychol Sci* 12, 43–47.

- 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 Cogn Neurosci* 18, 1488–1497.
- Tanaka, J.W., Farah, M.J., 1993. Parts and wholes in face recognition. *Q J Exp Psychol A* 46, 225–245.
- Tanaka, J.W., Taylor, M., 1991. Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23, 457–482.
- Tanaka, K., 1996. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.
- Tanskanen, T., Näsänen, R., Montez, T., Päällysaho, J., Hari, R., 2005. Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cereb. Cortex* 15, 526–534.
- Tate, A.J., Fischer, H., Leigh, A.E., Kendrick, K.M., 2006. Behavioural and neurophysiological evidence for face identity and face emotion processing in animals. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 361, 2155–2172.
- Thierry, G., Martin, C.D., Downing, P., Pegna, A.J., 2007a. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat. Neurosci.* 10, 505–511.
- Thierry, G., Martin, C.D., Downing, P.E., Pegna, A.J., 2007b. Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nature Neuroscience* 10, 802–803.
- Thompson, P., 1980. Margaret Thatcher: a new illusion. *Perception* 9, 483–484.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520–522.
- Thorpe, S.J., Crouzet, S., Kirchner, H., Fabre-Thorpe, M., 2006. Ultra rapid face detection in natural images: implications for computation in the visual system, in: *First French Conference on Computational Neurosciences (Abbaye Des Premontres, PontaMousson, France)*. Citeseer, pp. 124–127.
- Todd, R.M., Lewis, M.D., Meusel, L.-A., Zelazo, P.D., 2008. The time course of social-emotional processing in early childhood: ERP responses to facial affect and personal familiarity in a Go-Nogo task. *Neuropsychologia* 46, 595–613.
- Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. *Annu Rev Psychol* 63, 483–509.
- Torralba, A., Oliva, A., 2003. Statistics of natural image categories. *Network* 14, 391–412.

- Trautner, P., Dietl, T., Staedtgen, M., Mecklinger, A., Grunwald, T., Elger, C.E., Kurthen, M., 2004. Recognition of famous faces in the medial temporal lobe: an invasive ERP study. *Neurology* 63, 1203–1208.
- Tsao, D.Y., Livingstone, M.S., 2008. Mechanisms of face perception. *Annu Rev Neurosci* 31, 411–437.
- Tsao, D.Y., Moeller, S., Freiwald, W.A., 2008. Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19514–19519.
- Tulving, E., 1985. Memory and consciousness. *Canadian Psychology/Psychologie Canadienne* 26, 1.
- Valentine, T., 1988. Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology* 79, 471–491.
- Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 43, 161–204.
- Valentine, T., 2001. Face-space models of face recognition. In: Wenger M.J. and Townsend J.T. (eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Mahwah: LEA, 83–113.
- Valentine, T., Bruce, V., 1986. The effect of race, inversion and encoding activity upon face recognition. *Acta Psychol (Amst)* 61, 259–273.
- Van Essen, D.C., Anderson, C.H., Felleman, D.J., 1992. Information processing in the primate visual system: an integrated systems perspective. *Science* 255, 419–423.
- Van Rullen, R., Gautrais, J., Delorme, A., Thorpe, S., 1998. Face processing using one spike per neurone. *BioSystems* 48, 229–239.
- VanRullen, R., 2006. On second glance: still no high-level pop-out effect for faces. *Vision research* 46.
- VanRullen, R., Thorpe, S.J., 2001a. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception* 30, 655–668.
- VanRullen, R., Thorpe, S.J., 2001b. The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci* 13, 454–461.
- Vinje, W.E., Gallant, J.L., 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276.

- Waite, C., Gerald, M.S., Little, A.C., Krauselburd, E., 2006. Selective attention toward female secondary sexual color in male rhesus macaques. *Am. J. Primatol.* 68, 738–744.
- Walton, G.E., Bower, N.J.A., Bower, T.G.R., 1992. Recognition of familiar faces by newborns. *Infant Behavior and Development* 15, 265–269.
- Weiner, K.S., Grill-Spector, K., 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52, 1559–1573.
- Wild-Wall, N., Dimigen, O., Sommer, W., 2008. Interaction of facial expressions and familiarity: ERP evidence. *Biol Psychol* 77, 138–149.
- Wixted, J.T., Squire, L.R., 2011. The familiarity/recollection distinction does not illuminate medial temporal lobe function: response to Montaldi and Mayes. *Trends Cogn. Sci. (Regul. Ed.)* 15, 340–341.
- Wright, D.B., Sladden, B., 2003. An own gender bias and the importance of hair in face recognition. *Acta Psychol (Amst)* 114, 101–114.
- Yin, R.K., 1969. Looking at upside-down faces. *Journal of Experimental Psychology* 81, 141–145.
- Yonelinas, A.P., Aly, M., Wang, W.-C., Koen, J.D., 2010. Recollection and familiarity: examining controversial assumptions and new directions. *Hippocampus* 20, 1178–1194.
- Yonelinas, A.P., Jacoby, L.L., 1995. The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language* 34, 622–643.
- Yonelinas, A.P., Levy, B.J., 2002. Dissociating familiarity from recollection in human recognition memory: different rates of forgetting over short retention intervals. *Psychon Bull Rev* 9, 575–582.
- Young, A.W., De Haan, E.H., Newcombe, F., 1990. Unawareness of impaired face recognition. *Brain Cogn* 14, 1–18.
- 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.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E.D., Grafman, J., 2009. The neural basis of human social values: evidence from functional MRI. *Cereb. Cortex* 19, 276–283.

- Zárate, M.A., Sandoval, P., 1995. The effects of contextual cues on making occupational and gender categorizations. *Br J Soc Psychol* 34 (Pt 4), 353–362.
- Zarate, M.A., Smith, E.R., 1990. Person Categorization and Stereotyping. *Social Cognition* 8, 161–185.
- Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., Desimone, R., 2011. Object decoding with attention in inferior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8850–8855.
- Zheng, X., Mondloch, C.J., Segalowitz, S.J., 2012. The timing of individual face recognition in the brain. *Neuropsychologia* 50, 1451–1461.

Résumé

Reconnaître *rapidement* les visages familiers est une fonction fondamentale qui joue un rôle important dans nos interactions sociales. D'un point de vue évolutionniste, il semble en effet crucial de pouvoir rapidement déterminer si la personne qui nous fait face est amie ou ennemie pour adapter son comportement en conséquence. C'est sans doute pour cela que, dans l'opinion publique, mais également dans la littérature scientifique, la reconnaissance des visages est supposée être un processus très rapide, efficace et exécuté sans effort. Cependant, la reconnaissance des visages familiers est-elle réellement si rapide ? Ne serait-ce pas simplement une idée reçue ? Rapide, admettons, mais à quelle vitesse ?

En adaptant des protocoles de catégorisation visuelle rapide (tâche de *go/no-go*) développés initialement pour étudier la rapidité du système visuel et en mettant en place de nouveaux protocoles de catégorisation ultra-rapide (« Speed and Accuracy Boosting procedure » ; SAB), nous avons pu déterminer les latences comportementales et électrophysiologiques les plus précoces pour reconnaître explicitement un visage célèbre. Nous nous sommes intéressés d'une part à la reconnaissance de type « bottom-up » (reconnaître plusieurs visages célèbres sans savoir au préalable de qui il s'agit) et d'autre part à la reconnaissance « top-down » (reconnaître une personne en particulier parmi des inconnus). Le temps de réaction minimum pour reconnaître des visages célèbres parmi des inconnus (reconnaissance « bottom-up ») est d'environ 360-390 ms, ce temps de réponse ne pouvant être amélioré ni par un apprentissage intensif des stimuli (Article 1), ni par un protocole de catégorisation ultra-rapide (Article 2). Ce temps de réaction est environ 100 ms plus tardif que lors d'une tâche de détection de visage (Article 1) ou de genre (Article 1). Ces latences sont très différentes lorsque la procédure SAB est appliquée à une reconnaissance de type « top-down », descendant à environ 300 ms contre 270 ms dans une tâche de détection de visages (Article 3).

De plus, en appliquant une méthode de MVPA (Multi-Variate Pattern Analysis) à des données d'EEG de surface, nous avons montré que l'activité neuronale liée à la reconnaissance des visages célèbres était disponible dès 230 ms après la présentation du stimulus (voire 200 ms pour le sujet les plus rapides) alors que l'activité neuronale liée à la détection d'un visage humain parmi des visages d'animaux était disponible dès 80 ms (Articles 4 et 5). L'activité neuronale était de plus fortement corrélée aux temps de réponses minimales en reconnaissance, confirmant ainsi son rôle dans la prise de décision.

Nous discutons ces latences au regard des modèles de la voie visuelle ventrale et des modèles de la reconnaissance des visages. Nous distinguons trois modèles différents pouvant théoriquement être à l'origine de la familiarité et en favorisons un en particulier.

Abstract

Recognizing familiar faces rapidly seems crucial in everyday life. The actual speed at which a familiar face can be recognized remains however unknown. The current thesis aimed at tracking down the minimal behavioral and neural processing time necessary to recognize known faces.

To address this issue, we used different *go/no-go* paradigms and a new task relying on highly time-constraining task (the Speed and Accuracy Boosting procedure, "SAB"). Relying on minimum reaction times analyses, we report that 360-390 ms are needed to recognize famous faces among unknown ones when bottom-up recognition task is required (subjects did not know the identity of the celebrities that they had to recognize before the test; this situation can be compare to the ecological situation of unexpectedly bumping into someone in the street) (Article 1). This latency could not be decreased even after extensive training (Article 1), or using the SAB (Article 2). Overall, this is 100 ms more than when subjects have to detect human faces in natural scene or process gender (Article 1). Bottom-up recognition is much slower than top-down recognition (recognizing somebody whom you know you are going to meet, corresponding to the ecological situation of looking for someone in particular in a crowd), which takes about 300 ms (Article 3).

Additionally, MVPA (Multivariate pattern analysis) was applied on EEG data recorded from the scalp surface to determine at which latency familiarity could be read-out. We report that famous faces could be robustly distinguished from unknown faces as soon as 230 ms after stimulus onset. This familiarity-selective signal was directly linked to the subject's recognition speed (Article 5). Such latency was again much longer than the latencies observed in face categorisation task, in which case category could be read out starting around 80 ms post-stimulus (Article 4).

These latencies are with respect to the different models of visual ventral stream and models of face recognition. Three main models are identified and one is favored in particular.