



Université catholique de Louvain
Faculté de Psychologie et des Sciences de l'Éducation
Unité de Neurosciences Cognitives

Etude des corrélats cérébraux sous-tendant les processus associatifs impliqués dans l'identification des personnes.

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

Promoteurs: Pr. Raymond Bruyer
Salvatore Campanella
Jury: Pr. Xavier Seron
Pr. Pierre Philippot
Pr. Jean-Michel Guérit
Pr. Pascal Belin

A Laure-Anne, Clara et G.

« Ce n'est pas dans la Science qu'est le bonheur, mais dans l'acquisition de la Science ». (E. Poe)

Au terme de ce doctorat, j'aimerais tout d'abord remercier mes promoteurs, le professeur Raymond Bruyer et Salvatore Campanella pour leur soutien, leurs conseils, leur disponibilité, leur sympathie. Salva, je tiens à t'assurer que mes précédents remerciements, couchés sur papier il y a maintenant 6 ans, restent plus que jamais d'actualité.

Je remercie également les membres de mon comité d'encadrement, les professeurs Xavier Seron, Marc Crommelinck et Jean-Michel Guérit, ainsi que les professeurs Pierre Philippot et Pascal Belin qui m'ont fait le plaisir d'accepter de faire partie de mon jury. Vos conseils lors de la rédaction de cette thèse m'ont été précieux et vos encouragements me sont allés droit au cœur.

Merci aussi à Gaëlle et Emilie qui ont grandement collaboré à ce travail en tant qu'étudiantes. Travailler avec vous fut un réel plaisir.

Merci à Mauro et à Pierre, qui m'honorent chaque jour de leur amitié. Vous m'avez donné envie de venir travailler le lundi matin, et je pense que vous devinerez tout ce que la pudeur m'empêche d'écrire ici.

Mes pensées se tournent vers mes proches qui me témoignent chaque jour leur affection. « L'amitié double les joies et réduit de moitié les peines ». J'ai une pensée particulière pour mes parents, pour Vincent, mon frère, pour Houda, Arno, Daniel, Bernadette, et Jérôme. Laure-Anne, Clara, les mots me manquent pour vous exprimer mon amour. « Sans petits bisous, les rêves sont flous ». Les vôtres autant de soleils dans ma vie. G., sache que depuis huit mois, je n'ai cessé de penser à toi.

Table des matières

Introduction générale	1
1ere partie : contexte théorique	15
Chapitre 1 : corrélats cérébraux du traitement des visages	17
Chapitre 2 : corrélats cérébraux du traitement des noms propres	49
Chapitre 3 : corrélats cérébraux du traitement des voix	65
Chapitre 4 : les processus associatifs entre visages, noms de famille et voix	75
Conclusions de la partie théorique et introduction à la partie expérimentale	105
2eme partie : approche expérimentale	119
Etude 1 : the electrophysiological correlates sustaining the retrieval of face-names associations : an ERP study.	121
Introduction à l'étude 2	163
Etude 2 : the associative processes involved in face- proper names vs. objects-common names binding : a comparative ERP study.	169

Introduction aux études 3 et 4	213
Etude 3 : when audition alters vision : an ERP study of the cross-modal interactions between faces and voices.	219
Etude 4 : cross-modal integration of human faces and voices in the left angular gyrus.	241
Conclusions des études 3 et 4	275
Conclusions générales et perspectives	283
Références bibliographiques	313
Travaux annexes et collaborations	353

Introduction générale



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

L'homme, comme beaucoup d'autres organismes, est équipé de multiples canaux sensoriels au travers desquels il expérimente son environnement (Stein & Meredith, 1993). Chaque sens procure des impressions subjectives du monde qualitativement distinctes. Néanmoins, nous sommes capables de détecter les relations qui existent entre des informations de nature distincte et de les intégrer pour former une perception cohérente et unifiée de notre environnement (Calvert, 2001). Non seulement nous sommes capables de créer des liens entre des informations de nature distincte, mais nous sommes également capables de stocker ces associations en mémoire à long terme et de les récupérer par la suite.

Contrairement aux idées constructivistes développées par Piaget (1952, in Lewkowicz, 2002), il apparaît que les capacités d'intégration intermodale émergent de manière très précoce dans le développement humain. Ainsi par exemple, Bower, Broughton et Moore (1970) ont observé que dès l'âge de 6 jours, les nourrissons sont capables de projeter leurs mains vers des cibles visuelles, attestant une coordination précoce de l'œil et de la main. De même, Spelke (1976), et après elle Bahrick (1983, 1988), ont montré que les bébés âgés

de 4 mois sont capables d'établir des relations entre les objets visuels et les sons spécifiques que ces objets produisent. De plus, il semble qu'il n'existe pas de différences réelles entre les enfants (dès 5-6 ans) et les adultes quant aux capacités de transfert intermodal, l'amélioration des performances avec le vieillissement étant due à l'amélioration concomitante des performances propres à chaque modalité (Hatwell, 1994, pp. 578). L'organisation des modalités et leurs interactions se mettent donc en place très tôt et demeurent ensuite stables au cours de la vie.

Ces capacités associatives sont essentielles à notre adaptation, et tout particulièrement à notre adaptation sociale. En effet, chaque jour, nous sommes amenés à interagir avec un grand nombre de personnes différentes. Ces interactions sociales nécessitent sans cesse l'apprentissage et la récupération d'associations entre informations de nature différente. Ainsi, nous sommes sans cesse amenés à associer visages et noms propres, afin de pouvoir identifier chaque personne que nous connaissons. Cette « labellisation » est nécessaire, tant pour s'adresser aux personnes à qui l'on parle que pour faire référence à une personne particulière.

Mais les associations entre visages et noms ne sont pas les seules à intervenir au sein de nos interactions quotidiennes. Nous sommes également amenés à constamment intégrer les informations visuelles et auditives qui nous parviennent des personnes avec qui nous interagissons. Tout d'abord, au niveau perceptif, nous associons les mouvements observables des lèvres de notre interlocuteur avec le discours perçu auditivement afin d'améliorer nos capacités de compréhension (Calvert, Campbell, & Brammer, 2000) et d'identifier les personnes qui nous parlent, notamment au sein d'une foule. Ensuite, nous avons des personnes une représentation unifiée en mémoire à long terme, contenant notamment les représentations des visages, des voix et des noms particuliers à chaque individu. Par exemple, à l'évocation de Charles de Gaulle, nous sommes capables de récupérer en mémoire une représentation de ses caractéristiques faciales, suffisante pour pouvoir le reconnaître, ainsi que le souvenir de sa voix, a minima proclamant le désormais célèbre « je vous ai compris ».

Notre mémoire des personnes nécessite donc de lier ensemble les trois principales informations nous permettant d'identifier les individus, à savoir le nom, les caractéristiques

faciales et la voix (Paller, Ranganath, Gonsalves, LaBar, Parrish, Gitelman, Mesulam, & Reber, 2003).

La question qui se pose dès lors est de savoir comment le cerveau opère pour lier ensemble ces différentes représentations pour en former un tout cohérent. En effet, il est largement admis de nos jours que ces informations sont traitées par des régions corticales distinctes et distribuées au sein du cerveau. Dès lors, il apparaît que la séparation des différentes représentations relatives à un même « objet » soulève la possibilité d'un « binding problem » (Wolfe & Cave, 1999). Bien que ce terme soit souvent employé pour qualifier les problèmes soulevés par l'unification des différents aspects d'une perception, Roskies (1999) souligne qu'il n'existe pas un « binding problem » mais des « binding problems ». Il distingue ainsi les problèmes de binding perceptif, tels qu'évoqués ci-dessus des problèmes de binding cognitif, au sein desquels se situent (1) les processus d'identification cross-modale, c'est-à-dire la capacité d'identifier un même objet sur base d'informations provenant de modalités sensorielles distinctes, et (2) la « reconstruction mnésique », qui se réfère à la liaison des informations précédemment

encodées afin d'en former une représentation unifiée et structurée.

Deux théories ont vu le jour pour résoudre ces problèmes. La première, proposée par Von der Malsburg (1981), postule que les différentes informations sont liées via la décharge synchronisée des assemblées de neurones traitant chaque type d'information. Cette hypothèse de corrélation temporelle a reçu différents appuis, notamment au travers des études ayant examiné la synchronisation des décharges neuronales dans la bande de fréquence gamma (30-70 Hz), tant chez l'animal (Engel, König, Kreiter, & Singer, 1991 ; Gray, & Singer, 1989 ; Brosch, Bauer, & Eckhorn, 1997) que chez l'homme (Tallon-Baudry, & Bertrand, 1997a, 1997b, 1998, 1999). Toutes ces études présupposent que les activités oscillatoires, enregistrées au sein de régions sensorielles distinctes, sont liées à la génération de représentations des objets et permettent l'émergence d'un percept unifié (Varela, 1995 ; Lachaux, Rodriguez, Martinerie, & Varela, 1999).

La deuxième théorie postule que le « binding problem » serait résolu via la convergence des inputs en provenance des régions traitant les différentes caractéristiques des objets vers des régions corticales de plus haut niveau, chargées de

l'intégration de ces différentes caractéristiques en une représentation unifiée et structurée des objets (Olhausen, Anderson, & Van Essen, 1995 ; Shadlen & Movshon, 1999). Cette théorie a également reçu différents soutiens empiriques. Ainsi, des neurones répondant à des stimulations en provenance de plusieurs modalités sensorielles ont été trouvés dans les couches profondes du colliculus supérieur ainsi que dans le cortex de chats (Meredith & Stein, 1983 ; Wallace, Meredith, & Stein, 1992) et de singes (Jay & Sparks, 1984 ; Duhamel, Colby, & Goldberg, 1991 ; Graziano & Gross, 1998). Chez l'homme, différentes études ont également montré l'implication du colliculus supérieur et de l'insula dans les processus d'intégration visuo-auditive (Bushara, Grafman, & Hallett, 2001 ; Calvert, Hansen, Iversen, & Brammer, 2001), ainsi que de la scissure temporale supérieure (STS) dans le binding crossmodal d'informations linguistiques visuelles et auditives (Calvert et al., 2000, Raij, Uutela, & Hari, 2001). Dans cette optique, la formation d'un concept unifié à partir de percepts distincts serait réalisée à la fois par les régions « unimodales » et par des régions « hétéromodales » intégrant le fruit du travail de chaque région.

Shadlen et Movshon (1999) suggèrent aussi que le lobe pariétal joue un rôle prépondérant dans les processus de binding. Ils se basent pour cela sur l'observation clinique de déficits d'intégration suite à des lésions pariétales (Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997). Par ailleurs, il semble que les régions pariétales ne sont pas impliquées que dans les processus d'associations perceptives, mais également dans les processus associatifs permettant l'identification des objets et en particulier des visages. Ainsi, Campanella, Joassin, Rossion, De Volder, Bruyer et Crommelinck (2001) ont réalisé une étude par PET-scan examinant les substrats neuraux des processus impliqués dans la reconnaissance d'associations entre visages et noms propres. Cette expérience, qui sera détaillée plus loin, a mis en évidence une activation du lobule pariétal inférieur gauche, interprétée comme reflétant les processus associatifs entre les représentations distinctes des visages et des noms. Cette étude, s'accordant avec la théorie de l'existence de régions intégratrices, a constitué la pierre angulaire des différentes recherches constituant le corps de cette thèse, et qui ont eu pour but principal d'approfondir nos connaissances des mécanismes cérébraux permettant d'associer en une seule et

même représentation mentale les visages, les noms et les voix des personnes familières.

Cette thèse s'articule en trois parties distinctes et complémentaires. Dans une première partie, nous tenterons de dresser un état des lieux des connaissances relatives aux bases neurales du traitement respectif des visages, des noms propres et des voix. Nous aborderons ces différents processus de traitement de l'information sous l'angle de la neuropsychologie cognitive, de l'imagerie cérébrale fonctionnelle et de l'électrophysiologie. Cet état des lieux ne se veut pas exhaustif, mais a pour objectif de montrer que le traitement des visages, des noms propres et des voix est soutenu par des régions cérébrales distinctes. Nous terminerons cette introduction théorique en passant en revue différentes recherches ayant eu pour but d'examiner les corrélats cérébraux des processus associatifs entre informations de nature distincte, et verrons que si les bases neurales du traitement des trois types d'informations sont de mieux en mieux circonscrites, nous ne savons encore que peu de choses de la manière dont le cerveau opère pour les lier en mémoire.

La deuxième partie de cette thèse sera consacrée à l'approche expérimentale de ces processus associatifs entre visages, noms propres et voix. Quatre études, se basant sur les résultats de l'étude de Campanella et al. (2001), seront exposées.

La première étude, répliquant celle de Campanella et al. (2001) au moyen de la technique des potentiels évoqués, aura pour but de dresser un modèle anatomo-temporel des activités cérébrales liées à la récupération en mémoire d'associations entre visages et noms propres. L'utilisation conjointe de différentes techniques d'imagerie cérébrale offrant une haute précision à la fois anatomique et temporelle, permettra de contraindre temporellement les activations observées dans l'étude pré-citée.

La deuxième étude, utilisant également les potentiels évoqués, tentera de répondre à la question de savoir si les activités cérébrales mises en lumière par Campanella et al. (2001) sont spécifiques aux associations entre visages et noms propres ou si elles reflètent des mécanismes associatifs plus généraux permettant de lier tout objet visuel à son nom. C'est donc la question de la généralité des processus associatifs qui sera abordée ici.

Les études 3 et 4 examineront la question de la possible amodalité des régions cérébrales impliquées dans les processus associatifs permettant la reconnaissance des personnes. En effet, dans l'étude de Campanella et al. (2001), toutes les informations étaient présentées dans la modalité visuelle (visages et noms propres écrits). Néanmoins, sachant que les processus d'intégration visuo-auditive requièrent l'activation de régions hétéromodales particulières, comme la scissure temporale supérieure (STS, Calvert et al., 2001), il était intéressant d'étudier les corrélats cérébraux des interactions entre visages et voix dans une tâche de reconnaissance, et de pouvoir comparer directement les activités cérébrales produites par la reconnaissance de visages et de voix présentés isolément ou conjointement. Afin de répondre à la nécessaire complémentarité entre techniques anatomiques et temporelles, l'étude 3 a examiné cette question via la technique des potentiels évoqués, l'étude 4 via l'imagerie par résonance magnétique fonctionnelle (IRMf).

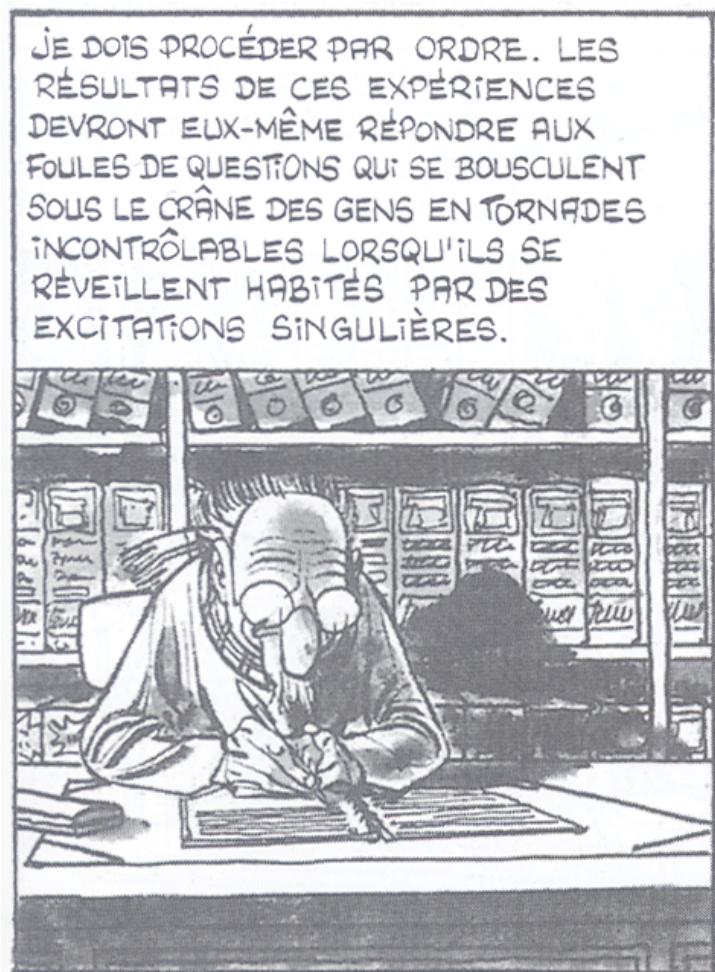
Enfin, dans la troisième et dernière partie de cette thèse, nous tenterons de tirer des conclusions générales des

différentes recherches menées autour de la thématique des processus associatifs impliqués dans l'identification des personnes, et nous envisagerons différentes perspectives de recherches nous permettant de continuer à approfondir nos connaissances de cet aspect particulier et fondamental du fonctionnement cérébral.

Nous aimerions conclure cette introduction générale en soulignant que cette thèse, utilisant différentes techniques d'imagerie cérébrale, électrophysiologiques et anatomo-fonctionnelles, s'inscrit délibérément dans le champs des neurosciences cognitives. Elle n'en reste toutefois pas moins un travail de psychologie, puisque, comme l'affirme Marc Jeannerod (2002), « Les neurosciences cognitives ne cherchent pas à analyser le fonctionnement d'un élément isolé de l'ensemble. Elles cherchent au contraire à comprendre comment le fonctionnement ordonné du cerveau dans son ensemble contribue à notre pensée, notre langage, notre mémoire, en un mot à notre activité cognitive ».

1ere partie

Contexte théorique



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

Chapitre 1:

corrélats cérébraux du traitement des visages.

Les visages ont un statut particulier au sein des objets perçus visuellement. En effet, non seulement ils jouent un rôle fondamental dans nos interactions sociales, mais de plus ils exigent des capacités discriminatives extrêmement fines nous permettant d'identifier un exemplaire particulier au sein de la catégorie. La littérature scientifique concernant les processus de traitement des visages est abondante. Dans ce chapitre, nous nous contenterons de passer en revue différentes études, couvrant les champs de la neuropsychologie cognitive, de l'imagerie cérébrale fonctionnelle et de l'électrophysiologie, tendant toutes à montrer que le traitement des visages – perception et reconnaissance – implique l'activation de régions cérébrales particulières. Néanmoins, avant toute chose, il sera judicieux de rappeler succinctement les différents modèles théoriques sous-tendant les processus d'identification des visages.

Modèles cognitifs sériels et parallèles des processus d'identification des visages

Le premier modèle, proposé par Bruce & Young (1986) envisage l'identification des visages familiers en quatre étapes distinctes (Figure 1). La première étape d'encodage structural représente le produit final de l'analyse perceptive d'un visage menant à une représentation abstraite en trois dimensions indépendante de l'orientation, de l'expression, du contexte et du point de vue de l'observateur. C'est au niveau de cette étape que sont également analysées les autres informations contenues dans le stimulus, telles que l'expression émotionnelle ou la direction du regard. Dans une deuxième étape, cette représentation abstraite sera comparée et appariée à une unité de reconnaissance faciale (Face Recognition Unit, FRU) correspondant à la représentation de ce visage contenue en mémoire à long terme. Dès cet instant, le visage sera reconnu comme familier. Cette FRU donnera ensuite accès au PIN (Person Identity Node), nœud sémantique permettant la récupération de l'ensemble des informations sémantiques (biographiques et épisodiques) liées à la personne reconnue. Notons que les PIN peuvent être

activés par d'autres stimulations que le visage, comme par exemple la voix, le nom ou encore le visage d'une personne fortement associée. Enfin, la quatrième étape permet la récupération du nom de la personne.

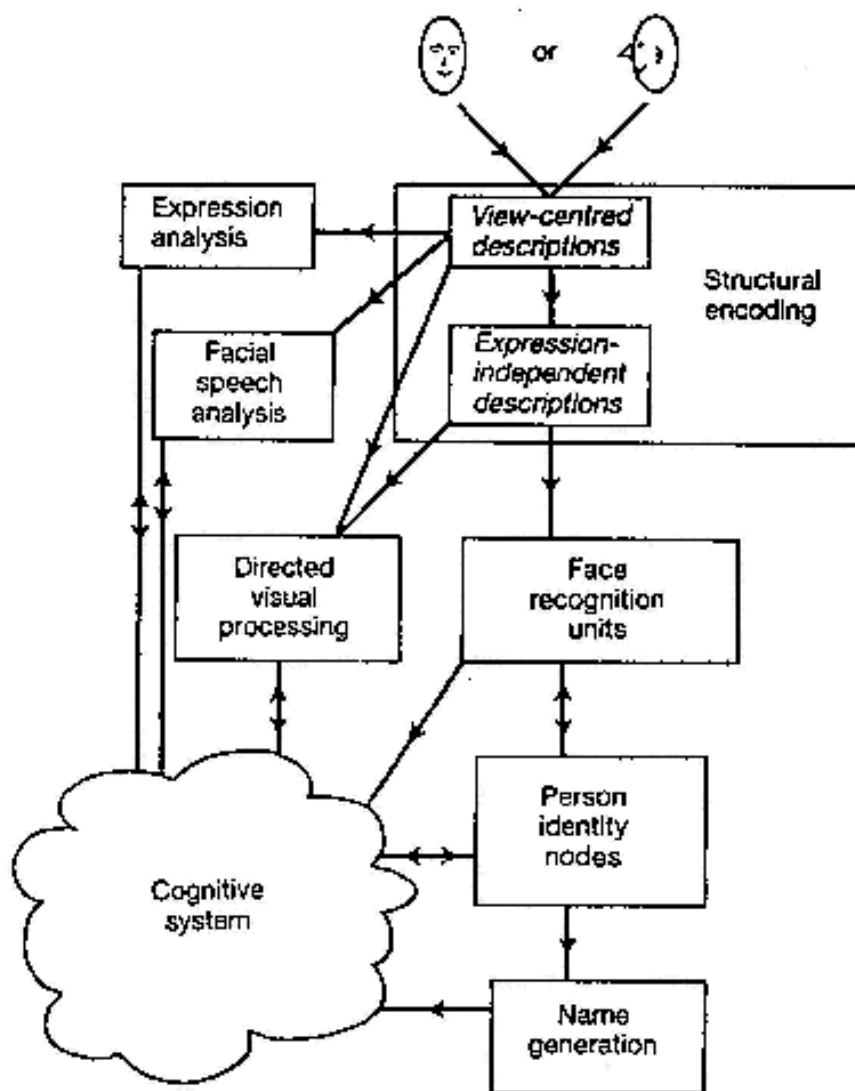


Figure 1 : Modèle sériel d'identification des visages de Bruce & Young (1986)

Ce qu'il est important de souligner par rapport à ce modèle, c'est son caractère strictement séquentiel. Chaque étape du traitement doit être achevée avant que ne débute la

suivante, et une activation simultanée de deux étapes différentes de l'identification est impossible. Une deuxième remarque concerne le statut des noms propres au sein de ce modèle, puisqu'ils sont considérés comme constituant l'étape ultime du processus d'identification des visages, stockés de manière distincte des informations sémantiques et accessibles uniquement si les informations sémantiques concernant la personne ont été préalablement récupérées.

Burton & Bruce (1992) proposeront un modèle alternatif dont les modifications porteront essentiellement sur les deux points soulevés plus haut. Par rapport à la séquentialité des étapes de traitement, Burton & Bruce (1992) développent un modèle connexionniste comprenant trois sous-systèmes entretenant entre eux des relations excitatrices (entre sous-systèmes) et inhibitrices (à l'intérieur des sous-systèmes).

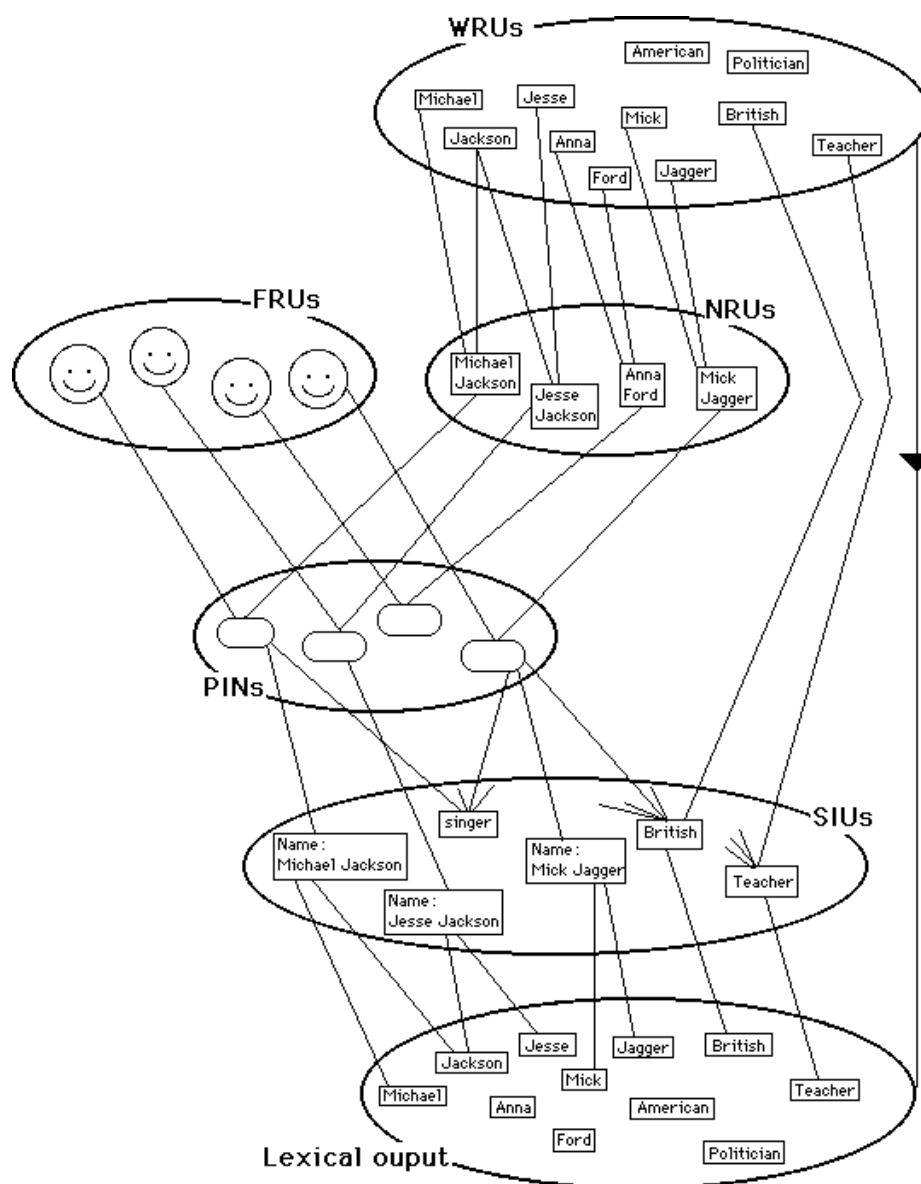


Figure 2 : Modèle parallèle d'identification des visages de Burton & Bruce (1992)

Ces trois sous-systèmes sont les FRU (dont le rôle est identique à celui proposé par Bruce & Young, 1986), les PIN, considérés comme un lieu d'accès au système sémantique, et les SRU (Semantic Recognition Units, Figure 2). Le changement majeur par rapport au modèle précédent concerne l'organisation du système sémantique. En effet,

l'activation d'un PIN entraînerait l'activation d'une série de SRU correspondant chacune à une information sémantique particulière. Ici, les noms propres sont envisagés comme une information sémantique parmi d'autres et non plus comme étant liés à un niveau ultérieur de traitement.

Les deux modèles ont reçu différents supports empiriques. Par exemple, Young, McWeeny, Hay, & Ellis (1986) montrent que la récupération du nom d'une personne suite à la présentation du visage prend plus de temps que la récupération des données biographiques. Cette observation appuie une vision hiérarchique du processus d'identification. Néanmoins, le modèle connexionniste de Burton & Bruce (1992) permet également d'expliquer ces résultats. En effet, dans ce modèle où les connexions entre unités sont bi-directionnelles, les noms, uniques à chaque personne, sont « isolés » au sein du système sémantique et donc moins facilement activés. Burton & Bruce (1992) tiendront également compte de la dimension temporelle dans l'organisation du système sémantique et postuleront que les connexions entre les différentes unités sémantiques peuvent évoluer et se réorganiser en fonction des interactions avec l'environnement. Ce postulat permet d'ailleurs d'expliquer les

observations de Scanlan et Johnston (1997) selon lesquelles les enfants récupèrent plus rapidement le nom d'une personne que sa profession. Cet effet serait dû à une immaturité du système sémantique de l'enfant, chez qui le nom tiendrait lieu d'information centrale. Au fil de la construction du système sémantique et de l'arrivée d'autres informations biographiques, ce lien privilégié perdrait sa prépondérance au profit d'autres SRU comme la profession ou la nationalité.

Bien qu'on ne puisse pas encore actuellement les départager de manière définitive, les deux modèles d'identification des visages sommairement présentés ici restent de solides points de repère pour mieux comprendre les processus permettant d'identifier les personnes familières. Il est à noter que plusieurs études ont tenté de départager les deux modèles. En particulier, Schweinberger, Burton et Kelly (2001) ont utilisé une méthode d'indiçage associatif de visages célèbres à partir d'indices phonologiques ou sémantiques pour départager les deux modèles d'identification des visages. Ils ont obtenu des résultats leur permettant de se positionner en faveur du modèle parallèle. Néanmoins, nous avons répliqué leur étude en étudiant

particulièrement l'influence de la spécificité des indices sémantiques et phonologiques sur le processus d'indiciage associatif. Nos résultats, divergeant de ceux obtenus par Schweinberger et al. (2001) ne nous ont pas permis de favoriser un modèle plutôt que l'autre, et ont mis en question la validité de tels paradigmes expérimentaux dans l'étude comparative des modèles cognitifs d'identification des visages (Joassin, Maurage, Campanella, & Bruyer, sous presse). Le lecteur intéressé pourra trouver l'abstract de cet article en annexe.

D'autres modèles de perception des visages ont également été proposés. En particulier, Haxby, Hoffman et Gobbini (2000) ont élaboré un modèle rendant compte des bases neurales sous-tendant la perception des visages (Figure 3). Leur modèle fait la distinction entre les représentations invariantes, permettant la reconnaissance des visages et les représentations changeantes de ces visages, tels que la direction du regard, l'expression et les mouvements des lèvres.

Leur modèle est hiérarchique en ce sens qu'il est divisé en deux systèmes distincts : un système central et un système étendu. Le système central comprend les régions occipito-

temporales des cortex visuels extra-striés, sous-tendant l'analyse visuelle des visages.

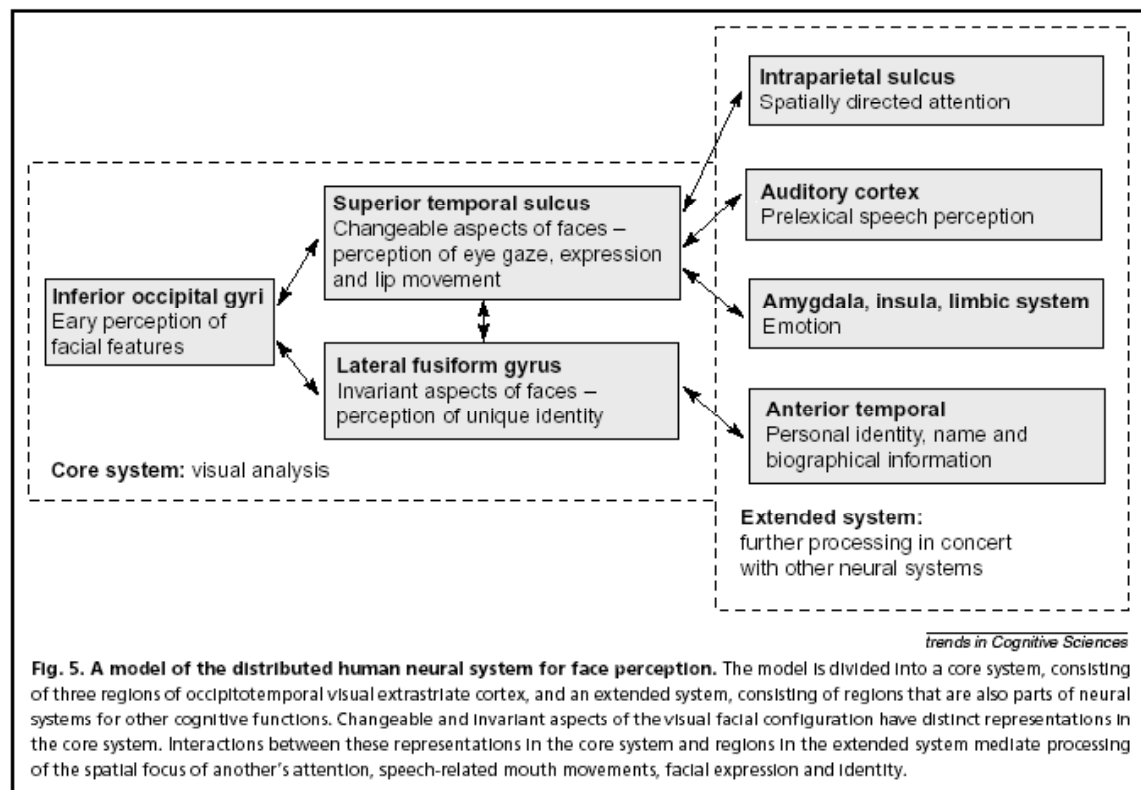


Figure 3 : modèle de perception des visages d'Haxby et al. (2000)

Dans ce système, la représentation des aspects invariants des visages est assurée par les régions répondant de manière privilégiée aux visages (FFA) alors que la représentation des aspects changeants des visages est assurée par des régions répondant aux visages localisées dans le STS. Les système étendu, quant à lui, comprend les régions assurant d'autres fonctions cognitives et qui peuvent être recrutées pour extraire du visage perçu des informations telles que l'émotion,

la perception du discours ou les informations biographiques liées à la personne.

Données lésionnelles – la prosopagnosie

La prosopagnosie, définie pour la première fois par Bodamer (1947), est un type d'agnosie visuelle marquée par un déficit de reconnaissance des visages familiers. Il est à noter que les patients souffrant de ce type d'agnosie restent à même d'identifier les personnes sur base de leurs autres attributs, tels que la voix, le nom, la démarche, etc.

La prosopagnosie n'est pas un déficit unitaire, ni dans son étiologie ni dans ses formes. On peut ainsi distinguer la prosopagnosie acquise suite à une lésion cérébrale de la prosopagnosie congénitale, présente dès l'enfance en l'absence de toute lésion acquise ou d'autre atteintes neurologiques concomitantes (pour une revue, voir Behrmann & Avidan, 2005). Par ailleurs, on distingue classiquement deux formes de prosopagnosie, à savoir la prosopagnosie aperceptive et la prosopagnosie associative. La première forme serait caractérisée par une altération du traitement des caractéristiques physiques du visage et de leur

intégration en un percept unifié. La deuxième forme de prosopagnosie correspondrait à un déficit d'accès aux représentations des visages stockés en mémoire à long terme.

Différentes hypothèses ont vu le jour pour expliquer la prosopagnosie. Ainsi, Farah (1990) propose l'hypothèse d'un traitement visuel s'établissant selon un continuum allant du traitement holistique des visages au traitement analytique des mots. Les objets seraient traités de manière intermédiaire, tantôt holistiquement, tantôt analytiquement en fonction de leur structure. De cette hypothèse découlent les prédictions suivantes : (1) un patient souffrant de prosopagnosie et d'alexie devrait impérativement présenter également des troubles de la reconnaissance des objets, (2) un patient présentant un trouble de la reconnaissance des objets devrait toujours souffrir également soit d'un trouble de la reconnaissance des visages, soit d'un trouble de la reconnaissance des mots. Ces prédictions ont été toutes deux infirmées, par De Renzi et di Pellegrino (1998) et Buxbaum, Glosser et Coslett (1999) qui ont décrit les cas de patients souffrant de prosopagnosie et d'alexie sans agnosie pour les objets, et par Rumiati, Humphreys, Riddoch et Bateman (1994)

qui ont étudié le cas d'un patient souffrant d'agnosie pour les objets sans signe d'alexie ni de prosopagnosie.

Une autre hypothèse postule que ce trouble serait dû à une atteinte spécifique du module de traitement des visages. Elle s'appuie sur l'observation d'une double dissociation entre reconnaissance des visages et d'autres catégories de stimuli visuels. Ainsi, plusieurs cas de prosopagnosie sans trouble de reconnaissance d'autres types de stimuli visuels ont été décrits (De Renzi, Faglioni, Grossi, & Nichelli, 1991 ; McNeill & Warrington, 1993 ; Duchaine & Nakayama, 2005). A l'inverse, des cas d'agnosie sévère pour les objets sans trouble de la reconnaissance des visages ont également été décrits (Moscovitch, Winocur, & Behrmann, 1997).

Enfin, Damasio, Damasio et Van Hoesen (1982) proposent l'hypothèse selon laquelle la prosopagnosie serait due à une légère agnosie visuelle ne se marquant que pour les visages, plus complexes à traiter que les autres stimulations visuelles. En effet, alors que la plupart des objets sont reconnus à un niveau basique (c'est-à-dire comme représentant un exemplaire d'une catégorie donnée d'objets), chaque visage doit être identifié en tant qu'exemplaire particulier parmi l'ensemble des représentations de visages hautement

similaires stockées en mémoire à long terme. Appuyant cette hypothèse de déficit d'expertise visuelle, plusieurs études ont montré que des patients prosopagnosiques présentaient également des troubles de la reconnaissance d'objets pour lesquels ils avaient acquis une expertise visuelle (Damasio et al., 1982 ; Gauthier, Behrmann, & Tarr, 1999). Néanmoins, le patient prosopagnosique WJ, décrit par McNeil et Warrington (1993), qui devint éleveur de moutons après son accident cérébral, apprit à reconnaître et nommer la plupart de ses moutons. De même, le patient prosopagnosique décrit par Sergent et Signoret (1992a) était tout-à-fait capable de discriminer différents modèles de voitures d'une même marque pour laquelle il avait acquis une expertise visuelle. Ces différents cas suggèrent que la prosopagnosie ne résulte pas simplement d'une perte des capacités de discrimination visuelle entre différents exemplaires d'une même catégorie d'objets. La question de savoir si la reconnaissance des visages est assurée par un module spécifique de traitement ou répond à des capacités élevées d'expertise visuelle, n'est toujours pas résolue à l'heure actuelle.

Outre l'étude des déficits proprement dits, nombre de chercheurs ont tenté de localiser les lésions cérébrales

responsables de la prosopagnosie, sans parvenir toutefois à un consensus. Certaines études de patients prosopagnosiques, utilisant des techniques d'imagerie cérébrale telles que le CT-scan, la PET-scan ou l'IRMf, mettent en évidence des lésions des cortex occipito-temporaux bilatéraux (Damasio et al., 1982 ; De Renzi & di Pellegrino, 1998 ; Buxbaum, Glosser, & Coslett, 1999), alors que d'autres n'observent que des lésions du gyrus temporo-occipital droit (De Renzi, Peroni, Carlesimo, Silveri, & Fazio, 1994 ; Takahashi, Kawamura, Hirayama, Shiota, & Isono, 1995). Face à cette apparente contradiction, De Renzi et al. (1994) émettent l'hypothèse que (1) dans la plupart des cas, les deux hémisphères cérébraux pourraient être impliqués dans le traitement des visages et qu'une lésion bilatérale des gyri occipito-temporaux serait nécessaire pour provoquer une prosopagnosie ; (2) dans une minorité de cas, les processus de traitement des visages seraient tellement latéralisés dans l'hémisphère droit qu'une lésion des régions occipito-temporales de cet hémisphère serait suffisante pour provoquer un déficit de reconnaissance des visages. Néanmoins, même si ces études s'accordent à reconnaître la prépondérance des régions postérieures droites dans la

reconnaissance des visages, il n'en reste pas moins que les régions homologues de l'hémisphère gauche semblent également y jouer un rôle important. Ainsi, Mattson, Levin et Grafman (2000) décrivent le cas d'un patient souffrant de prosopagnosie associative et dont l'examen neurologique n'a révélé qu'une lésion des régions temporo-occipito-pariétales gauches.

Au niveau intra-cortical, il semble que des lésions des gyri fusiformes et linguaux soient déterminantes dans l'apparition des troubles prosopagnosiques (Takahashi et al., 1995). Plus spécifiquement, de récentes études ont montré un rôle majeur d'une région du gyrus occipital inférieur droit, la OFA (Occipital Face Area) dans l'apparition des troubles de reconnaissance des visages. Rossion, Caldara, Seghier, Schuller, Lazeyras et Mayer (2003) ont ainsi décrit le cas d'une patiente prosopagnosique présentant une lésion de cette partie du gyrus occipital inférieur droit sans atteinte du gyrus fusiforme moyen droit correspondant à la FFA (Face Fusiform Area). Steeves, Culham, Duchaine, Pratesi, Valyear, Schindler, Humphrey, Milner et Goodale (2006) observent le même profil de lésion chez un patient présentant une agnosie visuelle et une prosopagnosie associée. Ces observations suggèrent que

la FFA et l'OFA de l'hémisphère droit constituent un réseau d'aires cérébrales dont l'intégrité est nécessaire aux processus de reconnaissance des visages.

Toutefois, ces hypothèses localisatrices doivent être nuancées à la lumière de l'étude de Clarke, Lindemann, Maeder, Borruat et Assal (1997). Ces auteurs ont montré que deux patients, souffrant de prosopagnosie associative, présentaient des lésions situées dans des régions cérébrales distinctes dont la superposition est minimale. En outre, la comparaison de ces sites lésionnels avec les résultats d'études d'imagerie cérébrale antérieures, a indiqué que (1) les aires impliquées dans la reconnaissance des visages chez les sujets sains ne correspondaient pas à celles lésées chez les deux patients, (2) les régions lésées n'étaient pas activées chez les sujets sains. Ainsi, il semble qu'il existe une certaine variabilité inter-individuelle dans la localisation des lésions responsables des troubles prosopagnosiques, leur hétérogénéité renforçant d'autant plus les difficultés de localisation cérébrale.

Données d'imagerie cérébrale fonctionnelle

Le recours aux techniques d'imagerie cérébrale fonctionnelle - que ce soit la tomographie par émission de positrons (PET-scan) ou l'imagerie par résonance magnétique fonctionnelle (IRMf) - pour étudier les régions cérébrales impliquées dans le traitement des visages est assez récent. Une des toutes premières études à examiner ces activations cérébrales fut celle de Sergent, Otha et McDonald (1992b), qui montrent, dans une tâche d'identification de visages, l'activation d'un réseau incluant le gyrus lingual droit, les gyri fusiformes des deux hémisphères, le cortex temporal antérieur bilatéral et le gyrus parahippocampique. Dès cet instant, la littérature ne cessera de s'enrichir d'études ayant pour but de localiser les aires cérébrales responsables de la perception et de la reconnaissance des visages (Haxby, Horwitz, Ungerleider, Maisog, Pietrini, & Grady, 1994 ; Andreasen, O'Leary, Arndt, Cizadlo, Hurtig, Rezai, Watkins, Ponto, & Hichwa, 1996 ; Puce, Allison, Asgari, Gore, & McCarthy, 1996, Kanwisher, McDermott, & Chun, 1997). En particulier, l'étude de Kanwisher et al. (1997) marquera un tournant dans l'étude du traitement des visages, puisqu'ils seront les premiers à

mettre en évidence l'activation spécifique d'une région particulière du gyrus fusiforme, principalement à droite, qu'ils appelleront « Fusiform Face Area » (FFA). Tentant de définir les propriétés de réponses de cette région, ils établiront que son activité est plus importante lors de la perception passive de visages humains que lors de la perception d'autres objets visuels ou d'autres parties du corps humain. D'autres études montreront que la FFA est préférentiellement activée par la perception de visages humains en comparaison à la perception de suite de lettres (Puce et al., 1996), d'animaux (Kanwisher, Stanley, & Harris, 1999), de têtes humaines vue de dos ou de maisons (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). Sur base de la conception modulaire du fonctionnement cérébral de Fodor (1983), les auteurs postuleront l'existence d'un module spécifique au traitement des visages et envisageront la perception et la reconnaissance des visages comme étant un domaine spécifique de la cognition humaine (Kanwisher, 2000).

Néanmoins, s'appuyant sur l'observation que les objets ne sont souvent identifiés que comme appartenant à une catégorie donnée (Jolicoeur, Gluck, & Kosslyn, 1984), alors les visages doivent être identifiés comme exemplaires particuliers

et spécifiques au sein de la catégorie (Tanaka, 2001), d'autres auteurs mettront en doute cette conception modulaire du traitement des visages. Ainsi, Gauthier, Tarr, Anderson, Skudlarski et Gore (1999) proposent l'hypothèse selon laquelle la FFA ne serait pas spécifiquement dédiée au traitement des visages mais refléterait plutôt le traitement de stimuli visuels complexes nécessitant un haut niveau d'expertise. Différentes études, appuyant cette hypothèse, montreront que la FFA est aussi activée par la perception de visages que par la perception de stimuli (voitures, oiseaux, greebles, c'est-à-dire des objets virtuels non-réels et non-faciaux générés par ordinateur et dont chaque exemplaire peut être identifié) pour lesquels les sujets avaient préalablement acquis une expertise visuelle (Tarr, & Gauthier, 2000 ; Gauthier, Skudlarski, Gore, & Anderson, 2000 ; Xu, 2005). Notons cependant que le débat est loin d'être clos. Grill-Spector, Knouf et Kanwisher (2004) ont ainsi récemment montré que l'activation de la FFA de l'hémisphère droit était corrélée à la perception de visages, mais aussi à l'identification de visages particuliers, alors que la perception et l'identification d'autres stimuli (dans le cas présent, des voitures chez des sujets experts) n'était pas corrélée à une

activation préférentielle de la FFA. De même, Rhodes, Byatt, Michie et Puce (2004) ont montré que la FFA était plus activée par la reconnaissance individuelle de visages que par la reconnaissance individuelle de papillons, et ceci même chez des entomologistes experts en papillons. Leurs résultats d'imagerie indiquent cependant que la reconnaissance des papillons activait 15% des voxels situés dans la FFA. Ils en concluent que certains neurones du gyrus fusiforme pourraient être impliqués dans les processus d'identification d'exemplaires particuliers des différentes classes d'objets, mais que la FFA serait spécifiquement dédiée aux processus d'identification des visages. A notre connaissance, il semble donc qu'à ce jour aucune étude ne soit parvenue à confirmer définitivement une hypothèse plutôt qu'une autre.

Bien que le gyrus fusiforme droit joue un rôle prépondérant dans la perception et la reconnaissance des visages humains, d'autres régions corticales semblent également impliquées dans ces processus. Turk, Rosenblum, Gazzaniga et Macrae (2005), contrastant les jugements faciaux à différents niveaux de spécificité sémantique (catégorisation selon l'identité ou la profession), ont mis en évidence une activation plus importante du gyrus fusiforme moyen gauche

lors de l'identification des visages présentés aux sujets. Cette activité préférentielle du gyrus fusiforme gauche supporte l'idée que les régions visuelles associatives sont impliquées dans l'accès aux représentations sémantiques relatives aux personnes. D'autres chercheurs corroboreront cette interprétation et mettront en lumière le rôle du gyrus fusiforme gauche dans l'accès aux représentations sémantiques des personnes, stockées en mémoire à long terme (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005).

Une région du gyrus occipital inférieur, dénommée Occipital Face Area (OFA, Gauthier, Tarr, Moylan, Skudlarski, Gore, & Anderson, 2000), semble également étroitement impliquées dans la perception des visages. Elle pourrait constituer, avec la FFA, un réseau complexe d'aires cérébrales dédié aux traitements configurationnels de haut niveau nécessaires à la reconnaissance des visages. Ainsi, l'activation de la OFA pourrait résulter de « feedback projections » en provenance de la FFA et servir à une analyse visuelle fine des visages (Rossion et al., 2003 ; Steeves et al., 2005).

Enfin, bien que moins systématiquement mise en évidence que la FFA ou la OFA, une activité du STS a

également été observée (Halgren, Dale, Sereno, Tootell, Marinkovic, & Rosen, 1999 ; Haxby, Ungerleider, Clark, Schouten, Hoffman, & Martin, 1999 ; Allison, Puce & McCarthy, 2000 ; Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001 ; Puce, Syngeniotes, Thompson, Abbott, Wheaton, & Castiello, 2003). Le rôle fonctionnel du STS reste controversé. Pour certains auteurs, son activation reflèterait les processus de traitement des stimuli sociaux (Allison, Puce, Spencer & McCarthy, 1999), pour d'autres, il serait principalement impliqué dans l'analyse émotionnelle des visages (Narumoto et al., 2001) ou dans l'intégration des formes et des mouvements faciaux (Puce et al., 2003).

Données électrophysiologiques

Si les études d'imagerie fonctionnelle cérébrale ont mis au jour l'existence de régions cérébrales spécialisées dans la perception et la reconnaissance des visages, les études électrophysiologiques utilisant la technique des potentiels évoqués ont quant à elles aussi examiné les activités cérébrales liées au traitement des visages, mettant ainsi en

évidence des ondes spécifiques à ces processus de traitement de l'information.

Allison et ses collaborateurs (1994, 1999), comparant les activités électrophysiologiques intracrâniennes enregistrées en réponse à la présentation de différents stimuli visuels chez des patients épileptiques, détecteront une activité négative apparaissant aux environs de 200 millisecondes (ms) au niveau du cortex occipito-temporal ventral et du gyrus temporal moyen. Ils remarqueront que cette N200 présente une amplitude nettement augmentée suite à la présentation de visages humains que tout autre stimulation visuelle, suggérant que cette activité est spécifique au traitement des visages.

L'enregistrement de l'activité électrique cérébrale à la surface du scalp fera également apparaître différentes ondes liées au traitement des visages. La première d'entre elles est le complexe d'ondes N170/VPP (Occipito-temporal Negativities/Vertex Positive Potential). La N170 est une onde négative enregistrée de manière maximale sur les électrodes temporales aux alentours de 170 millisecondes (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Ce potentiel est classiquement couplé à une onde positive, la VPP,

apparaissant au même moment sur les électrodes centrales (Jeffreys, 1989; Rossion, Campanella, Delinte, Debatisse, liard, Dubois, Bruyer, Crommelinck, & Guérit, 1999). Alors que la N170 est dans la majorité des cas plus ample dans l'hémisphère droit, la VPP ne semble pas latéralisée. Ces deux activités ont récemment été interprétées comme reflétant l'activité d'un même générateur neural et donc comme représentant deux facettes d'un unique processus cérébral (Joyce & Rossion, 2005). La N170 apparaît suite à la présentation de visages, qu'ils soient humains ou non (par exemple des visages de singes, Carmel & Bentin, 2002) et est considérablement plus ample pour les visages que pour d'autres catégories d'objets, tels que des oiseaux, des mains humaines, des voitures, des fleurs ou des papillons (Bentin et al., 1996). Il a également été démontré que cette onde n'est pas sensible à la familiarité des visages (Bentin & Deouell, 2000, Eimer, 2000a), qu'elle peut être générée par des parties de visages, en particulier les yeux (Bentin et al., 1996), et qu'elle est sensible à l'effet d'inversion, sa latence étant retardée et son amplitude augmentée lorsque les visages subissent une rotation de 180° (Rossion, Delvenne, Debatisse, Goffaux, Bruyer, Crommelinck, & Guérit, 1999 ; Rossion,

Gauthier, Tarr, Despland, Bruyer, Linotte, & Crommelinck, 2000 ; Eimer, 2000b). L'absence de sensibilité à la familiarité pousseront les auteurs à émettre l'hypothèse que la N170 reflète des processus d'encodage structurel des visages prenant place avant l'identification, que ces processus d'encodage concernent les différentes parties du visage (Eimer, 1998), les yeux (Bentin et al., 1996), ou une combinaison de traitements holistiques et analytiques (Sagiv & Bentin, 2001). L'effet d'inversion quant à lui peut être expliqué par une surcharge de traitement pour les visages inversés. Toutefois, d'autres données suggèrent que la N170 pourrait également être modulée par des processus de plus haut niveau impliqués dans le traitement de l'identité des visages. Ainsi, Tanaka, Luu, Weisbrod et Kiefer (1999) ont remarqué que l'amplitude de la N170 était plus grande lorsque les objets, dont des visages, étaient traités à un niveau sémantique sub-ordonné plutôt qu'au niveau général de la catégorie. Campanella, Hanoteau, Dépy, Rossion, Bruyer et Crommelinck (2000) ont quant à eux observés un effet de répétition sur la N170 (diminution de l'amplitude suite à la répétition de visages) dépendant de l'identité des visages mais pas de leurs caractéristiques physiques. Ils n'ont par

contre pas observé cet effet de répétition lorsque les visages n'avaient pas la même identité. Cet effet a été répliqué par Itier et Taylor (2002), et par Heisz, Watter et Shedden (2006). Signalons qu'un effet similaire a également été observé sur la FFA par Rothstein, Henson, Treves, Driver et Dolan (2005), qui ont montré une diminution de l'activité de la FFA suite à la répétition de visages. Leur étude a permis de démontrer que la FFA droite est sensible à l'identité des visages mais pas à leurs propriétés physiques.

Différentes études ont tenté de localiser les générateurs neuraux du complexe N170/VPP, qui semble pouvoir être expliqué par une modélisation dipolaire comprenant une seule paire de dipôles localisés bilatéralement dans le cortex occipito-temporal et pointant vers les régions centro-frontales (Itier & Taylor, 2002 ; Schweinberger, Pickering, Jentsch, Burton, & Kaufmann, 2002). Plus particulièrement, Schweinberger et al. (2002) mettront en évidence une source occipito-temporale postérieure et latéralisée pour la N170, située en dehors du gyrus fusiforme et de la FFA.

Si le complexe N170/VPP semble lié à l'encodage structurel des visages, d'autres ondes cérébrales ont été mises en évidence, qui pourraient refléter les processus de

reconnaissance individuelle de chaque visage. Ainsi, Schweinberger et al. (2002) ont montré, dans une tâche de reconnaissance, que la répétition de visages familiers provoquait l'apparition d'une onde négative entre 180 et 290 ms sur les électrodes temporales inférieures, avec une prédominance dans l'hémisphère droit. Cette onde N250r était diminuée, voire annulée suite à la présentation de visages inconnus, mais était par contre maximale lors de la répétition de visages célèbres. Ces résultats ont amené les auteurs à postuler que cette onde refléterait l'accès aux représentations des visages familiers stockés en mémoire. Une modélisation dipolaire a en outre montré que la source de la N250r se trouvait dans le gyrus fusiforme droit, pouvant correspondre à la FFA.

De même, Bentin & Deouell (2000) et Eimer (2000a) ont mis en évidence, lors de tâches de reconnaissance de visages célèbres, l'existence d'un complexe d'activités cérébrales constitué d'une onde négative apparaissant entre 250 et 500 ms de manière fronto-centrale (Bentin & Deouell, 2000) ou pariéto-centrale (Eimer, 2000a), suivie d'une onde positive enregistrée à partir de 500 ms. Ce complexe N400f/P600f était sensible à la familiarité, n'apparaissant pas suite à la

présentation de visages inconnus. Une positivité fronto-centrale apparaissant à partir de 550 ms et semblable à la P600f a également été observée par Henson, Goshen-Gottstein, Ganel, Otten, Quayle, & Rugg (2003) dans une tâche de reconnaissance de visages. Il est dès lors possible que ce complexe d'ondes soit lié au traitement sémantique des visages familiers. Plus spécifiquement, Eimer (2000a), sur base du modèle d'identification des visages de Bruce & Young (1986), émet l'hypothèse que la N400f reflèterait l'accès aux représentations faciales stockées en mémoire, alors que la P600f reflèterait l'accès aux informations sémantiques liées aux visages.

Enfin, signalons l'étude de Paller, Gonsalves, Grabowecky, Bozie, & Yamada (2000), qui ont étudié les différences électrophysiologiques liées à la récupération de visages connus ou non. Trois types de visages étaient présentés : des visages inconnus, des visages précédemment étudiés, avec ou sans informations biographiques associées. Leurs résultats montrent que les visages familiers étaient associés à des potentiels évoqués positifs entre 300 et 600 ms par rapport aux visages non-familiers. De plus, ils ont mis en évidence des différences entre les visages familiers lorsque

ceux-ci étaient ou non accompagnés d'informations biographiques. En effet, pour cette dernière catégorie de visages, la positivité était observée sur les électrodes antérieures et postérieures, alors que les visages sans informations biographiques n'ont produit de négativité que sur les électrodes postérieures. Ces résultats laissent supposer deux types de traitement des visages connus : un premier type de traitement permettant l'accès aux représentations faciales stockées en mémoire et sous-tendues par des activités cérébrales postérieures, un deuxième type de traitement permettant l'accès aux représentations sémantiques et biographiques liées aux personnes familières et sous-tendues par des activités cérébrales frontales et préfrontales. Certains auteurs prétendent néanmoins que la reconnaissance des visages est plus précoce et peut être réalisée bien avant 100 millisecondes. Ainsi, Seeck, Michel, Mainwaring, Cosgrove, Blume, Ives, Landis et Schomer (1997) ont montré, en comparant les ondes électrophysiologiques produites par la perception de visages familiers et de nouveaux visages, que la reconnaissance des visages entraînait des réponses cérébrales distinctes dès 50 millisecondes après la présentation des stimuli. Debruille, Guillem et Renault (1998) contesteront

toutefois ces résultats. Selon eux, ces ondes précoces pourraient refléter des effets de répétition des stimuli plutôt que des processus de reconnaissance. Utilisant la même méthodologie que Seeck et al. (1997), ils observeront que les premiers effets liés à la reconnaissance des visages ne prennent cours qu'aux environs de 100 millisecondes.

Conclusions

Les données lésionnelles, anatomiques et électrophysiologiques indiquent qu'il existe un vaste réseau d'aires cérébrales répondant préférentiellement à la perception de visages humains. Ce réseau implique les régions occipito-temporales bilatérales, et particulièrement le gyrus fusiforme droit, tant dans la perception que dans la reconnaissance des visages familiers. Néanmoins, il nous faut souligner qu'il existe toujours à l'heure actuelle un débat quant à la spécificité de ces régions cérébrales dans la perception et la reconnaissance des visages. Certains auteurs penchent pour l'hypothèse de l'existence d'un module spécifique au traitement des visages, d'autres se positionnent en faveur de l'existence de mécanismes cérébraux spécifiques

aux objets pour lesquels nous aurions acquis une expertise visuelle particulière.

Chapitre 2:

corrélats cérébraux du traitement des noms propres.

Les noms de famille permettent de faire référence aux individus que nous connaissons, et tiennent à ce titre un rôle tout aussi primordial que les visages dans nos interactions sociales. Le but de ce chapitre sera de démontrer, via l'étude de patients cérébro-lésés et les études d'imagerie cérébrale fonctionnelle et électrophysiologiques, que le traitement des noms propres requiert l'activation de régions cérébrales spécifiques et distinctes de celles impliquées dans le traitement des visages.

Toutefois, avant d'aborder ces différentes études et par souci de cohérence, nous présenterons succinctement le modèle de reconnaissance des noms de famille le plus répandu.

Modèle cognitif des processus de reconnaissance des noms de famille

Valentine, Brédart, Lawson et Ward (1991) proposent un modèle séquentiel de la reconnaissance des noms de famille, qui complète le modèle de reconnaissance des visages de Bruce & Young (1986) présenté précédemment (Figure 4). La première étape de ce modèle correspond à la création d'une représentation perceptive des noms. Cette représentation est ensuite mise en correspondance avec une unité de reconnaissance des noms (Name Recognition Unit, NRU), déclenchant le sentiment de familiarité. Cette NRU donne accès aux informations biographiques concernant la personne, qui peuvent également être activées par le visage de la personne concernée. Enfin, le nom peut être produit via l'activation du lexique phonologique de sortie.

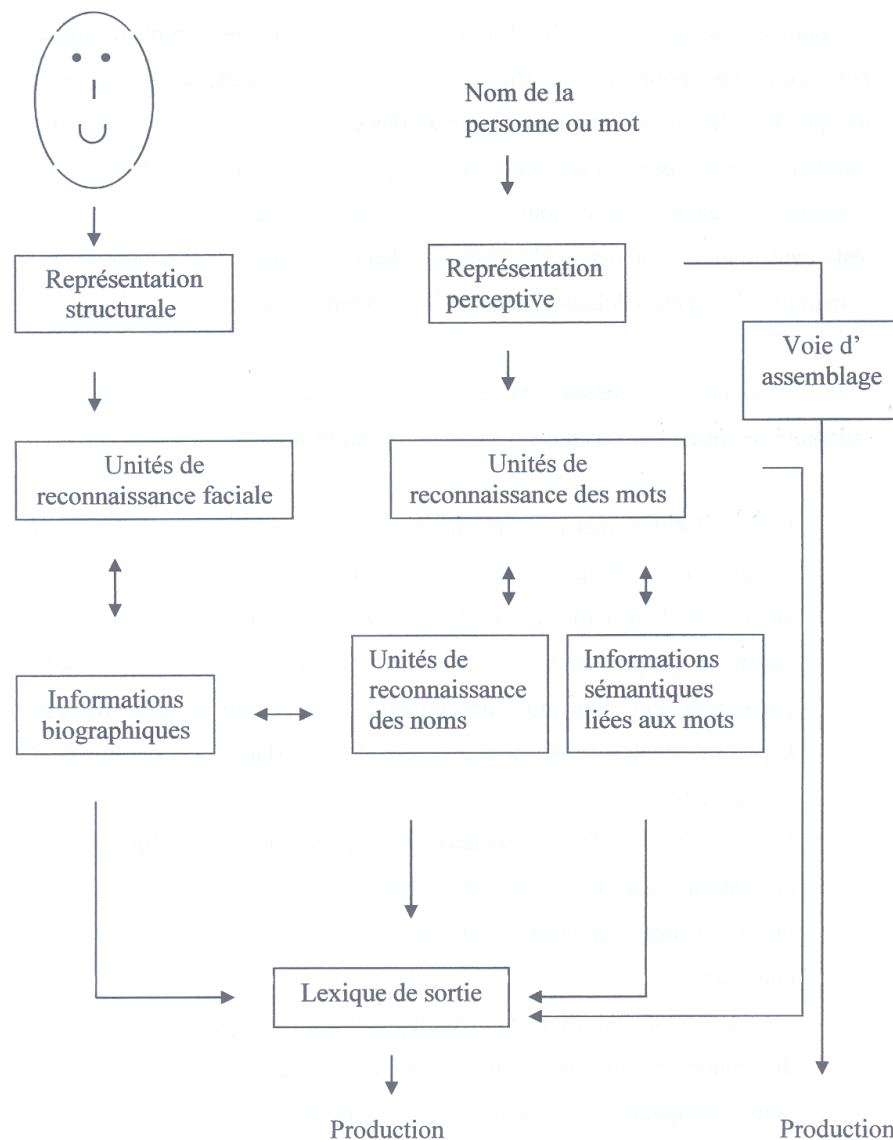


Figure 4 : Modèle cognitif de reconnaissance et de production des noms de personnes de Valentine et al. (1991)

Données lésionnelles – L'anomie pour les noms propres

L'anomie pour les noms propres peut se définir comme étant un trouble sélectif de la production de noms propres, la production de noms communs étant préservée. L'anomie pour les noms propres peut avoir différents niveaux de sélectivité :

anomie pour toutes les classes de noms propres (noms de famille, noms de lieux, de monuments, noms géographiques), anomie pour les noms de famille, voire anomie pour une catégorie spécifique de noms de famille (noms de personnes célèbres, de personnages historiques, de proches).

Différents cas de patients, présentant un déficit de production des noms propres (de famille et/ou géographiques) avec une préservation de la production des noms communs ont été décrits (Semenza & Zettin, 1988 ; Flude, Ellis, & Kay, 1989 ; Carney & Temple, 1993), menant à l'hypothèse que le traitement des noms propres, à l'instar du traitement des visages par rapport aux objets, pourrait être spécifique et dissociable du traitement des noms communs. Néanmoins, d'autres auteurs mettront en doute cette hypothèse et proposeront plutôt que les troubles observés chez les patients anomiques pour les noms propres seraient dus à un déficit d'apprentissage et/ou de récupération d'associations arbitraires entre deux informations (Luchelli & De Renzi, 1992 ; Hittmair-Delazer, Denes, Semenza, & Mantovan, 1994 ; Saetti, Marangolo, De Renzi, Rinaldi, & Lattanzi, 1999).

Des patients souffrant du pattern de déficits inverse, à savoir une altération de la production des noms communs avec une préservation de la production de noms propres, ont également été décrits (Semenza & Sgaramella, 1993 ; Cipolotti, McNeil, & Warrington, 1993). Cette apparente double dissociation, en faveur d'une spécificité de la production des noms propres par rapport aux noms communs, doit cependant être nuancée. Il faut tout d'abord noter que tous les patients décrits dans ces études présentaient un trouble principalement circonscrit à la production de noms propres, avec une préservation – relative ou entière - des connaissances sémantiques et biographiques relatives aux personnes qu'ils ne parvenaient plus à dénommer. De plus, Brédart, Brennen et Valentine (1997) soulignent qu'il règne une certaine confusion entre troubles de la production et troubles de la compréhension, et que le cas d'anomie spécifique aux noms communs décrits par Cipolotti et al. (1993) pourrait également s'expliquer par une atteinte du système sémantique.

Il semblerait néanmoins que l'on puisse s'acheminer vers une double dissociation quant aux systèmes sémantiques propres aux noms de famille et aux noms communs. Miceli,

Capasso, Daniele, Esposito, Magarelli et Tomaiuolo (2000) décrivent le cas du patient APA, souffrant d'une anomie spécifique aux noms de famille. L'examen neuropsychologique révélera un déficit dans la récupération des informations biographiques relatives aux personnes que le patient ne peut dénommer. Lyons, Hanley et Kay (2002) décriront le cas inverse d'un patient présentant une anomie sévère pour les noms communs et les noms propres géographiques, associée à un déficit des connaissances sémantiques liées à ces différentes catégories, sans toutefois de déficit d'accès aux connaissances biographiques ou de dénomination des personnes. Les auteurs interprètent ces données et celles de Miceli et al. (2000) comme reflétant une double dissociation en faveur de l'hypothèse d'une organisation du système sémantique en catégories spécifiques (Caramazza & Shelton, 1998). Il est à noter également que Yasuda & Ono (1998), étudiant trois groupes de patients souffrant d'aphasie globale, remarqueront que ces patients restaient meilleurs en compréhension de noms de personnes qu'en compréhension de noms géographiques ou de noms communs, ce qui laisse penser que des distinctions

peuvent également exister au sein même de la catégorie des noms propres.

Peu de ces études se sont intéressées à la localisation précise des lésions cérébrales responsables des troubles décrits. Notons toutefois que les examens par CT-scan réalisés auprès des patients ont majoritairement révélé des lésions localisées dans l'hémisphère gauche et incluant principalement la jonction occipito-pariétale (Semenza & Zettin, 1988, Yasuda & Ono, 1998), le gyrus frontal (Lyons, Hanley, & Kay, 2002), le gyrus temporal médian (Saetti et al., 1999) et le thalamus (Luchelli & De Renzi, 1992). Reinkemeier, Markowitsch, Rauch et Kessler (1997) ont examiné quant à eux de manière plus détaillée les sites lésionnels associés à l'anomie pour les noms de famille. Leur examen d'un patient anomique par PET-scan a montré un hypométabolisme circonscrit à l'hémisphère gauche du cortex temporal médian, du cortex occipito-temporal et du thalamus. Un examen par IRMf a également révélé une lésion du gyrus parahippocampique et du thalamus gauche. Papagno & Capitani (1998) ont observé une lésion du gyrus temporal antérieur gauche. Cette hétérogénéité des lésions au sein de

l'hémisphère gauche sera confirmée par Semenza, Mondini et Zettin (1995) dans une méta-analyse des régions cérébrales lésées chez douze patients anomiques pour les noms propres.

Données d'imagerie cérébrale fonctionnelle

L'enregistrement de l'activité cérébrale de sujets sains dans des tâches de récupération de noms propres confirme l'implication majeure du gyrus temporal gauche dans ce type de tâche. Damasio, Grabowski, Tranel, Hichwa et Damasio (1996) examinent par PET-scan les régions cérébrales impliquées dans la récupération de noms de personnes, d'animaux et d'outils. Chez des sujets sains, ils observent que la dénomination de chaque catégorie d'objets était associée à l'activation de régions distinctes du lobe temporal, respectivement le pôle temporal, le gyrus temporal inférieur et le gyrus temporal postérieur. La même tâche est réalisée auprès de patients cérébro-lésés. Ils observent une corrélation entre les performances à la tâche et le site lésé. Ainsi, ils notent une corrélation significative entre déficit en dénomination de personnes et lésion du pôle temporal gauche, déficit en dénomination d'outils et lésion de la partie

postérieure du gyrus temporal inférieur, et entre déficit de dénomination des animaux et lésion de la partie antérieure du gyrus temporal inférieur. Graboswki, Damasio et Damasio (1998) se focaliseront sur l'activité des lobes frontaux dans ce type de tâche. Ils observeront une activation du gyrus frontal inférieur gauche, présente pour la dénomination des trois catégories d'objets, mais n'observeront l'activation de la partie antérieure du gyrus frontal médian gauche que pour la dénomination de visages. Ainsi, il semble que la récupération lexicale de noms de personnes implique un réseau cortical distribué au sein des gyri temporaux et frontaux de l'hémisphère gauche, avec une participation spécifique du gyrus temporal antérieur.

Gorno-Tempini, Price, Josephs, Vandenberghe, Cappa, Kapur et Frackowiack (1998) ont comparé quant à eux la perception des noms de familles et des noms communs. Ils ont remarqué que, comparés à des suites de lettres non-significatives, les noms propres et les noms communs activaient le sillon temporal supérieur et la partie postérieure du gyrus temporal moyen gauches. De plus, par rapport aux noms communs, les noms propres activaient le gyrus temporal moyen antérieur (aire 21 de Brodmann). Contrairement à

Damasio et al. (1996), ils interpréteront cette activité du gyrus temporal antérieur gauche comme reflétant les différences de complexité de traitement entre noms communs et noms propres, ces derniers représentant des entités uniques, plus difficiles à traiter. Cette interprétation est en accord avec celle de Sergent, McDonald et Zuck (1994), qui ont réalisé une étude par PET-scan afin de déterminer les régions cérébrales impliquées dans le traitement des noms de famille. Trois tâches étaient proposées : une catégorisation des noms en fonction du sexe, un jugement de familiarité et une catégorisation en fonction de la profession. Une activité du pôle temporal gauche a été observée dans les tâches de jugement de familiarité et de catégorisation selon la profession, qui nécessitaient une identification des personnes et étaient donc plus complexes à réaliser que la catégorisation selon le sexe.

Tsukiura, Fujii, Fukatsu, Otsuki, Okuda, Umetsu et al. (2002) ont également mis en évidence par IRMf, chez des patients lobectomisés et des chez des sujets sains, l'importance du pôle temporal gauche dans la récupération des noms de famille.

Données électrophysiologiques

Peu d'études se sont intéressées au décours temporel des activités cérébrales liées au traitement des noms propres. Citons néanmoins l'étude de Dehaene (1995), qui a comparé les potentiels évoqués dans une tâche de catégorisation de mots appartenant à différentes catégories (animaux, verbes d'actions, nombres, noms de célébrités et suites de consonnes). Des divergences significatives pour chaque catégorie de mots sont apparues entre 250 et 280 ms après l'apparition des stimuli, les noms de célébrités produisant spécifiquement une composante négative temporelle inférieure gauche. Dehaene (1995) interprète ces différences électrophysiologiques comme reflétant une organisation catégorielle des mots au sein des régions temporo-pariétales gauches. Proverbio, Lilli, Semenza et Zani (2001) ont également tenté d'isoler les activités cérébrales liées à la récupération de noms communs et de noms propres. Ils ont enregistré une onde négative apparaissant vers 150 millisecondes sur les électrodes antérieures du lobe temporal gauche et spécifique à la récupération de noms de personnes, et ils ont également mis en évidence une augmentation de

l'amplitude de la P300 pour les noms de personnes, qu'ils interprètent comme étant le reflet de la plus grande complexité de traitement des noms propres, due à leur caractère unique.

Sommer, Komoss et Schweinberger (1997) ont enregistré les potentiels évoqués dans deux tâches distinctes d'apprentissage et de reconnaissance de visages et de prénoms. Calculant la différence entre potentiels évoqués lors de l'apprentissage et lors de la reconnaissance des stimuli, ils observent que les visages et les prénoms ont produit des activités cérébrales distinctes entre 500 et 800 millisecondes: le système mnésique responsable de la reconnaissance des visages a produit une activité positive soutenue sur les électrodes pariétales, alors que la reconnaissance des prénoms a produit une activité positive sur les électrodes frontales. Ces données les amènent à conclure à l'existence de systèmes mnésiques différents pour les visages et les noms propres.

Ces différentes études ont toutes présenté les noms propres dans la modalité visuelle. Or, ceux-ci peuvent également être présentés en modalité auditive. Müller et

Kutas (1996) montreront d'ailleurs que les différences de traitement existant entre noms communs et noms propres présentés en modalité visuelle existent également lorsque les noms sont présentés auditivement. Leur étude en potentiels évoqués consistait en une écoute de phrases commençant par un nom de famille, un nom commun ou le propre nom du sujet. Ils ont montré que les composantes N100 (sur les électrodes centrales) et P200 (sur les électrodes temporales et occipitales) étaient plus amples lors de la perception des noms de famille que des noms communs. La perception par les sujets de leur propre nom entraînait par ailleurs une onde N400 pariéto-centrale, reflétant possiblement la surprise des sujets. Les différentes catégories de noms étant appariées quant à leurs caractéristiques physiques, les auteurs en concluent que ces résultats apportent la preuve d'une organisation du lexique en domaines spécifiques.

Conclusions

Les différentes études abordées dans ce chapitre tendent toutes à montrer que le traitement des noms propres, que ce soit lors d'épreuves de perception, de catégorisation, ou de

dénomination, requiert l'activation d'un réseau d'aires cérébrales latéralisé dans l'hémisphère gauche et incluant principalement les gyri temporaux antérieurs et frontaux médians. Différents débats restent cependant d'actualité, tant sur le statut des noms propres que sur la signification fonctionnelle des aires activées. Ainsi, l'étude de patients cérébro-lésés nous apprend que si la spécificité des traitements des noms propres en général – et des noms de famille en particulier – par rapport aux noms communs n'a pas à ce jour été clairement établie au niveau de la production des noms, il semble néanmoins que les systèmes sémantiques de ces deux catégories de noms doivent être envisagés de manière distincte.

Les études d'imagerie cérébrale et électrophysiologiques ont quant à elles porté le débat sur la question des processus sous-tendus par les régions temporelles antérieures. Certains auteurs soutiennent une vision modulaire des processus de traitement des noms propres par rapport aux noms communs, et donc des systèmes sémantiques et lexicaux organisés en catégories de noms ; d'autres pensent plutôt que l'activation de ces régions reflèterait les différences de complexité entre les deux catégories de noms, les noms propres étant plus

complexes à traiter que les noms communs de par leur caractère unique.

Quoi qu'il en soit, alors que dans le chapitre précédent, les différentes études passées en revue tendaient à montrer l'implication majeure du gyrus fusiforme, principalement droit, dans la perception et la reconnaissance des visages, les études décrites dans ce chapitre nous indiquent que les régions cérébrales sous-tendant le traitement des noms propres, bien que plus diffuses au sein de l'hémisphère gauche, sont clairement distinctes de celles impliquées dans le traitement des visages.

Chapitre 3:

corrélats cérébraux du traitement des voix humaines.

La voix humaine est probablement le son le plus important de notre environnement, à tout le moins de notre environnement social. A l'instar des visages, nos capacités d'analyse et de catégorisation des informations contenues dans les voix jouent un rôle fondamental dans nos interactions sociales. Ces informations peuvent être divisées en trois catégories : les informations langagières, les informations relatives à l'identité du locuteur, et les informations relatives à l'état émotionnel du locuteur. Si beaucoup d'études neurolinguistiques se sont penchées sur les processus de traitement des informations langagières contenues dans le discours, les processus de traitement de l'identité et de l'émotion de la voix elle-même ont été nettement moins examinés. Pourtant, comme le souligne Belin, Fecteau et Bédard (2004), la voix peut être considérée comme « un visage auditif », puisque les deux stimulations

contiennent une combinaison unique de caractéristiques physiques propres à chaque individu. Nous ne nous intéresserons ici qu'aux études ayant tenté d'identifier les régions cérébrales sous-tendant les processus d'identification des voix.

Modèle cognitif des processus d'identification des voix

Tout comme pour les noms propres, Belin et al. (2004) proposent d'utiliser le modèle d'identification des visages de Bruce & Young (1986) pour rendre compte des processus cognitifs en jeu lors de l'identification des voix (Figure 5). Tout comme dans le modèle de Bruce & Young (1986), l'identification des voix commencerait par une phase d'encodage structurel permettant la création d'une représentation de la voix, qui sera ensuite comparée à une des représentations de voix – ou VRU (Voice Recognition Unit) stockées en mémoire à long terme. L'activation d'une VRU donne ensuite accès à un PIN, système amodal d'accès aux informations sémantiques relatives aux personnes. Il est à noter que l'activation d'une VRU est indépendante du contenu langagier ou du ton émotionnel de la voix entendue.

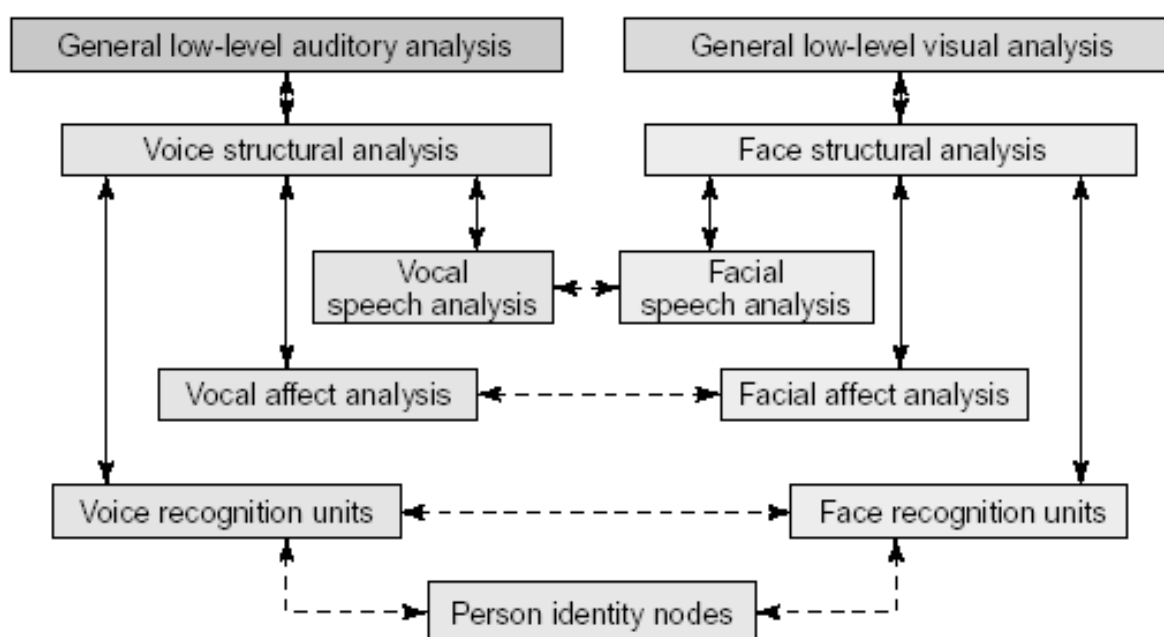


Figure 5 : Modèle cognitif de la reconnaissance des voix de Belin et al.

(2004)

Données lésionnelles – la phon(o)agnosie

Quelques études cliniques ont rapporté l'existence de cas de patients cérébro-lésés présentant un déficit de reconnaissance des voix, appelé « phonoagnosie » ou « phonagnosie » (Van Lancker & Canter, 1982 ; Van Lancker & Kreiman, 1987 ; Peretz, Kolinsky, Tramo, Labrecque, Hublet, Demeurisse, & Belleville, 1994). Ces études montrent que le déficit de reconnaissance des voix familières peut être distingué d'un déficit de discrimination des voix parmi d'autres stimuli auditifs, bien que ces deux types d'altérations

semblent dues à des lésions du cortex temporal droit. Van Lancker, Kreiman et Cummings (1989) observeront également une corrélation entre lésion du cortex pariétal droit et trouble de la reconnaissance des voix. Une double dissociation existe aussi entre trouble de la perception du langage et trouble de l'identification vocale. Ainsi, des cas de patients présentant un déficit de reconnaissance vocale avec préservation de la perception du langage ont été décrits, tout comme des cas de patients présentant un déficit de perception du langage tout en restant à même d'identifier les voix qui parlaient (Assal, Aubert & Buttet, 1981 ; cité dans Belin et al., 2004).

Donnée d'imagerie cérébrale fonctionnelle

Ce n'est que récemment que les neurosciences cognitives se sont intéressées aux substrats neuraux impliqués dans la reconnaissance des voix. Une des premières études sur sujets sains fut celle de Nakamura, Kawashima, Sugiura, Kato, Nakamura, Hatano et al. (2001), qui ont examiné par PET-scan les régions cérébrales activées durant une tâche de discrimination de voix familières parmi des voix inconnues. Leurs résultats montrent une activation de plusieurs régions

corticales durant ce jugement de familiarité, dont le pôle frontal gauche, le pôle temporal droit, le cortex enthorinal droit et le précunéus gauche. De plus, ils noteront une corrélation positive entre l'activation des régions temporales antérieures droites et les performances des sujets, confirmant ainsi les premières observations de Belin, Zatorre, Lafaille, Ahad et Pike (2000). Ainsi, plus les participants identifiaient correctement les voix familières, plus l'activité du pôle temporal droit était importante, ce qui suggère que cette région cérébrale pourrait être impliquée dans la mise en correspondance des voix entendues avec les représentations de voix stockées en mémoire.

Von Kriegstein, Eger, Kleinschmidt et Giraud (2003), utilisant la fMRI, ont comparé l'activité cérébrale dans une tâche d'identification dirigée sur le contenu verbal des stimuli ou sur l'identité du locuteur. Ils ont observé une activité de la partie antérieure de la scissure temporale supérieure (STS) et du précunéus droit lorsque les sujets devaient identifier le locuteur, et une activité du STS moyen gauche lorsque l'attention était dirigée sur le contenu verbal des stimuli plutôt que sur l'identité du locuteur. Il apparaît donc que l'activité de la partie antérieure du STS droit soit dédiée aux

processus d'identification des voix, indépendante des traitements verbaux réalisés sur ces voix. Belin et Zatorre (2003) confirmeront ces observations dans une étude par IRMf comparant l'activité cérébrale en réponse à l'écoute passive de syllabes énoncées soit par la même personne, soit par des personnes différentes. Ils montreront que l'écoute de syllabes prononcées par une même personne conduisait à une diminution de l'activité du STS droit. Cette diminution d'activité pourrait représenter un mécanisme d'adaptation des populations neuronales du STS aux caractéristiques acoustiques des voix.

Ces données constituent des preuves convergentes du rôle important des régions temporeles de l'hémisphère droit, et en particulier de la partie antérieure du STS, dans les processus d'identification des voix humaines. Fecteau, Armony, Joanette et Belin (2004) ont d'ailleurs montré que le STS est significativement plus activé lors de l'écoute de vocalisations humaines que lors de l'écoute de vocalisations animales ou de sons non-vocaux (instruments de musique), suggérant une spécificité du STS aux voix humaines. Récemment, la même équipe de chercheurs a montré, qu'outre le STS droit, la perception de voix humaines était

corrélée à une activation du cortex préfrontal inférieur gauche (pars orbitalis, Fecteau, Armony, Joanette, & Belin, 2005). Cette augmentation d'activité a été observée aussi bien pour la perception de stimuli linguistiques que de vocalisations, et était plus importante pour les vocalisations émotionnelles que neutres. Il semble donc que la perception et l'identification des voix humaines repose sur l'activation d'un réseau d'aires cérébrales incluant notamment les cortex temporaux et pariétaux droits et le gyrus frontal inférieur gauche.

Données électrophysiologiques

Rares sont les études ayant examiné les corrélats électrophysiologiques liés au traitement des voix humaines. Citons néanmoins les études de Levy, Granot et Bentin (2001, 2003). Ces auteurs ont observé que la présentation de voix humaines sous forme de chants produisait une onde positive apparaissant après 320 millisecondes et latéralisée dans l'hémisphère droit. Cette VSR (Voice-Specific Response) avait des caractéristiques permettant de la distinguer d'une P300 liée à la nouveauté, et n'était pas produite pour d'autres stimuli sonores de même fréquence fondamentale. Cette

observation a amené les auteurs à postuler l'existence d'un système neural spécifique au traitement des voix humaines. Ils ont également montré que lorsque les sujets ne prêtaient pas attention aux stimulations, cette VSR n'était pas observée. La composante attentionnelle serait donc primordiale dans l'apparition d'activités cérébrales spécifiques aux voix humaines. Gunji, Koyama, Ishii, Levy, Okamoto, Kakigi et Pantev (2003), utilisant les mêmes stimuli, ont observé par MEG (magnéto-encéphalographie) un champ magnétique dont la source était la plus forte suite à la perception de voix humaines, entre 300 et 500 ms.

Conclusions

Bien que les voix humaines puissent être considérées comme des « visages auditifs » (Belin et al., 2004), l'étude des corrélats neuraux du traitement de la voix n'en est qu'à ses débuts et ce champ des neurosciences cognitives n'a pas encore atteint l'ampleur de celui consacré à l'étude des processus de traitement des visages. Néanmoins, les diverses recherches consacrées à l'identification des voix mettent toutes en évidence l'implication importante du cortex

temporal droit, et plus particulièrement de la partie antérieure du STS. Comme le soulignent Belin et al. (2004), ces découvertes sont en total accord avec les modèles récents d'organisation du cortex auditif, dans lequel il existerait une voie ventrale homologue à celle existant dans le cortex visuel (Kaas & Hackett, 1999) et spécialisée dans la reconnaissance de stimulations auditives, et en particulier des voix humaines.

Chapitre 4:

Les processus associatifs entre visages, noms de famille et voix.

Les trois précédents chapitres ont été consacrés à l'étude des substrats neuraux sous-tendant les processus de traitement des principaux attributs au moyen desquels nous sommes capables d'identifier les personnes que nous connaissons. Nous avons vu que chaque attribut pris isolément requiert l'activation de régions neurales particulières et distribuées au sein du cerveau, dont notamment: les gyri fusiformes antérieurs, principalement droit pour les visages, le cortex temporal antérieur principalement gauche pour les noms de famille, et la partie antérieure du STS droit pour les voix.

La question qui se pose dès lors est de savoir comment le cerveau opère, à partir de ces activités neurales distinctes, pour créer et garder en mémoire une représentation unifiée et cohérente des personnes familières. Il est en effet communément admis que les différents canaux sensoriels

dont nous sommes pourvus interagissent pour maintenir une perception cohérente de notre environnement (Calvert, 2001), et, plus spécifiquement, que les différentes informations relatives aux personnes familières (apparence physique, visage, nom, caractéristiques personnelles, etc.) sont liées en mémoire à long-terme (Paller et al., 2003).

Depuis une quinzaine d'années, la littérature relative aux processus intégratifs s'est considérablement étoffée et a abordé une multitude de thèmes différents, parmi lesquels nous pouvons citer l'étude des processus intégratifs visuo-auditifs et somesthésiques chez l'animal (par exemple, Stein, Meredith, & Wallace, 1993 ; Stein, 1998 ; Wallace & Stein, 2000), chez l'homme, l'étude de leur ontogenèse et de leur développement de l'enfance à l'âge l'adulte (Bahrick, 1992 ; Lewkowicz, 1996, 2000a, 2000b), ainsi que les processus d'intégration visuo-tactile (Hadjikani & Roland, 1998 ; Banati, Goerres, Tjoa, Aggleton, & Grasby, 2000), ou encore les phénomènes d'attention spatiale cross-modale (Spence & Driver, 1996 ; Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999 ; Macaluso, Frith, & Driver, 2000 ; Macaluso & Driver, 2001 ; Shafritz, Gore, & Marois, 2002 ; Macaluso, Driver, & Frith, 2003). Il serait fastidieux, et non forcément judicieux, de

tenter de dresser un état des lieux exhaustif de toutes ces thématiques. Par souci de clarté, nous suivrons donc l'exemple de Calvert (2001) et nous contenterons d'abord la question des processus associatifs sous deux angles distincts: l'étude des processus associant des stimulations non-linguistiques, et l'étude des processus associant des stimulations linguistiques.

Avant de passer en revue les différentes études s'étant intéressées aux processus associatifs entre informations de nature distincte, il nous faut aborder succinctement les aspects comportementaux de ces interactions.

Effets comportementaux : facilitation vs. interférence crossmodale

Notons tout d'abord que les effets comportementaux relatés ci-dessous ont été dans la majorité des cas observés lors de processus d'intégration crossmodale visuo-auditive.

Deux effets comportementaux différents peuvent être observés lors d'interactions entre stimulations visuelles et auditives. Le premier est un effet de facilitation crossmodale, à savoir une diminution des temps de réaction lorsque les

participants sont confrontés à des stimulations bimodales par rapport à des stimulations unimodales (Frens, Van & Van-der, 1995 ; Calvert et al., 2001 ; Teder-Sälejärvi, McDonald, Di Russo & Hillyard, 2002). Il est important de signaler que dans ces études, on considère qu'il y a facilitation si et seulement si les temps de réaction en situation bimodale sont significativement plus courts que ceux des deux situations unimodales, et non pas simplement qu'au moins une des deux situations unimodales.

Calvert et al. (2001) soulignent l'importance de deux facteurs dans ces effets de facilitation crossmodale, à savoir la proximité spatiale et la proximité temporelle des stimulations visuelles et auditives. A l'inverse, une désynchronisation temporelle et/ou une disparité spatiale peuvent entraîner une diminution des performances en situation bimodale (Sekuler, Sekuler & Lau, 1997), menant à un effet d'interférence crossmodale.

Ce deuxième effet comportemental, l'interférence d'une modalité sur une autre, peut non seulement mener à une détérioration des performances, mais peut également prendre la forme d'une perception illusoire. Ainsi, la vision peut interférer sur l'audition et créer une perception auditive

erronée. Les deux exemples les plus célèbres sont l'effet McGurk, qui sera détaillé plus loin dans ce chapitre (McGurk & McDonald, 1976) et l'effet « du ventriloque » que nous expérimentons tous chaque fois que nous regardons un film à la télévision, puisqu'il s'agit de l'illusion que les voix émanent de la bouche des acteurs perçus sur l'écran alors que les sources auditives et visuelles sont spatialement éloignées (Howard & Templeton, 1966, in Shimojo & Shams, 2001). Ces deux illusions s'appuient sur l'ambiguïté (en termes de congruence ou de localisation spatiale) entre les stimulations visuelles et auditives.

L'audition peut également altérer la vision, et ceci même en dehors de toute situation ambiguë. Ainsi, Shams, Kamitani et Shimojo (2002) ont découvert l'illusion du « double flash ». Dans cette illusion, la présentation d'un seul flash lumineux accompagné de bips sonores multiples entraîne la perception de flashes lumineux multiples (par exemple, deux bips sonores entraînent la perception de deux flashes lumineux). Cette illusion amènera les auteurs à postuler que les interactions entre stimulations visuelles et auditives dépendent, au moins en partie de la structure des stimuli, le stimulus le plus saillant

(dans ce cas-ci le plus discontinu, à savoir les bips sonores) influençant la perception du stimulus dans l'autre modalité.

Quoi qu'il en soit, ces différents effets crossmodaux nous prouvent que les stimulations en provenance de modalités sensorielles différentes ne sont pas traitées de manière isolée et indépendante, mais interagissent entre elles, soit en permettant une amélioration des performances, soit en créant une illusion perceptive, dans le but de maintenir un comportement adapté face aux stimulations de l'environnement.

Etudes impliquant des stimulations non-linguistiques

Plusieurs études électrophysiologiques ont été consacrées aux processus d'intégration visuo-auditive. Une des études pionnières fut celle de Giard et Peronnet (1999). Les potentiels évoqués ont été enregistrés chez leurs sujets durant une tâche de reconnaissance d'objets préalablement appris soit sur base de leurs caractéristiques auditives (ton A ou B), soit sur base de leurs caractéristiques visuelles (déformation d'un cercle en ellipse verticale ou horizontale), soit sur base de la combinaison de ces deux caractéristiques

(objet A = ton A + déformation horizontale, objet B = ton B + déformation verticale). Cette étude est une des premières à avoir utilisé la méthode soustractive, c'est-à-dire à soustraire de la condition bimodale les deux conditions unimodales, afin d'isoler les activités cérébrales spécifiques aux processus d'intégration visuo-auditive. Les résultats en potentiels ont révélé (1) une activité apparaissant dans les aires visuelles dès 40 millisecondes après la présentation des stimuli, (2) une augmentation de l'amplitude de la N1 dans le cortex auditif, et (3) une activité fronto-temporale apparaissant entre 140 et 165 millisecondes. De plus, les auteurs ont séparé leurs participants en deux groupes selon leur modalité sensorielle dominante (auditive ou visuelle). Si l'activité fronto-temporale restait présente dans les deux groupes, l'activité enregistrée dans les régions unimodales par contre différait significativement entre les groupes. Ainsi, dans le groupe à dominance visuelle, l'ajout d'une information visuelle à l'information auditive n'avait pas d'effet sur le cortex visuel. Par contre, l'ajout d'une information auditive à l'information visuelle augmentait l'activité du cortex auditif. A l'inverse, le groupe à dominance auditive montrait le pattern d'activités opposé. Les auteurs en concluent que l'intégration multi-

sensorielle a lieu très tôt dans la chaîne de traitement de l'information, et qu'elle implique aussi bien des régions sensorielles spécifiques à chaque modalité que des régions hétéromodales d'intégration, comme le cortex fronto-temporal.

L'idée que les régions sensorielles puissent contribuer au processus d'intégration a été renforcée par Foxe, Morocz, Murray, Higgins, Javitt et Schroeder (2000). Utilisant les potentiels évoqués, ces auteurs ont examiné les corrélats cérébraux des interactions audio-somesthésiques. Leurs participants étaient confrontés soit à des tonalités de 1000 Hz, soit à une impulsion électrique, soit à la combinaison des deux stimulations. L'analyse des potentiels évoqués et des topographies a révélé des interactions prenant la forme d'une positivité apparaissant dès 50 millisecondes dans le cortex somato-sensoriel et vers 70 millisecondes dans le cortex auditif.

La même équipe de recherche a récemment examiné les corrélats électrophysiologiques de l'intégration visuo-auditive (Molholm, Ritter, Murray, Javitt, Schroeder & Foxe, 2002). Utilisant un paradigme expérimental proche de celui de Giard et Peronnet (1999), ils ont montré que les premières

interactions entre audition et vision prenaient place dans le cortex pariéto-occipital droit dès 46 millisecondes après l'apparition des stimulations. Ils ont également observé des interactions au niveau occipito-temporal vers 165 millisecondes, et centrales vers 180 millisecondes. Leurs résultats sont assez similaires à ceux de Giard et Peronnet (1999). La rapidité des premières interactions visuo-auditives amènera les auteurs à postuler (1) soit une liaison directe entre le cortex auditif et le cortex visuel, (2) soit une liaison indirecte entre les deux régions unisensorielles, passant par des régions hétéromodales d'intégration, comme le STS.

Teder-Sälejärvi et al. (2002) ont également examiné les processus d'intégration visuo-auditifs au moyen des potentiels évoqués, nuancant quant à eux les observations des études précédentes. Les participants de leur étude étaient confrontés à une stimulation soit auditive (bip, condition A), soit visuelle (flash, condition V), soit visuo-auditive (combinaison des deux, condition AV), dans une tâche « oddball » consistant en la détection de stimulations rares parmi une série de stimulations fréquentes. Ces stimuli rares étaient soit un bip plus intense, soit un flash plus lumineux, soit une combinaison des deux au sein d'un train de stimuli

fréquents. A la suite de Giard et Peronnet (1999), ils appliqueront le principe soustractif à leur paradigme, en soustrayant les potentiels évoqués produits par les conditions unimodales visuelle et auditive de ceux produits par la condition bimodale ($AV - (A + V)$). Tout comme celles de Giard et Peronnet (1999) et de Molholm et al. (2002), cette étude a révélé un effet de facilitation crossmodale, à savoir que les participants détectaient les stimulations bimodales déviantes plus rapidement et plus correctement que les stimulations déviantes unimodales, qu'elles soient visuelles ou auditives. De plus, la soustraction a montré plusieurs interactions visuo-auditives, la plus précoce apparaissant sous la forme d'une négativité frontale dès 20 millisecondes, accompagnée d'une positivité postérieure dès 30 millisecondes. Ils ont également observé des interactions plus tardives, prenant la forme (1) d'un complexe fronto-central positif-négatif-positif entre 100 et 160 millisecondes, suivi d'une déflexion négative vers 250 millisecondes, (2) d'un complexe postérieur négatif-positif entre 130 et 170 millisecondes, suivi d'une négativité entre 250 et 270 millisecondes. Toutefois, les auteurs montreront que les activités précoces sont le reflet de processus anticipatifs antérieurs à l'apparition des stimuli et

disparaissant avec l'utilisation d'un filtre passe-haut. Ils concluent donc que si les interactions visuo-auditives ne correspondent pas à la simple somme des activités visuelles et auditives, celles-ci ne semblent pas démarrer avant 100 millisecondes et impliquent principalement les régions unisensorielles. Une analyse de localisation de source a d'ailleurs révélé une implication du gyrus occipital visuel (aux environs de 160 millisecondes) et du cortex temporal antérieur, dans sa région péri-sylvienne (aux environs de 250 millisecondes), cette dernière région pouvant toutefois correspondre aussi bien au cortex auditif associatif qu'à une région hétéromodale comme le STS (dont nous avons noté plus haut l'étendue et l'hétérogénéité).

Néanmoins, dans toutes ces études, il est difficile de déterminer si les interactions observées sont le fruit de connections directes entre populations neuronales unisensorielles, ou si elles sont réalisées par l'intermédiaire de régions intégratrices profondes dont l'activité ne peut être enregistrée par des électrodes de surface. Dans une étude par PET-scan, Bushara et al. (2001) ont d'ailleurs montré, dans une tâche de détection synchronisée de stimuli visuels et auditifs, l'activation de l'insula droite, du gyrus pariétal et de régions

préfrontales. En outre, ils ont observé que l'augmentation de l'asynchronie entre les deux stimulations, produisant une augmentation des exigences attentionnelles, était corrélée à une augmentation de l'activité dans l'insula droite, le thalamus postérieur et le colliculus supérieur. Ces résultats suggèrent que les processus de traitement synchrones de stimulations en provenance de modalités sensorielles distinctes pourraient être réalisés via l'activation de régions sous-corticales. Cette implication de l'insula et du colliculus supérieur dans la détection de correspondances temporelles entre événements a été répliquée par Calvert et al. (2001).

Citons enfin l'étude par IRMf événementielle de Bushara, Hanakawa, Immisch, Toma, Kansaku et Hallett (2003), qui ont examiné les réponses cérébrales liées au « binding » visuo-auditif. Ils se basent sur l'observation que (1) lorsque deux objets visuels se croisent et poursuivent leur route, l'observateur peut avoir l'illusion qu'ils se heurtent et repartent en sens inverse, (2) l'ajout d'un son au moment du croisement biaise fortement la perception en faveur d'une collision. Leurs stimuli consistaient donc en barres verticales mouvantes se rapprochant l'une de l'autre, se croisant, et s'éloignant ensuite. Au moment du croisement, un son était

produit. Pour chaque essai, les participants devaient dire s'ils avaient perçu les deux barres comme se croisant et continuant leur route (essais sans intégration) ou comme se heurtant et repartant en sens inverse (essais avec intégration visuo-auditive). Leurs résultats ont révélé que les essais classés comme « collision », par rapport aux essais classés comme « croisement », étaient associés à une augmentation de l'activité de régions hétéromodales incluant l'insula, les cortex préfrontaux et pariétaux postérieurs, le thalamus et le colliculus supérieur, et à une diminution des régions unimodales visuelles et auditives. Les auteurs interprètent ce pattern d'activations comme étant une preuve d'interactions réciproques et compétitives entre régions hétéro- et unimodales, interactions dans lesquelles les régions hétéromodales intégratrices ne sont pas simplement la dernière étape d'une voie hiérarchique menant des régions unimodales vers les régions d'intégration, mais sont impliquées dans les processus perceptifs précoces et peuvent moduler l'activité des régions unimodales.

Etudes impliquant des stimulations linguistiques

Nous aborderons dans cette section différentes études qui se sont intéressées aux processus de synthèse audio-visuelle du discours. Les études pionnières dans ce domaine ont montré une amélioration des capacités de compréhension auditive lorsque les lèvres du locuteur et les mouvements de bouche sont visibles (Reisberg, 1987), ainsi qu'une meilleure localisation de la personne qui parle lorsque la voix et les mouvements de bouche peuvent être associés (Driver, 1996). Suite à ces découvertes, plusieurs études ont montré que la lecture labiale produisait une activation du cortex auditif, et que lorsque le locuteur pouvait être vu en train de parler, une plus forte activation des cortex visuels et auditifs était enregistrée que lorsque les mouvements de lèvres ou la voix était perçus isolément (Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen, & David, 1997 ; Calvert, Brammer, Bullmore, Campbell, Iversen, & David, 1999). Dans une étude ultérieure en IRMf, Calvert et al. (2000) ont tenté de cerner les régions cérébrales activées lors de la détection d'associations congruentes entre voix et mouvements de lèvres, ou inhibées lors d'associations incongruentes. Ils ont

comparé les activités cérébrales dans cinq conditions différentes : repos, discours entendu isolément (A), mouvements de lèvres seuls (V), associations congruentes entre discours et mouvements de lèvres (AV congruent), et associations incongruentes entre discours et mouvements de lèvres (AV incongruent). Leur but était de cerner les régions impliquées dans l'intégration visuo-auditive du discours selon (1) qu'elles sont activées par chaque modalité isolément, (2) qu'elles montrent une activation pour les associations congruentes ($AV \text{ congruent} > A + V$), ou (3) qu'elles montrent une inhibition pour les associations incongruentes ($AV \text{ incongruent} < A + V$). Ils ont observé que les régions visuelles et auditives montraient une activation pour les associations congruentes, mais que seule la partie ventrale postérieure du STS gauche montrait à la fois une activation pour les associations congruentes et une inhibition pour les associations incongruentes, ce qui les amènera à postuler que le STS joue un rôle fondamental dans le binding des information linguistiques visuelles et auditives.

Olson, Gatenby et Gore (2002) contesteront ces découvertes. Dans leur étude en IRMf, ils ont utilisé l'effet McGurk pour examiner les régions intégratrices impliquées

dans la perception du langage. Dans ce phénomène (McGurk & McDonald, 1976), la présentation simultanée d'une syllabe auditive (par exemple « bat ») et d'un mouvement de lèvres incongruent (par exemple « vet ») produit la perception subjective d'une autre syllabe (par exemple « vat »). Deux conditions d'intérêt étaient présentées aux participants, dans lesquelles les stimulations visuelles et auditives étaient temporellement synchronisées ou désynchronisées. La synchronisation temporelle étant un facteur déterminant de l'intégration sensorielle (Meredith, Nemitz, & Stein, 1987), seule la condition synchronisée devrait déclencher un processus d'intégration visuo-auditif, se marquant comportementalement par un effet McGurk. C'est ce que les données comportementales ont confirmé. De plus, les résultats d'imagerie ont montré que les deux conditions étaient associées à une activation des aires liées à la perception et la production du langage, et incluaient le gyrus frontal inférieur gauche, l'aire de Broca gauche et une large activation du gyrus temporal supérieur bilatéral s'étendant jusqu'au STS. Par contre, seule la condition durant laquelle les associations étaient synchronisées était associée une activation du claustrum gauche. Pour les auteurs, il semble

que le STS ne soit pas responsable des processus d'intégration cross-modale *per se*, bien qu'il puisse être impliqué dans le traitement des informations visuelles et auditives. A la suite de Bushara et al. (2001), ils insistent sur l'importance des structures sous-corticales dans ces processus intégratifs, et en particulier du claustrum, pouvant servir de relais entre les régions corticales traitant chaque modalité et permettant à chacune de ces régions de communiquer avec les autres.

Sekiyama, Kanno, Miura et Sugita (2003) ont également utilisé l'effet McGurk pour examiner les processus de perception visuo-auditive du discours. Ils ont contrasté deux conditions d'associations, dont la composante auditive était hautement ou faiblement intelligible. Ce contraste s'explique par le fait que leurs participants étaient japonais et qu'il a été prouvé que ceux-ci sont moins sensibles à l'effet McGurk (Sekiyama & Tohkura, 1993), les japonais utilisant peu les indices visuels pour comprendre le langage. Aucun effet n'était donc prédit dans la condition hautement intelligible, alors qu'un effet McGurk était attendu dans la condition faiblement intelligible où les participants devaient s'appuyer sur la lecture labiale pour comprendre les syllabes entendues. Les résultats de leur étude ont montré une activation des aires

auditives dans les deux conditions, avec une extension de cette activation vers des aires visuelles lorsque le discours était difficile à comprendre. En outre, la comparaison directe des deux conditions a permis d'observer une activation spécifique de la partie postérieure du STS gauche dans la condition faiblement intelligible, qui comme prédit était associée à un fort effet McGurk.

Comme nous pouvons le constater, différentes études, utilisant des techniques et des méthodologies différentes, ont mis en lumière une participation du STS gauche dans le binding visuo-auditif des informations langagières. D'autres auteurs, par contre, ne sont pas parvenus à observer cette activation, mais ont plutôt mis en évidence l'importance des régions sous-corticales dans ces processus intégratifs. A notre connaissance, aucune étude jusqu'à présent n'est parvenue à faire la part des choses ni à délimiter le rôle fonctionnel de chaque structure corticale dans l'intégration de visages et de voix.

Il semble également qu'outre la proximité temporelle et/ou spatiale, les facteurs sémantiques influencent également la perception multisensorielle. Pourtois et de Gelder (2002) ont utilisé la technique de stimulation

magnétique trans-crânienne (TMS) pour comparer les processus d'intégration perceptive entre stimulations visuo-auditives porteuses ou non de sens (visages-voix ou forme géométrique-son). Les participants devaient catégoriser l'émotion de visages (joie ou peur), de voix, ou de leur combinaison congruente, ou catégoriser l'orientation (verticale ou horizontale) d'ellipses, la fréquence (500 ou 520 Hz) de sons, ou la combinaison des deux (ellipse verticale et son de 500 Hz ou ellipse horizontale et son de 520 Hz). Ils ont observé, via l'analyse des temps de réaction, que l'application d'une stimulation perturbante sur le cortex pariétal postérieur à 200 ms empêchait l'intégration des associations de type « forme-son », mais pas l'intégration des associations « visage-voix ». Ils en conclueront que la contigüité spatio-temporelle des informations auditives et visuelles n'est pas le seul déterminant des processus de binding crossmodaux, mais que le contenu sémantique des associations joue également un rôle déterminant dans ces processus. Ils émettront également l'hypothèse qu'il n'existe sans doute pas un seul site neuronal responsable de l'intégration visuo-auditive mais bien différents sites, interagissant et prenant en compte les différents aspects de l'intégration.

Les associations visages-noms et visages-voix

Toutes les recherches passées en revue dans la section précédente étaient dédiées à l'étude des processus intégratifs liant vision et audition d'un point de vue perceptif (tâches de détection d'informations présentées dans les deux modalités, perception passive de stimulations visuo-auditives, etc.). Néanmoins, la question centrale de cette thèse concerne les processus associatifs impliqués dans les processus d'identification des personnes. Peu d'études se sont penchées à ce jour sur les processus intégratifs autres que perceptifs, tels que la récupération d'associations ou l'identification crossmodale de personnes.

Sperling, Bates, Cocchiarella, Schacter, Rosen et Albert (2001) ont étudié par IRMf les processus d'encodage d'associations entre visages et noms propres (prénoms). Comparant l'encodage de nouvelles paires de visages et de prénoms à la vision répétée des paires précédemment apprises, ils ont montré que l'activation cérébrale lors de la tâche d'encodage était principalement localisée dans les parties antérieures de l'hippocampe, le thalamus, le gyrus fusiforme et les cortex préfrontaux dorsolatéraux. La même

équipe de recherches observera, dans une étude ultérieure comparant les activités cérébrales selon que l'encodage est fructueux ou non (c'est-à-dire lorsque les sujets se souviennent ultérieurement des paires encodées ou les oublient), que l'activation de l'hippocampe et des gyri préfrontaux était significativement plus importante lorsque l'encodage d'associations visages-noms était fructueux que lorsque les associations étaient oubliées (Sperling, Chua, Cocchiarella, Rand-Giovannetti, Poldrack, Schacter, & Albert, 2003). Cette implication de l'hippocampe, et plus généralement du lobe temporal médian dans l'encodage fructueux de ce type d'associations a été confirmée par différentes autres recherches (Kirwan & Stark, 2004 ; Jackson & Schacter, 2004).

Herholz, Ehlen, Kessler, Strotmann, Kalbe et Markowitsch (2001) ont également examiné par PET-scan les processus d'encodage d'associations entre visages et prénoms. Ils ont comparé cet encodage à une tâche de catégorisation selon le genre. Ils ont montré que l'encodage de paires de visages et de prénoms était associé à une activité du cortex visuel bilatéral, incluant les gyri fusiformes et linguaux. Une activation supplémentaire du cortex temporal inférieur et du gyrus orbitofrontal gauches fut également observée. Pour ces

auteurs, ces données suggèrent le rôle prépondérant du cortex visuel associatif et des aires préfrontales dans les processus mnésiques. Toutefois, ils n'ont observé aucune activation des aires temporales médianes ni de l'hippocampe, ne corroborant pas les données observées dans les études citées précédemment, et ne donnent aucune explication quant à ces différences de patterns d'activations dans des tâches pourtant fortement similaires.

Guo, Voss et Paller (2005) ont quant à eux utilisé les potentiels évoqués pour aborder la question de l'encodage en mémoire d'associations entre visages et noms propres. Trois types d'encodage ont été examinés : l'encodage de visages, de voix, et d'associations entre les deux types de stimuli. Ils ont également utilisé une méthodologie permettant de comparer les activités cérébrales produites par les stimuli reconnus ultérieurement de celles produites par les stimuli oubliés. Les processus d'encodage fructueux étaient associés à différents patterns d'activités bilatérales entre 200 et 800 millisecondes : une négativité fronto-centrale pour les noms, une positivité centro-pariétale pour les visages, et une positivité fronto-centrale pour les associations, suggérant que différents mécanismes neuraux sont à l'œuvre lors de

l'encodage des informations isolées ou associées, même si ces résultats ne permettent pas de se prononcer sur les différents mécanismes mnésiques en jeu.

Ces études ont porté sur les processus d'encodage des associations visages-noms. D'autres, peu nombreuses, ont tenté de cerner les régions cérébrales responsables de la reconnaissance de telles associations. Il nous faut tout d'abord citer l'étude de Shah, Marshall, Zafiris, Schwab, Zilles, Markowitsch et Fink (2001), qui ont comparé la perception de visages et de voix familiers à la perception de visages et de voix non-familiers, afin d'examiner les corrélats neuraux de la familiarité aux personnes. Leur résultat principal fut d'observer une activation du gyrus cingulaire postérieur et du cortex rétrospinal en réponse aux stimuli familiers (visages et voix) par rapport aux stimuli non-familiers. Ils ont également mis en lumière l'activation des régions auditives classiques en réponse à la perception des voix (familières ou non) et des régions visuelles en réponse à la perception des visages (familiers ou non). Ils interprètent leurs résultats comme étant la preuve que reconnaître une personne implique l'activation de régions spécifiques à chaque modalité, ainsi que de régions « multimodales » comme le cortex rétrospinal.

Cependant, les auteurs n'ont pas la possibilité de déterminer si cette activation du cortex rétrospinal, classiquement impliquée dans les processus de mémoire épisodique et de saillance émotionnelle, serait impliquée dans les processus d'identification des personnes ou dans l'évaluation de la familiarité des stimuli. De plus, ils n'ont pas directement examiné les processus d'intégration crossmodaux des visages et des voix, chaque type de stimulation étant présenté isolément.

Tsukiura, Namiki, Fujii et Iijima (2003) ont comparé la reconnaissance d'associations entre visages et noms propres préalablement apprises en fonction du délai entre apprentissage et reconnaissance (délai de 5 minutes ou de 2 semaines), et en présentant des visages avec ou sans expression émotionnelle (joie ou neutralité). Les participants étaient confrontés en IRMf avec les visages précédemment appris et devaient choisir le nom leur correspondant parmi deux possibilités. Les activations liées à la reconnaissance des associations étaient localisées dans les lobes temporaux antérieurs bilatéraux pour la tâche de reconnaissance immédiate, et uniquement dans le lobe temporal gauche lors de la reconnaissance différée. La reconnaissance des visages

présentant une émotion, comparée à la reconnaissance des visages neutres, était quant à elle associée à une activation des aires péri-amygdaliennes et du pôle temporal gauche. Cependant, la présence d'une expression émotionnelle n'était pas corrélée avec les performances de reconnaissance des associations. Ces données sont intéressantes en ce sens qu'elles montrent l'importance du lobe temporal gauche dans la récupération lexicale des noms propres à partir des visages, et qu'elles suggèrent que l'implication du lobe temporal droit dans ces processus pourrait décroître avec le temps.

Dans une étude ultérieure, Tsukiura, Mochizuki-Kawai et Fujii (2005) manipuleront les informations sémantiques liées aux personnes. Dans cette étude en IRMf, 2 types d'associations étaient apprises par les participants : soit des associations visages-noms propres contenant également une profession, soit des associations sans autres informations sémantiques. De plus, la moitié de ces associations étaient apprises deux fois, tandis que l'autre moitié était apprise sept fois. Durant l'acquisition des images en IRMf, les sujets devaient reconnaître les noms à partir de la présentation des visages, ou reconnaître les visages à partir de la présentation des noms. Comparant les associations contenant une

information sémantique à celles n'en contenant pas, ils ont remarqué que le lobe temporal antérieur gauche était plus activé par la reconnaissance des noms à partir des visages alors que le lobe temporal antérieur droit était plus activé pour la reconnaissance des visages à partir des noms. De plus, ils ont noté que l'activation du lobe temporal antérieur gauche était significativement réduite pour les associations apprises 7 fois par rapport à celles apprises seulement 2 fois. Il semble donc que les lobes temporaux antérieurs des deux hémisphères soient impliqués dans les différents processus de traitement des informations contenues dans les associations entre visages, noms propres et informations sémantiques, et que l'activité cérébrale liée à ces processus puisse être modulée en fonction du degré de consolidation mnésique de ces associations.

Von Kriegstein, Kleinschmidt, Sterzer et Giraud (2005) ont quant à eux examiné les interactions entre régions auditives et visuelles lors de la reconnaissance de voix. Les participants de leur expérience en IRMf devaient reconnaître les voix de personnes familières (personnes de leur entourage) ou peu familières (personnes rencontrées juste avant l'expérience). La soustraction des activités cérébrales

liées à la reconnaissance de voix non-familiales des activités liées à la reconnaissance des voix familières a révélé une activation d'un réseau d'aires cérébrales incluant les régions pariéto-temporo-occipitales bilatérales, les régions antérieures des gyri temporaux inférieurs, et le gyrus fusiforme bilatéral, avec une prédominance de l'hémisphère droit. Le fait que la reconnaissance de voix familières puisse activer le gyrus fusiforme correspondant à la FFA, région dédiée au traitement des visages, sera interprété par les auteurs comme étant une preuve d'interactions crossmodales entre régions visuelles et auditives, et ils suggèrent même que ces interactions ne nécessitent pas l'implication de régions supramodales d'intégration, mais sont soutenues par un partage direct des informations entre les régions traitant les visages et les voix humaines. Toutefois, les auteurs n'excluent pas que cette activation de la FFA puisse refléter des processus d'imagerie mentale implicite, tels que l'ont montré Ishai, Haxby et Ungerleider (2002). Il est à noter que cette interprétation n'est qu'indirecte, puisqu'à aucun moment ils n'ont directement comparé les activités cérébrales liées à la reconnaissance de voix, de visages et d'associations entre visages et voix.

Conclusions

Les différentes études exposées ci-dessus ont toutes comme point commun d'examiner les corrélats cérébraux des processus intégratifs entre stimuli de nature différente. Elles ont mis en lumière l'existence d'un réseau étendu d'aires cérébrales sous-tendant ces processus intégratifs, incluant différentes régions corticales et sous-corticales. Cependant, il nous paraît difficile de tirer des conclusions claires et concises de ces découvertes, au vu de l'hétérogénéité des techniques employées et des paradigmes expérimentaux utilisés. En effet, les techniques d'examen sont diverses (IRMf, potentiels évoqués, PET-scan, TMS) ; la gamme de stimuli utilisés rend toute comparaison hasardeuse (associations arbitraires et non-porteuses de sens entre stimuli visuels et auditifs simples vs. associations entre visages et voix de personnes familières) ; toutes les études n'examinent pas les mêmes processus (perception passive d'associations visuo-auditives, encodage d'associations visages-noms, reconnaissance de telles associations) ; les méthodes d'identification des régions crossmodales diffèrent également d'une étude à l'autre, et vont de l'analyse de conjonctions entre diverses conditions

unimodales à l'analyse soustractive entre conditions unimodales et bimodales.

Malgré ces différences méthodologiques et les divergences théoriques opposant les auteurs défendant l'idée que l'intégration s'opérerait via l'activation commune d'aires unimodales – via ou non un relais sous-cortical – et ceux défendant l'hypothèse de l'existence de régions corticales hétéromodales d'intégration, certaines régions semblent néanmoins être régulièrement activées dans les tâches intégratrices. C'est le cas du colliculus supérieur, de l'insula et du claustrum, régions sous-corticales considérées soit comme des régions intégratrices, soit comme des relais d'informations entre régions unisensorielles, mais aussi de régions corticales autres que les régions unisensorielles, telles que le STS (semblant fortement impliqué dans les processus intégratifs visuo-auditifs et identitaires) et les gyri préfrontaux.

Conclusions de la partie théorique et introduction à la partie expérimentale.

Au terme de cette introduction théorique, nous ne pouvons que constater l'écart qui existe entre le nombre d'études ayant examiné les mécanismes cérébraux du traitement des différents types d'informations nous permettant d'identifier les personnes familières (visages, noms et voix), et le nombre d'études ayant tenté de cerner les mécanismes cérébraux permettant de créer des liens entre ces différentes informations.

En effet, d'une part, les premiers chapitres nous ont permis de nous rendre compte de l'abondance de la littérature consacrée aux mécanismes cérébraux sous-tendant le traitement des visages, des noms propres et des voix. La plupart de ces études, qu'elles abordent ces processus sous l'angle neuropsychologique ou via l'utilisation de techniques d'imagerie cérébrale fonctionnelle ou électrophysiologiques, convergent vers la mise en évidence de régions cérébrales et/ou de mécanismes électrophysiologiques spécifiquement dédiés au traitement de l'un ou l'autre type d'information.

Ainsi, il semble établi que le traitement des visages est assuré par des régions particulières du cortex occipito-temporal, et plus particulièrement par le gyrus fusiforme de l'hémisphère droit. Le traitement des noms propres par contre semble être sous-tendu par les régions temporales antérieures de l'hémisphère gauche, et le traitement des voix par la partie antérieure de la scissure temporale supérieure droite.

D'autre part, bien qu'il nous faille garder à l'esprit la relative jeunesse des méthodes d'imagerie cérébrale fonctionnelle, force est de constater que comparativement, peu d'études se sont à ce jour penchées sur la manière dont le cerveau opère pour lier ces différentes représentations des personnes familières afin de créer une représentation unifiée des individus sur base de leur visage, de leur nom et de leur voix. De plus, nous avons vu que la question des processus associatifs entre les différentes informations relatives aux personnes souffre d'une grande hétérogénéité, tant au niveau des méthodes expérimentales employées qu'au niveau des méthodes d'analyses utilisées pour identifier les mécanismes intégratifs. Ce dernier point est particulièrement marquant dans le cadre des études ayant examiné les processus associatifs entre visages et noms propres. A notre

connaissance, aucune étude n'a utilisé la méthode soustractive pour isoler les régions cérébrales responsables de l'intégration en une représentation cohérente des visages et des noms, alors que cette méthode est une des méthodes d'analyse classique pour l'investigation des processus intégratifs, notamment crossmodaux (Giard & Peronnet, 1999 ; Calvert et al., 2000 ; Teder-Sälejärvi et al., 2002 ; Molholm et al., 2002). De manière générale, la méthode soustractive consiste à soustraire les conditions « unimodales » (ou ne contenant qu'un seul type d'information, généralement deux conditions) d'une condition multimodale dans laquelle sont simultanément présentés les deux types d'informations. Le postulat de base de cette méthode est que la création de représentations multimodales est plus que la simple somme des activités cérébrales liées au traitement de chaque information (Calvert, 2001 ; Teder-Sälejärvi et al., 2002). La soustraction des mécanismes cérébraux dédiés au traitement de chaque information a donc pour but d'isoler les activités cérébrales spécifiquement liées aux processus intégratifs entre ces informations. Cette méthode permet de pallier les limitations de l'analyse de conjonction, qui consiste « simplement » à superposer les activités liées aux traitements

unimodaux, avec l'hypothèse que les zones d'intersection dans les activités enregistrées correspondent aux activités d'intégration. Néanmoins, il nous faut souligner que la méthode soustractive connaît certaines limitations, notamment du fait qu'elle consiste à enregistrer les activités cérébrales de vastes populations neuronales. La réponse enregistrée reste donc toujours une moyenne de réponses d'un vaste groupe de cellules nerveuses.

A notre connaissance, la seule étude ayant utilisé cette méthode soustractive dans une tâche de reconnaissance d'associations entre visages et noms propres est celle de Campanella, Joassin, Rossion, De Volder, Bruyer et Crommelinck (2001). Cette étude, qui fut l'objet de mon mémoire de licence (Joassin, 2000), est à la base des différentes recherches effectuées dans le cadre de cette thèse. Nous allons donc la décrire plus amplement avant de détailler les différents objectifs des études exposées dans la partie expérimentale de cette thèse.

Cette expérience, utilisant la technique de PET-scan, avait pour but d'examiner les populations neuronales impliquées dans la récupération des représentations visuelles des visages suite à la présentation d'un nom associé et

inversément. Sept participants neurologiquement sains et droitiers ont pris part à cette étude. Elle comportait une phase d'apprentissage et une phase d'acquisition des images fonctionnelles du cerveau. Durant la phase d'apprentissage, les participants devaient se familiariser avec 24 associations entre un visage et un nom de famille, groupées en 12 couples homme-femme. Durant la phase expérimentale, quatre conditions étaient proposées, différant selon le type d'associations : soit une paire de noms (NN), soit une paire de visages (VV), soit un nom et un visage (NV), soit un visage et un nom (VN). Pour chaque essai, les deux stimuli étaient présentés successivement pendant 700 millisecondes et séparés par un intervalle de 500 millisecondes. La tâche était identique dans toutes les conditions : il s'agissait de juger, en appuyant sur un bouton, si la paire de stimuli présentée correspondait ou non à un des couples précédemment appris (50 % des paires étaient correctes). Chaque condition était répétée 3 fois, et deux scans au repos, servant de référence, ont également été enregistrés.

Cette méthodologie a permis d'utiliser une méthode d'analyse soustractive en vue d'isoler les régions cérébrales responsables de l'association entre visages et noms propres.

La soustraction $\{ (NV + VN) - (VV + NN) \}$ a révélé l'activation de trois régions cérébrales distinctes localisées dans l'hémisphère gauche : le gyrus frontal inférieur (aire de Brodmann 45), le gyrus frontal médian (aire de Brodmann 6) et le lobule pariétal inférieur (aire de Brodmann 40, Figure 6). L'analyse des temps de réaction et des performances n'a d'autre part pas révélé de différences significatives entre conditions, si ce n'est un effet d'habituation (diminution des latences de réponses correctes) entre la première et la deuxième présentation de chaque condition. Il ne semble donc pas que ces activations ne puissent être expliquées par une différence de complexité entre conditions.

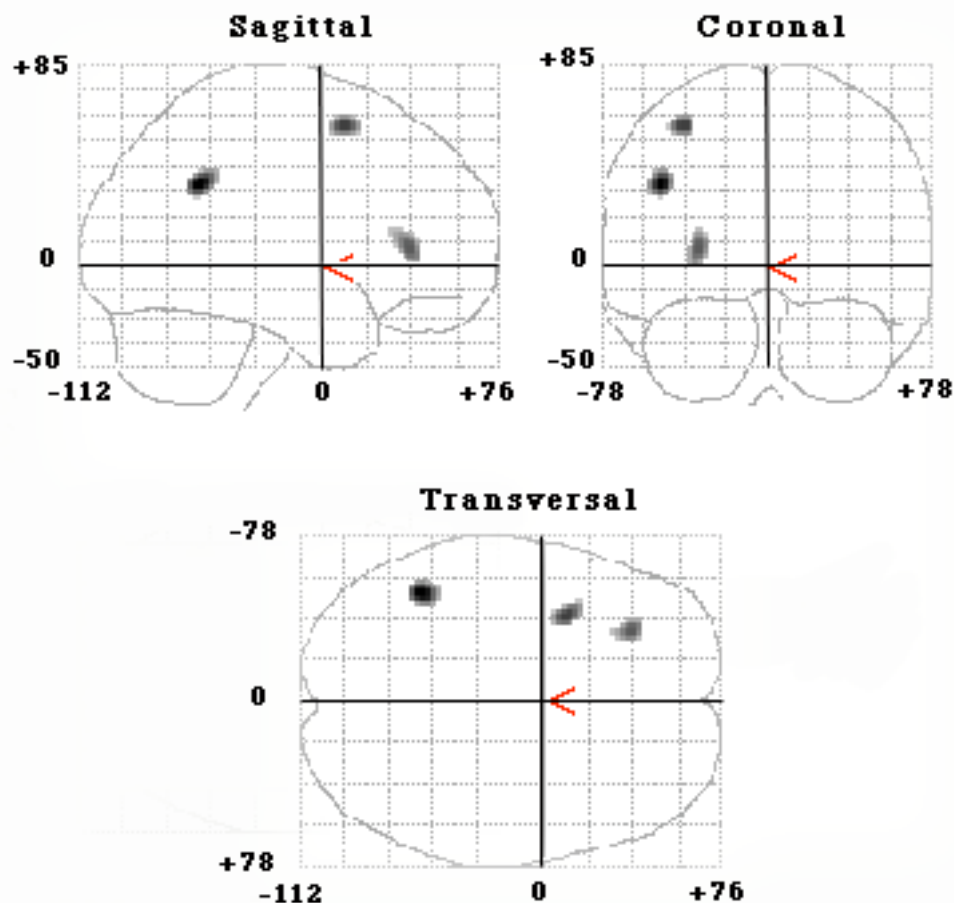


Figure 6 : cartes statistiques paramétriques (SPM{Z}) illustrant les régions cérébrales activées par la soustraction (VN + NV) – (NN + VV)

L'activation du gyrus frontal médian gauche, appelé également « aire motrice supplémentaire », est traditionnellement associé aux processus de mémoire exécutive, et a été observée dans des tâches telles que la Tour de Londres (Morris, Ahmed, Syed, & Toone, 1993) ou le Wisconsin Card Sorting Test (Nagahama, Okada, Katsumi, Hayashi, Yamauchi, Sawamoto et al., 1999), épreuve testant les processus de shift attentionnel. Cette activation du gyrus frontal médian a été interprétée comme reflétant les

processus de shift attentionnel permettant de passer, dans les conditions VN et NV, d'un type de représentation (par exemple un visage) à un autre (un nom de famille). Le gyrus frontal inférieur gauche semble par contre être impliqué dans les processus de mémoire épisodique. Ainsi, Rugg, Fletcher, Chua et Dolan (1999) ont montré son activation dans une tâche de mémoire de source (par exemple, se souvenir si un mot présenté était apparu à gauche ou à droite du point de fixation lors d'une phase préalable d'apprentissage) par rapport à une tâche de reconnaissance (par exemple, reconnaître les mots ayant été encodés préalablement). Cette région est en fait considérée non seulement comme une région permettant la récupération en mémoire de différentes informations, mais surtout comme une région permettant de récupérer les liens sémantiques entre diverses informations (Fletcher, Shallice, Frith, Frackowiak, & Dolan, 1996 ; Nolde, Johnson, & Raye, 1998 ; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Cette région reflète donc, selon les auteurs, les processus de récupération en mémoire à long terme des associations visages-noms apprises précédemment, en vue de les comparer avec les associations présentées durant l'expérience.

Enfin, de nombreuses études utilisant du matériel linguistique ont mis en évidence une activation du lobule pariétal inférieur gauche, et comme cette région reçoit de multiples inputs en provenance de régions sensorielles distinctes, plusieurs auteurs ont proposé l'hypothèse selon laquelle elle serait une région de convergence multimodale permettant le « binding » de différentes représentations sensorielles en vue d'en former une représentation unique (Damasio, 1989 ; Clark, Egan, McFarlane, Morris, Weber, Sonkkilla, Marcina, & Tochon-Tanguy, 2000). Nous pensons donc que l'activation du lobule pariétal inférieur gauche observée dans cette étude reflète l'association de représentations visuelles distinctes de visages et de noms, en vue d'en former une représentation unifiée qui puisse être comparée avec les représentations stockées en mémoire à long terme.

Ces résultats ont ouvert la voie à de nombreuses questions. Trois questions principales relatives à cette première étude seront développées dans la partie expérimentale de cette thèse :

1. l'étude de Campanella et al. (2001) a montré l'implication de trois aires cérébrales distinctes distribuées dans l'hémisphère gauche dans les processus associatifs entre visages et noms propres. Cependant, la technique de tomographie par émission de positrons, si elle offre une précision anatomo-fonctionnelle importante, ne peut permettre d'inférer aucune hypothèse quant au déroulement temporel de ces différentes activations. La première étude proposée dans cette thèse tentera de répondre à cette question, en répliquant l'étude de Campanella et al. (2001) au moyen de la technique des potentiels évoqués. L'utilisation conjointe de différentes techniques d'imagerie cérébrale offrant à la fois une précision anatomique et temporelle, permettra de contraindre temporellement les activations observées dans l'étude pré-citée ;

2. les deux premiers chapitres nous ont montré que le traitement des visages et des noms de personnes est assuré par des régions cérébrales particulières, telles que par exemple le gyrus fusiforme antérieur de l'hémisphère droit et qu'il répond préférentiellement à la présentation de visages qu'à toute autre stimulation visuelle. De plus, même si le statut des noms propres - et particulièrement des noms de

personnes - pose encore question, plusieurs recherches ont montré que les différentes parties des lobes temporaux et frontaux gauches répondaient de manière distincte aux différentes classes de noms (noms communs naturels, artificiels, noms de personnes). Nous sommes donc en droit de nous poser la question de la spécificité des processus liant visages et noms propres au sein même des mécanismes associatifs plus généraux liant un objet et son nom. En d'autres termes, les activités cérébrales mises en lumière par Campanella et al. (2001) et dans l'étude 1 de cette thèse sont-elles spécifiques aux associations visages-noms ou reflètent-elles des mécanismes associatifs plus généraux ? L'étude 2 examinera cette question au moyen des potentiels évoqués et comparera les activités cérébrales liées à la reconnaissance d'associations entre visages et noms propres, ou entre animaux et noms communs ;

3. les études 3 et 4 concernent la question de la possible amodalité des régions cérébrales impliquées dans les processus associatifs permettant la reconnaissance des personnes. En effet, dans l'étude de Campanella et al. (2001) et dans la première étude de cette thèse, toutes les informations étaient présentées dans la modalité visuelle

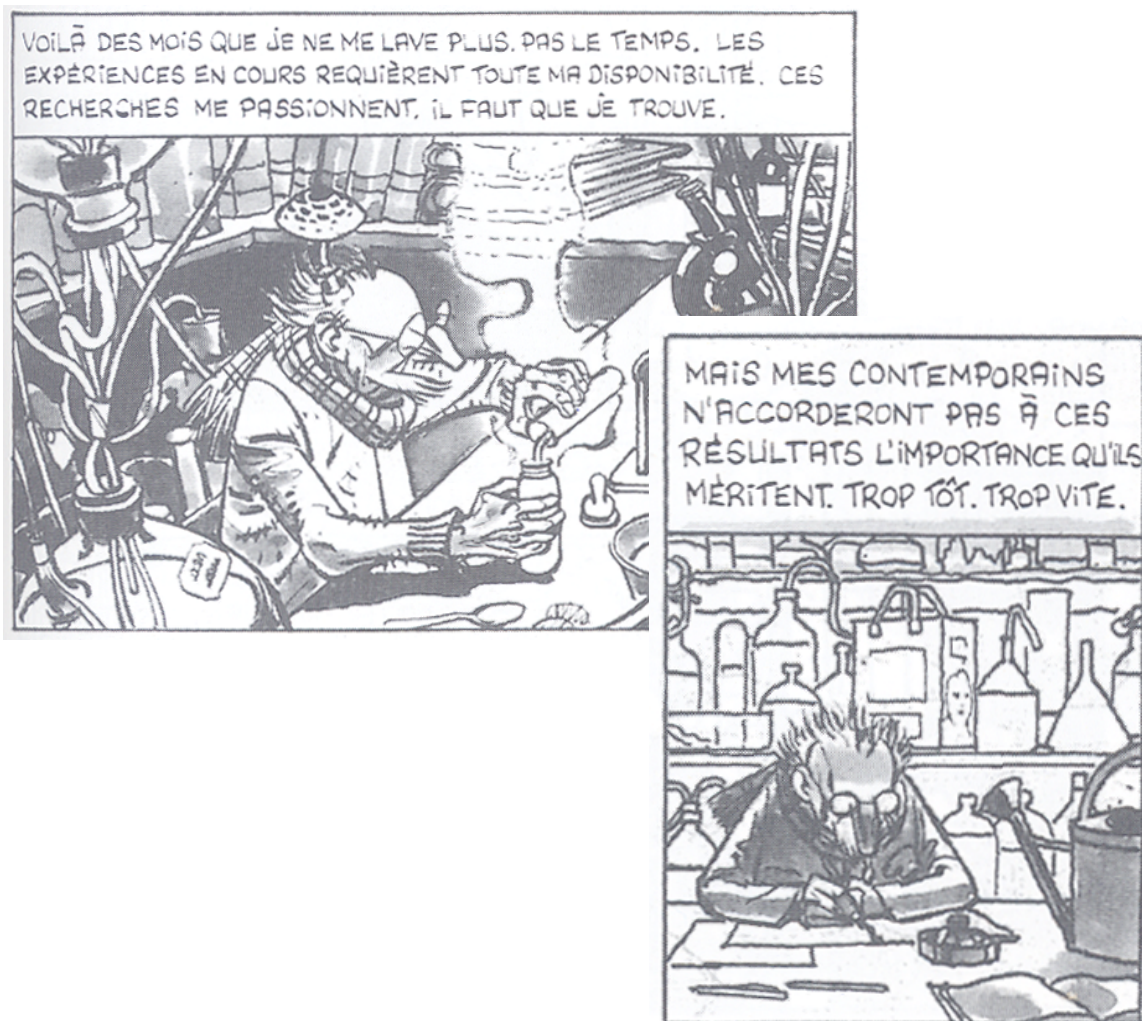
(visages et noms propres écrits). Or, nous avons vu dans le chapitre précédent que les processus d'intégration visuo-auditive requièrent également l'activation de régions hétéromodales particulières, comme le STS. De plus, plusieurs études ont montré une activation du gyrus pariétal inférieur gauche dans les interactions entre informations linguistiques visuelles et auditives (Booth, Burman, Meyer, Gitelman, Parrish, & Mesulam, 2002, 2003). Il était donc intéressant d'étudier les corrélats cérébraux des interactions entre visages et voix dans une tâche de reconnaissance, et de pouvoir comparer directement les activités cérébrales produites par la reconnaissance de visages et de voix présentés isolément ou conjointement. Afin de répondre à la nécessaire complémentarité entre techniques anatomiques et temporelles, l'étude 3 examinera cette question via la technique des potentiels évoqués, l'étude 4 via la fMRI.

En conclusion, les études réalisées dans le cadre de cette thèse, utilisant différentes techniques (imagerie fonctionnelle et électrophysiologie), ont tenté de mieux cerner les processus associatifs permettant de lier ensemble les différents attributs caractérisant chaque personne, et se sont

penchées plus spécifiquement sur trois questions : le déroulement temporel des processus associatifs entre visages et noms propres (étude 1), la spécificité de ces processus au sein des mécanismes associatifs liant un objet et son référent (étude 2), et la possible amodalité de ces processus via l'étude des interactions crossmodales entre visages et voix (études 3 et 4).

2eme partie

approche expérimentale



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

Etude 1

Publiée dans *Psychophysiology*, 41 (2004), 625-635.

**The Electrophysiological Correlates Sustaining
the Retrieval of Face-Name Associations : an
ERP Study.**

F. Joassin¹, S. Campanella^{1,2}, D. Debatisse¹, J.M. Guérit³, R.
Bruyer¹, and M. Crommelinck²

¹ Unité de Neurosciences Cognitives (NESC), ² Laboratoire de
Neurophysiologie (NEFY), ³ Unité de Neurophysiologie Clinique,
Cliniques universitaires St Luc et Unité de Neurochimie (NCHM),
Faculté de Médecine, Université catholique de Louvain, Louvain-la-
Neuve, Brussels, Belgium

Corresponding author :

F. Joassin, Ph. D student
Unité de Neurosciences Cognitives (PSP / NESC)
Place Cardinal Mercier, 10
B 1348 Louvain-la-Neuve
Belgium
Tél. : +32(10) 47.90.62
Fax : +32(10) 47.37.74
e-mail : frederic.joassin@psp.ucl.ac.be

Abstract

An ERP study on nine healthy participants was carried out in order to temporally constrain the neural network proposed by Campanella et al. (2001) in a PET study investigating the cerebral areas involved in the retrieval of face-name associations. Three learning sessions served to familiarize the participants with 24 face-name associations grouped in 12 male/female couples. During EEG recording, participants were confronted with four experimental conditions, requiring the retrieval of previously learned couples on the basis of the presentation of name-name (NN), face-face (FF), name-face (NF) or face-name (FN) pairs of stimuli. The main analysis of this experiment consisted in the subtraction of the non-mixed conditions (NN and FF) from the mixed conditions (NF and FN). It revealed two main ERP components : a negative wave peaking at left parieto-occipital sites around 285 msec and its positive counterpart recorded at left centro-frontal electrodes around 300 msec. Moreover, a dipole modeling using 3 dipoles whose localization corresponded to the 3 cerebral areas observed in the PET study (left inferior frontal gyrus, left medial frontal gyrus, left inferior parietal lobe) explained more than 90% of the variance of the results. The complementarity between anatomical and neurophysiological techniques allowed us to discuss the temporal course of these cerebral activities and to propose an interactive and original anatomo-temporal model of the retrieval of face-name associations.

Keywords : Event-Related Potentials, faces, names, associative processes, dipole localization

Introduction

The ability to recognize familiar people and to recall specific information about them, such as their name, plays a fundamental role in all our social interactions. Because these processes are highly reliable and automatic, one could imagine that they are quite simple (Sergent et al., 1992b). However, several recent studies have shown, for instance, that the ability to recognize a person and to recall his name is sustained by distinct brain areas. As distinct neural regions appear to be engaged in these different recognition processes, how the outputs of these diverse processes are integrated in a single everyday life experience is still matter of debate (Damasio, 1989).

Face processing is a well-documented research area. Studies of brain-damaged patients have shown that prosopagnosia, a specific impairment of face recognition (Bodamer, 1947), is often associated with lesions of occipito-temporal regions, and in particular with lesions of right lingual and fusiform gyri (De Renzi et al., 1994; Takahashi et al., 1995). PET and fMRI activation studies investigating the perception of faces support these results. They generally show

an enhanced activation of bilateral extrastriate regions, including the inferior temporal gyrus (Puce et al., 1995), the lingual and fusiform gyri (Sergent et al., 1992b; Haxby et al., 1994; Puce et al., 1996; Dubois et al., 1999), with a predominance of the right hemisphere (Kanwisher et al., 1997; Kim et al., 1999).

The processes involved in proper name retrieval are less clearly circumscribed. Studies of patients suffering of proper name anomia, an inability to recall and produce people's name whereas semantic and biographical information about people remain accessible (Semenza & Zettin, 1988), have revealed lesions of left medial and anterior temporal lobe (Damasio et al., 1990; Reinkemeier et al., 1997; Papagno et Capitani, 1998). Brain imaging studies of healthy subjects also support these results, as people's names retrieval is associated with a greater activation of left middle temporal gyrus (Damasio et al., 1996), and more specifically with activation of its anterior portion, corresponding to Brodmann area (BA) 21 (Gorno-Tempini et al., 1998).

Overall, there is clear evidence that distinct cortical regions mediate face and proper name processing. These observations lead us to ask the question of how the brain

operates to integrate these distinct faces and names representations to form a unified representation of familiar people. Firstly, one can wonder how people are able to encode and store such an integrated face-name representation in long-term memory. Several studies seem to outline the key role of the hippocampus as mediating the relational bound created between face and name representations (Sperling et al., 2001; Cohen et al., 1999; Leveroni et al., 2000). Secondly, one can also wonder how people are able to retrieve such face-name associations. This question has recently been examined by Campanella et al. (2001), by means of a PET study. Participants had to learn 24 unknown faces associated with 24 family names and to group the 24 so-created people in 12 male/female couples (e.g., male CORNET/ female BODART). By this way, subjects were able to retrieve the representation of a face (face CORNET) when presented with the name of the associated person (name BODART), and conversely. During PET recording, the task was to decide whether each presented pair of stimuli was a previously learned couple or not. Results showed that recognition of face-name associations, relative to recognition of face-face and name-name associations, activated a neural

network lateralized in the left hemisphere and including the inferior frontal gyrus (BA 45), the medial frontal gyrus (BA 6) and the inferior parietal lobe (BA 40). In accordance with other studies (Clark et al., 2000), the authors suggested that the activation of left BA40 could be interpreted as the locus where the association of the distinct visual representations presented to subjects took place.

Nevertheless, the PET technique, due to its poor temporal resolution, does not allow to infer any information about the *timing of occurrence* of these activations. As a major aim of neurocognitive studies is to delineate *when* the different processes of a cognitive function take place in the brain, it was therefore relevant to replicate the study of Campanella et al. by using Event Related Potentials (ERPs) (Rugg & Coles, 1995). ERPs are defined as an imaging technique characterized by a high temporal resolution, which would allow us to neurophysiologically index the cerebral processes involved in the retrieval of a face-name association.

Indeed, if the psychophysiology of face perception and proper names processing have already been studied, we know of no study that has examined the temporal course of the retrieval of face-name associations. On the one hand, face

perception is classically referred to two early electrophysiological components, known as the P1/N1 and the N170/VPP. The first complex is composed by a positive peak (P1) culminating around 100 milliseconds (msec) at posterior electrodes (Oz) and a centro-frontal (Cz) negativity (N1) occurring with a similar latency. P1/N1 complex is typically described in visual ERP studies to reflect early striate and extrastriate visual processing (Gomez et al., 1994; Heinze et al., 1994; Campanella et al., 2002). The N170/VPP complex is composed by a negativity peaking bilaterally around 150 msec (N170) at temporal sites (T5 and T6) (Bentin et al., 1996; George et al., 1996; Campanella et al., 2000). This negativity reverses polarity at the level of Cz and gives rise to the VPP (Rossion et al., 1999; Jeffreys, 1989; Bötzel et al., 1995). The N170/VPP is considered as the process indexing the structural analysis of facial information in order to obtain a configurational face representation (Jeffreys, 1996).

On the other hand, ERP studies concerning proper names processing are less numerous than for faces. Nevertheless, it seems that the perception of proper names elicited a bilateral P1 on temporo-occipital sites, a left-lateralized N1, and a large P2 at vertex (Müller & Kutas, 1996; Dehaene, 1995; Proverbio

et al., 2001). Moreover, Dehaene (1995) observed ERP differences beginning to appear around 260 msec after stimulus onset, with a larger left inferior temporal negativity for proper names relative to other types of written words (animal names, numerals and verbs).

With all these considerations in mind, the main issue of this research was to assess the existence of specific electrophysiological waves associated with the retrieval of face-name associations, in order to temporally constrain the neuroanatomical model proposed by Campanella et al. (2001). In the present experiment, we used the same methodology as in the PET study. Participants were first submitted to a learning phase. Then, during EEG recording, they were confronted with four experimental conditions : (1) a "name-name" condition (NN), in which they were visually presented with pairs composed by a name followed by another name; (2) a "face-face" condition (FF); (3) a "name-face" condition (NF); and (4) a "face-name" condition (FN), in which they were presented with pairs composed by a name and a face. Their task always consisted in deciding whether each presented couple corresponded to one of the previously learned male/female couples. It is worth noting that all displayed

stimuli were familiar. So, when a new couple was presented, it was nevertheless formed with two names, two faces or a name and a face previously learned but issuing from two different couples. Also, the pairs were always constituted by a male and a female stimulus.

Using the same material and task across all conditions will ensure an identical balance over task conditions, which allowed us to make three different assumptions. First, we should observe the early electrophysiological peaks classically described in the literature and associated with the perceptual processing of faces (P1/N1 and N170/VPP) and proper names (P1 / N1, P2). Moreover, if, as expected, conditions shared the same level of difficulty and required identical attentional resources, we shouldn't observe any significant latency or amplitude differences between waves associated with the first and second stimuli of each pair (first names of NN and NF conditions, first faces of FF and FN conditions, second names of NN and FN conditions, second faces of FF and NF conditions) Second, we use a paradigm in which two experimental conditions were defined by two stimuli presenting the same visual format (NN and FF), while two other conditions were defined by two stimuli presenting two

different visual formats (NF and FN). In other words, subjects had to make a judgment (correct male/female couple or not) when confronted with the second stimuli of these pairs, and this second stimuli shared (FF and NN) or not (FN and NF) the visual format of the first one. It can therefore be hypothesized that the subtraction of the second stimuli of NN and FF conditions from the second stimuli of NF and FN conditions will reveal the neurophysiological mechanisms specifically implicated in the retrieval of an association composed by distinct faces and names visual representations. In the sight of the results of the PET study of Campanella et al. (2001), we can further assume that these neurophysiological activities (1) will be generated in the left hemisphere, and (2) will reflect specific integrative processes (of visual facial representation with visual name representation), distinct from those implied in the processing of either a face or a name. It is worth noting that both NN and FF conditions could lead to implicit activations of the cerebral regions sustaining these specific integrative processes, but it has been shown that the explicit processing of information gives rise to enhanced cerebral activations relative to the implicit processing of these information (Price et al., 1996, 1997). Moreover, this

subtraction method has already received strong empirical support by using ERPs (e.g. Dehaene et al., 2001; Giard & Peronnet, 1999). Thirdly, even if the spatial resolution of the ERP technique is poorer than the one of PET or fMRI, a dipole localization analysis of the specific integrative waves revealed by the subtraction could (1) give prominence to neural generators spatially close from the three activated neural regions found by Campanella et al. (2001); and (2) reveal a precise timing of their order of occurrence.

Methods

Subjects.

Nine paid volunteers (5 males), who did not take part in the previous PET study, participated in the present experiment. They were aged between 20 and 29 years (mean : 22.8) and were all right-handed but one. They all had a normal or corrected-to-normal vision and no one had previous or present history of significant medical illness. All participants gave their written informed consent.

Stimuli.

We used the same stimuli as in the PET study, i.e. twenty-four unknown faces (12 females) selected from the “Psychological Image Collection at Stirling (PICS)” at the website address “<http://www.pics.psych.stir.ac.uk>”. All faces were presented in frontal position with a neutral expression. They were downloaded into a Macintosh computer and were edited by Adobe Photoshop 5.0 (Adobe Systems Incorporated). Gray-scale images were created and scaled to 274 x 350 pixels (123 x 96 mm, corresponding to a visual angle of 7.04° x 5.5 °). A set of 24 proper names (those used in our previous study) were chosen from the “Belgian National Institute of Statistics” in order that each name has nearly the same frequency of occurrence in the Belgian population and was constituted by 6 letters and 2 syllables. Gray-scale images of 274 x 350 pixels (123 x 96 mm, visual angle of 7.04° x 5.5°) were also created with Adobe Photoshop 5.0 for these names. Then the 24 faces were randomly paired with the 24 names and each male “face-name association” was coupled with a female one (Figure 1).

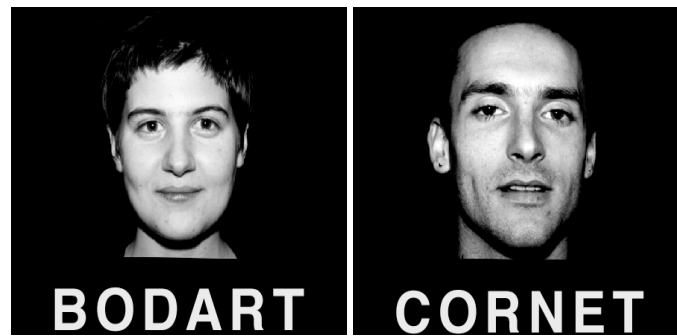


Figure 1: illustration of a couple (female/male) face-name association

Procedure.

Participants were presented with the same learning phase, defined by 3 sessions in the 3 consecutive days preceding the EEG recording session. Each learning session lasted about 1 hour. The first day, they were shown 12 faces (6 females) associated with their respective name on a monitor. Subjects were asked to encode these associations, with no time pressure. Then, they were given sheets of paper with the names and faces just encoded and they had to re-create the face-name pairs as accurately and quickly as possible. If they performed the task correctly, the 12 learned face-name associations were then presented in 6 different male / female couples, in order to encode these associations. Following that, sheets of paper with the names and faces encoded were again given to subjects who had to form the 6 couples as accurately and quickly as possible. The session ended when they were

able to complete this last task with no error or hesitation at all (if they weren't, a new encoding phase was performed and the task was again proposed). The second learning session was dedicated to the encoding of the 12 other face-names associations and the 6 resulting couples. The same procedure than for the first session was applied. On the third day, subjects were presented with the 24 face-name associations and the 12 male / female couples. This recapitulation session used the same tasks as previously and was ended when subjects were able to retrieve all the couples quickly and totally accurately.

On the fourth day, EEG was recorded in 4 different experimental conditions. These were matching tasks after name-name (NN), face-face (FF), face-name (FN) and name-face (NF) presentations. The stimuli were presented as male-female couples. A trial was defined by a small white cross during 300 msec, the first stimulus presented 700 msec, a black screen as inter-stimulus for 500 msec, the second stimulus for 700 msec and a final black screen for 1000 msec. In the non-mixed conditions, the 2 stimuli were 2 names (NN) or 2 faces (FF), whereas in the mixed conditions, subjects were confronted with either a face followed by a name (FN) or by a

name followed by a face (NF). Subjects had to decide whether the pairs of stimuli referred correctly to a previously learned couple (for instance, if the name Bodart was followed by the face Cornet, the correct response is "yes"; however, a "no" response would have been required if, for instance, the name "Bodart" had been presented with the male face "Jansen"). To do so, they had to press one of 2 pre-defined keys (counterbalanced across participants) on a response button box with the right finger. They were confronted with 576 trials (144 per condition), randomly presented in 24 blocks. Half of the trials were incorrect. None of the incorrect couples appeared more than once and an incorrect couple was never defined by 2 stimuli representing the same person or the same sex (e.g. 2 males). During EEG recording, subjects were sitting on a chair in a dark and quiet room with their head restrained by a chin rest. Their head was placed 100 cm from a high-resolution computer screen where stimuli were presented. Written instructions were given to subjects, with emphasis on speed and accuracy. Before the first block, some practice trials were given to subjects in order to familiarize them with the task.

By contrast with the PET study where conditions were presented by block, here we chose to mix the four conditions in the same proportions within each block, in order to maintain identical levels of attention and prevent some anticipatory effects across trials.

EEG recordings.

Sixty-four electrodes mounted in an electrode cap (Quick-Cap, NeuroMedical Supplies®, Neurosoft, Inc.) recorded EEG. Electrode positions included the standard 10-20 system locations and additional intermediate positions. Recordings were made with the 2 ears as the physical reference. The EEG was amplified by battery-operated SYNAMPS® amplifiers with a gain of 30,000 and a band pass of 0,015-30 Hz. EEG was continuously recorded (sampling rate 500 Hz, Neuroscan) and stored on disk for further analyses. Trials with EOG artifacts were automatically eliminated and epochs beginning 150 msec prior to stimulus onset and continuing for 850 msec were created. A recalculation was made in order to obtain common average reference recordings (Bertrand, Perrin et Perrier, 1985). Codes synchronized with stimulus delivery were used to average

selectively epochs associated with different stimulus types. Three parameters were coded for every stimulus : (1) the condition (FF, NN, FN or NF); (2) the position of the stimulus in the pair (first of second) and (3) the kind of pair (correct or incorrect pair). This coding allowed us to compute different averages of ERP target stimuli. These averages were made for each subject individually. A sample grand average was obtained by averaging across the subjects first names of NN, first names of NF, second names of NN, second names of FN, first faces of FF, first faces of FN, second faces of FF and second faces of NF conditions, second stimuli of NN and FF conditions, and second stimuli of FN and NF conditions. Only correct trials were included in these averages.

Statistical analyses.

Latencies and percentages of correct responses were computed and analyzed with Systat 5.2.1® (Systat, Inc.). In order to ensure that the mixed (NF and FN) and the non-mixed (FF and NN) conditions shared identical levels of difficulty and attention, we compared electrophysiological waves elicited by the visual processing of the first (and second) faces of FF and FN (NF) conditions and the first (and second) names of NN and NF (NF) conditions. At selected electrodes, individual peak

amplitudes of different components were obtained individually. These values were tested using paired Student *t* tests and repeated measures ANOVAs.

We also compared the first names and faces of mixed and non-mixed conditions at later latencies to check if one kind of condition led to preparatory and anticipative mechanisms relative to another, that could explain some of the results of the subtraction. For each participant, mean amplitudes were computed at Oz, Pz, Cz and Fz within four time windows (300-398 ms, 400-498 ms, 500-598 ms, 600-698 ms, one measure every two milliseconds). Repeated measures ANOVAs were then calculated between names of NN and NF conditions and faces of FF and NF conditions for each time interval.

The major analysis of this experiment was the subtraction of the second stimuli of NN and FF (non-mixed) conditions from the second stimuli of FN and NF (mixed) conditions, in order to assess specific integrative processes. Mixed and non-mixed conditions were averaged and subtracted in each subject, and peak amplitudes and latencies were measured for each electrode of interest. Significant effects were then calculated on the grand average at each

selected electrode using Student t-tests (amplitude of the difference wave compared to zero). The spatio-temporal patterns that had a significant amplitude ($p < .01$) at least one electrode for 15 msec (Rugg, Doyle & Wells, 1995) were considered as significant.

The analysis of intracranial dipoles was performed with the ASA software (A.N.T. Software BV, Enschede, The Netherlands), which determines the position and orientation of intracranial dipoles and their time-varying strength by using a three-layer spherical head model. We used a rotating dipole method in which the position of the dipoles was fixed but the orientation remained free.

Results

Behavioral data.

An ANOVA with kind of conditions (mixed conditions : mean of NF and FN, non-mixed conditions : mean of NN and FF) and kind of pairs (correct, i.e. learned Vs incorrect) as within-factors was performed on percentages of correct responses (Table 1). There were significant effects of the kind of condition ($F(1,8) = 23.008$; $p = .001$), the kind of pair ($F(1,8)$

= 10.028; $p = .034$) and interaction between kind of condition and kind of pair ($F(1,8) = 10.028$; $p = .034$). The main effect of the kind of condition indicates that the mixed conditions were better performed than the non-mixed conditions. But, as shown in Table 1, all conditions were highly performed (above 94 %) and no difference between conditions exceeded 3.1 %. Moreover, the segregation of the four conditions reveals that all but FF conditions were identically performed (correct NN : 97%, incorrect NN : 97%; correct FF : 96%, incorrect FF : 92%; correct NF : 97 %, incorrect NF : 99%; correct FN : 98%, incorrect FN: 96%) The main effect of response type show that subjects performed the task more efficiently when the presented pairs corresponded to a previously learned couple. However, paired Student t tests showed that this difference was only significant for the non-mixed conditions ($T(8) = 2.53$, $p = .04$), which explains the significant interaction.

The same ANOVA was performed on correct response latencies (Table 1). There was a main effect of the kind of pair ($F(1,8) = 13.43$; $p = .01$) but no significant main effect of the kind of condition ($F(1,8) = 0.06$; *ns*) and no significant interaction ($F(1,8) = 0.71$; *ns*). Thus, subjects answered faster

to the previously learned pairs in mixed and non-mixed conditions.

Overall, behavioral measures showed that subjects answered better and faster for the previously learned pairs, but there was no clear evidence of major differences of complexity level or attentional resources between mixed and non-mixed conditions.

Table 1

Mean correct responses and latencies (SD) for the correct and incorrect pairs of mixed (FN and NF) and non-mixed (NN and FF) conditions.

	Mixed		Non-mixed	
Response type	correct	incorrect	correct	incorrect
Mean correct responses	97.4	97.4	96.4	94.3
(%)	(1.42)	(1.13)	(1.88)	(1.73)
Mean correct latencies	522	577	526	569
(msec)	(154.9)	(155.9)	(137.4)	(146.9)

Electrophysiological data

1) Visual processing of faces and names

The purpose of this first set of analyses was to ensure that all faces and names were encoded in the same way across mixed and non-mixed conditions.

Faces

For faces, we considered the P1/N1 and the N170/VPP complexes (Figure 2a, left part), reflecting the structural encoding processes as described in the literature. The P1/N1 complex was characterized for all faces by a positivity over posterior electrodes culminating (Oz) at 118 msec and a polarity reversal at central sites (Cz) at the same latency. The N170 wave was observable bilaterally at parieto-occipital sites. It culminated at PO7 and PO8 at 161 msec and was synchronized with a Vertex Positive Potential maximally recorded at Cz at 162 msec.

At first, ERPs elicited by the first faces of FF condition were compared to ERPs elicited by the first faces of FN condition. An ANOVA on peak voltage amplitude of P1 with conditions (FF, FN) and lateralization (O1, O2) as factors was carried out. We report here only the significant effects. The complete results of the ANOVAs are summarized in Table 2. It showed a significant effect of lateralization ($F(1,8) = 7.15$; $p = .03$) and a significant interaction ($F(1,8) = 5.78$; $p = .04$). Paired Student *t*-test showed that amplitude of P1 was larger in the right hemisphere (O2) only in FF condition ($T(8) = 3.71$; $p = .01$). The same analysis was performed on the negative

counterpart (N1) at central electrodes. An ANOVA with conditions (FF, FN) and lateralization (C1, C2) revealed a significant effect of lateralization ($F(1,8) = 11.41$; $p = .01$). The interaction between conditions and lateralization did not reach a significant statistical level ($F(1,8) = 0.32$; *ns*). The amplitude of N1 was therefore larger in the left hemisphere (C1) in both conditions.

An ANOVA was performed on peak voltage amplitude of N170 with conditions (FF, FN) and lateralization (PO7, PO8). Factors failed to show any significant effect of condition, lateralization or interaction. A similar analysis was applied on peak voltage amplitude of VPP. An ANOVA with condition (FF, FN) and lateralization (C1, C2) showed no significant effects.

Then, we compared the ERPs elicited by the second faces of FF and NF conditions (Figure 2a, right part). An ANOVA on peak voltage amplitude of P1 with conditions (FF, NF) and lateralization (PO7, PO8) showed a significant effect of lateralization ($F(1,8) = 6.19$; $p = .04$) without significant effect of condition ($F(1,8) = 0.28$; *ns*) or interaction ($F(1,8) = 0.08$; *ns*), indicating that P1 was larger in the right hemisphere in both conditions. The ANOVAs performed on N1, N170 and VPP failed to show any significant effect.

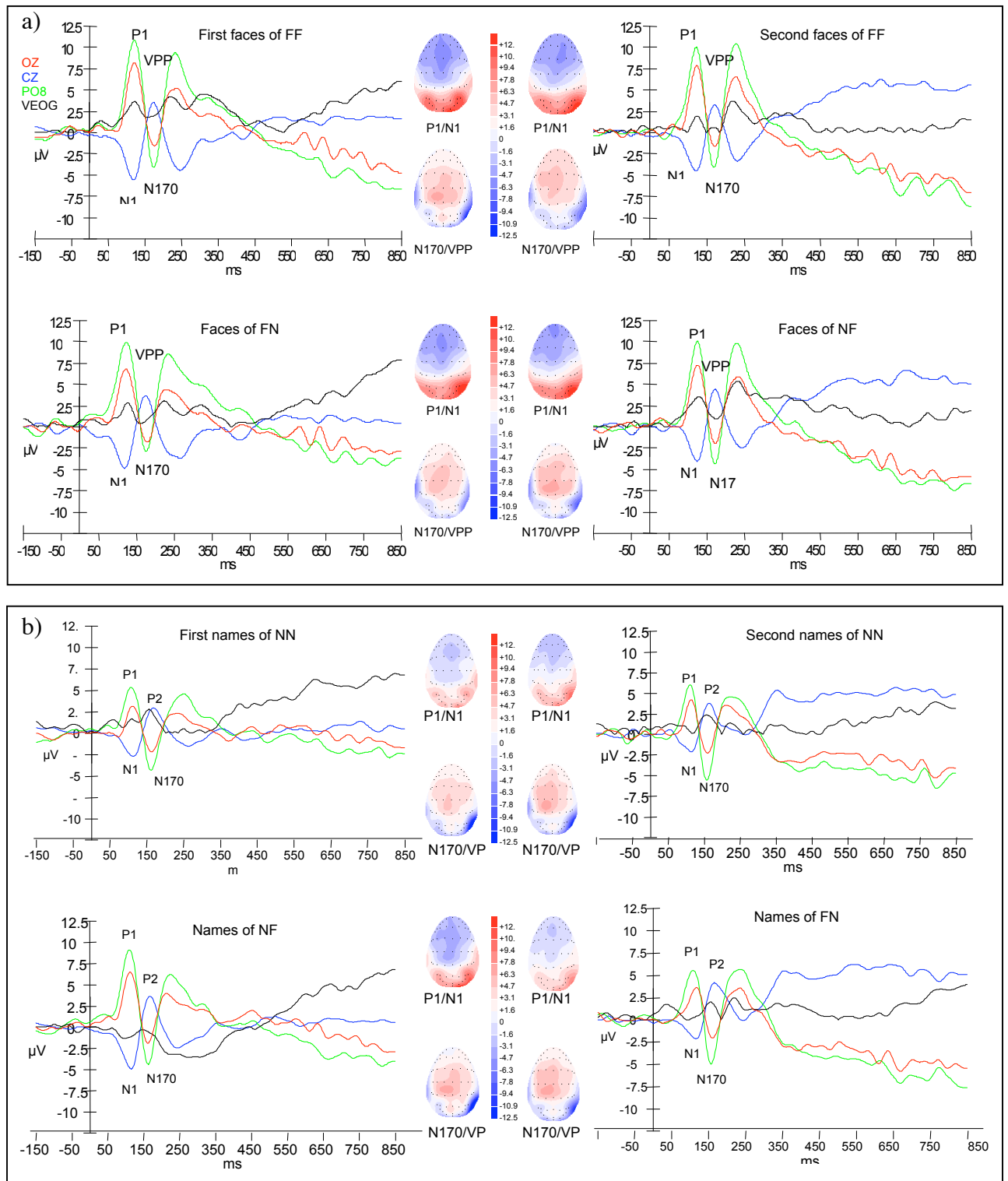


Figure 2 : illustration of grand-average ERPs of P1/N1 and N170/P2-VPP complexes elicited by : a) first and second faces of FF, FN and NF conditions; b) first and second names of NN, NF and FN conditions. For each figure, the upper topographies correspond to P1/N1 complexes and the lower to N170/P2 (or VPP) complexes.

Names

We observed the same electrophysiological waves for names than for faces, i.e. a P1/N1 and a N170/P2 complex (Figure 2b), corresponding to the encoding processes of orthographical stimuli (Bentin et al., 1999). The P1 positivity peaked at Oz electrode between 101 and 105 msec and the reversal polarity was maximum at Cz between 107 and 109 msec. The N170 wave appeared bilaterally at parieto-occipital electrodes (PO7, PO8) at 155 msec and was synchronized with a P2 wave culminating at central electrodes (Cz) at 158 msec.

As for faces, we compared firstly the ERPs elicited by the first names of NN and NF conditions (Table 2). ANOVAs on peak voltage amplitudes of P1, N1, N170 and P2 with conditions (NN, NF) and lateralization (P1 : O1, O2; N1 : C1, C2; N170 : PO7, PO8; P2 : C1, C2) as factors didn't reveal any significant effect of condition, lateralization or interaction. The comparison of the second names of NN and FN conditions failed also to show any significant differences between conditions or any effect of lateralization.

Table 2

Results of the ANOVA's performed on peak voltage amplitudes of P1, N1, N170 and VPP (or P2) with conditions (first faces: FF, FN; second faces: FF, NF; first names: NN, NF; second names: NN, FN) and lateralization (P1: O1, O2; N1: C1; C2; N170: PO7, PO8; VPP-P2: C1, C2) as factors. p values above .05 were considered as non-significant (ns).

a)	First faces (FF – FN)			Second faces (FF – NF)		
	Condition	lateralization	interaction	Condition	lateralization	interaction
P1	$F(1,8) = 2.86$ ns	$F(1,8) = 7.15$ p = .03	$F(1,8) = 3.71$ p = .01	$F(1,8) = 0.28$ ns	$F(1,8) = 6.19$ p = .04	$F(1,8) = 0.08$ ns
N1	$F(1,8) = 0.1$ ns	$F(1,8) = 11.41$ p = .01	$F(1,8) = 0.32$ ns	$F(1,8) = 0.25$ ns	$F(1,8) = 0.89$ ns	$F(1,8) = 0.06$ ns
N170	$F(1,8) = 2.82$ ns	$F(1,8) = 0.59$ ns	$F(1,8) = 1.14$ ns	$F(1,8) = 1.63$ ns	$F(1,8) = 1.80$ ns	$F(1,8) = 1.21$ ns
VPP	$F(1,8) = 0.27$ ns	$F(1,8) = 0.28$ ns	$F(1,8) = 1.3$ ns	$F(1,8) = 0.03$ ns	$F(1,8) = 2.18$ ns	$F(1,8) = 2.34$ ns
b)	First names (NN – NF)			Second names (NN – FN)		
	Condition	lateralization	interaction	Condition	lateralization	interaction
P1	$F(1,8) = 1.2$ ns	$F(1,8) = 3.04$ ns	$F(1,8) = 4.99$ ns	$F(1,8) = 0.07$ ns	$F(1,8) = 3.95$ ns	$F(1,8) = 4.16$ ns
N1	$F(1,8) = 0.59$ ns	$F(1,8) = 0.34$ ns	$F(1,8) = 4.49$ ns	$F(1,8) = 0.18$ ns	$F(1,8) = 0$ ns	$F(1,8) = 2.06$ ns
N170	$F(1,8) = 3.66$ ns	$F(1,8) = 0.06$ ns	$F(1,8) = 0.12$ ns	$F(1,8) = 0.86$ ns	$F(1,8) = 0.15$ ns	$F(1,8) = 0.63$ ns
P2	$F(1,8) = 0$ ns	$F(1,8) = 0.57$ ns	$F(1,8) = 0.67$ ns	$F(1,8) = 0.11$ ns	$F(1,8) = 0.52$ ns	$F(1,8) = 0.05$ ns

To sum up, the comparison of the electrophysiological waves associated with the processing of the first stimuli (faces or names) of mixed and non-mixed conditions did not reveal any significant differences between conditions, except some lateralization effects of P1/N1 in conditions requiring the perception of faces. Nevertheless, the mean amplitudes of the

P1/N1 and the N170/VPP (or P2 for names) complex were identical for the first faces of FF and FN conditions and for the first names of NN and NF conditions. Therefore, it seems that names and faces were processed in the same way in all conditions.

2) Comparison of the first stimuli of mixed and non-mixed conditions between 300 and 700 ms

We compared mean amplitudes recorded at Oz, Pz, Cz and Fz in four 100 ms-time windows (one measure every 2 ms) after the onset of the first names of NN and NF conditions and the first faces of FF and FN conditions, to ensure that one kind of condition did not create anticipatory mechanisms relative to another. Repeated measures ANOVAs with conditions (names : NN and NF; faces : FF and FN) and electrode position (Oz, Pz, Cz, Fz) as factors were performed for each time window. They failed to find any significant effect at any time for the names or for the faces (Table 3).

Table 3

Results of the ANOVAs performed on the mean voltage amplitudes measured between 300 and 398 ms, 400 and 498 ms, 500 and 598 ms, 600 and 698 ms after the onset of the first names and faces of mixed and non-mixed conditions. Factors : electrode position (Oz, Pz, Cz, Fz) and condition (NN and NF for the names, FF and FN for the faces). p values above .05 were considered as non-significant (ns).

	ms	Electrode position	condition	interaction
First names of NN and NF	300-398	$F(3,24) = 0.74$	$F(1,8) = 2.52$	$F(3,24) = 0.94$
		ns	ns	ns
	400-498	$F(3,24) = 0.42$	$F(1,8) = 0.18$	$F(3,24) = 0.09$
		ns	ns	ns
	500-598	$F(3,24) = 0.03$	$F(1,8) = 0.77$	$F(3,24) = 0.06$
		ns	ns	ns
	600-698	$F(3,24) = 1.83$	$F(1,8) = 0.94$	$F(3,24) = 1.07$
		ns	ns	ns
First faces of FF and FN	300-398	$F(3,24) = 0.13$	$F(1,8) = 2.10$	$F(3,24) = 1.80$
		ns	ns	ns
	400-498	$F(3,24) = 0.88$	$F(1,8) = 0.01$	$F(3,24) = 1.29$
		ns	ns	ns
	500-598	$F(3,24) = 0.35$	$F(1,8) = 0.12$	$F(3,24) = 0.93$
		ns	ns	ns
	600-698	$F(3,24) = 2.14$	$F(1,8) = 0.86$	$F(3,24) = 0.63$
		ns	ns	ns

3) Subtraction of NN and FF from NF and FN

This subtraction revealed two major electrophysiological components : a bilateral negativity spreading over all posterior electrodes around 284 msec in the left hemisphere and 293 msec in the right hemisphere, and a left centro-frontal positivity peaking around 300 msec (Figure 3a, b, c). For the negativity, a significant negative amplitude was

recorded at Oz and POz; O1, PO3, PO5, PO7, P3 and P5 in the left hemisphere, PO4, PO6, PO8, P4 and P2 in the right hemisphere (for amplitudes, latencies and significance levels, see Table 4). The centro-frontal positivity was maximally observed at C1 and FC1, and to a lesser extent at Cz and FCz. At the corresponding electrodes of the right hemisphere (C2 and FC2), the positive amplitude did not reach a significant statistical level (Table 5).

Table 4

Negativity in (NF + FN) – (NN + FF). (1) hemispheres, (2) electrodes, (3) peak amplitudes, (4) latency of maximum amplitude difference from zero (Student t-tests) and significance level.

<i>Electrodes</i>		<i>Amplitude (μV)</i>	<i>Latence (Msec)</i>
Left	O1	-1.70	287 **
	PO3	-1.79	284 **
	PO5	-1.45	283 **
	PO7	-1.42	284 **
	P3	-1.38	283 **
	P5	-1.35	284 ***
Midline	Oz	-1.48	287 ***
	POz	-1.52	287 *
Right	PO4	-1.68	290 **
	PO6	-1.45	293 *
	PO8	-1.47	293 *
	P4	-1.49	294 **
	P2	-1.29	297 **

* = $p < .01$. ** = $p < .005$. *** = $p < .0025$

Table 5

Positivity in (NF + FN) – (NN + FF). (1) hemispheres, (2) electrodes, (3) peak amplitudes, (4) latency of maximum amplitude difference from zero (Student t-tests) and significance level.

<i>Electrodes</i>		<i>Amplitude (μV)</i>	<i>Latence (Msec)</i>
Left	C1	1.06	303 ****
	FC1	1.32	300 ****
Midline	Cz	1.01	310 ***
	FCz	1.48	300 ***
Right	C2	0.56	302 *
	FC2	1.09	297 **

* = $p < .1$. ** = $p < .05$. *** = $p < .01$. **** = $p < .005$

An ANOVA with lateralization (left or right hemisphere) and electrode position (occipital : O1, O2; occipito-parietal : PO5, PO6; parietal : P5, P6) as factors was performed on mean amplitudes and latencies of the negativity. It failed to find any significant difference of amplitude between hemispheres (lateralization : $F(1,8) = 2.67$; *ns*) nor any effect of electrode position ($F(2,16) = 1.35$; *ns*), nor any interaction ($F(2,16) = 3.27$, *ns*). On latencies, the analysis revealed that the negativity appeared first in the left hemisphere (hemisphere : $F(1,8) = 8.41$, $p = .02$) but didn't reveal any significant effect of electrode position ($F(2,16) = 0.31$; *ns*) or interaction ($F(2,16) = 0.38$; *ns*). As the centro-frontal positivity was only significant in the left hemisphere, no amplitude or latency analyses of lateralization were performed on this component.

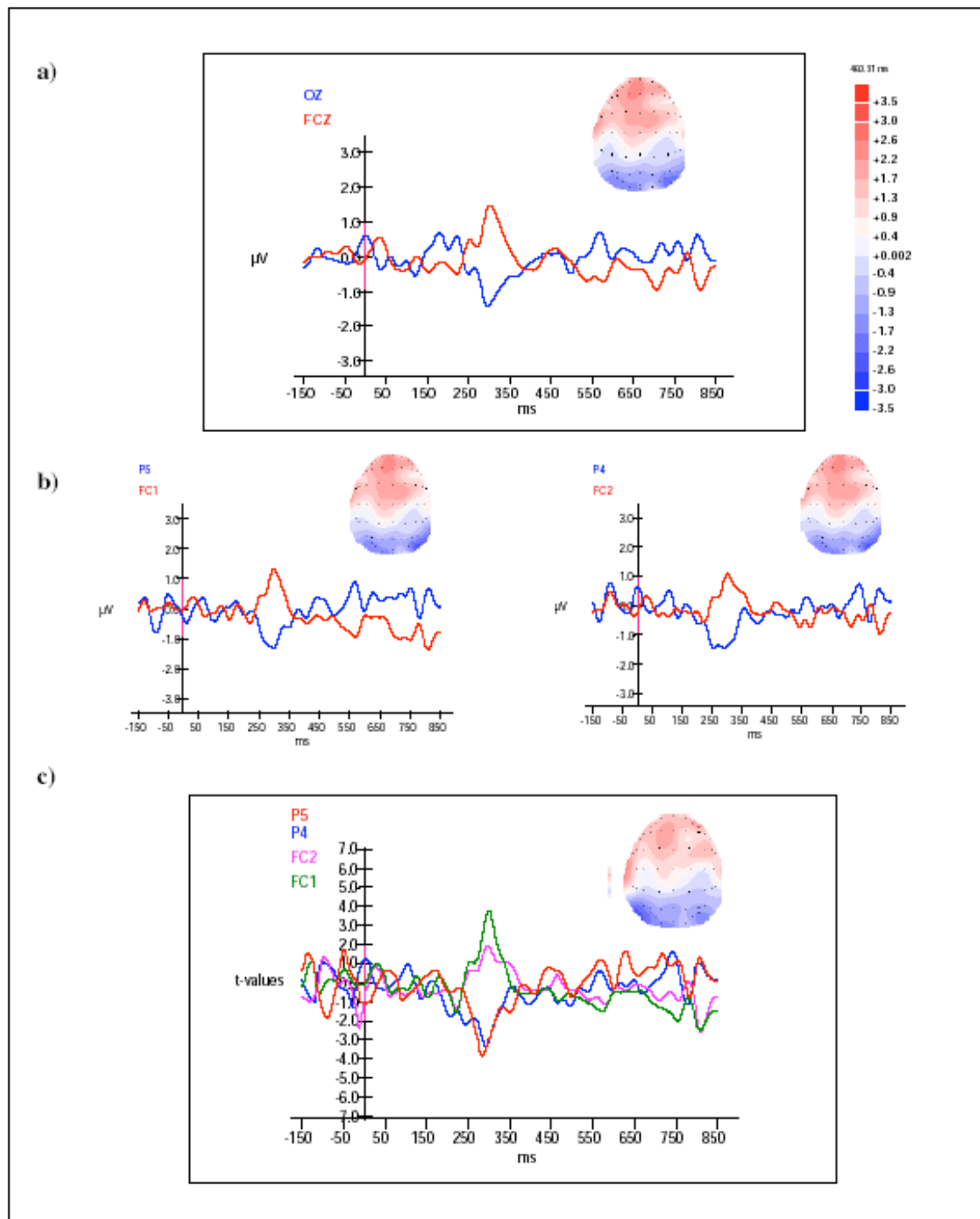


Figure 3 : Illustration of grand average ERPs and topographies elicited by the subtraction of NN and FF conditions from FN and NF conditions. a) superposition of the negative wave (in μV) recorded at Oz around 296 msec and the positive wave recorded at FCz around 301 msec ; b) comparison of the ERP components (in μV) recorded in the left (P5 and FC1) and the right (P4 and FC2) hemispheres, c) comparison (in t-values) of the positive wave recorded in the left (FC1 : Maximum Amplitude Difference (MAD) at 300 msec, $p < .005$) and right (FC2 : MAD at 297 msec, $p < .05$) hemispheres and the negative wave recorded in the left (P5 : MAD at 284 msec, $p < .0025$) and right (P4 : MAD at 294 msec, $p < .005$) hemispheres.

4) Dipole modeling

Dipole localization was carried out on the grand average file of the subtraction described above. Our goal was to temporally constrain the anatomical network whose activation was observed by Campanella et al. (2001). In a simultaneous fit of all source analyses, we constrained the model with 3 rotating dipoles whose localization corresponded to the 3 cerebral areas observed in the previous PET study. In the chosen time window (260 - 320 msec), the 3 dipoles explained 90.26% of the variance (residual variance (RV) : 9.74%, correlation : 95%) (Figure 4). We then performed separate fit analyses for every 1 msec intervals between 260 and 320 msec for each of the 3 dipoles in order to determine when each neural generator maximally contributed to the observed electrophysiological activity. The dipole corresponding to the left parietal lobe (BA 40; x : -50, y : 44, z : 32) showed a lower residual variance (RV) than the other 2 for all intervals with a minimum RV at 298 msec (RV : 14.1%, correlation : 92.68%). The dipole localized in the left medial frontal gyrus (BA 6; x : 8, y : 34, z : 56) showed a minimum RV at 286 msec (RV : 27.01%, correlation : 85.43%) and the dipole corresponding to

the left inferior frontal gyrus (BA 45; $x : 36, y : 28, z : 6$) had a minimum RV at 310 msec (RV : 24.64%, correlation : 86.81%).

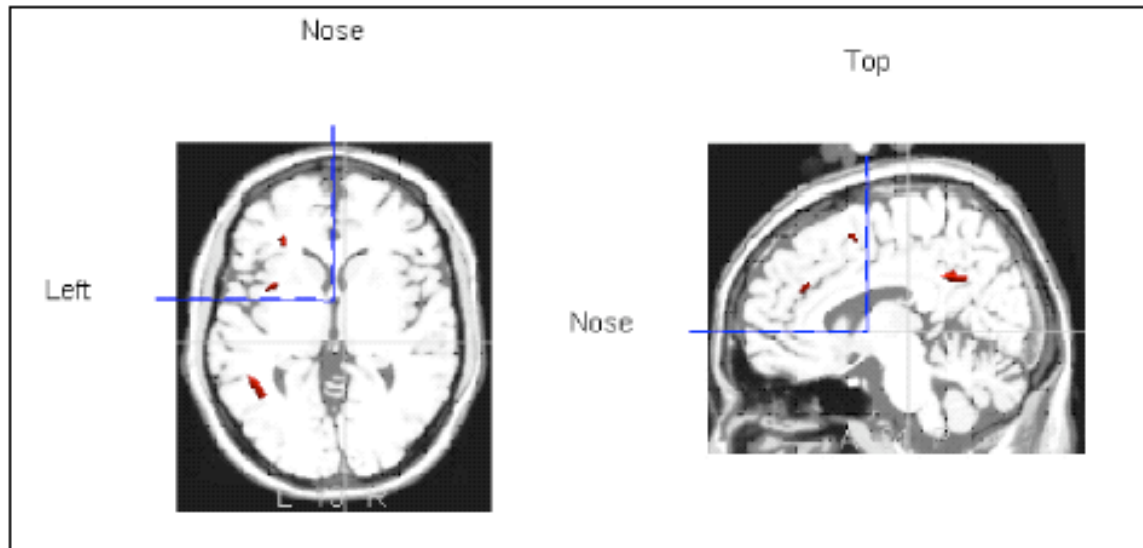


Figure 4 : illustrations of the dipole modelization applied on the grand average of the subtraction (NF+FN)-(NN+FF). Length of the dipoles represents their strength. Dipole position (on the down part of the figure, from left to right) : IFG 45 ($x : 36, y : 28, z : 6$), MFG 6 ($x : 8, y : 34, z : 56$) and IPL 40 ($x : -50, y : 44, z : 32$). Note the most prominent importance of the parietal dipole relative to the 2 others.

Discussion

The aim of the present study was to investigate the temporal course of the cerebral activity implied in the retrieval of face-name associations, when subjects had to recognize couples formed by two stimuli belonging to the same visual format (two names or two faces) or not (a face and a name). We first assumed that we should observe the classical electrophysiological waves elicited by the perceptual

processing of faces and names (P1/N1 and N170/VPP or P2 complexes), and that, if the four conditions were matched for difficulty and attention, no amplitude or latency differences should exist between the stimuli of mixed (NF and FN) and non-mixed (NN and FF) conditions. These early ERP components are known to be sensitive to modulations of attentional processes (Coull, 1998; Itier & Taylor, 2002). But, as suggested by statistical analyses, no difference of amplitude or latency was found between the first faces of FF and FN conditions and the first names of NN and NF conditions, and between the second faces of FF and NF conditions and the second names of NN and FN conditions, nor on the P1/N1 complex nor on the N170/VPP (or P2 for names) complex. The only effect we observed was a lateralization effect on the P1/N1 complex in conditions requiring perception of faces, P1 being right-located in all conditions but FN and N1 being left-located for the first faces of FF and FN conditions. We cannot completely explain these lateralization effects. Although P1 is generally found to have a bilateral origin (Rossion et al., 1999, Allison et al., 1999, Schweinberger et al., 2002), interhemispheric differences have also been described in the literature (Small, 1988). On the other hand, it has been

suggested in the literature that P1 and N1 are two ERP components reflecting the same neural activity (Campanella et al., 2002). Thus, it could suggest that the interaction observed between lateralization and conditions for the P1 and the absence of such interaction for the N1 don't really reflect functional differences in the cerebral activities involved in the perception of faces between FF and FN conditions.

Whatever, the absence of amplitude or latency differences between mixed and non-mixed conditions for ERP components elicited by the encoding processes of faces and proper names lead us to conclude that all conditions were matched for difficulty and attention. This interpretation is strengthened by the analysis of behavioral data. Firstly, as expected, subjects answered faster and more accurately when the presented pairs of stimuli corresponded to a previously learned couple. Secondly, latencies of correct responses were identical between mixed and non-mixed conditions. But all percentages of correct responses remained above 92 %, probably reflecting a ceiling effect and, moreover, only correct trials (whose performances were identical between mixed and non-mixed conditions) were used in our main subtraction. Finally, mean percentages of correct responses

and mean correct responses latencies were very close to the values observed in the PET study of Campanella et al. (2001). Thus, it seems that the main result of the subtraction cannot be explained by an additional requirement of attentional resources in mixed conditions in comparison with non-mixed conditions.

The main results of this study derive from the subtraction of FF and NN conditions from FN and NF conditions. First, this analysis was performed to put in light the precise timing of the electrophysiological mechanisms implied in the integrative processes of representations of faces and names, necessary to provide a unified representation of people. Second, on the basis of the PET study of Campanella et al. (2001), we also hypothesized that these neurophysiological activities would be generated in the left hemisphere, and that a dipole localization analysis of the specific integrative waves revealed by the pre-cited subtraction could give prominence to neural generators spatially close from the activated neural regions found by this PET study. The subtraction revealed two main ERP components : (1) a negative wave spreading bilaterally over all posterior electrodes between 280 and 300 msec, but

appearing more precociously and with higher statistical significance levels in the left hemisphere (maximum peak at left parietal electrode P5 at 284 msec); and (2) its positive counterpart maximally peaking at left centro-frontal sites (FC1 / C1) around 300 msec. Statistical analysis performed on the negativity showed that even if there were no difference in the wave amplitude between left and right hemispheres, there were a significant latency effect between the two hemispheres, the negativity appearing significantly earlier in the left hemisphere. Thus, we can consider that the main result of this study is that we have showed that processes involved in the integration of visual name and face representations took place in the brain around 280 and 300 ms, with a preferential involvement of the left hemisphere.

Moreover, a dipole modeling showed that 90% of the variance of these results can be explained by 3 dipoles corresponding to the neural generators proposed by Campanella et al. (2001). They showed, using the PET technique, the involvement of left parietal (BA 40) and frontal (BA 6 and BA 45) regions in the association of different representations of faces and names when subjects had to retrieve a face on the basis of a related name and conversely

(Campanella et al., 2001). BA 40 was interpreted as the brain region where the association of the distinct representations presented to subjects operates. The activation of the frontal regions was seen as reflecting executive and non-spatial attentional processes (BA 6) and retrieval of the learned associations stored in semantic memory (BA 45).

The present findings seem also to temporally constrain this anatomical model. Indeed, it was possible by considering each of these 3 dipoles individually to define their precise time of maximal activity. In sight of the results of this dipole modeling, we could propose that (1) the left medial frontal activity, appearing first in time (around 270 ms), could reflect the attentional processes necessary to shift from one visual processing (for instance the processing of a face) to another (the processing of a name) (Nagahama et al., 1999); (2) when the perceptual processing of the second stimulus is achieved, its representation could then be integrated to the representation of the first stimulus in the left inferior parietal lobe (around 285 ms), seen as a multimodal sensory convergence zone, relevant to perceptual binding and memory retrieval by providing a unique representation of combined sensory features (Damasio et al., 1990; Clark et al.,

2000) ; finally (3) this activity could spread, around 300ms, towards left inferior frontal regions, known to be involved in complex reflexive memory processes (Nolde et al., 1998 ; Rugg et al., 1999; Paller et al., 2000) such as the selection of information among competing alternatives in semantic memory (Thompson-Schill et al., 1997). The comparison of the presented face-name association to one of the couples stored in long-term memory could allow subjects to make a correct decision. As a matter of fact, we have thus a model of face-name retrieval implying sequentially the left frontal medial gyrus, the left inferior parietal lobe and the left inferior frontal gyrus.

We are aware that (1) the relative limitation of anatomical resolution of the ERP technique; and (2) the relative importance of the residual variance for each of our 3 dipoles in the separate fit analysis does not allow us to claim categorical and definitive conclusions. Moreover, matching rCBF activations and ERPs waves is a real challenge. On the one hand, rCBF activations result from the averaging of a multiple cognitive processes taking place over an entire block of trials. On the other hand, ERP waves reflect cerebral activities averaged in the time range of single trials. But,

overall, it is important to note that the 2 experiments (the previous PET study and the present ERPs one), using different techniques and different participants, lead to very similar results, i.e the existence of a network of cortical areas lateralized in the left hemisphere and producing cerebral activities spreading from posterior to anterior sites when subjects have to retrieve the representation of a face when presented with an associated name and conversely.

If the neuroanatomical and the neurophysiological correlates of face perception begin to be well known, our experiments are however among the rare studies that have investigated the integration and retrieval of face-name associations. Our results raise many interesting questions and further research will be necessary to provide a better comprehension of these processes, notably about the implication of our neural network in cross-modal integration processing (such as visuo-auditive face-name integration) or the specificity of this network relative to more general binding processes (such as objects-common names associations).

Acknowledgments

This study was supported by FRSM No. 3.4550.02. The first and second authors are supported by the Belgian Fund of Scientific Research (FNRS).

Introduction à l'étude 2

L'étude de Campanella et al. (2001) et la première étude de cette thèse, utilisant des techniques d'imagerie différentes mais une méthodologie identique, ont permis (1) d'isoler les régions cérébrales impliquées spécifiquement dans les processus associatifs entre visages et noms propres, (2) de contraindre temporellement l'activité de ces régions cérébrales.

La conclusion majeure de ces deux études est qu'il semble exister un réseau d'aires corticales latéralisé dans l'hémisphère gauche et produisant une activité cérébrale, aux alentours de 300 millisecondes après la présentation des stimuli, se propageant des régions postérieures vers les régions antérieures lorsque les participants devaient retrouver la représentation d'un visage à partir de la présentation d'un nom propres associé et inversement.

Il semble en outre que le lobe pariétal inférieur gauche joue un rôle particulier au sein de ces processus associatifs, puisque c'est cette région même qui pourrait sous-tendre l'intégration des représentations visuelles des visages et des

voix en une seule représentation unifiée et cohérente des personnes.

Ces études posent de nombreuses questions. Une d'entre elles au moins a trait à la spécificité des processus associatifs unissant visages et noms propres au sein des mécanismes plus généraux nous permettant de lier entre eux les éléments pertinents de notre environnement. En effet, il existe différentes preuves suggérant notamment que le traitement des visages et des objets pourrait être sous-tendu par des régions cérébrales distinctes. La neuropsychologie cognitive a mis au jour une double dissociation entre d'une part altération de la reconnaissance des visages / préservation de la reconnaissance des objets (Buxbaum et al., 1999; De Renzi et al., 1991; McNeil & Warrington, 1993), et d'autre part altération de la reconnaissance des objets / préservation de la reconnaissance des visages (Feindberg et al., 1994; Moscovitch et al., 1997). Plusieurs études d'imagerie cérébrale fonctionnelle ont également mis en évidence différents patterns d'activation cérébrale selon la catégorie d'objets à identifier (Kanwisher et al., 1997, 1999). Mitchell, Heatherton et Macrae (2002) ont quant à eux suggéré que le réseau d'informations sémantiques relatives aux personnes pourrait

être fonctionnellement distinct de celui lié aux autres catégories d'objets. Les résultats de leur étude par IRMf, durant laquelle les participants devaient produire des jugements sémantiques sur des visages ou des objets, montrent une activation d'un réseau d'aires cérébrales incluant notamment le cortex préfrontal, le cortex temporal supérieur, la scissure intrapariétale et le gyrus fusiforme pour les jugements sur les visages, et une forte désactivation de ces régions pour les jugements sur les objets.

Enfin, l'électrophysiologie a également mis en évidence des activités cérébrales spécifiques aux visages, aussi bien via des méthodes d'enregistrement intra-crânien (Allison et al., 1999) que via la méthode des potentiels évoqués (Bentin et al., 1996). En particulier, l'onde N170 montre une amplitude plus grande en réponse aux visages qu'en réponse à tout autre type de stimulation visuelle, suggérant des mécanismes spécifiques de traitement des visages humains (Bentin et al., 1996; Carmel & Bentin, 2002). Rappelons néanmoins que la question de savoir si ces activités reflètent des processus de traitement spécifiques ou des processus dépendant du niveau d'expertise acquis pour une catégorie particulière de stimulations visuelles, reste d'actualité.

Le statut des noms propres au sein de la catégorie lexicale des noms fait également débat, et plusieurs études ont montré des différences dans les patterns d'activités cérébrales liés à la récupération lexicale de noms communs et de noms propres (Damasio et al., 1996; Grabowski et al., 1998; Gorno-Tempini et al., 1998). En particulier, il semble que la récupération lexicale de noms de personnes implique l'activation de régions particulières au sein des gyri temporaux et frontaux de l'hémisphère gauche, avec une activation particulière des régions antérieures du lobe temporal.

L'électrophysiologie confirmera cette implication du lobe temporal gauche, puisque Proverbio et al. (2001) ont observé une onde négative apparaissant vers 150 millisecondes sur les électrodes temporales gauches lorsque leurs participants devaient retrouver en mémoire des noms de personnes par rapport à des noms communs.

Puisqu'il semble que le traitement des visages et le traitement des noms propres font l'objet de processus cérébraux distincts de ceux impliqués dans le traitement des autres objets visuels et des autres catégories de noms, on

peut se poser la question de savoir si les activités cérébrales associatives observées dans les études précédentes sont spécifiques aux associations entre visages et noms propres, ou si elles reflètent des mécanismes de "binding" plus généraux entre tout objet visuel et son nom.

L'étude décrite ci-dessous aura donc pour but de comparer les activités cérébrales associatives unissant visages et noms propres d'une part, et objets et noms communs d'autre part, afin de déterminer si ces processus associatifs sont ou non spécifiques aux informations relatives aux personnes. Tout comme dans l'étude précédente, la méthode soustractive, appliquée aux potentiels évoqués, sera utilisée.

Etude 2

Soumise pour publication

**The associative processes involved in faces-
proper names vs. animals-common names
binding: a comparative ERP study.**

Frederic Joassin, Gaëlle Meert, Salvatore Campanella, &
Raymond Bruyer

Unité de Neurosciences Cognitives, Université catholique de Louvain,
Louvain-la-Neuve, Belgium

Corresponding author:

Frédéric Joassin
Université catholique de Louvain
Faculté de Psychologie et des Sciences de l'Education
Unité de Neurosciences Cognitives (NESC)
Place du Cardinal Mercier, 10
B-1348 Louvain-la-Neuve
Belgium
Tel.: +32 (0)10/47.90.62
Fax: +32 (0)10/47.37.74
e-mail: Frederic.Joassin@psp.ucl.ac.be

Abstract

The aim of this ERP study was to investigate the electrophysiological correlates of the associative processes between human faces and proper names and to compare these mechanisms to those involved in the associations of objects and common names. For this purpose, twelve healthy participants had to perform a categorization task in six conditions while ERPs were recorded: famous proper names (P), famous faces (F), associations of faces and proper names (FP), animals (A), common names (C) and associations of animals and common names (AC). This allowed us to perform the subtractions $[FP-(F+P)]$ and $[AC-(A+C)]$ to isolate the specific binding processes involved in each kind of association. Results showed that both kinds of associations produced two main posterior negative / anterior positive complexes, with a more frontal distribution for AC than for FP associations, and bilateral temporal activities. These findings confirm that the general associative processes – independently of the kind of association – are not simply the sum of the activities elicited by each stimulation and suggest that they could imply both unimodal sensory and multimodal convergence regions.

Keywords: associative processes, faces, names, ERPs

Introduction

The ability to recognize person's faces and recall their names is a key feature of our social life. Indeed, faces and proper names are two major attributes by which we can identify people. For healthy individuals, identifying a person is an operation that often seems easy, fast and automatic (Sergent et al., 1992). However, recognizing a face requires fine discriminative perceptive processes to extract an abstract representation of the face, independent for instance from its orientation or emotional expression (at least according to the classical cognitive model of Bruce & Young, 1986). Moreover, it is well known that face and proper names processing are sustained by distinct brain areas, as faces processing is mediated by posterior occipito-temporal cortical regions, with a preferential activity in the right hemisphere (Sergent et al., 1992; Kanwisher et al., 1997 ; Campanella et al., 2001) whereas proper names processing is correlated with more anterior left temporal cerebral activity (Damasio et al., 1996; Gorno-Tempini et al., 1998). One can thus wonder how the brain operates to create a single coherent representation of people on the basis of the different attributes of a person, such as

their faces and their names, processed by distinct cortical regions.

One solution to solve this « binding problem » (Von der Malsburg, 1981), is to postulate the existence of heteromodal regions, which receive inputs from the cerebral areas processing each kind of information and which bind them together in order to create coherent and unified representations of the objects of our surroundings. This hypothesis has received strong supports, notably by several studies which have shown that some particular regions of the temporal cortex are dedicated to the auditory-visual speech integration (Calvert et al., 2000 ; Sekyama et al., 2003). Moreover, it seems that lesions of heteromodal regions, such as the parietal cortex or the medial temporal lobe, can lead to severe cognitive impairments (Valenza et al., 2004 ; Lancelot et al., 2005). For instance, Lancelot et al. (2005) have shown that patients with medial temporal lobe damage were impaired in learning auditory-spatial associations, whereas they were still able to process auditory and spatial information presented in isolation. These results suggest that (1) associative regions could be crucial for the integration of the different features of an object, (2) integration

impairments can exist without impairment of the processing of each kind of information.

We recently investigated the question of the associative processes between faces and proper names in two studies using the PET-scan and the ERP (Event Related Potentials) techniques. They were both aimed at examining the cerebral activities elicited by the retrieval of face-name associations (Campanella et al., 2001 ; Joassin et al., 2004). After a learning phase of 24 face-name associations forming 12 male/female couples, participants were confronted with 4 experimental conditions, requiring the recognition of the learned couples on the basis of the presentation of name-name (NN), face-face (FF), name-face (NF) or face-name (FN) pairs of stimuli. As classically described in this kind of literature (e.g. Giard & Peronnet, 1999), the main analysis consisted in subtracting the non-mixed conditions (NN and FF) from the mixed ones (NF and FN), to isolate the specific associative processes between faces and proper names. The PET-scan study showed that the recognition of face-name associations was associated with the activation of 3 cerebral areas located in the left hemisphere: the left inferior frontal gyrus (Brodman area BA 45), the left median front gyrus (BA 6) and the left inferior

parietal lobe (BA 40). This last region was interpreted as the brain region where the association of the distinct representations of faces and names operates. The ERP study revealed two main ERP components: a negative wave peaking at left parieto-occipital sites around 285 milliseconds (ms) and its positive counterpart maximally recorded at left centro-frontal electrodes around 300 ms.

Thus, these two experiments showed that the retrieval of the representations of face-name associations is sustained by a network of cortical areas lateralized in the left hemisphere and producing cerebral activities spreading from posterior to anterior sites around 300 ms (Campanella et al., 2001; Joassin et al., 2004). A particular emphasis was made on the left inferior parietal cortex, classically considered as a multimodal convergence region (Booth et al., 2002, 2003 ; Niznikiewicz et al., 2000) and interpreted as reflecting the place where the distinct representations of faces and names are bound together.

Nevertheless, both studies raised several questions amongst one at least concerns the specificity of the network of cortical regions involved in face-name associations relative to more general binding processes, such as the association

between objects and common names. One can hypothesize that face-name associations would involve cerebral regions different from those implied in the binding of objects and common names. Indeed, objects and faces have been shown to be processed by distinct brain areas. On the one hand, studies of brain-damaged patients led to a double dissociation between preserved processing of faces / altered processing of objects and conversely. For instance, several cases of prosopagnosic patients have been described, who were unable to recognize familiar people whereas still able to recognize objects (Buxbaum et al., 1999; De Renzi et al., 1991; McNeil & Warrington, 1993). The inverse case, that is, agnosic patients deficient for object recognition but whose face recognition was preserved have also been described (Feindberg et al., 1994; Moscovitch et al., 1997).

On the other hand, several cerebral imaging studies have shown that, within the ventral and dorsal visual pathways, perception and identification of different categories of objects such as faces, animals and houses activate different cortical regions (Damasio et al., 1996; Grabowski et al., 1998; Kanwisher et al., 1997; Mitchell et al., 2002). Nevertheless, there is still a debate to know whether face perception

requires a specific module of processing (Bentin et al., 1996; Kanwisher et al., 1997) or whether their recognition depends on a high level of expertise enabling subjects to differentiate each exemplar of this particular category of visual objects (Gauthier et al., 1999, 2000; for a recent review on this debate, see Grill-Spector, 2003). At the electrophysiological level, this debate focused in particular on the N170 wave. This component is a negative potential peaking around 150 – 170 ms after the stimulus onset on lateral occipito-temporal electrodes, and appears to reflect structural analysis of facial information in order to obtain a configurational face representation (Bentin et al., 1996; Jeffreys, 1996). The observation of a larger amplitude of the N170 in response to faces than to meaningful non-face objects such as birds, cars, or human hands led several scientists to conclude that this component reflects specific visual mechanisms devoted to the processing of human faces (Bentin et al., 1996; Carmel & Bentin, 2002). Other authors challenged this idea by showing that a large N170 can be elicited by other objects than faces (Tanaka & Curran, 2001; Rousselet et al., 2004) and that it could rather depend on the level of expertise of the visual

mechanisms recruited to identify a given exemplar of a category of objects (Gauthier et al., 1999, 2000).

If face processing seems to be related to particular brain mechanisms (in terms of specific domain or of expertise abilities) as compared to other visual objects, the status of proper names relative to common names is less clear, at least from a neuropsychological point of view. It is here necessary to distinguish a production impairment for one or another category of words from a deficit in the access to the semantics related to these words. Concerning the production of proper names, Semenza and Zettin (1988) have described the case of a brain-damaged patient who was anomie for proper names (family names and names of geographical places) but not for common names. It has been hypothesized that this specific trouble could reflect difficulties in learning arbitrary associations (Luchelli & De Renzi, 1992; Hittmair-Delazer et al., 1994; Saetti et al., 1999). Nevertheless, no patient suffering from the reverse pattern of impairment, that is, an altered production of common names with a preserved production of proper names, have up to now been described. On the contrary, such a double dissociation exists for the semantic access to common and proper names as (1) the patient APA,

examined by Caramazza and Shelton (1998), showed a deficit in the semantic access to people who he could not name without any semantic deficit for other objects; (2) Lyons et al. (2002) studied a patient, FH, who was severely impaired for the production of common and geographical names with a semantic impairment for the corresponding entities, while he was still able to name people and recall semantic information about them.

Cerebral imaging studies focused also on the specific cerebral regions supporting the processing of proper and common names. For instance, Damasio et al. (1996) investigated by PET-scan the naming of faces, animals and tools in healthy and brain-damaged participants. In both groups, they showed distinct cerebral activities within the left temporal pole and inferior temporal gyri for the three kinds of objects. In particular, the middle part of the left temporal gyrus seems to play an important role in accessing family names relative to common names (Sergent et al., 1994; Gorno-Tempini et al., 1998; Tsukiura et al., 2002). This temporal preferential activity for family names was also observed with ERP's: Dehaene (1995) showed that celebrity names, relative to animals names and verbs, elicited a left negative inferior

temporal activity between 250 and 280 ms. Proverbio et al. (2001) observed that proper names produced a negative left anterior temporal activity around 150 ms whereas a negative posterior activity around 275 ms was recorded for common names. They also showed that P300 was modulated by the lexical category: it was more ample for proper names, which could reflect the greater complexity in recalling them as they represent unique entities.

On the basis of the literature, the main question of the present study was to know whether associative activities put in light in our previous experiments (Campanella et al. 2001; Joassin et al., 2004) are specific to face-name associations or whether the binding of faces and proper names is sustained by more general associative mechanisms, independent from the category of objects. For this purpose, we performed an ERP experiment in which participants were confronted either with common names / animals associations either with family names / faces associations. Six conditions were proposed (3 for each kind of association): common name alone (C), animal alone (A), simultaneous presentation of an object and its name (AC), family name alone (P), face alone (F), or

simultaneous presentation of a face and his (her) name (FP). As classically described in previous studies (Giard & Peronnet, 1999), this methodology allowed us (1) to perform the subtraction $[AC - (A + C)]$ and $[FP - (F + P)]$ in order to isolate the associative cerebral activities associated with the retrieval of a single and coherent representation of each association, and (2) to compare the results of each subtraction, in order to put in light a possible specificity for face-name associative processes relative to object-name associations.

Methods

Participants.

Thirteen healthy undergraduate volunteers (9 females) took part to this experiment (mean age: 22.5 years, SD: 1.33). All were French native speakers, right-handed and had a normal-to-corrected vision. They gave their written informed consent before the beginning of the experiment.

Stimuli.

Four kinds of stimuli were used: photos of famous faces (actors), their corresponding written family names, photos of animals, and their corresponding written common names.

Sixteen photos of celebrities were selected on internet. People were represented in frontal view with, as far as possible, a neutral emotional expression. Eight celebrities were Americans, the remaining being Europeans (see Appendix). Sixteen photos of animals were also selected on Internet. Half represented complete animals originating from North America, the remaining originating from Europe. All stimuli were presented in black and white on a grey background and all other visual information than the faces or the animals were erased (Figure 1). Each stimulus covered approximatively a surface of 220 pixels height x 130 pixels width (9 x 5 cm, visual angle 8° x 4,6°).

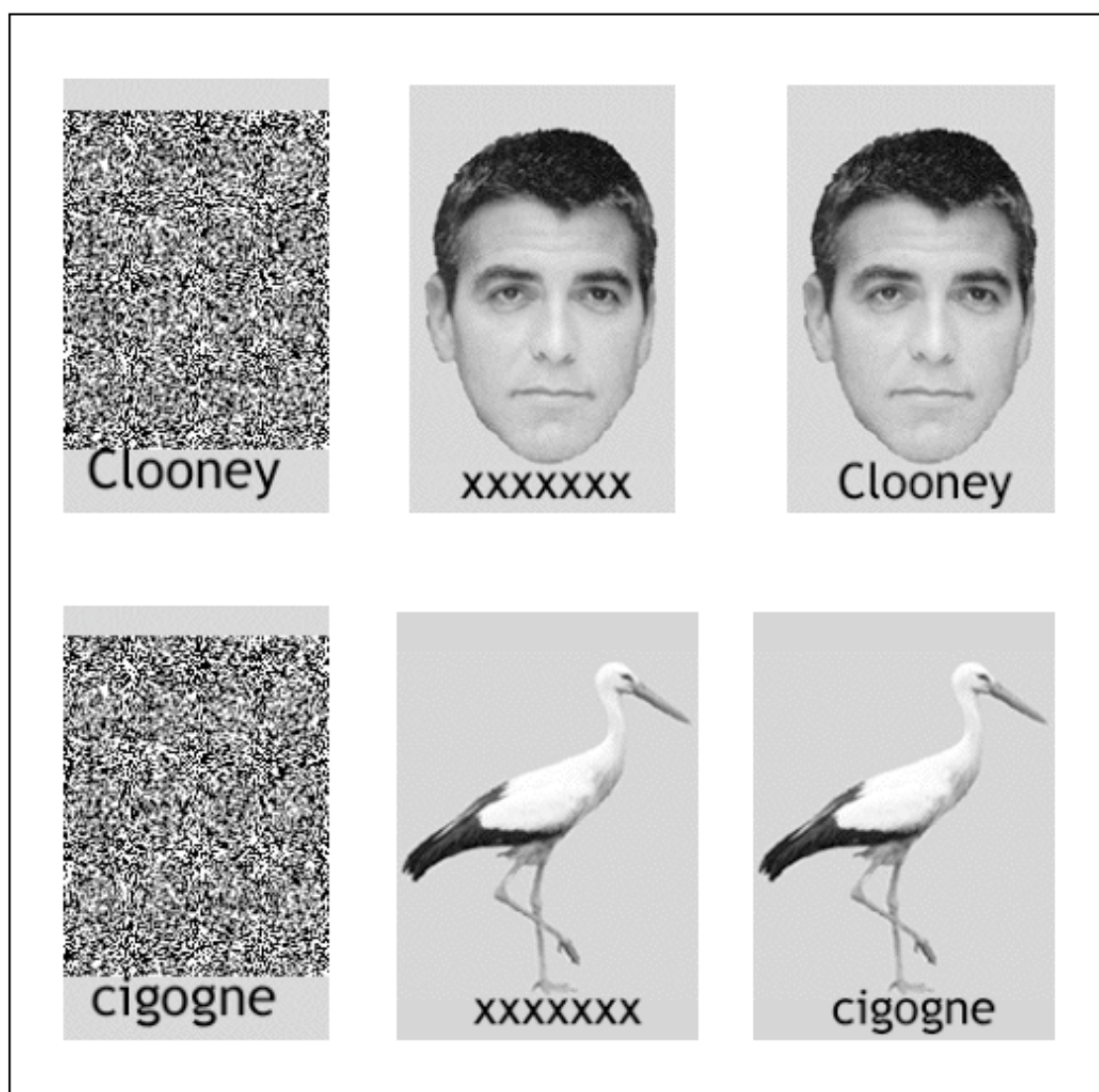


Figure 1.

Examples of stimulations. Upper part (from left) : proper name (P), face (F), face/proper name (FP). Lower part : common name (C), animal (A), animal/common name (AC).

The corresponding proper and common names appeared (alone or with their corresponding face / animal) in black Trebuchet MS 18. When faces and animals were presented alone, their names were replaced by x strings. When names were presented alone, the face or animal was replaced by a

« noise » square with the same physical characteristics than the real stimuli. The names or the x strings appeared always below the face or the animal.

Procedure.

Firstly, participants performed a familiarisation with the 16 faces and the 16 animals. For animals, they were presented with each stimulus along with its name and geographical origin (American or European) and asked to memorize these information in order to be able to retrieve them later. There was no time pressure. After they viewed each stimulus, participants were asked to recall the name and the geographical origin of each animal. Errors were immediately corrected. The same procedure was applied for the 16 faces. The familiarisation was continued until accuracy reached 100 %. The experimental phase began just after.

During the EEG recording session, six conditions were presented to participants, containing either one information (proper names (P), common names (C), faces (F), animals (A)) or two (simultaneous presentation of a face and his (her) family name (FP), simultaneous presentation of an animal and its name (AC)). The task was identical across conditions:

participants had to categorize each stimulus according to its nationality (or geographical origin), by pressing with right fingers one of 2 pre-defined keys (counterbalanced accross participants) on a paddle.

768 trials were presented (128 for each condition), mixed in 24 blocks of 32 trials. Twelve blocks contained only human stimuli and 12 blocks only animal stimuli. Animal and human blocks were mixed within the experiment and their order was balanced across participants. Within each block, the three kinds of conditions (1/3 of names, 1/3 of figures and 1/3 of associations) were mixed. Each trial began with a fixation cross appearing at the center of the screen for 300 ms and replaced by an empty interval of 500 ms. Then the stimulus was presented for 700 ms and was followed by an empty intertrial interval of 1500 ms.

Participants were installed in a dark and quiet room with their head placed at 50 cm of the screen. They were given written instructions with a particular emphasis on accuracy and speed. A training block was performed as a familiarization with the task.

EEG was recorded by 32 electrodes mounted in an electrode Quick-Cap. Electrode positions included the

standard 10-20 system locations and intermediate positions. Recordings were made with a linked mastoid physical reference. The EEG was amplified by battery-operated A.N.T.® amplifiers with a gain of 30 000 and a band-pass of 0.01 - 30 Hz. The impedance of all electrodes was kept below 20 k Ω . EEG was continuously recorded (sampling rate 512 Hz, Eeprobe, A.N.T.®) and trials contaminated by EOG artifacts (observed on the prefrontal electrodes) were manually eliminated off-line. A recalculation was made to obtain common average reference recordings (Bertrand, Perrin et Perrier, 1985). Epochs beginning 200 ms prior to stimulus onset and continuing for 1500 ms were created. Codes synchronized with stimulus delivery were used to average selectively the epochs associated with different stimulus types. Two parameters were coded for every stimulus: (1) the type of the stimulus (P, C, F, A, FP, AC), (2) the type of trials (American or European). This coding allowed us to compute different averages of ERP target stimuli. The averages were computed for each subject individually. Only trials that were correctly performed were included in these averages. Then, grand-averaging were computed for each of the 6 conditions.

Note that correct european and american trials were averaged together.

Results

Behavioral data.

Percentages and latencies of correct responses can be seen in Table 1. An ANOVA with the categories of objects (people or animals) and the kind of conditions (names alone, figures alone, associations of names and figures) as within factors, was performed on latencies of correct responses. Please note that all and only significant effects ($p \leq .05$) will be described throughout the results section.

Table1

Mean latencies (in ms) and percentages of correct responses (standard deviation between parentheses) for the 6 conditions: F = faces condition, P = proper names condition, FP = faces/proper names condition, A = animals condition, C = common names condition, AC = animals/common names condition.

	F	P	FP	A	C	AC
Latencies	675,4 (71,57)	731,1 (84,72)	658,1 (76,55)	766,2 (117,6)	797,5 (107,48)	720,9 (103,11)
Percentages	93,1 (7,33)	96,8 (3,85)	96,6 (2,61)	89 (5,49)	88,8 (7,03)	92,2 (4,75)

The 2x3 ANOVA revealed a significant main effect of the category of objects ($F(1,12) = 33.25, p < .001$), indicating that animals stimuli were categorized significantly more slowly than people stimuli. It also revealed a significant main effect of the kind of conditions ($F(2,24) = 67.9, p < .001$) and a significant interaction between the category of objects and the kind of conditions ($F(2,24) = 4.99, p < .02$). Subsequent paired Student t-tests comparing the latencies of each condition in each category of object, showed that (1) the associative conditions (FP and AC) were performed more quickly than the other conditions (FP vs. F: $t(12) = 8.94,$

$p < .001$, FP vs. P: $t(12) = 3.67$, $p < .001$; AC vs. C: $t(12) = 13.16$, $p < .001$, AC vs. A: $t(12) = 4.99$, $p < .001$) (2) participants answered more quickly for animals than for common names ($t(12) = -3.84$, $p < .001$), and for faces than for proper names ($t(12) = -5.13$, $p < .001$).

The same ANOVA was performed on percentages of correct responses. It also showed a significant main effect of the category of objects ($F(1,12) = 22.69$, $p < .001$), indicating that A, C and AC conditions were less well performed than F, P and FP conditions. It failed to show any significant main effect of the kind of conditions ($F(2,24) = 2.92$, ns) but showed a significant interaction between the category of objects and the kind of conditions ($F(2,24) = 4.99$, $p < .05$). Student paired t-tests were also carried out. For the category of people, they failed to show any significant differences between conditions. For the category of animals, they showed that the associative condition was better performed than the common names condition ($t(12) = -2.68$, $p < .02$) and the animals condition ($t(12) = -4.98$, $p < .001$), these last two conditions not differing from each other. No speed-accuracy tradeoff was found.

Electrophysiological data.

Firstly, we analyzed the classical ERP components associated with the processing of visual objects, in order to examine the effect on the cerebral activities of the effect of observed in the behavioral section. Within the same conditions of each category of objects (i.e., faces vs. animals, proper names vs. common names, faces/proper names, vs. animals/common names), we compared the latencies and peak voltage amplitudes of four electrophysiological components: the P1 (O1, Oz, O2) / N1 (C3, Cz, C4) complex, the N170 (T5 and T6) / VPP (C3, Cz, C4) complex, the N250 (C3, Cz, C4) and the P300 (P3, PZ, P4). Mean amplitudes and latencies of these ERP waves can be seen in Table 2.

Table 2

Mean latencies in ms (bold) and amplitudes in μV of P1, N1, N170 (or N2), VPP (or P2), N250 and P300 components for the 6 conditions at midline (z electrodes), left (odd electrodes) and right (even electrodes) hemispheres. F = faces, P = proper names, FP = associations of faces and proper names, A = animals, C = common names, AC = associations of animals and common names.

		F	P	FP	A	C	AC
Component	Electrode						
P1	O1	122	118	124	125	121	126
		6,04	5,46	6,66	6,17	5,63	6,17
	Oz	125	117	126	124	122	127
		5,58	4,52	6,05	5,15	4,53	5,47
	O2	123	118	126	123	120	124
		6,51	6,18	6,18	5,92	6,19	5,95
N1	C3	123	122	126	128	118	123
		-2,08	-2,25	-2,12	-1,95	-2,22	-2,23
	C0z	124	127	127	122	120	126
		-3,18	-3,15	-3,32	-2,94	-3,01	-3,15
	C4	128	110	119	126	123	123
		-2,05	-1,98	-2,17	-1,96	-1,95	-2,08
N170 / N2	T5	180	188	182	187	186	189
		-6,75	-3,12	-6,27	-4,78	-2,54	-4,52
	T6	180	190	181	189	177	185
		-6,78	-3,37	-6,87	-4,55	-3,51	-4,95
	C3	179	188	181	189	194	185
		2,77	3,04	2,15	2,01	2,32	1,82
VPP / P2	Cz	179	175	179	189	181	184
		3,47	2,47	3,09	2,53	2,38	2,77
	C4	173	177	157	188	180	189
		2,69	1,46	2,62	1,89	1,38	1,7
	C3	280	298	292	314	291	295
		-3,83	-3,65	-3,72	-3,73	-3,72	-3,85
N250	Cz	284	272	281	292	271	295
		-5,41	-4,95	-5,14	-5,42	-4,96	-5,43
	C4	279	266	271	292	274	283
		-2,38	-2,26	-2,22	-1,82	-2,18	-1,95
	P3	588	511	522	527	521	534
		6,16	6,6	6,46	7,08	6,08	7,82
P300	Pz	556	505	524	544	519	538
		7,87	8,22	8,37	8,46	7,63	9,61
	P4	495	395	483	521	435	512
		7,31	7,83	7,76	7,93	7,83	8,59

For each comparison and each component separately, an ANOVA with the category of objects (animals, people) and electrodes position (left, central, right) as within factors was computed for latency and amplitude separately. The comparison of A and F conditions revealed a significant main effect of the category of objects on the latency and the amplitude of the N170 (latency: $F(1,12) = 25.38, p < .001$; amplitude: $F(1,12) = 8.04, p < .02$) and the VPP (latency: $F(1,12) = 15.79, p < .001$; amplitude: $F(1,12) = 12.88, p < .001$). The N250 was larger in the left hemisphere for both kinds of stimuli ($F(2,24) = 30.44, p < .001$). The P300 was also larger in the animal condition than in the face condition ($F(1,12) = 6.81, p < .02$) whereas it peaked at the same latency in both conditions (Figure 2). An effect of lateralization was also observed on the amplitude of the P300 ($F(2,24) = 3.55, p < .04$), indicating that the wave was more ample on the central parietal electrode. No effect was observed on the P1/N1 complex.

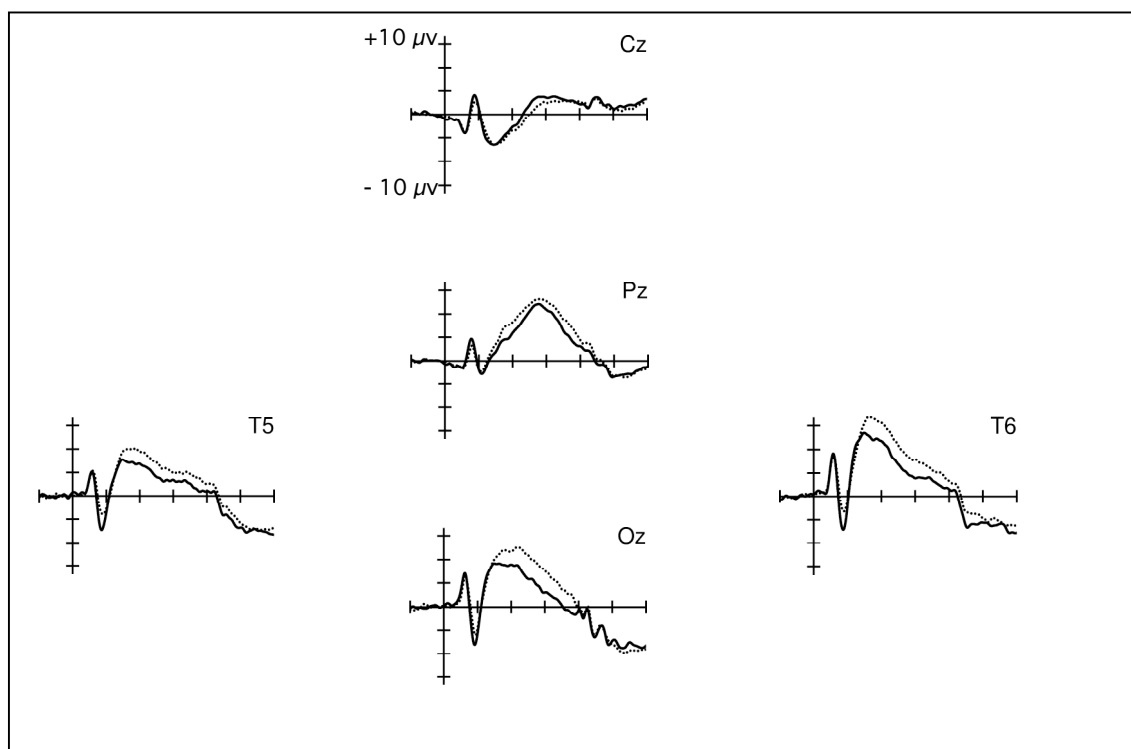


Figure 2.

Grand-average ERPs of Animals (dash lines) and Faces (solid lines) conditions on occipital (Oz), temporal (left : T5, right : T6), parietal (Pz) and central (Cz) electrodes.

In the comparison of P and C conditions, we observed that the P1 was significantly larger in the right hemisphere for both types of names ($F(2,24) = 4.95, p < .02$) whereas the N1 was significantly larger on the central and left electrodes ($F(2,24) = 18.37, p < .001$). The N250 was also significantly left-lateralized in both conditions ($F(2,24) = 23.6, p < .001$). The amplitude of the P300 was significantly larger for the common names than the proper names ($F(1,12) = 4.63, p < .05$) and was larger on Pz and P4 than in P3 ($F(2,24) = 4.08, p < .03$) (Figure 3).

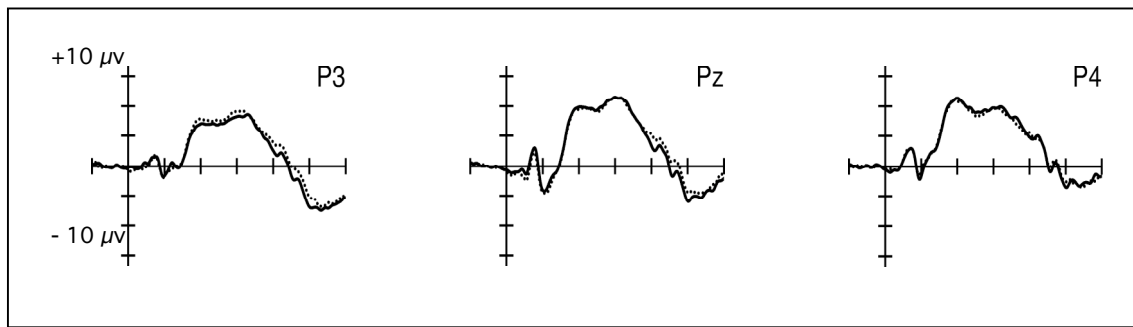


Figure 3.

Grand-average ERPs of Common Names (dash lines) and Proper Names (solid lines) conditions on parietal electrodes (left hemisphere : P3, middle : Pz, right hemisphere : P4).

Finally, the comparison of FP and AC conditions showed that, as for faces and animals, the N170 appeared significantly faster and with a greater amplitude in FP condition (N170 latency: $F(1,12) = 19.11$, $p < .001$; N170 amplitude: $F(1,12) = 12.24$, $p < .001$), and that the VPP had a shorter latency in FP than in AC ($F(1,12) = 6.21$, $p < .03$). The N250 was again significantly left-lateralized in both conditions ($F(2,24) = 30.28$, $p < .001$), and the P300 was significantly larger in AC condition ($F(1,12) = 9.32$, $p < .01$) and right-lateralized in both conditions ($F(2,24) = 4.57$, $p < .02$) (Figure 4).

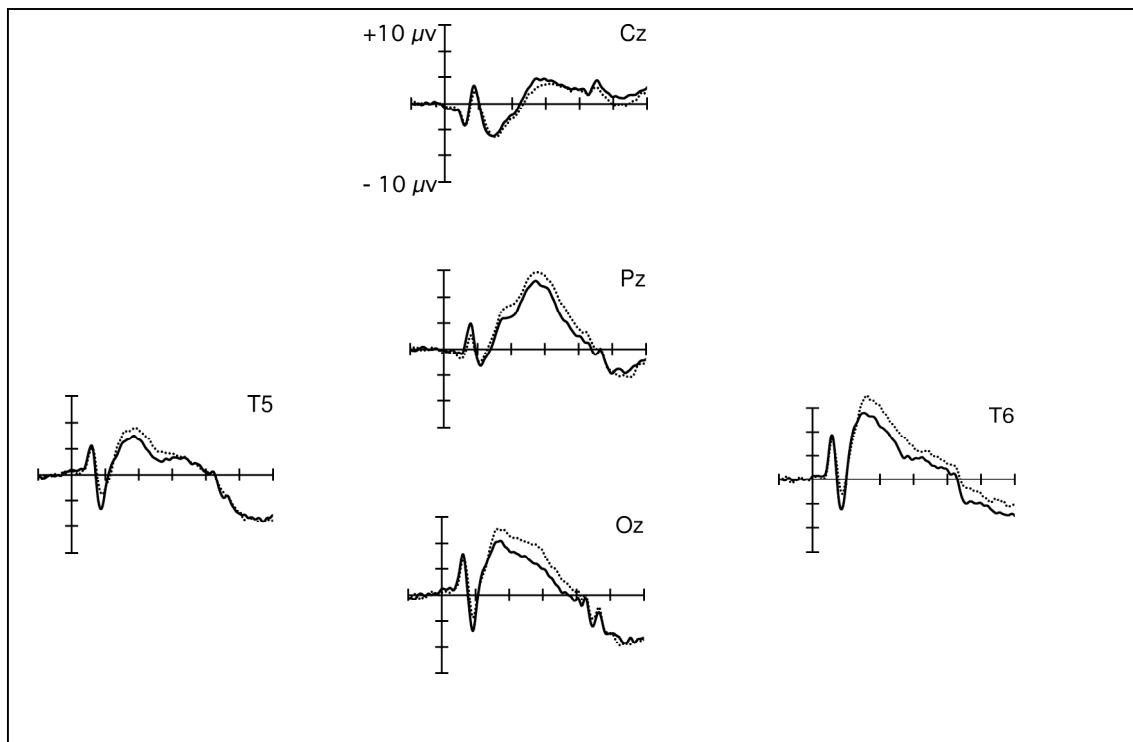


Figure 4.

Grand-average ERPs of Animals/Common Names (dash lines) and Faces/Proper Names (solid lines) conditions on occipital (Oz), temporal (left : T5, right : T6), parietal (Pz) and central (Cz) electrodes.

The main analyses of this study consisted in subtracting, for each category of objects, the visual and names conditions from the associative conditions. Subtractions $FP - (F + P)$ and $AC - (A + C)$ were computed for each participant and were then grand-averaged. Significant effects were computed on the grand-averages at each selected electrode by using Student t-test (amplitude of the difference waves compared to zero). Spatio-temporal patterns whose t-values were above 3.11 or under -3.11 ($p < .01$) for at least 20 ms were considered as significant (Rugg, Doyle & Wells, 1995).

The subtraction $FP - (F + P)$ revealed a posterior negative (O1, O2, T5 and T6) / central positive (C3, Cz) component significantly different from zero between 70 and 130 ms, a second posterior negative (O1, Oz, O2, T5 and T6) / fronto-central positive (C3, Cz, F3, Fz, F4) component whose t-values were significant between 240 and 880 ms. It also revealed a significant negativity on T4 between 220 and 290 ms, a significant positivity on T3 between 340 and 430 ms, and a significant parietal positivity (Pz) / negativity (P3, Pz, P4) respectively between 190 and 230 ms, and 320 and 860 ms (Table 3 and Figure 5).

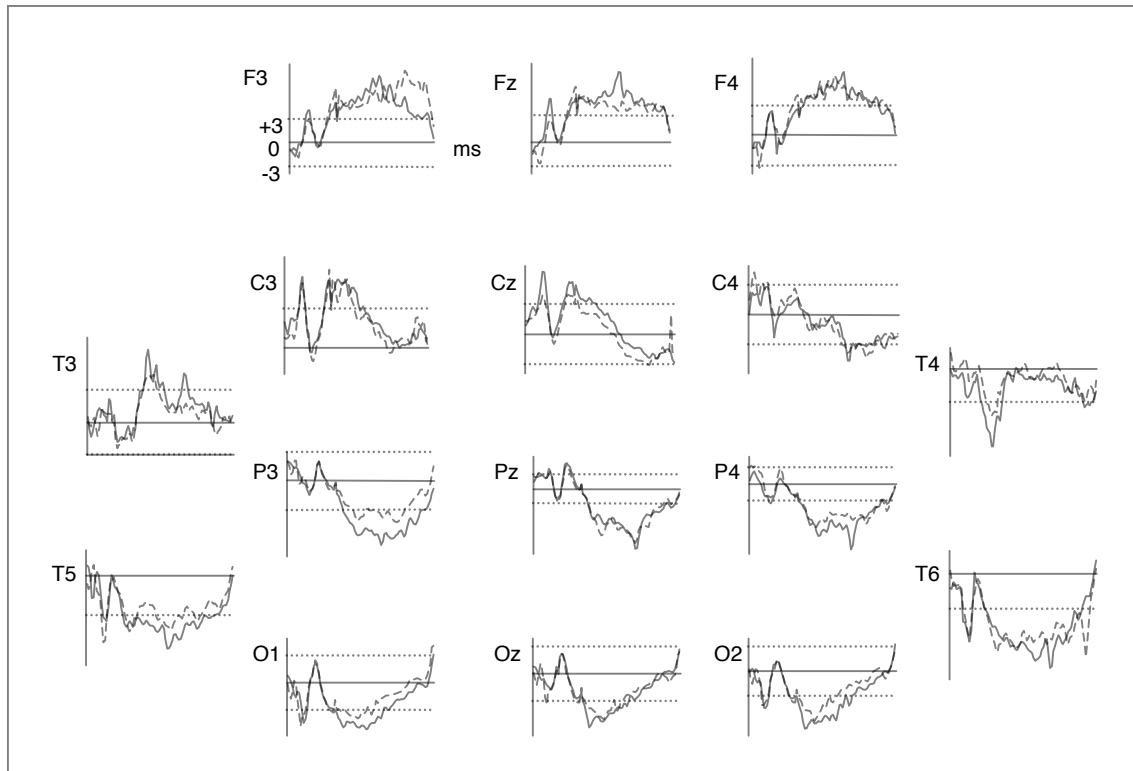


Figure 5.

Representation of the t-values for the subtractions $FP - (F + P)$ (solid lines) and $AC - (A + C)$ (dash lines). X-axis : t-values; Y-axis : time in ms. The statistical threshold ($t = \pm 3.11$, $p < .01$) are represented by red lines.

Table 3

Significant activities elicited by the subtractions AC – (A + C) and FP – (F + P). Latencies = intervals in ms during which t-values were above 3.11 ($p < 0.01$) or under -3.11. Max t-values = maximum t-value for each interval (latency in parentheses).

Electrode	AC – (A + C)		FP – (F + P)	
	latencies	Max t-values	latencies	Max t-values
O1	90-120	4,56 (100)	70-120	4,92 (80)
	290-590	5,09 (340)	260-490	4,05 (430)
Oz	300-580	6,06 (340)	260-540	5,12 (430)
O2	90-130	4,65 (100)	80-120	4,25 (90)
	250-630	7,16 (340)	260-540	6,18 (400)
C3	100-130	-5,44 (110)	90-130	4,98 (110)
	250-470	-5,24 (380)	250-440	-5,27 (390)
Cz	80-130	-6,34 (100)	90-120	-3,96 (100)
	230-420	-5,29 (260)	240-320	-4,02 (250)
C4	580-720	4,72 (600)	560-760	4,42 (700)
F3	100-130	-4,11 (120)		
	240-870	-8,5 (550)	230-880	-9,22 (710)
Fz	90-140	-5,47 (110)		
	230-880	-8,92 (560)	220-880	-7,07 (280)
F4	320-730	-6,58 (570)	250-840	-6,25 (510)
T5	100-130	3,54 (120)	90-130	5,21 (100)
	220-780	5,69 (490)	240-310	4,31 (260)
			420-700	4,22 (510)
T6	80-140	5,89 (120)	90-140	5,52 (120)
	220-810	8,14 (610)	220-860	6,72 (290)
T3	350-480	-6,81 (370)	340-430	-4,48 (370)
T4	190-320	7,16 (250)	220-290	4,3 (250)
P3	340-850	6,78 (570)	400-720	4,38 (670)
Pz	190-240	-5,57 (200)	190-230	-5,21 (200)
	330-870	12,93 (620)	340-800	11,13 (610)
P4	300-860	12,13 (620)	320-860	8,09 (400)

The subtraction $AC - (A + C)$ showed the same general pattern of activities, i.e. a posterior negativity (O1, O2, T5, T6) / fronto-central positivity (C3, Cz, F3, Fz) statistically significant between 90 and 140 ms, a second posterior negative (O1, Oz, O2, T5, T6) / fronto-central positive (C3, Cz, F3, Fz, F4) component significant between 250 and 880 ms, a significant positivity on T3 between 350 and 480 ms, a significant negativity on T4 between 190 and 320 ms, and a significant positive (Pz) / negative (P3, Pz, P4) wave from 190 to 240 ms and from 300 to 870 ms. Nevertheless, the first positive activity between 90 and 140 ms was here more frontally distributed whereas we did not observe any significant frontal activity in the former subtraction (Table 3 and Figure 6).

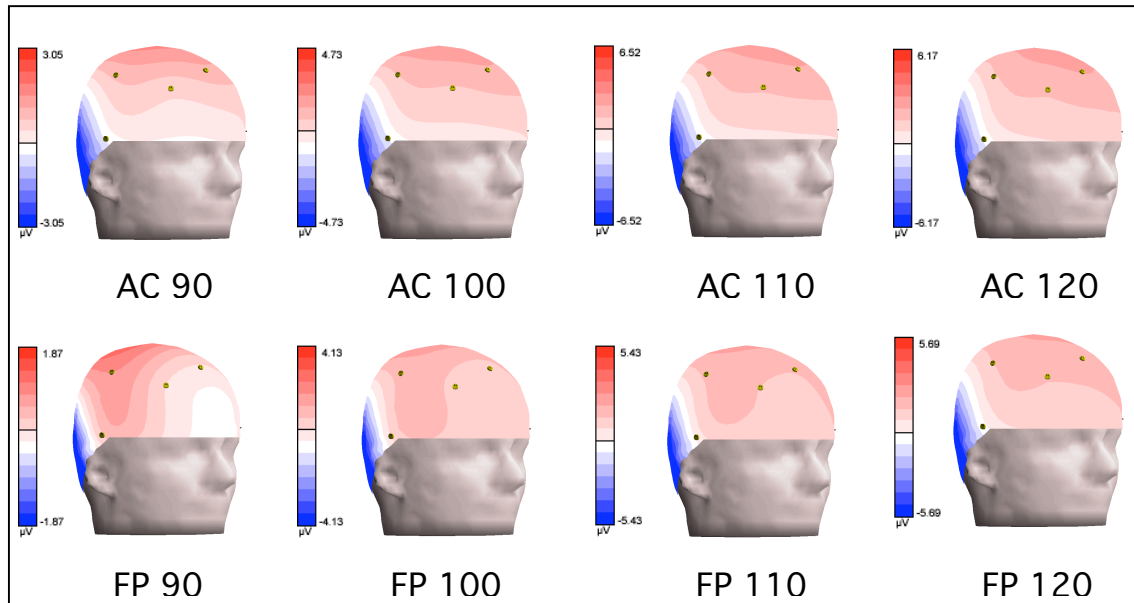


Figure 6.

Topographies of the subtractions AC - (A + C) (upper part) and FP - (F + P) (lower part), illustrating the more frontal activity for AC between 90 and 120 ms relative to the more central activity for FP.

Ten ANOVAs with kind of subtraction (2) and electrode position (3: left, middle, right) as within factors were computed on peak amplitudes and latencies of each significant activity. A significant main effect of the kind of subtraction appeared only once, for the latency of the parietal negativity ($F(1,11) = 6.91, p = .02$), indicating that this activity appeared earlier for FP - (F + P) than for AC - (A + C). However, four significant effects of electrode position emerged: amplitude of the first occipital negativity was significantly larger in O1 and O2 than in Oz ($F(2,24) = 6.06, p = .01$), amplitudes of both negativities were significantly

larger on the right temporal electrode (T6) than on the left temporal one (T5; first negativity: $F(1,11) = 13.99$, $p=0.001$; second negativity: $F(1,11) = 6.66$, $p=.03$), and amplitude of the frontal positivity was significantly lateralized in the left hemisphere ($F(2,22) = 7.48$, $p=.001$). No significant interaction between the kind of subtraction and the electrode position was observed.

Discussion

This ERP experiment was aimed at examining whether the cerebral mechanisms that allow us to retrieve in memory unified representations of people are specific to face-name associations or whether they can be involved in more general associative processes. For this purpose, we compared the cerebral electrophysiological activities of healthy participants asked to semantically categorize either faces / proper names associations or animals / common names associations as American or European.

Firstly, at the behavioral level, the conditions relative to animals (A, C and AC) were less well and more slowly performed than the conditions relative to people (F, P and FP),

indicating that the former ones were more difficult to perform than the last ones. This complexity effect could be explained by the semantic task used. We needed to have identical tasks throughout the different conditions to ensure the validity of the comparisons between conditions, but while the nationality is one of the major semantic information about people, the geographical origin could not be a crucial piece of information used to define animals. In other words, it could be more common to categorize people according to their nationality rather than categorizing animals according to their geographical origin.

Another interpretation relies on the differences of expertise levels between faces and non-faces stimuli. It has been claimed that humans are experts in face processing and that faces, more than other visual objects, are processed holistically, that is in using relatively less part decomposition than other types of objects (Tanaka & Farah, 1993; Farah et al., 1998). This holistic processing of faces could lead to the observed differences in our experiment as in this view, less time is needed to process faces than animals.

In spite of these qualifications, let us recall that the learning stage was achieved when accuracy reached 100% for

both materials. In addition, it is worth noting that these behavioral differences did not influence the early ERP components associated with the visual processing of the stimuli, while several electrophysiological studies have shown a modulation of the P1/N1 complex according to attention or discrimination difficulties (Senkowski & Herrmann, 2002; Di Russo & Spinelli, 1999; Johnson & Olhausen, 2005). We did not observe any differences between the processing of faces and animals on the P1/N1 component, reflecting the early striate and extrastriate visual processing (Gomez et al., 1994; Heinze et al., 1994; Campanella et al., 2002). These findings ensure us that the behavioral differences observed were not due to a difference of perceptual processes between the two kinds of visual objects, but rather to more difficult later decisional processes.

In support of that, ERP differences were observed on the P300. Indeed, the P300 was larger in the three animals conditions. This ERP wave is classically interpreted as reflecting the decision and closure processes (Polich & Kok, 1995), and it has been suggested that its amplitude increases with task difficulty (Johnson Jr, 1988; Johnson and Olhausen, 2005). The difference of amplitude we observed could be due

to the fact that animal conditions were more difficult to perform than the people conditions, requiring more elaborated decisional processes that caused a larger P300.

ERP differences were also observed on the N170/VPP (or N2/P2) complex. The N170/VPP is thought to reflect early structural encoding before face recognition take place (Eimer, 2000; Sagiv & Bentin, 2001). The present results are in accordance with previous findings showing that the N170 appears faster and is considerably larger in response to human faces than to non-face objects such as birds, human hands, cars, flowers or butterflies (Bentin et al., 1996; George et al., 1996; Allison et al., 1999; McCarthy et al., 1999).

A N250 wave, maximally peaking at left central electrodes between 270 and 300 ms, was also elicited by both human and animal stimuli. This component could be related to the N400f (Bentin & Deouell, 2000), corresponding to the recognition or identification of individual faces. Schweinberger et al. (2002) recently suggested that this wave, observed in a repetition priming paradigm on famous faces, could depend on the access to stored representations of familiar faces. A dipole source location also suggested that, while the N170 originated in posterior lateral temporal areas,

the N250 source was located in a more anterior area of the ventral temporal lobe corresponding to the fusiform gyrus. We are aware that the present results are not sufficient to claim that people and animals share a common semantic store but, as we did not observe any significant difference of latency or amplitude between conditions for this component, we can hypothesize that the access to semantics take place at the same moment for animals and people.

The main purpose of this study was to investigate the associative processes involved in the retrieval of face/proper name or animal/common name associations. At first, behavioral data revealed that both associative conditions (FP and AC), in which two pieces of information were simultaneously presented to participants, led to facilitation: they were performed more rapidly (and better in the case of animals) than the conditions in which only one kind of information was presented. This phenomenon can be referred to the spreading-activation theory in semantic memory (Collins & Loftus, 1975), in which a concept will be activated faster when it receives activation from two related representations than from only one, due to additivity of

activations. This facilitation ensured that both stimuli were effectively processed in the associative conditions.

Then, the electrophysiological correlates of these associative processes were examined by the computation of two subtractions aimed at isolating the cerebral activities specifically elicited by the associations of faces and proper names, and of animals and common names. Both subtractions revealed several cerebral activities.

Firstly, we observed two complexes of posterior negativities / anterior positivities around 115 ms, and between 300 and 400 ms. Such posterior negative / central positive complexes have already been observed by Joassin et al. (2004) in a recent ERP study investigating the electrophysiological correlates of cross-modal interactions between human faces and voices. These activities, around 110 ms, were hypothesized to reflect the activity of the visual brain regions, and notably the Fusiform Face Area (FFA) dedicated to the processing of human faces. The two posterior negative / central positive waves observed in the present study could also reflect the activation of the associative visual areas involved in the processing of visual objects. Nevertheless, we observed here a more extended

activity of the associative visual areas (until 300 ms), probably due to the fact that visual stimulations rather than auditory-visual stimulations were used.

Secondly, associative processes were correlated with a right temporal negativity around 220 and 290 ms, and a left temporal positivity around 400 ms. Several studies of brain-damaged patients suffering from progressive impairments in familiar people recognition have shown that these defects were associated with temporal lobes atrophy (Gentileschi et al., 1999; Miceli et al., 2000; Gainotti et al., 2003). In particular, Gainotti et al. (2003) hypothesized that this deficit could be due to impaired associative processes between the different modes of recognition of people and that the temporal lobes could be seen as convergence regions supporting the integration of the different information relative to individuals, such as faces, names and voices. Moreover, several neuroimaging studies showing an involvement of the right superior temporal sulcus (STS) in auditory-visual speech perception (Calvert et al., 2001) or speaker recognition (Von Kriegstein et al., 2005) support this hypothesis. The temporal activities observed in the present study are in line with this interpretation and could reflect the activation of areas

dedicated to the association of the different features related to individuals and objects.

Alternatively, they could also reflect the activation of the semantic system for people and objects. Several authors have suggested that semantic representations of objects are distributed across left and right temporal regions, with stronger connections between the semantic units and phonological representations in the left hemisphere (Simons et al., 2001; Lambon Ralph et al., 2001; Kellenbach et al., 2005). For faces, there are also some evidence that both temporal lobes are involved in person-semantic representations and that the contributions of each hemisphere are asymmetrical. The left hemisphere would have a preferential access to names while the right hemisphere would have a preferential access to faces (Eslinger et al., 1996; Giovanello et al., 2003). It is also well-known that the left temporal pole is closely involved in the retrieval of words for objects belonging to various conceptual categories (Grabowski et al., 2001), with a particular involvement in recalling words denoting unique entities (Gorno-Tempini et al., 1998). Our study does not allow us choose one hypothesis or the other. Further studies, using different tasks (for

instance the direct comparison of a semantic vs. a purely perceptual or phonological task), could be useful to better understand the implication of the temporal regions in the associative processes.

Finally, we observed that associations between animals and common names produced a more frontally distributed activity around 120 ms compared to associations between faces and proper names. Graboswki et al. (1998) showed that retrieval of words belonging to different conceptual categories involved distinct patterns of activation in the left frontal lobe, and particularly the activation of the left middle frontal gyrus, that could reflect some working memory processes (Petrides et al., 2002 ; Owen et al., 1996). Behavioral data showed that faces/proper names associations were easier to categorize than animals/common names associations. It is then possible that this more frontal activity for animals associations reflect these differences of complexity between the two categories of objects: if participants were less familiarized to categorize the animals according to their geographic origin, they could have used more working memory resources to maintain the association active until the decision was made.

On the whole, these results are in accordance with the theory of interactions between unimodal sensory regions and heteromodal integrative regions developed by Driver & Spence (2000). In this model, multimodal convergence regions receive inputs from unimodal regions, integrate them into a coherent representation and then send back feedback projections to unimodal regions in order to allow the sensory regions to modulate their activities according to the other stimulations from the environment. In the present case, the electrophysiological activities revealed by the subtractions could reflect the activation of both “unimodal” regions, that is regions processing particular aspects of each stimulation, such as the associative visual areas or the cerebral regions involved in the phonological processing of common and proper names, and “multimodal” convergence regions, such as the temporal cortex, that creates unified representations of people and objects. Moreover, Contrary to Giard & Peronnet (1999) or Molholm et al. (2002), we did not observe any early associative activities (around 50 ms). The associative cerebral activities elicited in the present study spread in the time window of the ERPs dedicated to the processing of the stimuli, from 90 to 400 ms. This observation reinforces the hypothesis

of parallel activations of uni- and hetero-modal regions. The use of more precise anatomo-functional neuroimaging techniques like functional Magnetic Resonance Imaging (fMRI), could be useful to better investigate this question and confirm the activation of both sensory and multimodal convergence regions in the binding processes between the objects and their names.

Nevertheless, the slight differences observed between the two kinds of associations can not lead us to firmly conclude that the associative processes implied in the recollection of the unified representations of individuals are distinct from those implied in the binding mechanisms of the other objects of our environment. One possible explanation is that both associative processes required the same general binding mechanisms. Alternatively, this lack of major differences could be due to the fact that we choose animals as objects to be compared to human people. This choice was guided by previous neuroimaging experiments and neuropsychological case studies showing distinct cerebral activities for living vs. non-living objects (Leube et al., 2001; Caramazza & Shelton, 1998), or natural vs. hand-made objects (for a review, see Joseph, 2001). In sight of these observations,

animals, that is living / natural objects seemed to be the closest non-human objects to be compared with human faces, in order to avoid that any differences could be simply explained by the living / non-living or natural / hand-made features. It would be interesting to apply the present experimental paradigm with other kinds of objects like hand-made objects, with the hypothesis that the associative processes between natural and artificial objects and their names could involve different brain regions, according to the sensory / functional theory (Camarazza & Shelton, 1998).

In conclusion, this study showed that the associative processes between faces and proper names in the one hand, and animals and commons names in the other hand, are not simply the sum of the cerebral activities involved in the processing of each stimulation. Indeed, they specifically produced two negative posterior / positive anterior complexes, suggesting an enhanced activity of the visual areas, and specific temporal activities that could reflect the semantic processing of the associations or the associative processes between visual and linguistic materials *per se*. We also showed that these associative processes differ according

to the kind of associations, with a more frontal cerebral activity for the animals / common names associations.

Aknowledgments

Frederic Joassin is a Research Assistant at the National Fund for Scientific Research (F.N.R.S., Belgium).

Appendix

List of the 16 celebrities and the 16 animals used in the experiment.

People		animals	
American	European	American	European
Georges Clooney	Benoît Poelvoorde	Bald eagle (Haliaeetus leucoc.)	White stork (ciconia ciconia)
Michael Douglas	Thierry Lhermitte	American bison (bison bison)	Chamois (rupicapra rupicapra)
Clint Eastwood	Gerard Jugnot	Beaver (castor canadensis)	Mole (talpa europaea)
Sylvester Stallone	Gerard Depardieu	Grizzly bear (Ursus arctos horrib.)	Boar (sus scrofa L.)
Jodie Foster	Nathalie Baye	Caribou (rangifer tarandus)	European stag (cervus elaphus)
Julia Roberts	Catherine Deneuve	Cougar (felis concolor)	European lynx (lynx lynx)
Sharon Stone	Sophie Marceau	Coyote (canis latrans)	Fox (vulpes vulpes)
Cameron Diaz	Audrey Tautou	Raccoon (procyon lotor)	Hedgedog (ericaneus europaeus)

Introduction aux études 3 et 4

Les deux études décrites précédemment ont toutes deux utilisé du matériel uniquement visuel (visages ou objets et noms écrits). Néanmoins, la modalité visuelle n'est pas la seule modalité pertinente nous permettant d'identifier les personnes familières. Ainsi, nous avons vu dans le troisième chapitre introductif que la voix est un canal privilégié d'identification, à tel point qu'elle peut être considérée comme un "visage auditif" nous permettant de reconnaître les individus (Belin et al., 2004). Il semble d'ailleurs qu'à l'instar du gyrus fusiforme pour les visages, certaines régions du lobe temporal de l'hémisphère droit, et plus particulièrement la partie antérieure du STS, jouent un rôle majeur dans l'identification des voix humaines (Nakamura et al., 2001; Von Kriegstein et al., 2003; Belin & Zatorre, 2003).

Les visages et les voix, perçus simultanément, ne sont pas traités de manière isolée mais interagissent l'un avec l'autre et affectent les processus de traitement de chacune des modalités. L'effet McGurk, consistant en une perception biaisée sur base de la présentation simultanée de mouvements de lèvres et de syllabes auditives incongruents

(McGurk & McDonald, 1976), en est un exemple flagrant. Les études d'imagerie cérébrale fonctionnelle examinant les corrélats cérébraux de cet effet ont mis en évidence une activation des régions unisensorielles visuelles et auditives, ainsi que l'activation du STS gauche, interprété comme étant une région d'intégration multimodale où les représentations visuelles et auditives des visages et des voix perçues sont intégrées en vue d'en former un percept unifié (Calvert et al., 2000; Sekiyama et al., 2003).

La manière dont les visages et les voix interagissent au niveau de l'identification des personnes a été bien moins étudiée que leurs interactions perceptives. Shah et al. (2001) ont comparé par IRMf les activations cérébrales liées à la reconnaissance de visages et de voix. Ils ont observé que, quel que soit la modalité de présentation des stimuli, la familiarité produisait une activation du cortex cingulaire postérieur, incluant le cortex rétrospinal. Toutefois, ils n'ont jamais examiné les interactions crossmodales directes entre visages et voix, chaque stimulation étant présentée de manière isolée.

C'est ce que feront également Von Kriegstein et al. (2005) dans une étude IRMf comparant la reconnaissance de voix très familières à la reconnaissance de voix peu familières.

Leurs résultats indiquent que la reconnaissance des voix familières était associée à une activation des aires temporales auditives ainsi que des gyri fusiformes, de manière plus marquée à droite. Cette activation du gyrus fusiforme dans une région correspondant à la FFA a amené les auteurs à faire l'hypothèse que les interactions visuo-auditives entre visages et voix humaines ne nécessitent pas d'activation de régions hétéromodales d'intégration mais peuvent se faire via un transfert direct des informations entre régions unisensorielles visuelles et auditives.

Néanmoins, comme nous l'avons déjà signalé, aucune de ces études n'a examiné de manière directe les interactions crossmodales entre visages et voix humaines. Les deux études décrites ci-dessous, utilisant la méthode soustractive déjà détaillée dans les études précédentes, auront pour but de déterminer quelles activités cérébrales sous-tendent spécifiquement les interactions crossmodales entre visages et voix impliquées dans l'identification des personnes. Elles tenteront de déterminer si les processus associatifs mis en lumière dans l'étude de Campanella et al. (2001) et dans la première étude de cette thèse, sont amodaux, c'est-à-dire

sont identiques quelles que soient les modalités sensorielles dans lesquelles sont présentées les stimulations, ou sont spécifiques aux associations présentées dans la modalité visuelle.

La première de ces deux études utilisera la technique des potentiels évoqués. Lors d'un pré-test comportemental, qui ne sera pas détaillé ici, nous avons pu observer que, contrairement aux études ayant utilisé des stimulations simples et sans signification (Calvert et al., 1999; Teder-Sälejärvi et al., 2002), la présentation simultanée de visages et de voix n'entraînait pas d'effet de facilitation crossmodale: la catégorisation selon l'identité des associations présentées était significativement plus lente que la catégorisation des visages présentés isolément. Plutôt qu'une facilitation, c'est une interférence des voix sur la reconnaissance des visages qui a donc été mise en lumière.

Il a été montré précédemment que les voix sont plus difficiles à identifier que les visages (Ellis et al., 1997; Schweinberger et al., 1997), et il est possible que ce soit ces différences de complexité de reconnaissance qui soient responsables de l'effet d'interférence observé. Néanmoins, qu'il s'agisse de facilitation ou d'interférence, ces effets

témoignent des interactions qui existent entre les informations visuelles et auditives relatives aux personnes. Faisant l'hypothèse que cette situation d'interférence peut se rapprocher des situations que nous rencontrons quotidiennement, nous avons examiné les activités cérébrales impliquées dans ce type d'interactions.

Dans la deuxième étude, nous avons utilisé la fMRI pour examiner les régions cérébrales spécifiquement activées dans les processus d'intégration crossmodale entre visages et voix, en utilisant le même paradigme expérimental que dans l'étude précédente. Toutefois, il nous semblait intéressant de comparer les substrats neuraux impliqués dans ces processus intégratifs lorsque les associations proposées amènent ou non à une facilitation des jugements d'identité. A cette fin, nous avons mis au point une méthodologie expérimentale dans laquelle la moitié des associations visages-voix comprenaient des visages transformés par morphing afin de les rendre aussi difficiles à identifier que les voix. Un nouveau pré-test nous a permis de constater que dans un tel cas, les associations produisaient une facilitation crossmodale, à savoir que ces associations étaient identifiées plus rapidement que les visages et les voix présentés isolément. Cette étude nous

permettra donc (1) d'isoler les régions cérébrales intégratrices entre visages et voix, (2) selon que les associations présentées produisent ou non une facilitation des réponses

Etude 3

Publiée dans *Neuroscience Letters*, 369 (2004), 132-137.

When audition alters vision : an ERP study of the cross-modal interactions between faces and voices.

F. Joassin¹, P. Maurage¹, R. Bruyer¹, M. Crommelinck², and S. Campanella¹

¹Unité de Neurosciences Cognitives (NESC), Faculté de Psychologie et des Sciences de l'Education, ²Unité de Neurophysiologie Clinique (NEFY), Faculté de Médecine, Université catholique de Louvain, Louvain-la-Neuve, Brussels, Belgium.

Corresponding author :

Frédéric Joassin
Ph D student – aspirant F.N.R.S.
Université catholique de Louvain
Faculté de Psychologie et des Sciences de l'Education
Unité de Neurosciences Cognitives (NESC)
Place du Cardinal Mercier, 10
1348 Louvain-la-Neuve
Belgium
Tél. : +32 (0)10/47.90.62
Fax : +32 (0)10/47.37.74
e-mail : frederic.joassin@psp.ucl.ac.be

Abstract

Ten healthy volunteers took part in this ERP study aimed at examining the electrophysiological correlates of the crossmodal audio-visual interactions in an identification task. Participants were confronted either to the simultaneous presentation of previously learned faces and voices (audio-visual condition ; AV), either to the separate presentation of faces (visual ; V) or voices (auditive ; A). As expected, an interference effect of audition on vision was observed at a behavioral level, as the bimodal condition was performed more slowly than the visual condition. At the electrophysiological level, the subtraction ($AV - (A + V)$) gave prominence to three distinct cerebral activities : (1) a central positive/posterior negative wave around 110 ms, (2) a central negative/posterior positive wave around 170 ms, (3) a central positive wave around 270 ms. These data suggest that crossmodal cerebral interactions could be independent of behavioral facilitation or interference effects. Moreover, the implication of unimodal and multisensory convergence regions in these results, as suggested by a source localization analysis, is discussed.

Cross-modal interactions between different cortical areas are one of the main cerebral processes contributing to our daily adapted behaviors. If the perception of stimuli within single sensory modalities begins to be well known, very little is known about the neural mechanisms by which the brain is able to establish relationships between sensory events and how it integrates them into a unified representation in order to interact properly with the objects of our environment.

Moreover, a great part of the studies examining cross-modal interactions has focused on facilitation processes : responses are facilitated (reduced reaction times and/or enhanced performances) when participants are confronted with two stimuli from two different sensory modalities relative to one unimodal stimulus [Molholm et al., 2002].

However, cross-modal interactions can also take the form of an *interference effect* in which bimodal stimulations alter the performances of the participants rather than facilitating them. The most studied interference effects are the McGurk and the ventriloquist effects in which vision biases the auditory perception [Olson et al., 2002 ; Alais et al., 2004]. But conversely, audition can also alter vision. Sekuler et al. have

shown that sounds can bias the perception of moving visual targets [Sekuler et al., 1997]. Moreover, the alteration of vision by audition is not limited to the cases in which the visual information is ambiguous or moving. Shams et al. have recently put in light that the synchronous presentation of multiple auditory beeps and a single visual flash induces the subjective perception of multiple flashes [Shams et al., 2002]. This sound-induced illusory flash phenomenon was also studied with neurophysiological techniques and revealed that auditory stimulations modulated the visual evoked potentials at short latencies [Shams et al., 2001] and produced significant induced gamma band responses in visual cortical regions [Bhattacharya, et al., 2002].

Therefore, the present study aimed at examining the electrophysiological correlates of the cross-modal interactions elicited by the interference of audition on vision with non-ambiguous complex stimuli. For this purpose, we used an identification task based either on the simultaneous perception of faces and voices, or on the isolated perception of one of the two kinds of information. We used faces and voices for their high ecological level. If many event-related potentials (ERPs) experiments were devoted to face

perception [Bentin et al., 1996 ; Campanella et al., 2002] and voice perception [Levy et al., 2001 ; Müller et al., 1996], very little is known about the combination of the two. But actually, we are confronted everyday with social situations in which we have to integrate visual and auditive information, at least to know who is speaking. We speculated that voices could alter identification of faces for several reasons. First, auditory stimulations have been shown to be more slowly processed than visual ones [Teder-Sälejärvi et al., 2002]. Second, the amount of available information at the onset of the stimulations is not identical for faces and voices : for faces, all the visual information is available immediately whereas, for voices, the auditory information spreads in time. This temporal asynchrony has been argued to be an important feature leading to an interference effect [Calvert et al., 2001]. In the same way, Shimojo & Shams [2001] showed that the direction of cross-modal interactions depends on the structure of the stimuli, that is the more discontinuous (or salient) signal becomes the influential signal. In this case, voices can be considered as more discontinuous than faces and could thus alter the processing of faces.

Ten healthy undergraduate subjects (eight males) took part to this experiment (mean age : 21,5 ; SD : 1,7). All but one were right-handed and all had a normal-to-corrected vision. A pre-experimental learning phase served to familiarize them with the twelve face-name-voice associations used in the experiment. Each association was composed by an unknown face picked from the Stirling face database (black/white photos, front view, neutral facial expression, 6 males/6 females, see at <http://pics.psych.stir.ac.uk>), a belgian family name controlled for its frequency of appearance in the belgian population, and a particular voice pronouncing the family name. At the end of the learning phase, all participants were able to identify each voice and to recuperate the name associated with each face.

During the experiment, participants were confronted to 3 different conditions mixed in 20 blocks of 30 trials (200 trials per condition), that differed in the kind of stimuli presented : either a face alone, a voice alone or the simultaneous presentation of a face and a voice. Each trial began with the presentation of a grey fixation cross appearing for 300 ms on a black screen. It was replaced by a family name presented visually for 500 ms. After an empty interstimulus interval

varying randomly from 1000 to 2000 ms, the stimulus(i) appeared for 1000 ms, followed by an empty intertrial interval of 1000 ms. For each trial, participants had to decide if the presented stimulus(i) was (were) correctly primed by the preceding visual family name or not (50% of the trials were correct, the remaining were primed by another family name from the set of the twelve proper names previously learned). This procedure ensured that (1) the participants were implied in a better controlled cognitive task than the simple passive perception of stimuli [Calvert et al., 2001 ; Olson et al., 2002], (2) the task was identical across all conditions, allowing the comparison of the conditions and the subtraction of the unimodal conditions from the bimodal one. The participants answered by pressing a key on a paddle. In the unimodal auditory condition, the incorrect trials were always constituted by a wrong voice pronouncing the good name whereas in the unimodal visual condition, the incorrect trials consisted in one of the five other faces from the same gender. In the bi-modal condition, the incorrect trials were half constituted by the good face but a wrong voice (CV/IA), half by the good voice but a wrong face (IV/CA). All incorrect visuo-auditive trials were regrouped in the ERP analyses.

Furthermore, it is worth noting that in both auditory conditions (auditive and visuo-auditive), the voice was always pronouncing the primed name so that participants had to identify the voice itself.

EEG was recorded by 32 electrodes mounted in an electrode Quick-Cap. Electrode positions included the standard 10-20 system locations and intermediate positions. Recordings were made with a linked mastoid physical reference. The EEG was amplified by battery-operated A.N.T. ® amplifiers with a gain of 30 000 and a band-pass of 0.01 - 30 Hz. The impedance of all electrodes was kept below 20 k Ω . EEG was continuously recorded (sampling rate 512 Hz, Eeprobe, A.N.T.®) and trials contaminated by EOG artifacts were manually eliminated off-line. Epochs beginning 200 ms prior to stimulus onset and continuing for 1500 ms were created. Codes synchronized with stimulus delivery were used to average selectively the epochs associated with different stimulus types. Two parameters were coded for every stimulus: (1) the type of the stimulus (voices alone, faces alone, simultaneous faces and voices), (2) the type of trials (correct or incorrect). This coding allowed us to compute different averages of ERP target stimuli. The averages were

computed for each subject individually. Then, grand-averaging were computed for the 6 conditions (correct visual (CV), incorrect visual (IV), correct auditive (CA), incorrect auditive (IA), correct visuo-auditive (CV/CA), incorrect visuo-auditive (mean of CV/IA and IV/CA)).

At a behavioral level, an ANOVA with type of stimulus (visual, auditory, visuo-auditive) and type of trial (correct, incorrect) as within-participants factors was performed on percentages of correct responses (table 1, lower part). Please note that incorrect visuo-auditive trials were grouped in this first and in the following analysis. It showed a significant main effect of the type of stimulus ($F(2,18) = 14.2, p < .01$), but no significant main effect of the type of trial ($F(1,9) = 1.65, n.s$). The interaction between the type of stimulus and the type of trial failed to reach a significant value ($F(2,18) = 3.58, n.s$). The main effect of the type of stimulus is due to the fact that auditory condition was less well performed than the two others, that did not differ significantly between each other.

The same ANOVA was performed on reaction times (table 1, upper part). We did observe a significant main effect of the kind of stimulus ($F(2,18) = 78.6, p < .0001$) but no significant main effect of the kind of trial ($F(1,9) = 0.11, n.s$) and no

significant interaction between the type of stimulus and the type of trial ($F(2,18) = 2.02$, n.s). The auditory condition was thus less quickly performed than the two others. Moreover, the visual condition was performed faster than the visuo-auditive condition.

	<i>Visual</i>		<i>Auditive</i>		<i>Visuo-auditive</i>		
	CV	IV	CA	IA	CV/CA	CV/IA	IV/CA
RTs (in ms)	853 (204)	826 (129)	1009 (222)	990 (176)	953 (211)	1043 (173)	907 (142)
					Mean of CV/IA and IV/CA: 975 (156)		
% of correct responses	95,2 (1,8)	96,3 (2,7)	91 (5,8)	87,9 (7,6)	95,7 (3,5)	82,8 (15,4)	96,5 (2,7)
					Mean of CV/IA and IV/CA: 89,3 (8,5)		

Table 1 : mean reaction times and percentages of correct responses (standard deviations in parentheses) for correct and incorrect trials of each condition. CV : correct visual, IV : incorrect visual, CA : correct auditive, IA : incorrect auditive, CV/CA : correct visuo-auditive, CV/IA : correct visual/incorrect auditive, IV/CA : incorrect visual/correct auditive.

Two further analyses were performed to make sure of the existence of an interference effect. First, we compared the trials in which faces were presented in isolation (CV and IV) or with a voice (CV/IA and IV/CA). An ANOVA with type of stimulus (visual or visuo-auditive) and type of trial (correct or incorrect face) was computed. It showed a main effect of the type of stimulus ($F(1,9) = 104,92$, $p < .001$) indicating that the

visual conditions were performed more quickly than the bimodal ones, a main effect of the type of trial ($F(1,9) = 17,94$, $p < .001$) and a significant interaction between the type of stimulus and the type of trial ($F(1,9) = 12,14$, $p < .01$). This interaction is due to the fact that CV and IV did not differ significantly between each other ($F(1,9) = 0,80$, n.s) whereas the IV/CA trials, in which the visual information was sufficient to take a decision, were performed more quickly than the CV/IA trials in which the auditory information was necessary to answer correctly ($F(1,9) = 63,20$, $p < .001$).

Secondly, we compared the trials containing only voices (CA and IA) or voices and faces (CV/IA and IV/CA). The same analysis was performed. It revealed a main effect of the type of trial ($F(1,9) = 13,04$, $p < .01$) but the main effect of the type of stimulus failed to reach a significant statistical level ($F(1,9) = 2,08$, n.s). The interaction between the type of stimulus and the type of trial was significant ($F(1,9) = 25,77$, $p < .001$) and due to the fact that CA and IA did not significantly differ between each other ($F(1,9) = 0,42$, n.s).

The behavioral results clearly show an interference effect of audition on vision as (1) the auditory and bimodal conditions were performed more slowly than the visual

condition, (2) the addition of an auditive information interfered with the processing of the visual information, (3) the addition of a visual information did not facilitate the processing of the auditive information.

At an electrophysiological level, we first checked the ERP complexes classically associated with the early processing of faces and voices (figure 1, left) [Bentin et al., 1996 ; Müller et al., 1996]. In the visual condition, the P1 wave was maximally recorded at posterior sites (Oz, O1, O2) around 120 ms (mean amplitude of 4 μ v) and the N1 wave was maximally recorded at central electrodes (Cz, C1, C2) around 130 ms (mean amplitude of -2,13 μ v). The N170/VPP complex was maximally recorded around 180 ms at Oz, O1 and O2 for the N170 component (mean amplitude of -7 μ v) and Cz, C1 and C2 for the VPP component (mean amplitude of 4 μ v).

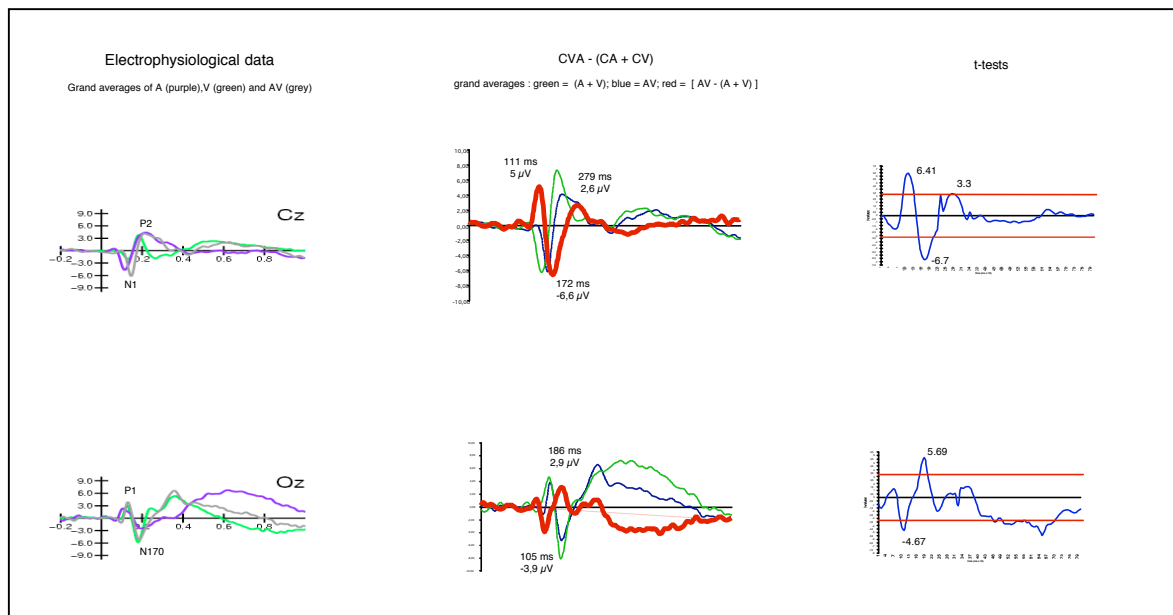


Figure 1 : left : ERP waves elicited by the 3 conditions (A in purple, V in green and AV in grey) ; Middle : ERP s elicited by the subtraction $[AV - (A + V)]$ (in red), the sum of $A + V$ (in green) and AV (in blue) ; right : t-values for the subtraction (significance levels are represented by red lines). Upper part : ERPs recorded at Cz, lower part : ERPs recorded at Oz.

The processing of voices also elicited two classical cortical ERP complexes. The P1/N1 complex peaked around 114 ms at posterior electrodes (Oz, O1, O2) for the P1 (mean amplitude of 3,12 μv) and at central electrodes (Cz, C1, C2) for the N1 (mean amplitude of -4,63 μv). The second ERP complex, the P2/N2, was maximally recorded at occipital electrodes around 198 ms (mean amplitude of -3,81 μv) for the N2 component, and at central electrodes around 212 ms (mean amplitude of 4,7 μv) for the P2 component.

These classical P1/N1 and N2/P2 complexes were also observed in the bi-modal condition. The P1 wave peaked at occipital electrodes around 132 ms (mean amplitude of 3,5 μv)

and the N1 wave at central electrodes around 145 ms (mean amplitude of $-6,13 \mu\text{v}$). The N2 wave was maximally recorded at posterior sites around 187 ms (mean amplitude of $-6,32 \mu\text{v}$) and the P2 was maximal at central sites around 215 ms (mean amplitude of $4,43 \mu\text{v}$).

To conclude these first observations, the classical ERP components associated with the perception and the early processing of stimulation have been elicited by our own faces and voices, which ensures us that our stimulations were correctly encoded by the participants.

The main analysis consisted in subtracting the unimodal averages from the bi-modal average $[\text{CV/CA} - (\text{CA} + \text{CV})]$, which is a classical method to put in light the electrophysiological correlates of the cross-modal interaction. Note that we performed this subtraction only on the correct trials, as it has been underlined that cross-modal interactions are more prone to appear in correct than incorrect trials because of their references towards representations in long term memory [Teder-Sälejärvi et al., 2002]. We first computed the subtraction for each participant individually. We then performed the grand-average of the results of the subtraction.

It revealed three main complexes of electrical activities at distinct latencies (figure 1, middle) : (1) around 100 ms, a positive-negative complex recorded respectively at central (Cz : peak latency at 111 ms, maximum amplitude of 5 μv) and posterior sites (Oz : peak latency at 105 ms, maximum amplitude of -3,9 μv) ; (2) around 175 ms, a negative-positive complex, the negativity peaking at Cz around 172 ms (maximum amplitude of -6,6 μv) and the positivity peaking at Oz around 186 ms (maximum amplitude of 2,9 μv) ; (3) a positive wave recorded at central electrodes and maximally peaking at Cz around 279 ms (maximum amplitude of 2,6 μv).

Significant effects were then calculated at each selected electrode using Student t-tests (amplitude of the difference wave compared to zero for successive 10-ms intervals from 0 to 800 ms). The spatio-temporal patterns that had a significant amplitude ($p < .01$) at least on one electrode for two consecutive intervals were considered as significant (figure 1, right). For the central electrodes, the subtraction wave was significantly different from zero at Cz between 90 and 130 ms ($t \equiv 6,41$, $p < .001$), 140 and 200 ms ($t \equiv -6,7$, $p < .001$) and from 260 to 280 ms ($t \equiv 3,3$, $p < .01$). For the posterior electrodes, the subtraction wave significantly differed from

zero at Oz between 90 and 110 ms ($t = -4.67$, $p < .01$) and between 170 and 200 ms ($t = 5.69$, $p < .001$).

Finally, a source localization analysis was performed on the grand-average ERP waves of the subtraction. We used the ASA 3.0 software (A.N.T.®) which determines the position and the orientation of intracranial dipoles and their time-varying strength by using a three-layer spherical head model. A fixed dipole method was applied in which the position of the dipoles was constrained but their orientation and strength remained free. Dipole modeling was performed within the time window of each wave (table 2 and figure 2).

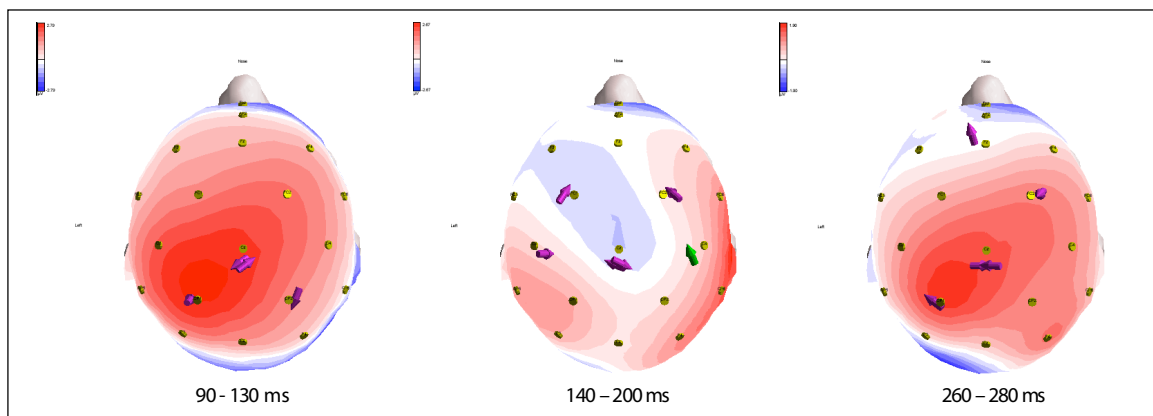


Figure 2 : top views of the dipole modeling and topographies of the ERPs waves elicited by the subtraction [AV – (A+V)]. Left : between 90 and 130 ms (two pairs of bilateral dipoles localized in the superior colliculi and the fusiform gyri) ; middle : between 140 and 200 ms (three pairs of bilateral dipoles localized in the superior colliculi, the medial temporal gyri and the inferior frontal gyri) ; right : between 260 and 280 ms (a pair of dipole localized in the superior colliculi, one dipole in the left fusiform gyrus, one dipole in the right inferior frontal gyrus, one dipole in the left superior frontal gyrus).

Time window (ms)	Best fit latency (ms)	RV (%)	Cerebral area	Coordinates (Talairach coordinate system)			Magnitude (NAm)
				x	y	z	
90 - 130	113	2,6	right SC	3	-28	-2	1297
			left SC	-3	-28	-2	1183
			right FG	40	-55	-10	100
			left FG	-40	-55	-10	43
140 - 200	174	5,8	left SC	-3	-28	-2	3326
			right SC	3	-28	-2	3243
			left MTG	-55	-20	1	168
			right MTG	55	-20	1	132
			left IFG	-41	26	13	105
			right IFG	41	26	13	50
260 - 280	275	8,8	left SC	-3	-28	-2	1085
			right SC	3	-28	-2	1136
			right SFG	11	72	19	164
			right FG	-40	-55	-10	65
			left IFG	-41	26	13	25

Table 2 : dipole modeling of the three significant ERP activities elicited by the subtraction $[AV - (A + V)]$. Cerebral areas: FG = fusiform gyrus (Brodmann Area 37), SC = superior colliculus, MTG = Middle temporal gyrus (BA 22), IFG = inferior frontal gyrus (BA 45), SFG = superior frontal gyrus (BA 10).

Between 90 and 130 ms, the electrophysiological results were best explained by two pairs of bilateral dipoles localized in the superior colliculi and the fusiform gyri (Brodmann Area 37), and more specifically in the Face Fusiform Area (FFA), a cerebral region highly responsive to human faces [Kanwisher et al., 1997]. For the 140-200 ms time window, the lowest residual variance was obtained with three pairs of bilateral dipoles localized in the superior colliculi, the superior temporal gyri (associative auditory cortex, BA 22) and the inferior frontal gyri (BA 45). Finally, between 260 and 280 ms,

the scalp cerebral activity was best explained by five dipoles, a pair of bilateral dipoles localized in the superior colliculi, one dipole in the superior frontal gyrus (BA 10), one in the inferior frontal gyrus (BA 45) and one in the fusiform gyrus (BA 37).

The present study aimed at examining the electrophysiological correlates of the interference effect of audition on vision with non-ambiguous complex stimuli in an identification task based either on the simultaneous perception of faces and voices, or on the isolated perception of one of the two kinds of information. At a behavioral level, we observed that audition actually altered vision, as the simultaneous presentation of a face and a voice slowed down the speed of response relative to the presentation of a face alone, confirming thus the hypothesized interference effect.

At the electrophysiological level, these visuo-auditive interactions were analyzed by subtracting the unimodal conditions from the bimodal condition. It revealed a positive-negative-positive ERP component recorded on the central electrodes, associated with a negative-positive wave recorded on the posterior electrodes. Furthermore, we observed that

the first central positive/posterior negative wave was best explained by a pair of dipoles originating in the associative visual cortex, whereas the central negative/posterior positive wave was best explained by a pair of dipoles localized in the associative auditory cortex. Finally, the second central positive wave was best explained by a network of cortical areas including the superior colliculi, the fusiform gyrus, the associative auditory cortex and the superior frontal gyrus. Then we suggest that the present cross-modal interactions can be modeled by distinct temporal interactions : (1) the auditory information could influence the processing of the visual information in the fusiform gyrus around 90-130ms, (2) the visual information could then influence the processing of the auditory information in the associative auditory areas (BA 22) around 140-200 ms, (3) finally, around 260-280 ms, the unimodal and cross-modal areas could interact with semantic areas, like the inferior and superior frontal gyri [Poldrack et al., 1999], in order to compare the created bimodal representation with one of the representations stored in long term memory.

It is worth noting that many studies focused on the neural correlates of cross-modal interactions producing a

facilitation of the responses, i.e. shorter RTs and/or higher performances when participants are confronted to bimodal stimulations relative to unimodal ones. Nevertheless, to our knowledge the present study is the first one to have examined the cerebral markers of the inhibitory processes of cross-modal interactions by using complex stimulations such as faces and voices. Interestingly, the present results are very similar to those obtained by Teder-Sälejärvi et al. [2002], who showed, in an ERP experiment using light flashes and noise bursts, that bimodal stimulations *facilitated* the task (detection of infrequent target stimuli) and produced a positive-negative-positive wave on frontal, central and parietal electrodes. Moreover, their dipole modelling gave results very similar to our present data. The similarity between the results of Teder-Sälejärvi et al. [2002] and the present results allow us to hypothesize that (1) these electrophysiological activities reflect general interactions between neural populations processing different sensorial stimulations, (2) these cross-modal interactions could be independent of facilitatory or inhibitory processes of one type of stimulation on another one.

It is also interesting to note that the results of our dipole modelling always included the superior colliculus, a mid-brain structure considered as a multimodal convergence region [Stein et al., 2001] and that has already been observed in an ERP study investigating the audio-visual object recognition processes [Fort et al., 2002]. When we excluded these regions from the source localization analyzes, the residual variance dramatically increased, which was not the case for visual or auditive stimulations presented in isolation. These results could give us new leads to better understand the mechanisms by which the different unimodal regions interact. It could be possible that the unimodal regions, processing simultaneously different sensory stimulations, send inputs to multimodal integration regions, such as the superior colliculi, where the representations of each stimulation are integrated. In turn, the multimodal convergence regions could send back projections to the unimodal brain areas, each « primary » sensory region being then modulated by the stimulations in the other modality [Driver & Spence, 2000].

Nevertheless, further experiments are needed, notably to compare the present data with the neural correlates of cross-modal interactions between complex and non-ambiguous

stimulations producing a behavioral facilitation, or to investigate the possible existence of synchronous cerebral activity between the unimodal regions within the range of the gamma band [Bhattacharya et al., 2002].

Aknowledgment

This study was made possible thanks to FRSM N° 3458303F. The first, second and last authors are supported by the Belgian Found for Scientific Research (F.N.R.S.).

Etude 4

Soumise pour publication

**Cross-modal integration of human faces and
voices in the left angular gyrus**

F. Joassin, M. Pesenti, E. Verreckt, R. Bruyer & S. Campanella

Université catholique de Louvain, Unité de Neurosciences Cognitives
(NESC), Louvain-la-Neuve, Belgium

Corresponding author:

Mauro Pesenti
Université catholique de Louvain
Faculté de Psychologie et des Sciences de l'Education
Unité de Neurosciences Cognitives (NESC)
Place Cardinal Mercier, 10
B-1348 Louvain-la-Neuve
Belgium

Tel.: +32 (0)10/47.88.22
Fax: +32 (0)10/47.37.74
Mauro.Pesenti@psp.ucl.ac.be

Abstract

Faces and voices are key features in human recognition. However, the way the brain links them together is still unknown. In a recent paper, we showed that voices are harder to identify than faces when presented alone. Moreover, their synchronized presentation disturbs recognition, as adding a voice to a face made face identification more difficult. We suggested that facilitative recognition effects occur when visual and auditory information are equally complex. In this report, we show that when faces were modified by morphing to become more difficult to identify, associating a voice to the face facilitated recognition, whereas associations between voices and unmorphed faces did not. We also measured brain activity using functional magnetic resonance imaging (fMRI) while subjects were recognizing voice–face associations which either did or did not facilitate identification. Voice–face associations, whatever they produced or not facilitation, activated both unimodal visual and auditory areas, and a multimodal convergence region located in the left angular gyrus. These findings demonstrate that cross-modal gains are modulated by the complexity of the visual and auditory information, and underline for the first time the cross-modal integration role played by the left inferior parietal cortex in creating unified representations of individuals from visual and auditory stimuli.

Introduction

Human social interactions are shaped by our ability to identify individuals, a process to which face and voice recognition contribute both separately and jointly. Much research has been devoted to unimodal face recognition. Neuroimaging studies have shown that human faces are mainly processed by temporo-occipital regions of the brain with a right hemispheric dominance, and in particular by the Fusiform Face Area (FFA, Sergent et al., 1992; Kanwisher et al., 1997; Rhodes et al., 2004). Studies of brain-damaged patients have revealed a selective impairment of face recognition, prosopagnosia, associated with damage of the right fusiform gyrus (De Renzi et al., 1994; Takahashi et al., 1995). Fewer studies have focused on voice recognition. Voice recognition takes place bilaterally in the superior temporal cortex, with a particular recruitment of the right superior temporal sulcus (STS, Belin et al., 2000, 2002; Von Kriegstein et al., 2003). Phonagnosia, the selective impairment of voice recognition, is classically associated with lesions of temporal lobes of either hemispheres (Van Lancker et al., 1989, Neuner & Schweinberger, 2000).

Although neuroanatomically segregated, faces and voices are assumed to interact during person recognition (Burton et al., 1990; Ellis et al., 1997) as we constantly integrate visual information from faces and auditory information from voices when interacting with others. It has been shown that such integration skills emerge very early in life (Bahrick et al., 2005). However, the mechanism by which voices and faces, which are processed by distinct cerebral regions, are integrated in the brain into a unique and coherent representation of a person is still unknown. This issue has usually been examined at the perceptual level (Calvert et al., 1999; Olson et al., 2002), and illusions have frequently been used to examine cross-modal perceptual interactions, such as the McGurck effect in which the perception of auditory syllables is altered by incongruent visual lips movements (Sekiyama et al., 2003).

So far, most studies did not examine cross-modal interactions between faces and voices in person identification. We investigated this question in a recent experiment in which participants had to perform identifications based on the perception of previously learned faces, voices, or synchronous presentation of faces and voices

together (Joassin et al., 2004). At the behavioral level, we did not observe the classical cross-modal facilitation effect, i.e. an improvement in performance in the bimodal condition relative to any of the unimodal conditions. Rather, we observed an interference effect: the bimodal identification was performed significantly more slowly than the visual identification. Adding auditory information did not facilitate the identification of faces but made it more difficult. It is known that cross-modal advantages depend not only on spatio-temporal proximity (Calvert et al., 2001) and semantic congruency (Pourtois & de Gelder, 2002) but also on the precision of the information (Wada et al., 2003). Therefore, faces may have been easier to identify than voices because, at their onset, the whole visual information was immediately available whereas auditory information was spread out in time. Voices could thus have been less informative or precise than faces, and hence more difficult to identify.

Such complexity in voice recognition relative to face recognition has been described (Ellis et al., 1997; Schweinberger et al., 1997), but knowing how faces and voices interact and influence each other is of particular relevance for a better understanding of problems in person

recognition, such as prosopagnosia, phonagnosia, or Capgras delusion, in which the patient believes that a close relative has been replaced by an imposter (Shah et al., 2001).

Here, we performed one behavioral and one functional magnetic resonance imaging (fMRI) study to investigate the cerebral correlates of voice–face interactions when associations between faces and voices do or do not produce cross-modal behavioral facilitation. The behavioral experiment was aimed at testing whether differences in the difficulty of recognizing voices and faces could account for the absence of cross-modal facilitation observed in our previous study. We predicted that voices and faces equally difficult to recognize would facilitate rather than impede recognition when presented together. With fMRI we investigated the cerebral correlates of the cross-modal interactions between faces and voices in a semantic task, and compared the cerebral activations when voice–face associations did and did not produce behavioral cross-modal facilitation. To the best of our knowledge, this question has never previously been directly examined (see for instance Von Kriegstein et al., 2005).

We expected that voices would elicit an activation of the bilateral temporal cortex and in particular the right STS, and that faces would elicit an activation of the right FFA, i.e. the classical areas dedicated to the processing of voices and faces respectively. On the basis of neuroimaging studies showing an involvement of unimodal and multimodal areas in cross-modal binding (Wada et al., 2003; Bushara et al., 2003), we also hypothesized that bimodal stimulations would activate both unimodal visual and auditory areas (as previously observed for auditory speech perception, Calvert et al., 1999), and multimodal areas, notably the left parietal cortex which can be considered as a heteromodal convergence zone involved in associative processes related to individuals (Campanella et al., 2001). Finally, we expected that the cerebral network activated by the specific voice–face association processes would differ according to the presence or absence of a behavioral facilitation effect, i.e. would be modulated by task complexity.

Methods

Participants

Behavioral study. Twenty-six undergraduate volunteers (23 females, mean age: 19.5, SD: 0.76) took part in this experiment. Twelve participated in a preliminary test, the remaining fourteen participated in the experiment itself. All were right-handed and had a normal visual and auditory acuity.

FMRI study. Twelve healthy volunteers (7 females, mean age: 23.5, SD: 3.99), who were not part of the behavioral experiment, participated in the fMRI study. All were native French speakers, right handed, had normal vision and audition, and gave their written informed consent. The experimental protocol was approved by the Biomedical Ethical Committee of the Catholic University of Louvain.

Stimuli

The stimuli consisted of four associations, each composed of a female face (black and white photos, front view, neutral expression), a Belgian family name, and a female voice saying the french word «bonjour» (mean duration: 685.75 ms, SD: 142.34). Two learning sessions served to

familiarize participants with these associations. The encoding of the associations was carried out by a computer presentation using Superlab 6.1 software (Cedrus Corporation). To ensure that the associations were correctly encoded, several recognition tests (face–name matching, voice–face matching, voice-name matching) and identification tests (recall of the name of each face and voice) were performed. Each learning session was continued until accuracy reached 100%.

For the experimental session, the faces were modified by morphing using MorphTM 2.5 (Gryphon software corporation). Each of the four faces was mixed with each of two other faces from the set in four different proportions, to form 32 new faces. These had either ninety percent of one face and ten percent of another (morphing at 10%), an eighty/twenty mixture (morphing at 20%), seventy/thirty (morphing at 30%), or sixty/forty (morphing at 40%).

The 12 participants in the preliminary test were asked to categorize previously learned unmorphed faces and new morphed faces (morphed at 10, 20, 30 and 40%, Figure 1a) according to their identity, by pressing one of two keys on a paddle.

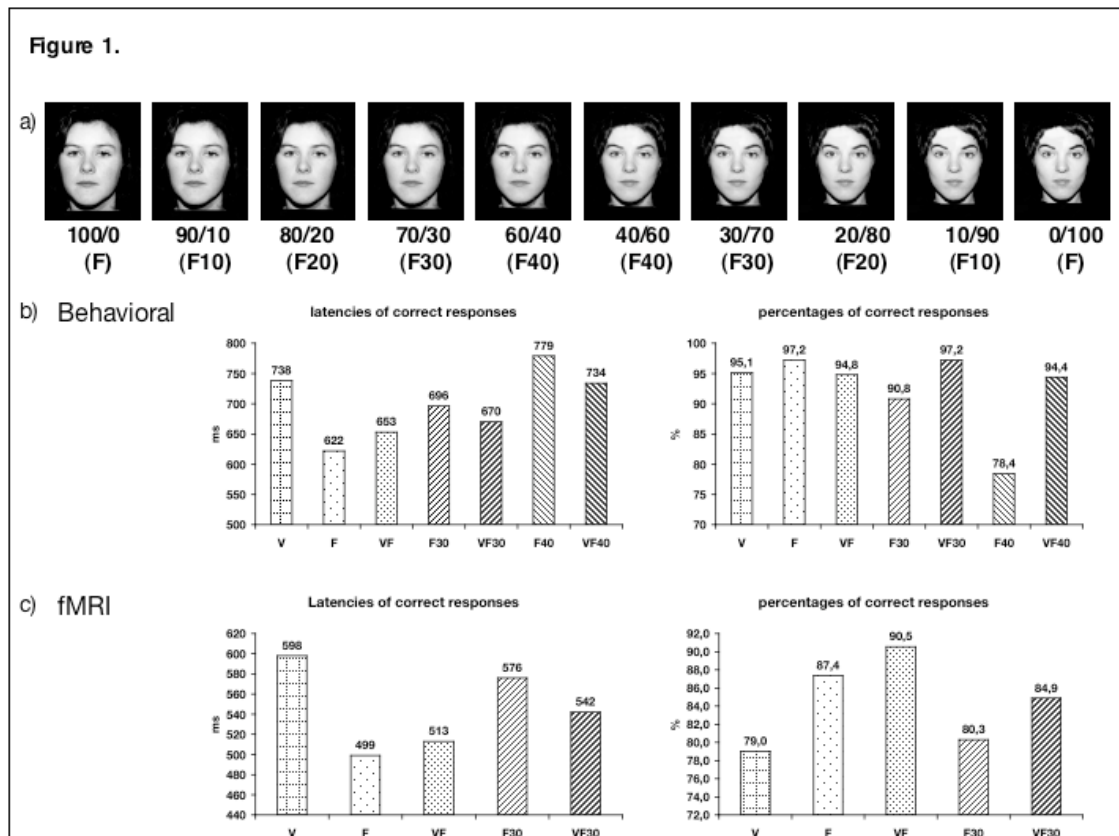


Figure 1. a) examples of faces with their percentages of morphing. b) mean latencies and percentages of correct responses for voices (V), faces (non-morphed = F, morphed = F10, F20, F30, F40) and voice-face associations in the behavioral experiment. c) mean latencies and percentages of correct responses for the categorization of faces (F and F30), face/voice associations (VF and VF30) and voices (V) in the fMRI experiment.

Each face was presented for 700 ms in four blocks each containing only two of the four identities. This test was undertaken to determine what proportion of morphing made the faces more difficult to identify than unmorphed faces. Recognition of faces morphed at 10% ($F(1,11) = 0.34$, ns, p-level set at 0.01 after Bonferonni correction for multiple comparisons) and 20% ($F(1,11) = 0.06$, ns, p-level set at 0.01 after Bonferonni correction for multiple comparisons) did not

differ significantly from that of unmorphed faces. However, faces morphed at 30% ($F(1,11) = 49.64$, $P < 0.001$, p-level set at 0.01 after Bonferonni correction for multiple comparisons) and 40% ($F(1,11) = 52.32$, $P < 0.001$, p-level set at 0.01 after Bonferonni correction for multiple comparisons) were identified significantly more slowly than unmorphed faces. These two last proportions of morphing were therefore used in the behavioral experiment. For the fMRI study, we used only unmorphed faces and faces morphed at 30%, as associations between voices and faces morphed at 40% did not produce behavioral facilitation (see Results).

Procedure

The behavioral experiment consisted of three conditions: voices alone (V), faces alone (F: unmorphed; F30: morphed at 30%; F40: morphed at 40%), and associations between a voice and the three kinds of face (VF, VF30, VF40). Each trial began with a fixation cross appearing for 300 ms. Then, a stimulus was presented for 700 ms, and was followed by an empty intertrial interval of 1250 ms. Six blocks of trials (each containing only two identities) were recorded, each block containing six trials in each modality (i.e. 6 voices, 6 faces and

6 associations). The participants had to categorize each stimulus according to its identity by pressing one of two keys on a paddle.

During the fMRI session, five different conditions were presented: unmorphed faces (F), morphed faces (F30), voices (V), voice-unmorphed face associations (VF), and voice-morphed face associations (VF30) (Figure 2).

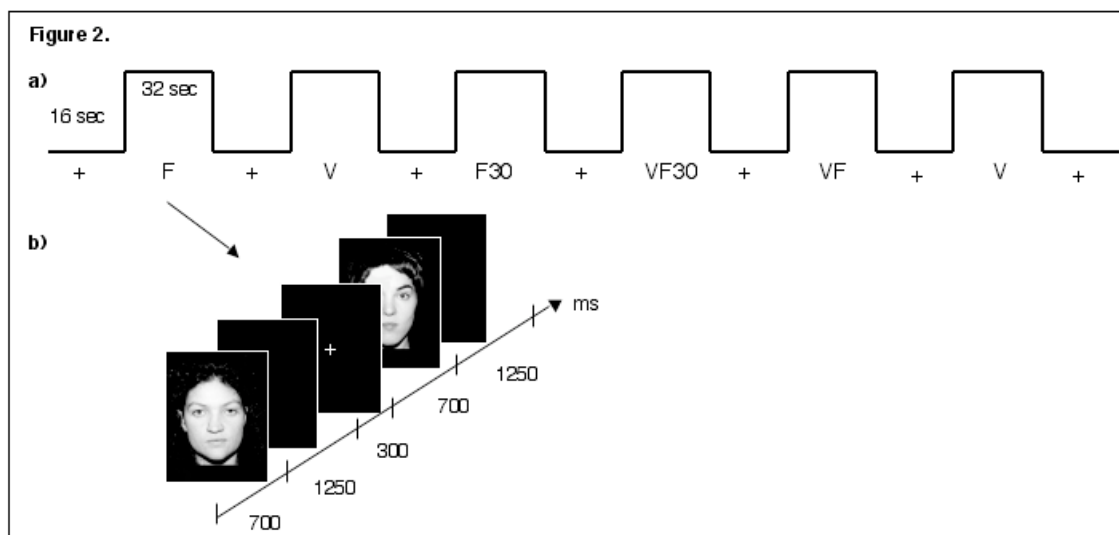


Figure 2. a) fMRI design : each run consisted in 6 alternances of a 16-sec fixation period (white cross on black background) and a 32-sec activation period. Each activation period corresponded to a different condition, presented in a pseudo-random order. b) examples of behavioral task : participants were presented with 14 trials in each condition. Each trial comprised a fixation cross for 300 ms, a stimulus - non-morphed faces, morphed faces, voices, non-morphed faces/voices associations, or morphed faces/voices associations - for 700 ms and a black intertrial interval for 1250 ms.

Only two identities were included in each run, and these were varied by run. The task consisted of categorizing every stimulation (face, voice or association) according to its identity by pressing one of two keys on a stimpad with the

right hand. The importance of speed and accuracy was emphasized. Each volunteer participated in six runs which each consisted of six experimental blocks of 32 seconds, interleaved with 16 second fixation periods (white cross on black background). Each experimental block (14 trials, identical to those in the behavioral experiment) corresponded to one of the five conditions, presented in a pseudo-random order, but the auditory condition was presented twice by run in each run to equalize the number of visual and auditory stimuli. Only one of the two auditory blocks, balanced across runs, was used in the analyses.

Apparatus and experimental set-up

In both experiments, stimulus presentation and response recording were controlled with ePrime (Schneider et al., 2002). For fMRI, back-projected images were viewed through a tilted mirror (Silent VisionTM System, Avotec, Inc., <<http://www.avotec.org>>) mounted on the head coil. Auditory stimuli were heard through headphones and the sound volume was adjusted for each participant so as to be clearly audible above the scanner noise.

Imaging procedure

Functional images were acquired with a 1.5 Tesla magnetic resonance imager and a standard head coil (Gyrosan, Philips Medical Systems) as series of blood-oxygen-sensitive T2*-weighted echo-planar image volumes (GRE-EPI). Acquisition parameters were: TE= 50 ms, TR= 2000 ms, Flip angle = 90°, Field of view = 210 x 210 mm, slice thickness = 6 mm with no interslice gap. Each image volume comprised 20 axial slices acquired in an ascending interleaved sequence. Each functional run comprised 152 volumes, the first 8 volumes being discarded to allow for T1 equilibration effects, which resulted in 16 volumes per condition per run. High-resolution T1-weighted 3D fast field echo anatomical images with 110 1.5 mm contiguous axial slices were also acquired for each participant (TE= 3 ms, TR= 30 ms, Flip angle= 30°, FOV= 220 x 175 mm; in-plane voxel size 0.859 x 0.859 x 1.5 mm³). Head movement was limited by a restraining band across the forehead.

Data processing and analysis

Latencies and percentages of correct responses recorded during the behavioral experiment were analyzed separately by

three distinct ANOVAs with the conditions (V, F, VF ; V, F30, VF30; and V, F40, VF40) as within-subjects factors. If necessary, each condition was subsequently compared to the two others with the same proportion of morphing (unmorphed, 30% and 40%) using paired Student t-tests. For the fMRI study, latencies and percentages of correct responses were analyzed separately using two distinct ANOVAs with the condition (V, F, VF ; V, F30, VF30) as within-subjects factor, followed by paired t-tests comparing each condition to the other two.

Neuroimaging data were processed (slice-time correction, realignment, coregistration, normalization to the MNI template implemented in SPM2, using an affine fourth degree β -spline interpolation transformation and a voxel size of $2 \times 2 \times 2 \text{ mm}^3$, smoothing with a Gaussian kernel of 8 mm FWHM) and analyzed using Statistical Parametric Mapping (SPM2, Wellcome Department of Cognitive Neurology, London, UK, <<http://www.fil.ion.ac.uk/spm>>), implemented in a Matlab 6.5.0 environment (The Mathworks, Inc.). Condition-related changes in regional brain activity were estimated for each participant by a general linear model in which the responses evoked by each condition of interest were modeled by a

standard hemodynamic response function. The contrasts of interest were computed at the individual level to identify the cerebral regions significantly activated by voices ($[V - \text{fix}]$), faces ($[F - \text{fix}]$ and $[F30 - \text{fix}]$) and bimodal stimuli ($[VF - \text{fix}]$ and $[VF30 - \text{fix}]$) relative to the fixation periods used as a baseline. The conjunction of $[VF - \text{fix}]$ and $[VF30 - \text{fix}]$ (using the Minimum Statistic compared to the Conjunction Null, MS/CN, Nichols et al., 2005) isolated the cerebral activations elicited by bimodal stimuli whatever they produced or not behavioral facilitation. The subtractions $[F30 - F]$ and $[F - F30]$ were computed to detect possible differences between morphed and unmorphed faces. Contrasts $[VF - (V + F)]$ and $[VF30 - (V + F30)]$ were computed to isolate the cerebral regions involved in the associative processes between faces and voices, with unmorphed faces in the first subtraction and morphed faces in the second. The conjunction (MS/CN) of the two subtractions was also computed to identify cerebral activations involved in the auditory-visual associative processes, regardless of whether or not they produced facilitation. Finally, the contrasts $[VF30 - VF]$ and $[VF - VF30]$ were computed to directly compare each associative condition to the other one and examine whether one

condition led to supplementary activations relative to the other one. Significant cerebral activations were then examined at the group level in random-effect analyses using one-sample t tests, with statistical threshold set at $p < .05$ (corrected) or $p < .001$ (uncorrected) and extending to at least 20 contiguous voxels.

Results

Behavioral study

Significant latencies differences were found between V, F and VF ($F(2,26) = 76.87, P < 0.001$), V, F30 and VF30 ($F(2,26) = 21.82, P < 0.001$), and V, F, VF40 ($F(2,26) = 4.79, P < 0.01$). Subsequent paired t -tests comparing each condition to the two others within each proportion of morphing revealed that associative conditions led to different patterns of responses (Figure 1b). Associations between unmorphed faces and voices did not produce facilitation as they were identified faster than voices alone ($t(13) = 7.82, P < 0.001$) but significantly more slowly than faces alone ($t(13) = -3.34, P < 0.01$). However associations between faces morphed at 30% and voices did lead to facilitation: they were identified more

rapidly than either voices alone ($t(13) = 7.87, P < 0.001$) or morphed faces alone ($t(13) = 2.34, P < 0.04$). Finally, associations between morphed faces at 40% and voices did not show this facilitation effect. These associations were identified significantly faster than faces alone ($t(13) = 2.79, P < 0.02$) but there was no significant difference between them and voices alone ($t(13) = 0.27, ns$).

The analyses of the percentages of correct responses also showed significant differences between V, F and VF ($F(2,26) = 3.79, p < 0.04$), V, F30 and VF30 ($F(2,26) = 23.36, p < 0.001$), and V, F40 and VF40 ($F(2,26) = 91.48, p < 0.001$). V was more error prone than F ($t(13) = -4.08, p < 0.001$), but there were no significant differences between V and VF ($t(13) = 0.28, ns$) or F and VF ($t(13) = 2.08, ns$). V was significantly better performed than F30 ($t(13) = 3.72, p < 0.01$) but VF30 was significantly better performed than F30 ($t(13) = 6.15, p < 0.001$) and V ($t(13) = 3.74, p < 0.01$). Finally, V was better performed than F40 ($t(13) = 11.06, p < 0.001$) but did not differ significantly from VF40 ($t(13) = 1, ns$), which was better performed than F40 ($t(13) = 9.15, p < 0.001$).

fMRI study

Behavioral data.

Our fMRI paradigm capitalizes on these behavioral results and the possibility of producing or not facilitation with voice–face associations by using morphed faces. At the behavioral level, we observed the same pattern of responses than in the behavioral experiment. Significant latencies differences were found between V, F and VF ($F(2,22) = 35.76$, $P < 0.001$) and V, F30 and VF30 ($F(2,22) = 8.77$, $P < 0.001$). Subsequent paired t-tests revealed that (1) voices were identified more slowly than unmorphed faces ($t(11) = -5.62$, $P < 0.001$) or voice–face associations ($t(11) = -4.93$, $P < 0.001$), although these two latter conditions did not differ significantly ($t(11) = 1.45$, ns, Figure 1c); (2) voices and morphed faces were identified at approximately the same speed ($t(11) = 1.16$, ns) and voice–face associations were identified more rapidly than morphed faces ($t(11) = 2.55$, $P < 0.02$) or voices alone ($t(11) = 3.04$, $P < 0.01$).

The percentages of correct responses showed the same patterns of results. Significant differences between V, F and VF ($F(2,22) = 16.72$, $p < 0.001$) were explained by the fact that voices were less well identified than unmorphed faces ($t(11) =$

-4.36, $p < 0.001$) and associations ($t(11) = -5.6$, $p < 0.001$), these two conditions being equivalent ($t(11) = 1.45$, ns). The significant differences between A, F30 and VF30 ($F(2,22) = 4.33$, $p < 0.03$) were explained by the fact that voices and morphed faces did not significantly differ ($t(11) = -0.51$, ns) whereas associations were better identified than morphed faces ($t(11) = 2.66$, $p < 0.02$) and voices ($t(11) = 2.88$, $p < 0.02$).

Brain imaging results.

Firstly, we identified areas significantly activated by voices ([V – fix]), unmorphed faces ([F – fix]), morphed faces ([F30 – fix]) and voice–face associations ([VF – fix] and [VF30 – fix]) relative to the fixation baseline. The results are shown in Table 1. Voices elicited an activation of the bilateral middle and superior temporal gyri (Table 1a); unmorphed faces activated the left calcarine sulcus, the bilateral fusiform gyri and the left supramarginal gyrus (Table 1b); morphed faces activated the bilateral fusiform gyri, the left inferior occipital gyrus, the right middle occipital gyrus, the left cuneus and the left supramarginal gyrus (Table 1c). However a direct comparison of morphed and unmorphed faces ([F30 – F] and [F – F30]) did not show any significant differences.

Associations between voices and morphed faces revealed activations of the bilateral temporal regions, the bilateral fusiform gyri and the left supramarginal gyrus (Table 1d). Associations between voices and unmorphed faces activated mainly the bilateral auditory temporal regions, the visual temporo-occipital regions including the left lingual gyrus, the left calcarine and the middle occipital gyrus, the bilateral putamen, the left thalamus, the left hippocampus and the left supramarginal gyrus. (Table 1e).

The conjunction of both kinds of associations compared to baseline was computed to examine the cerebral activations elicited by bimodal stimuli, irrespective of whether the faces were morphed or not. This revealed an activation of the right fusiform gyrus, the left calcarine sulcus and the bilateral middle and superior temporal gyri (Table 1f).

<u>Brain regions</u>	<u>x</u>	<u>y</u>	<u>z</u>	<u>L/R</u>	<u>k</u>	<u>t-statistic</u>
a <u>V - fix</u>						
Middle temporal gyrus	62	-26	0	R	116	5.91**
Superior temporal gyrus	-52	-32	6	L	62	5.67**
Superior temporal pole	-56	4	-2	L	24	5.40**
Supplementary motor area	6	6	60	R	89	5.84**
b <u>E - fix</u>						
Calcarine sulcus	-16	-92	-4	L	188	5.30**
Fusiform gyrus	-32	-56	-18	L	127	4.25*
Fusiform gyrus	38	-58	-12	R	965	4.89*
Supplementary motor area	6	6	60	R	410	5.03**
Precentral gyrus	58	10	42	R	64	4.36*
Precentral gyrus	-50	-2	46	L	97	4.60*
Supramarginal gyrus	-46	-36	46	L	141	4.69*
c <u>F30 - fix</u>						
Inferior occipital gyrus	-36	-82	-8	L	130	4.79*
Middle occipital gyrus	38	-84	6	R	228	4.57*
Cuneus	-18	-96	8	L	76	4.09*
Fusiform gyrus	-40	-52	-16	L	85	4.71*
Fusiform gyrus	36	-60	-12	R	185	4.60*
Supplementary motor area	-4	0	54	L	168	4.32*
Precentral gyrus	-52	-2	38	L	77	4.49*
Supramarginal gyrus	-40	-34	38	L	114	4.88*
d <u>VF30 - fix</u>						
Superior temporal gyrus	-52	-32	6	L	636	5.26**
Middle temporal gyrus	50	-20	-4	R	988	4.99*
Fusiform gyrus	-32	-56	-18	L	73	4.40*
Fusiform gyrus	36	-62	-12	R	834	4.99*
Supplementary motor area	8	-4	64	R	258	4.94*
Precentral gyrus	-50	-6	48	L	52	4.27*
Precentral gyrus	48	0	52	R	51	4.49*
Supramarginal gyrus	-44	-36	38	L	232	4.55*
cerebellum	-42	-72	-20	L	367	4.70*
e <u>VF - fix</u>						
Middle temporal gyrus	64	-26	2	R	2460	5.87**
Superior temporal gyrus	-64	-16	-2	L	1470	5.65**
Lingual gyrus	-10	-26	-8	L	229	5.81**
Calcarine sulcus	-16	-96	-6	L	1167	4.99*
putamen	-26	-10	14	L	281	4.97*
putamen	26	14	4	R	185	4.90*
Middle occipital gyrus	28	-76	-2	R	2463	5.29**
Supramarginal gyrus	-42	-36	36	L	290	4.88*
Precentral gyrus	48	4	54	R	107	4.81*
Precentral gyrus	-52	-4	42	L	101	4.45*
caudate	-10	20	-4	L	78	4.49*
thalamus	-10	-20	16	L	54	4.49*
Supplementary motor area	0	-2	60	R	292	5.11**

./..

f	Conj. (VF – fix) & (VF30 – fix)						
	Calcarine sulcus	-14	-98	2	L	660	4.28*
	Fusiform gyrus	34	-62	-18	R	751	4.29*
	Superior temporal gyrus	-54	-20	4	L	370	4.16*
	Middle temporal gyrus	60	-20	-4	R	383	4.07*

Table 1. Brain regions showing significant activation compared to baseline (fix) for voices (a), non-morphed faces (b), morphed faces (c), associations between voices and morphed faces (d) or unmorphed faces (e), and bimodal stimulations (f). x, y, z are Talairach coordinates of peak-height voxels. L = left hemisphere, R = right hemisphere. K = clusters size. * = P-values < 0.001 uncorrected, ** = P-values < 0.05 corrected for multiple comparisons.

Secondly, to isolate the cerebral regions specifically involved in the associative processes between faces and voices, the unimodal auditory and visual conditions were subtracted from the bimodal auditory-visual ones. Two subtractions were computed, with unmorphed faces in one case and morphed faces in the other contrasts ($[VF - (V + F)]$ and $[VF30 - (V + F30)]$ respectively). At first, regions activated during auditory-visual associative processes, regardless of the morphing of the faces, were assessed by a conjunction analysis of both subtractions. Activations were found in auditory regions (including the left middle temporal gyrus and the right superior temporal gyrus), visual regions (including bilateral inferior occipital gyri, the right middle occipital gyrus and the right fusiform gyrus), and in the left angular gyrus (Table 2a and Figure 3).

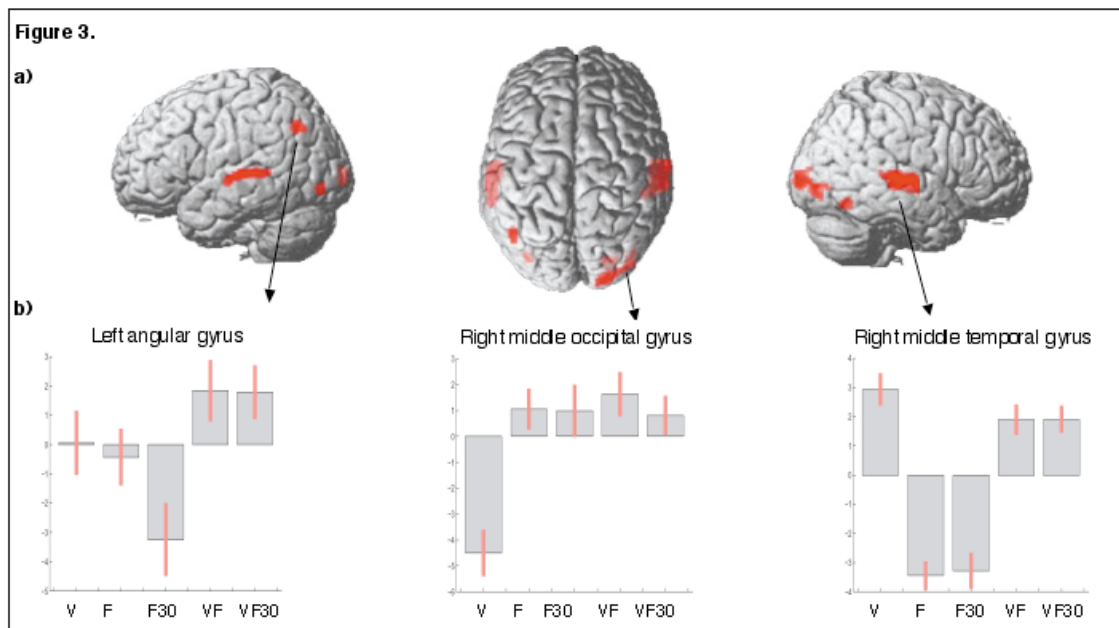


Figure 3. Brain regions activated in the conjunction of $[VF - (V + F)]$ and $[VF30 - (V + F30)]$. a) statistical parametric maps superimposed on MRI surface renders (left, top and right views). b) Size of effect for each condition in the left angular gyrus (left), the right middle occipital gyrus (middle) and the right middle temporal gyrus (right).

Then, compared to voices and unmorphed faces alone, non-facilitating voice–unmorphed face associations activated visual temporo-occipital and auditory temporal regions bilaterally, together with left frontal areas (Table 2b and Figure 4a). Finally, facilitating voice–morphed face associations enhanced the activity of the left middle temporal gyrus, the right superior temporal gyrus, the right middle occipital gyrus and the left angular gyrus (Table 2c and Figure 4b).

Brain regions	x	y	z	L/R	k	t-statistic
a $[VF-(V+F)]$ & $[VF30-(V+F30)]$						
Middle temporal gyrus	62	-26	2	R	351	5.27**
Superior temporal gyrus	-62	24	4	L	187	3.84*
Fusiform gyrus	40	-62	16	R	41	3.74*
Middle occipital gyrus	22	-98	0	R	160	4.28*
Inferior occipital gyrus	26	-84	-8	R	38	3.87*
Inferior occipital gyrus	-36	-80	-6	L	22	3.65*
Angular gyrus	-46	-66	38	L	43	3.60*
b $VF-(V+F)$						
Inferior occipital gyrus	36	-68	-8	R	2142	4.78*
Inferior occipital gyrus	-32	-92	-6	L	1182	4.66*
Middle occipital gyrus	-34	-70	32	L	58	4.70*
Lingual gyrus	12	-36	-4	R	257	4.21*
Middle temporal gyrus	48	-72	20	R	112	4.12*
Middle temporal gyrus	-64	-38	4	L	731	4.63*
Superior temporal gyrus	46	-36	10	R	357	4.10*
Inferior temporal gyrus	-48	-8	-28	L	92	3.75*
Inferior orbito-frontal gyrus	-46	44	-6	L	27	3.95*
Superior frontal gyrus	-24	68	4	L	34	3.37*
c $VF30-(V+F30)$						
Middle temporal gyrus	64	-22	2	R	207	4.09*
Superior temporal gyrus	-60	-28	8	L	114	3.73*
Middle occipital gyrus	18	-100	0	R	34	3.46*
Angular gyrus	-44	-68	40	L	92	3.76*
d $VF30-VF$						
No suprathreshold clusters						
e $VF-VF30$						
Inferior temporal gyrus	40	-12	-22	R	49	4.29*
Cuneus	-14	-78	40	L	27	3.75*
Fusiform gyrus	30	-36	-20	R	48	3.72*
Lingual gyrus	26	-58	-2	R	32	3.57*
cerebellum	8	-56	-4	R	34	4.31*
Olfactory gyrus	-2	12	-12	L	46	3.96*
Rolandic operculum	64	0	6	R	22	3.52*
Middle temporal gyrus	-48	-62	18	L	79	3.52*

Table 2. (a) Brain regions showing significant activation in the conjunction of both subtractions, (b) brain regions significantly activated by subtracting non-morphed faces (F) and voices (V) from non-moprhed faces / voices associations (VF), (c) brain regions significantly activated by subtracting morphed faces (F30) and voices (V) from moprhed faces / voices associations (VF30), (d) cerebral regions more activated in VF30 than in VF, (e) cerebral regions more activated in VF than in VF30. x, y, z are Talairach coordinates of peak-height voxels. L = left hemisphere, R = right hemisphere. K = clusters size. * = P-values < 0.001 uncorrected, ** = P-values < 0.05 corrected for multiple comparisons.

Thirdly, the direct comparison between VF and VF30 revealed no activation higher in VF30 than in VF (Table 2d), but an enhanced activation of the visual ventral pathway,

including the right inferior temporal, fusiform and lingual gyri, and the left cuneus was observed in VF compared to VF30 (Table 2e).

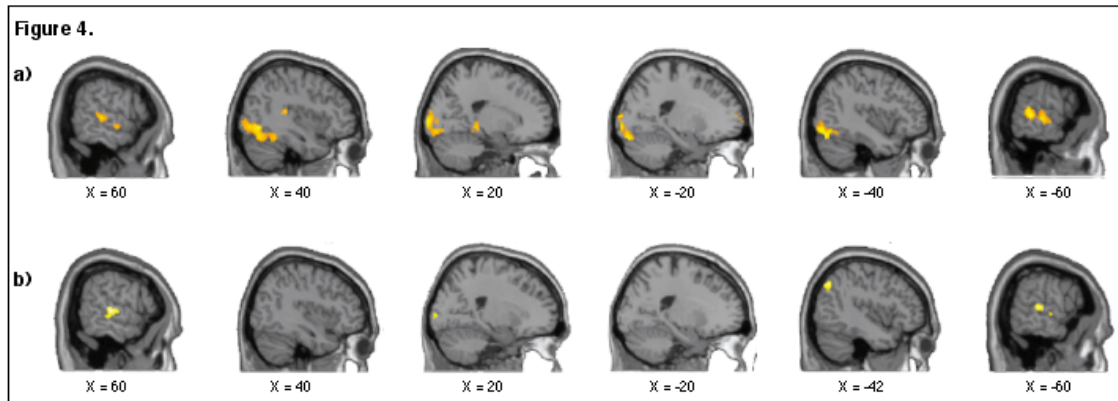


Figure 4. Statistical parametric maps superimposed on sagittal sections (x = distance from midline ; negative value : left hemisphere) representing the cerebral regions activated in a) the subtraction $[VF - (V + F)]$; b) the subtraction $[VF30 - (V + F30)]$.

Discussion

The aim of the present study was to investigate the cerebral correlates of voice–face interactions. The behavioral experiment was aimed at testing whether differences in the difficulty of recognizing voices and faces could account for the absence of cross-modal facilitation observed in our previous study. With fMRI we investigated the cerebral correlates of the cross-modal interactions between faces and voices in a semantic task, and compared the cerebral

activations when voice–face associations did and did not result in behavioral cross-modal facilitation.

Behavioral data

The behavioral experiment demonstrates that the ease with which voices and faces are recognized separately influences their cross-modal interactions. Indeed, we showed that voice–face associations in which unmorphed faces were easier to identify than voices did not facilitate identification, as they were not identified more rapidly than faces alone. On the contrary, when faces were modified by morphing to be as difficult to identify as voices, a clear facilitation effect appeared: voice–morphed face associations were identified significantly more quickly than either voices alone or morphed faces alone. This effect of complexity on the cross-modal interactions confirms our hypothesis that the voice helps in the rapid recognition of individuals when the face is difficult to identify (as in prosopagnosia; see for instance, Posamentier & Abdi, 2003), and that the simultaneous presentation of auditory and visual stimuli which are equally difficult to identify can improve the performance of participants.

fMRI data

The left Angular Gyrus as a “voice-face” integrator.

The neuroimaging data show that voice–face interactions, producing or not a behavioral facilitation, rely both (1) on the activation of unimodal visual and auditory areas, such as the right fusiform gyrus and in particular the Face Fusiform Area (Sergent et al., 1992; Kanwisher et al., 1997; Rhodes et al., 2004) and the right STS, known to be involved in voice identification processes (Giraud et al., 2004; Belin et al., 2004), and (2) on multimodal convergence regions such as the left angular gyrus which receives multiple inputs from modality specific sensory regions and provides a unique representation of the combined sensory features (Rämä & Courtney, 2005; Niznikiewicz et al., 2000; Booth et al., 2002, 2003). Unimodal activations during bimodal stimulations have already been described in non-human primates (Ghazanfar et al., 2005) and humans (Calvert et al., 1999), highlighting the importance of sensory brain regions within the auditory-visual integrative processes.

The present results underline the importance of the left parietal cortex, and in particular of the left angular gyrus, in cross-modal binding of visual and auditory information. The

left inferior parietal lobe is classically considered as a multimodal convergence region providing a unified concept of distinct sensory perceptions (Clark et al., 2000; Campanella et al., 2001). The activation of the left angular gyrus observed in this study is very similar to the activation of the left inferior parietal lobule we observed in a previous PET study (Campanella et al., 2001) which examined the cerebral activations elicited by the retrieval of face-name associations presented in the visual modality. The left parietal cortex thus seems to be closely implicated in the associative processes binding the different information relating to individuals, whatever this information is visual, verbal or auditory.

Modulation by task complexity.

In the previous section, we have shown that the left angular gyrus is activated in each voice-face combination, independently of their respective difficulties of recognition. This is in accordance with the hypothesis that faces and voices could be bound together automatically (Amedi et al., 2005), whatever the demands of the task to perform. This point was particularly important as it has been shown that parietal lesions can lead to severe multimodal integration

impairments (Valenza et al., 2004). Therefore, while faces alone activated FFA and voices alone recruited STS, our data clearly show that the association between these two separate representations is mediated by an independent, heteromodal, structure, i.e. the angular gyrus. However, besides the activation of "face areas", "voice areas" and their integrator, how is the complexity of such an integration implemented in the brain?

Our results suggest two main explanations. A first possibility is that participants confronted to associations between unmorphed faces and voices may have strategically based their decisions mainly on face processing rather than on the multimodal representation of voices and faces. Indeed, behavioral data showed that these associations were processed as fast as faces alone. Moreover, the direct comparison between the two associative conditions revealed that VF produced an enhanced activation of the visual ventral stream, including the inferior temporal gyrus, the fusiform gyrus, the cuneus and the lingual gyrus. It is then possible that in this case, though faces and voices were automatically integrated, a preferential processing of the visual information was sufficient to take a correct decision. Such attentional

bias towards a dominant modality has already been observed (Bushara et al., 2003), notably on the visuo-auditive categorization of emotional stimuli (Ethofer et al., 2006).

On the other hand, in the case of associations between voices and morphed faces, neither stimuli was easier to identify than the other and there was no reason for an attentional bias towards faces or voices to occur. The contrast between VF30 and VF did not reveal any supplementary activation in VF30 relative to VF, suggesting that the decisions were not based on one kind of processing rather than the other. Rather, it seems that responses were based on both stimuli and their integration into a unique representation, that allowed the identification to take place faster than with one stimulus alone.

An alternative possibility is that the crossmodal behavioral effects are reflected in the timing of occurrence of the different activations. For instance, the activation of the left angular gyrus could appear earlier in VF30 than in VF, as integration of voices and faces was crucial in the former condition relative to the second one, in which the different regions of the visual pathway could be activated earlier. Unfortunately, the block-design used in the fMRI experiment

does not allow us to infer any precise temporal information about the timing of occurrence of these activations. A replication of this study with the Event-Related Potentials (ERP) technique, or the simultaneous recording of fMRI and ERP data could help us to better understand how voices and faces interact and to draw up an anatomo-temporal model of such crossmodal interactions according to their respective difficulties of recognition.

Taken together, these results clearly show that cross-modal behavioral effects can be modulated by the degree of complexity in recognizing faces and voices information, and that these behavioral modulations could be reflected in different patterns of cerebral activity, and in particular by a differential involvement of the extrastriate visual areas, depending on the strategies of responses adopted by the participants. Finally, our results demonstrate the importance of unimodal and multimodal cerebral regions in cross-modal interactions and in particular of the left inferior parietal regions in binding together voice and face representations. They support a dynamic vision of cross-modal interactions in which heteromodal areas are not simply the final stage of a

hierarchical unimodal-to-multimodal processing model (Bushara et al., 2003). Rather, unimodal and multimodal regions work in parallel and influence each other.

These findings are of particular relevance for prosopagnosia, as it has been shown that brain-damaged patients rely on voices to identify friends and relatives, either overtly (Pallis, 1955) or covertly (de Haan et al., 1987). Our results, showing that voices help identification when face recognition is made more difficult, are, to our knowledge, the first to replicate these observations on healthy subjects. Moreover, some cases of patients with slowly progressive impairment of their recognition of familiar people from voices and faces have also been described (Gentileschi et al., 1999; Gainotti et al., 2003). Overall, these data on brain-damaged patients and healthy subjects are in line with the hypothesis that some people recognition disorders could be due to an impaired access from unimodal channels to a multimodal person-recognition system (Gainotti et al., 2003) and that some convergence point between different modes of recognition are needed to access semantics and to identify individuals efficiently.

Acknowledgments

This work was supported by grant No. 1.5.130.05F from the National Fund for Scientific Research (Belgium), and grant No. 01/06-267 from the Communauté Française de Belgique - Actions de Recherche Concertées (Belgium).

Frédéric Joassin is a Research Assistant and Mauro Pesenti a Research Associate at the National Fund for Scientific Research (Belgium). We thank the Radiodiagnosis Unit at the Cliniques St. Luc (Brussels) for its support, and Aurélie Mullier for her help during testing.

Conclusions des études 3 et 4

Ces deux études avaient pour but d'examiner les corrélats cérébraux des interactions crossmodales entre visages et voix, impliqués dans l'identification des personnes. L'étude 3, utilisant les potentiels évoqués, a examiné ces interactions lorsque les associations entre visages et voix ne produisaient pas de facilitation comportementale. L'utilisation d'une méthode soustractive, soustrayant les activités cérébrales produites par la présentation de visages et de voix isolés de celles produites par la présentation simultanée de ces stimulations, a permis d'isoler les activités spécifiquement impliquées dans les interactions crossmodales visuo-auditives. Nous avons principalement pu observer 3 complexes d'ondes cérébrales :

- un premier complexe, apparaissant vers 100 millisecondes après la présentation des associations, composé d'une positivité centrale et d'une négativité postérieure;
- un second complexe vers 175 millisecondes, composé d'une négativité centrale et d'une positivité postérieure;
- une activité positive centrale vers 280 millisecondes.

Une analyse de localisation de sources a en outre indiqué que ces activités pouvaient être expliquées par des activités de populations neuronales situées dans les cortex visuels et auditifs associatifs ainsi que dans le colliculus supérieur.

L'étude 4, reprenant la même méthodologie générale, nous a permis d'isoler les régions cérébrales impliquées dans l'intégration de visages et de voix. De plus, l'utilisation d'une technique de morphing rendant les visages aussi difficiles à identifier que les voix nous a permis de comparer ces processus intégratifs lorsque les associations présentées produisaient ou non une facilitation comportementale. Les principaux résultats de cette étude montrent que les associations entre visages non-morphés et voix ne produisaient pas de facilitation alors que les associations entre visages morphés et voix, aussi difficiles l'un que l'autre à identifier, produisaient une facilitation crossmodale, qui s'est marquée dans la condition bimodale par des temps de réaction plus courts et des performances plus élevées que dans les conditions unimodales. Nous avons également observé que ces associations, qu'elles produisent ou non une

facilitation comportementale, étaient associées à une activation des régions unisensorielles visuelles et auditives, ainsi qu'à une activation du gyrus angulaire gauche.

Ces deux études, faisant jouer la nécessaire complémentarité entre techniques électrophysiologiques et techniques d'imagerie cérébrale fonctionnelle, nous ont donc permis d'examiner à la fois la dimension temporelle et la dimension anatomo-fonctionnelle des interactions visuo-auditives entre visages et voix. Les conclusions que nous pouvons en tirer sont multiples:

1. il semble, en accord avec l'hypothèse d'Amedi, Von Kriegstein, van Atteveldt, Beauchamp et Naumer. (2005), que la présentation simultanée de visages et de voix implique des interactions automatiques entre ces stimulations, et que ces interactions sont indépendantes des effets comportementaux de facilitation ou d'interférence crossmodale. En effet, nous avons observé ces interactions aussi bien dans l'étude 3, dans laquelle le traitement des voix interférait avec le traitement des visages, que dans l'étude 4 où nous avons observé un effet de facilitation entre visages morphés et voix et une

absence d'effet facilitateur ou interférant entre visages naturels et voix. Signalons que les différences inter-études observées pour ce dernier type d'associations (interférence dans l'étude 3 vs. absence d'effet dans l'étude 4) peuvent être dues aux différences de paradigmes expérimentaux utilisés (conditions mixées dans l'étude 3 vs. conditions présentées "en blocs" dans l'étude 4);

2. ces interactions entre visages et voix prennent place entre 100 et 300 ms après la présentation des stimuli;
3. ces interactions impliquent l'activation conjointe des régions unisensorielles sous-tendant le traitement des différents types d'information, et de régions multimodales chargées de réaliser l'intégration des produits des différents traitements visuels et auditifs;
4. Néanmoins, sur base des résultats anatomo-fonctionnels, nous avons émis l'hypothèse que lorsque les visages et les voix sont présentés de manière naturelle (et sont donc plus faciles à identifier que les voix), les interactions visuo-auditives seraient dominées par le traitement visuel des visages. Par contre, lorsque les différentes informations constituant les associations

partageaient un même niveau de difficulté de reconnaissance, nous avons fait l'hypothèse que les jugements d'identité se basaient sur l'intégration des deux informations en une seule représentation unifiée.

Ces résultats doivent être mis en correspondance avec ceux observés dans l'étude de Campanella et al. (2001) et dans la première étude de cette thèse. D'un point de vue temporel, il est intéressant de constater que les deux études en potentiels évoqués ont donné des résultats assez similaires quant à la latence des activités cérébrales associatives. En effet, dans l'étude 1, nous avons principalement observé un complexe d'ondes cérébrales postérieure négative / antérieure positive prenant cours entre 285 et 300 millisecondes, c'est-à-dire dans la même fenêtre de latence que les complexes d'ondes observées dans l'étude 3.

Ces observations confirment celles de Teder-Sälejärvi et al. (2002), qui ont utilisé une méthodologie similaire pour isoler les processus d'intégration crossmodale entre stimulations simples (flashs lumineux et bips sonores). Ils ont réussi à prouver que les activités intégratrices apparaissant avant 100 millisecondes étaient dues à des artéfacts liés à la

méthode soustractive et reflétaient plutôt des mécanismes d'anticipation démarrant avant même l'apparition des stimuli que de réels mécanismes intégratifs. Selon eux, les véritables activités cérébrales intégratrices prenaient place à partir de 100 millisecondes.

La similarité de nos résultats au travers de deux études utilisant des méthodologies différentes (matériel uniquement visuel vs. visuo-auditif, sujets différents, matériel d'acquisition différent), à savoir l'observation dans les deux cas d'une onde centrale aux environs de 300 millisecondes spécifiques aux processus associatifs, confirme cette interprétation. Elle nous informe sur le timing d'apparition des activités associatives et nous montre que ces activités prennent place en même temps que les processus de traitement élaborés de chaque type d'information (comme l'encodage structural des visages ou l'accès au système sémantique). En outre, l'ensemble de nos observations concorde avec la vision des processus associatifs et crossmodaux développée par Driver & Spence (2000) et Bushara et al. (2003). Selon ces auteurs, les processus d'intégration ne doivent pas être vus comme le stade final d'une voie hiérarchique conduisant l'influx nerveux des régions unisensorielles vers les régions multimodales

d'intégration. Il semble plutôt que les traitements uni- et multimodaux prennent place ensemble au sein des processus de traitement de l'information, fonctionnent de manière parallèle et peuvent interagir et s'influencer mutuellement, notamment via des projections de type "feed-back" des régions hétéromodales vers les régions unimodales.

D'un point de vue anatomo-fonctionnel, il est intéressant de constater que les processus associatifs unissant les visages aux noms propres semblent agir de façon amodale, puisque tant l'étude PET-scan de Campanella et al. (2001) que l'étude 4 de cette thèse mettent en lumière une activation préférentielle de certaines régions du gyrus pariétal inférieur gauche, respectivement le lobule pariétal inférieur et le gyrus angulaire. Si l'on tient compte des différences entre les deux études (technique d'imagerie: PET-scan vs. fMRI, logiciels de traitement et d'analyses: SPM96 vs. SPM2, participants, paradigmes expérimentaux), la proximité spatiale des régions pariétales observées dans chaque étude peut nous laisser penser qu'une seule et même région a été activée dans les deux études.

Cette région pariétale est considérée comme une zone de convergence multimodale permettant de créer un concept unifié à partir de perceptions sensorielles distinctes (Damasio, 1989; Clark et al., 2000). De manière liée, Booth et al. (2002, 2003) ont suggéré que cette région cérébrale est impliquée dans l'intégration crossmodale des informations linguistiques visuelles et auditives.

L'ensemble de ces données nous pousse donc à conclure que les processus associatifs nous permettant d'unir en une représentation cohérente les différents attributs par lesquels nous identifions les personnes, impliquent l'activation des régions unimodales responsables du traitement de chaque type d'information. Ces régions interagissent, entre 100 et 300 millisecondes après la présentation des stimulations bimodales, avec une région particulière du lobe pariétal inférieur gauche correspondant au gyrus angulaire. Il semble également que cette région d'intégration puisse être considérée comme une région de convergence multimodale permettant d'unir les représentations des personnes familières présentées dans différents formats et modalités (visuelle, verbale et auditive).

Conclusions générales et perspectives



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

Le but principal de cette thèse était d'approfondir nos connaissances des mécanismes cérébraux permettant de lier entre elles les différentes représentations que nous avons des personnes que nous connaissons. Les trois attributs par lesquels nous identifions les personnes et auxquels nous nous sommes intéressés sont les visages, les noms de famille et les voix.

Etat des lieux

Dans la partie introductive, nous avons vu que les études consacrées à l'étude des corrélats cérébraux de ces trois types d'informations ne manquent pas. Elles ont principalement mis en évidence une implication de la partie antérieure du gyrus fusiforme droit (Puce et al., 1996; Kanwisher et al., 1997, 1999; Gauthier et al., 1999, 2000), des régions antérieures du lobe temporal inférieur gauche (Damasio et al., 1996; Gorno-Tempini et al., 1998; Tsukiura et al., 2002) et de la partie antérieure de la scissure temporale supérieure droite (Belin et al., 2000; Von Kriegstein et al., 2003; Fecteau et al., 2004) dans le traitement respectif des

visages, des noms propres et des voix. Néanmoins, si les processus de traitement de ces différents attributs commencent à être de mieux en mieux circonscrits, nous avons pu remarquer que la manière dont ils sont liés pour former une représentation unifiée des personnes familières reste assez méconnue, et que très peu d'études ont examiné directement les substrats neuraux permettant de réaliser cette liaison, ou ce "binding", à un autre niveau que le niveau perceptif (Calvert et al., 2000; Olson et al., 2002; Sekiyama et al., 2003).

Les quatre études réalisées dans ce cadre trouvent leur origine dans l'étude de Campanella et al. (2001). A notre connaissance, cette étude par PET-scan est la première à avoir utilisé la méthode soustractive, par ailleurs largement utilisée dans l'étude des processus associatifs entre stimuli simples (Giard & Peronnet, 1999; Calvert et al., 2001; Molholm et al., 2002; Teder-Sälejärvi et al., 2002), pour tenter d'isoler les régions cérébrales responsables du binding entre un visage et un nom de famille dans une tâche de reconnaissance de ce type d'associations. Cette étude a montré l'implication de trois régions cérébrales latéralisées dans l'hémisphère gauche dans ces processus associatifs, dont le lobule pariétal

inférieur, considéré comme étant la zone de convergence où les représentations des visages et des noms propres sont intégrées en une unique représentation qui sera comparée aux représentations stockées en mémoire à long terme.

Questions abordées

Ces résultats posent de nombreuses questions, dont trois ont constitué l'architecture de la présente thèse:

1. peut-on contraindre temporellement, au moyen des potentiels évoqués, les activations observées dans l'étude pré-citée?
2. Ces activités cérébrales sont-elles spécifiques aux processus associatifs entre visages et noms propres, ou témoignent-elles de processus associatifs plus généraux permettant de lier tout objet visuel au nom qui le désigne ?
3. Ces activités cérébrales sont-elles spécifiques à l'association de stimulations visuelles, ou sous-tendent-elles des processus associatifs amodaux, indépendants des modalités sensorielles dans lesquelles sont présentés les stimuli?

La première étude, répondant à la première question et utilisant une méthodologie identique à celle développée par Campanella et al. (2001), nous a montré que la récupération d'associations entre visages et noms de famille était associée à un pattern spécifique d'ondes cérébrales s'étendant des régions postérieures aux régions antérieures entre 280 et 300 millisecondes, et significativement latéralisé dans l'hémisphère gauche. Une analyse de localisation de sources a en outre révélé que les trois régions cérébrales activées dans l'étude par PET-scan expliquaient plus de 90 % de la variance des résultats. Ces données nous ont permis de dresser un modèle anatomo-temporel des processus de récupération d'associations visages-noms dans lequel (1) l'activation des régions frontales supérieures, apparaissant en premier lieu, reflèterait les processus de shift attentionnel nécessaires pour passer d'un type de représentations à un autre, (2) l'activation du lobe pariétal inférieur correspondrait à l'intégration en une représentation unifiée des visages et des noms présentés aux participants, et (3) l'activation du lobe préfrontal inférieur reflèterait la comparaison de la représentation ainsi formée avec celles stockées en mémoire à long terme.

La deuxième étude, répondant à la deuxième question, avait pour but de comparer les processus associatifs unissant les visages aux noms propres d'un part, et les objets aux noms communs d'autre part. Les objets sélectionnés pour cette étude étaient des animaux, ce qui nous a permis de comparer les processus associatifs relatifs à deux catégories "d'objets" vivants et animés. Les principaux résultats de cette étude découlaient de la soustraction des conditions contenant un seul type de stimulus des conditions associatives. Pour les deux types d'associations, nous avons observé différents complexes d'ondes électrophysiologiques (négativité postérieure / positivité antérieure) apparaissant aux alentours de 100 millisecondes et entre 300 et 400 millisecondes après la présentation des associations. Ces activités s'étendaient jusqu'aux régions frontales pour les associations entre animaux et noms communs, ce qui a été interprété comme reflétant des processus attentionnels et de mémoire de travail plus élaborés, probablement dus à la plus grande difficulté qu'ont eu les participants à catégoriser sémantiquement ces associations par rapport aux associations entre visages et noms propres. Nous avons également observé des activités temporelles bilatérales s'étendant jusque 400 millisecondes,

interprétées comme reflétant soit les processus associatifs unissant les objets (visages ou animaux) et leurs noms, soit l'accès au système sémantique.

Enfin, les études 3 et 4, répondant à la troisième question, avaient pour but d'examiner les processus associatifs entre les visages et les voix impliqués dans l'identification des personnes. Ces processus intégratifs crossmodaux ont été examinés via la technique des potentiels évoqués (étude 3) et via la fMRI (étude 4), les deux études utilisant à nouveau, et à notre connaissance pour la première fois, la technique soustractive en vue d'isoler les activités cérébrales spécifiques aux processus intégratifs. Ces deux études nous ont montré que (1) les interactions entre visages et voix impliquent l'activation des régions unimodales responsables du traitement de chaque type d'information, (2) ces régions interagissent entre 100 et 300 millisecondes avec une région particulière du lobe pariétal inférieur gauche correspondant au gyrus angulaire. D'un point de vue comportemental, ces deux études nous ont également permis d'observer que les effets de facilitation crossmodale entre visages et voix ne dépendent pas seulement de la

synchronisation spatiale et temporelle des informations (Sekuler et al., 1997; Calvert et al., 2001), mais également du niveau de complexité de leur identification.

Interprétation des résultats

Quelles conclusions peut-on tirer de l'ensemble de ces études ? Nous séparerons nos conclusions selon les trois aspects examinés tout au long de nos études, l'aspect comportemental, l'aspect électrophysiologique et l'aspect anatomo-fonctionnel.

1) D'un point de vue comportemental, nous avons appliqué dans toutes nos études un paradigme soustractif, comparant des conditions dans lesquelles un seul type d'informations étaient présenté aux participants à une condition dans laquelle deux types d'informations étaient présentés simultanément. L'hypothèse sous-jacente de cette méthode, d'un point de vue comportemental, est que la condition bimodale, présentant des paires d'informations congruentes, devrait être plus facile à réaliser et amener des temps de réaction plus courts que les conditions unimodales. Cette facilitation est en fait garante du fait que les deux

stimuli des conditions bimodales sont explicitement traités et intégrés. La présentation de matériel visuel (visages et noms propres écrits, ou images d'animaux et noms communs) semble effectivement remplir ces exigences. Ainsi, cet effet de facilitation a toujours été observé, aussi bien dans l'étude de Campanella et al. (2001) que dans nos études 1 et 2. Par contre, dans les études 3 et 4, nous avons pu observer que la présentation de stimuli dans des modalités différentes ne produisait pas de facilitation comportementale, voire produisait un effet d'interférence. Ces données vont à l'encontre de celles observées précédemment dans d'autres études sur les interactions crossmodales entre informations visuelles et auditives (Giard & Peronnet, 1999; Molholm et al., 2002; Teder-Sälejärvi et al., 2002). Deux facteurs peuvent expliquer ces divergences: (1) le fait que dans les études qui ont observé un effet de facilitation crossmodale, les stimuli utilisés étaient simples et abstraits (des ellipses ou des flashes lumineux et des bips sonores) et dénués de signification, (2) le fait que ces études ont toutes examiné des processus perceptifs de bas niveau (comme par exemple la détection de stimulations simultanées). A l'opposé, nos études ont utilisé des stimulations visuelles et auditives complexes, nécessitant

des capacités de discrimination fines permettant d'identifier chaque stimulation particulière. Corrolairement, les tâches envisagées ne concernaient pas la simple détection de synchronisations entre informations visuelle et auditive, mais impliquaient la mise en œuvre des capacités d'identification de chaque exemplaire particulier au sein de chaque modalité. Ces interprétations peuvent être rapprochées des résultats obtenus par Pourtois et de Gelder (2002). Ils ont observé, dans une étude utilisant la stimulation magnétique transcrânienne (TMS), que les interactions visuo-auditives dépendaient de réseaux de régions cérébrales distincts selon que les associations étaient ou non porteuses de sens (formes géométriques et sons *vs.* visages et voix).

Il a déjà été montré précédemment que les voix humaines sont plus difficiles à reconnaître que les visages (Ellis et al., 1997; Schweinberger et al., 1997). Néanmoins, à notre connaissance, nos études sont les premières à montrer que ces différences de complexité d'identification se marquent également dans les interactions crossmodales visuo-auditives. Différents facteurs interagissent dans la genèse des effets de facilitation crossmodale. Si la synchronisation temporelle et spatiale semblent être des

facteurs prépondérants dans l'apparition de ces effets, nos résultats montrent que les difficultés respectives avec lesquelles nous sommes capables d'identifier les visages et les voix influencent également leur intégration en une représentation unique permettant une identification plus rapide des personnes familières.

Il nous faut toutefois souligner que ces différences de difficulté d'identification ne peuvent être totalement détachées de l'influence de la synchronisation temporelle des stimulations sur les processus de facilitation. En effet, il est possible que cette différence de complexité entre identification des visages et identification des voix soit due au fait que la quantité d'informations disponibles à l'apparition des stimuli n'est pas la même pour les visages et les voix: dès l'apparition des visages, toute l'information visuelle nécessaire à l'identification est disponible instantanément, alors que pour les voix, l'information disponible augmente au fil du temps. Il est dès lors possible qu'il existe une désynchronisation temporelle entre le moment auquel un visage peut être identifié et le moment auquel une voix peut l'être.

Cette influence de la synchronisation temporelle devra être examinée dans nos recherches futures, notamment en faisant varier le moment d'apparition des voix par rapport aux visages.

2) D'un point de vue électrophysiologique, les différentes études réalisées dans le cadre de cette thèse nous apprennent que les processus associatifs entre informations de nature distincte et impliqués dans l'identification des personnes ne correspondent pas à la simple somme des activités cérébrales sous-tendant la reconnaissance de chaque type d'information. Au contraire, les soustractions des conditions unimodales des conditions bimodales nous ont révélé des activités cérébrales électrophysiologiques spécifiques aux processus associatifs, nous prouvant que les régions cérébrales impliquées dans le traitement de chaque information interagissent en vue de former des représentations unifiées des personnes familières.

Au travers de nos différentes études, nous avons pu remarquer que ces processus associatifs prenaient place principalement entre 100 et 400 millisecondes après l'apparition des stimuli. Ces observations vont à l'encontre de

celles de Giard & Peronnet (1999) et de Molholm et al. (2002), qui, utilisant également une méthodologie soustractive, ont observé des activités cérébrales associatives entre stimulations visuelles et auditives dès 45 millisecondes. Teder-Sälejärvi et al. (2002) contesteront ces observations et prouveront, par l'utilisation de filtres passe-haut que ces activités très précoces sont le reflet de processus d'anticipation apparaissant avant la présentation des stimuli. Selon ces auteurs, les activités réellement intégratrices ne prendraient place qu'aux alentours de 100 millisecondes. Nos données confirment cette interprétation et montrent qu'en effet, les processus associatifs, que ce soit entre visages et noms de famille ou entre visages et voix, opèrent dans la même latence que les processus de traitement de chaque type de stimulation. Cette observation est en accord avec la vision des processus intégratifs développée par Driver & Spence (2000) et Bushara et al. (2003) selon laquelle ces processus sont le fruit d'interactions entre régions unimodales traitant chaque type de stimulation et régions hétéromodales chargées de créer des représentations unifiées à partir des représentations distinctes élaborées par chaque système de traitement de l'information. Dans cette optique, les régions

hétéromodales ne sont pas vues comme le stade final d'une voie hiérarchique menant des régions traitant chaque stimulation isolément vers les régions hétéromodales intégrant le produit de ces traitements séparés. Il semble plutôt que les régions hétéromodales fonctionnent de manière parallèle aux régions unimodales et peuvent influencer les processus de traitement de chaque type d'information en fonction des autres traitements effectués en parallèle, à des fins d'adaptation à l'environnement.

Nous avons observé que les activités cérébrales associatives étaient principalement enregistrées au niveau des électrodes postérieures (occipito-temporales), fronto-centrales et temporales, et les analyses de localisation de source ont montré une implication des régions visuelles associatives, et notamment de la partie du gyrus fusiforme correspondant à la FFA, des régions auditives, et notamment de la partie antérieure du STS droit, de certaines régions préfrontales gauches, et de régions hétéromodales comme le lobe pariétal inférieur gauche ou le colliculus supérieur. Il nous faut garder à l'esprit que ces modélisations de dipôles ne sont qu'un moyen indirect de résoudre le problème inverse. Elles ne nous permettent que de faire des hypothèses

sur les générateurs neuraux responsables de l'activité électrique enregistrée à la surface du scalp. De plus, ces analyses dipolaires étaient contraignantes, c'est-à-dire que nous avons chaque fois contraint le programme à calculer la variance résiduelle des résultats à partir de dipôles placés virtuellement par nous-mêmes dans certaines régions précises du cerveau. Il est néanmoins frappant de constater que toutes ces analyses, aussi contraignantes soient-elles, ont toujours donné des résultats permettant d'expliquer plus de 90% de la variance des résultats observés. Nous avons donc de bonnes raisons de croire que, bien qu'hypothétiques, nos conclusions reflètent une certaine réalité des processus associatifs en jeu entre les différentes informations permettant d'identifier les personnes familières.

3) Les données anatomo-fonctionnelles récoltées dans l'étude 4 confirment l'implication des régions unimodales et hétéromodales dans les processus intégratifs liés à la reconnaissance des personnes. En effet, nous avons pu observer que les associations entre voix et visages généraient une activation des régions visuelles et auditives, mais également une activation du gyrus angulaire gauche. Outre le

fait que ces données sont en accord avec la vision des processus associatifs interactifs et réciproques entre régions sensorielles et régions d'intégration, elle nous renseigne également sur le rôle du lobe pariétal inférieur gauche dans ces processus. En effet, si l'on tient compte des différences de techniques et d'analyse, il apparaît que cette activation au sein du gyrus pariétal inférieur gauche est fortement similaire à celle observée par Campanella et al. (2001) dans leur étude par PET-scan examinant les processus associatifs entre visages et noms propres. Ils ont interprété cette activation comme étant le reflet des mécanismes permettant de créer une représentation unifiée des visages et des noms propres à partir des représentations élaborées par les régions cérébrales traitant chaque type d'information. Cette région cérébrale est classiquement considérée comme faisant partie du cortex associatif hétéromodal (Booth et al., 2002, 2003; Damasio, 1989; Clark et al., 2000), et la similitude des résultats entre les deux études, celle de Campanella et al. (2001) et l'étude 4 de cette thèse, nous pousse à conclure que cette région d'intégration peut être considérée comme une région de convergence multimodale permettant d'unir les

représentations des personnes familières présentées dans différents formats et modalités (visuelle, verbale et auditive).

Perspectives

Ces résultats ouvrent de nombreuses perspectives et nous amènent à nous poser différentes questions.

1) La première d'entre elles concerne la spécificité des processus associatifs entre visages et noms propres au sein des relations qu'entretiennent les objets et les noms qui les désignent. Tout d'abord, l'étude 2, examinant cette question au moyen des potentiels évoqués, n'a pas mis en lumière de différences majeures entre les associations visages-noms propres et les associations animaux-noms communs. Nous avons évoqué dans la discussion de cet article le fait que les animaux ont été sélectionnés pour leur caractère vivant et animé, qui les rapproche des visages humains. Néanmoins, nous avons vu que la récupération lexicale de différents types d'objets visuels recrute des régions contiguës au sein des lobes temporaux et frontaux gauches (Dehaene, 1995; Damasio et al., 1996; Graboswki et al., 1998, Gorno-Tempini et al., 1998). On ne peut dès lors pas exclure que des différences

existent entre les deux types d'associations, notamment en termes de régions cérébrales activées, différences que l'électrophysiologie ne peut mettre en lumière à cause de sa faible résolution spatiale. C'est pourquoi nous pensons que l'utilisation conjointe de techniques électrophysiologiques et anatomo-fonctionnelles, comme l'IRMf, pourrait nous permettre d'examiner plus précisément la question de la spécificité de ces processus associatifs.

Ensuite, si on peut faire l'hypothèse, sur base des résultats de l'étude 2, qu'il existe des processus associatifs généraux permettant d'unir un objet à son nom, il sera à l'avenir intéressant d'examiner si ces processus sont identiques quelle que soit la nature des objets présentés. En effet, il est classiquement admis qu'il existe une distinction dans l'organisation de nos représentations entre objets vivants et non-vivants. Les études de patients cérébrolésés ont permis d'établir une double dissociation entre (1) l'altération de la reconnaissance d'objets vivants (comme les animaux et les plantes) et la préservation de la reconnaissance d'objets non-vivants (comme les outils, Warrington & Shallice, 1984; McCarthy & Warrington, 1988), et (2) l'altération de la reconnaissance d'objets non-vivants et la préservation de la

reconnaissance d'objets vivants (Hillis & Caramazza, 1991). Une des explications de cette dissociation est que la mémoire sémantique est organisée en catégories. Selon cette théorie d'une organisation sémantique "domain-specific" (Caramazza & Shelton, 1998), des sous-systèmes neurologiques spécialisés pourraient s'être développés sous la contrainte de l'évolution pour catégoriser efficacement les différentes stimulations de l'environnement. Cette spécialisation se marquerait par des localisations cérébrales distinctes pour le traitement des objets de nature différente, permettant d'expliquer les troubles spécifiques de reconnaissance évoqués plus haut. Par exemple, il a été montré que la perception d'outils, mais pas la perception d'animaux, était associée à l'activation des régions cérébrales prémotrices, suggérant que les représentations visuelles des outils sont fortement corrélées avec la représentation de leur utilisation (Chao & Martin, 2000). Si l'hypothèse d'une organisation du système sémantique en catégories distinctes est fondée, il serait dès lors intéressant de comparer les processus associatifs entre visages et noms propres d'une part, et outils et noms communs d'autre part, afin d'examiner si les liens formés entre les différents attributs des objets de notre

environnement sont sous-tendus par des substrats neuraux communs ou s'ils sont spécifiques aux différentes catégories d'objets, notamment selon leur nature "naturelle" ou "artéfactuelle".

Enfin, nous avons vu dans les chapitres introductifs que certains patients présentaient une anomie spécifique aux noms de personnes, tout en étant toujours capables de dénommer les lieux géographiques (Yasuda & Ono, 1998; Miceli et al., 2000; Lyons et al., 2002). Ces observations suggèrent qu'au sein même de la classe des noms propres, il pourrait exister des mécanismes spécifiques au traitement des noms de famille.

Comme nous pouvons le constater, la question de la spécificité des associations entre visages et noms de famille est loin d'être résolue et les recherches futures, comparant les processus associatifs entre les objets visuels et les noms qui les désignent, que ce soit entre catégorie d'objets ou au sein même d'une catégorie lexicale, devrait nous permettre de mieux comprendre l'organisation des connaissances au sein du cerveau et la manière dont les différentes informations relatives aux stimulations de notre environnement sont liées au sein de cette organisation.

2) Un autre axe de recherches concerne les bases neurales de l'encodage et de la récupération des associations voix-visages. Il s'appuie sur les résultats de l'étude de Campanella et al. (2001) et les résultats de l'étude 4. Ces deux études nous ont amené à postuler que le lobe pariétal inférieur gauche est fortement impliqué dans les processus permettant d'associer en mémoire les différentes informations relatives aux personnes familières, que ces informations soient visuelles (comme les visages), auditives (comme les voix) ou verbales (comme les noms de famille). Néanmoins, ces résultats posent plusieurs questions, notamment quant aux mécanismes neuraux permettant d'associer visages et voix.

Ainsi, les études réalisées jusqu'à présent se sont penchées sur les processus crossmodaux impliqués dans la reconnaissance d'associations visages-voix. Or, pour une meilleure compréhension des mécanismes cérébraux permettant de lier entre elles différentes informations relatives aux personnes familières, il serait intéressant d'examiner quelles régions cérébrales sous-tendent l'encodage de telles associations. En effet, non seulement il a été suggéré qu'il existe des systèmes neuraux distincts sous-

tendant l'encodage et la reconnaissance des informations (Maguire, Frith, & Cipolotti, 2001), mais de plus, la capacité d'apprendre et de se remémorer des associations entre des informations d'origine distinctes est un aspect essentiel de notre système mnésique (Sperling et al., 2001).

Différentes études d'imagerie cérébrale fonctionnelle ont montré une implication de l'hippocampe, des gyri préfrontaux, et des régions visuelles occipito-temporales lors de l'encodage d'associations entre visages et noms propres (Sperling et al., 2001, 2003, Herholz, et al., 2001). Guo et al. (2005) ont utilisé quant à eux la technique des potentiels évoqués pour examiner les processus cérébraux mis en œuvre lors de l'encodage de visages, de voix ou d'associations entre visages et voix. Ils ont observé que l'apprentissage fructueux de chaque type d'information (c'est-à-dire un encodage menant ultérieurement à un souvenir précis) était associé à un pattern distinct d'activités cérébrales entre 200 et 800 millisecondes après la présentation des stimuli. Néanmoins, à notre connaissance, aucune étude n'a jamais comparé les corrélats cérébraux de l'encodage et de la reconnaissance d'associations visages-voix. En outre, la méthode soustractive n'a encore jamais été appliquée à cette problématique. Nous

pensons donc qu'il serait important, pour une meilleure compréhension des processus à l'œuvre lors de l'encodage et de la récupération de telles associations, de mettre au point des expériences permettant (1) d'isoler les régions cérébrales responsables de l'intégration de visages et de voix lors de l'encodage de telles associations, (2) de comparer les régions cérébrales activées lors de l'encodage et lors de la reconnaissance ultérieure de ces associations.

3) Une autre question concerne l'influence des facteurs sémantiques sur les processus associatifs entre visages et voix. Relativement peu d'études se sont à ce jour penchées sur les interactions crossmodales entre visages et voix humaines, la plupart des recherches dans ce domaine ayant utilisé des stimuli visuels et auditifs simples et sans réelle signification intrinsèque (comme par exemple des flash lumineux et des bips sonores, Giard & Peronnet, 1999 ; Teder-Sälejärvi et al., 2002 ; Bushara et al., 2003). Néanmoins, à la différence de ce type de stimulations, les visages et les voix humaines ont une forte signification, à la fois intrinsèque puisque ce sont deux types d'informations hautement écologiques et primordiaux dans nos interactions sociales, et

extrinsèque car de nombreuses informations biographiques, épisodiques et affectives y sont associées. Pourtois & de Gelder (2002) ont d'ailleurs montré, en utilisant la stimulation magnétique transcranienne (TMS), que les interactions crossmodales entre stimuli avec ou sans signification (paires composées de visages et de voix vs. de cercles et de sons) étaient sous-tendues par des régions cérébrales distinctes. On peut dès lors s'interroger sur la manière dont les informations sémantiques influencent ces processus intégratifs. Les recherches futures devraient nous aider à mieux comprendre le rôle de ces facteurs sémantiques, non seulement en comparant les processus crossmodaux entre stimulations porteuses ou non de sens, mais également en examinant, au sein même des associations visages-voix l'influence des informations sémantiques, par exemple en comparant les mécanismes neuraux mis en œuvre lors de la récupération d'associations apprises ou d'associations relatives à des personnes célèbres, voire familiales.

4) Enfin, une quatrième piste d'investigation pourrait concerner la question de la spécificité « intra-domaine » des processus associatifs impliqués dans la reconnaissance des

personnes au sein même des processus de traitement des visages. En effet, le modèle classique des processus d'identification des visages de Bruce & Young (1986) postule une indépendance fonctionnelle entre le traitement de l'identité et le traitement de l'expression faciale émotionnelle. Bien que de plus en plus mise en question, cette dissociation a été confirmée tant par l'étude de patients prosopagnosiques incapables d'identifier des visages familiers mais toujours à même de reconnaître les expressions faciales (Posamentier, 2002, in Posamentier & Abdi, 2003), que par différentes études électrophysiologiques (Munte, Brack, Grootheer, Wieringa, Matzke, & Johannes, 1998 ; Streit, Wulver, & Brinkmeyer, 2000, pour une revue, Posamentier et Abdi, 2003). En outre, certaines études suggèrent que la perception multisensorielle d'émotions implique l'activation de régions hétéromodales d'intégration (de Gelder & Vroomen, 2000 ; Pourtois et al., 2005)

La question qui se pose dès lors concerne la spécificité des activités intégratrices cross-modales mises en lumière dans nos études précédentes : sont-elles spécifiques au processus d'identification des personnes, ou s'étendent-elles aux différents processus de traitement des informations

relatives aux personnes, telles que la reconnaissance des émotions ? Dans le premier cas, les activités cérébrales d'intégration visuo-auditive devraient être différentes selon qu'elles portent sur l'identification d'associations visages-voix ou sur la reconnaissance visuo-auditive d'émotions. Dans le deuxième cas, ces activités intégratrices devraient être identiques quelle que soit la tâche.

Cette question trouve toute sa pertinence si elle est appliquée à l'étude des processus intégratifs visuo-auditifs en psychopathologie, et notamment auprès de patients atteints de schizophrénie. Il est en effet reconnu que les déficits de reconnaissance des informations affectives sont un des problèmes majeurs que peuvent rencontrer les patients schizophrènes dans leurs interactions sociales (Edwards, Jackson, & Pattison, 2002). Il semble en outre que l'intégration visuo-auditive soit un processus déficitaire chez les personnes souffrant de schizophrénie, tant pour l'intégration de stimuli complexes (de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003) que pour l'intégration de stimuli émotionnels (de Gelder, Vroomen, de Jong, Masthof, Trompenaars, & Hodiamont, 2005).

D'autre part, il a été montré que certains patients schizophréniques présentent une asymétrie des lobes pariétaux, et en particulier une atrophie du gyrus angulaire gauche (Niznikiewicz et al., 2000). Le lobe pariétal inférieur, dont fait partie le gyrus angulaire, est connu pour être une structure corticale impliquée dans les processus associatifs hétéromodaux (Clark et al., 2000 ; Campanella et al., 2001).

Il semble dès lors que l'étude de cette question auprès de participants neurologiquement sains et d'une population de patients schizophrènes pourrait nous aider à mieux comprendre (1) les mécanismes cérébraux déficitaires dans la schizophrénie, (2) le rôle du gyrus angulaire gauche dans les processus d'intégration visuo-auditif de visages et de voix impliqués dans l'identification des personnes et le traitement émotionnel des stimuli sociaux.

Conclusions

En conclusion, les études présentées dans cette thèse nous ont permis d'investiguer différents aspects des processus associatifs nous permettant quotidiennement d'identifier les personnes familières sur base d'informations

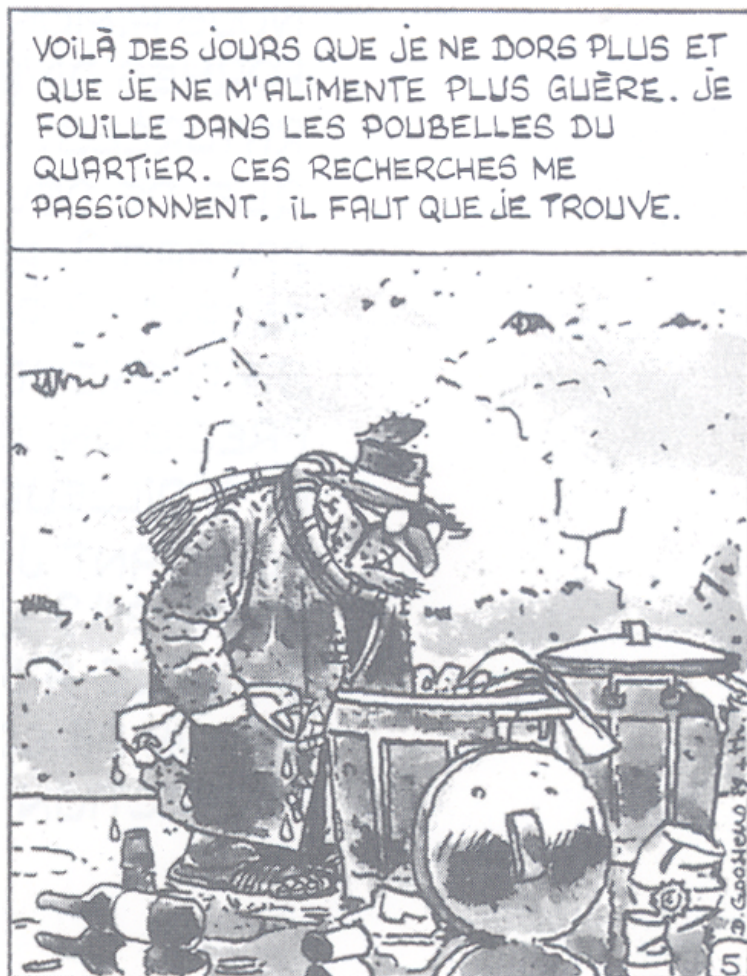
de nature distincte. Les résultats de nos expériences nous ont montré que ces processus sont sous-tendus par des mécanismes neuraux particuliers, pris en charge à la fois par les régions cérébrales traitant chaque type d'information et par des régions hétéromodales. Nous avons suggéré que ces différentes régions travaillent de manière interactive et s'influencent de façon réciproque. Néanmoins, l'étude de ces interactions n'en est qu'à ses débuts, et de nombreuses expériences sont encore à réaliser si l'on veut mieux comprendre comment nous sommes capables de créer des représentations unifiées des personnes sur base de ces activités cérébrales particulières. La compréhension de ces mécanismes associatifs est certes importante d'un point de vue fondamental, mais elle permettra également à l'avenir de mieux comprendre les processus cérébraux déficitaires dans diverses affections psychopathologiques telle que la schizophrénie.

Von Kriegstein et al. (2005) ont montré chez une patiente souffrant de prosopagnosie développementale, que la reconnaissance des voix familières entraînait une activation de la FFA et des régions auditives, tout comme chez les participants contrôles. A la différence de ces derniers, la

patiente montrait une activation diminuée des régions activées normalement par la familiarité des stimuli. Ces résultats indiquent que (1) la reconnaissance des voix bénéficie de la capacité à reconnaître les visages, (2) les associations entre visages et voix sont assurées – a minima – par l'activation conjointe des régions traitant chaque type d'information. Ces résultats suggèrent donc qu'une meilleure compréhension des mécanismes associatifs entre visages et voix pourrait nous permettre d'envisager de nouvelles manières d'intervenir plus efficacement dans la recherche de mécanismes compensatoires palliant les difficultés rencontrées dans les pathologies neuropsychologiques liées à la reconnaissance des personnes, comme la prosopagnosie ou la phonagnosie.

Références

bibliographiques



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

- Amedi, A., Von Kriegstein, K., van Atteveldt, N.M., Beauchamp, M.S., & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res.*, **166**, 559-571.
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optical bimodal integration. *Current Biology*, **14**, 257-62.
- Allison, T., Ginter, H., McCarthy, G., Nobre, A.C., Puce, A., Luby, M., & Spencer, D.D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, **71**(2), 821-825.
- 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.
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, **4**(7), 267-278.
- Andreasen, N.C., O'Leary, D.S., Arndt, S., Cizadlo, T., Hurtig, R., Rezai, K., Watkins, G.L., Ponto, L.B., & Hichwa, R.D. (1996). Neural substrates of facial recognition. *Journal of Neuropsychiatry and Clinical Neurosciences*, **8**(2), 139-146.
- Assal, G., Aubert, C., & Buttet, J. (1981). Asymétrie cérébrale et reconnaissance de la voix. *Revue Neurologique (Paris)*, **137**, 255-268.
- Bahrnick, L.E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, **6**, 429-451.

- Bahrick, L.E. (1988). Intermodal learning in infancy : learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, **59**, 197-209.
- Bahrick, L.E. (1992). Infant's perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology*, **53**(2), 180-199.
- Bahrick, L.E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Dev. Psych.*, **41**(3), 541-552.
- Banati, R.B., Goerres, G.W., Tjoa, C., Aggleton, J.P., & Grasby, P. (2000). The functional anatomy of visual-tactile integration in man : a study using positron emission tomography. *Neuropsychologia*, **38**(2), 115-124.
- Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia : face-blind from birth. *Trends in Cognitive Sciences*, **9**(4), 180-187.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P, & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**(6767), 309-312.
- Belin, P., Zatorre, R.J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* **13**, 17-26.
- Belin, P., & Zatorre, R.J. (2003). Adaptation to speaker's voice in right anterior temporal-lobe. *Neuroreport*, **14**, 2105-2109.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, **8**(3), 129-135.

- 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., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., & Pernier, J. (1999). ERP Manifestations of Processing Printed Words at Different Psycholinguistic Levels : Time Course and Scalp Distribution. *Journal of Cognitive Neuroscience*, **11**, 235-260.
- Bentin, S., & Deouell, L.Y. (2000). Structural encoding and identification in face processing : ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, **17**, 35-54.
- Bertrand, O., Perrin, F., & Perrier, J. (1985). A Theoretical Justification of the Average Reference in Topographic Evoked Potentials Studies. *Clinical Neurophysiology*, **62**, 462-464.
- Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception : role of gamma band responses. *Neuroreport*, **13**(14), 1727-30.
- Bodamer, J. (1947). Die Prosop-Agnosia (Die Agnosie des Physionomeerkennens). *Archiv für Psychiatrie und Nervenkrankheiten*, **179**, 6-33.
- Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., & Mesulam, M.M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage*, **16**, 7-22.
- Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., & Mesulam, M.M. (2003). Relation between brain activation and lexical performance. *Human Brain Mapping*, **19**, 155-169.

- Bötzel, K., Schulze, S., & Stodieck, R.G. (1995). Scalp Topography and Analysis of Intracranial Sources of Face-Evoked Potentials. *Experimental Brain Research*, **104**, 135-143.
- Bower, T.G.R., Broughton, J.M., & Moore, M.K. (1970). The coordination of visual and tactual inputs in infants. *Perception and Psychophysics*, **8**, 51-53.
- Brédart, S., Brennen, T., & Valentine, T. (1997). Dissociations between the processing of proper and common names. *Cognitive Neuropsychology*, **14**(2), 209-217.
- Brosch, M., Bauer, R., & Eckhorn, R. (1997). Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex. *Cerebral Cortex*, **7**(1), 70-76.
- Bruce, V., & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology*, **77**, 305-327.
- Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive model. *British Journal of Psychology*, **81**, 361-380.
- Burton, A.M., & Bruce, V. (1992). I recognize your face but I can't remember your name: a simple explanation. *British Journal of Psychology*, **83**, 45-60.
- Bushara, K.O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience*, **21**, 300-304.
- Bushara, K.O., Hanakawa, T., Immish, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience*, **6**(2), 190-195.

- Buxbaum, L.J., Glosser, G., & Coslett, H.B. (1999). Impaired face and word recognition without object agnosia. *Neuropsychologia*, **37**, 41-50.
- Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., & David, A.S. (1997). Activation of auditory cortex during silent lipreading. *Science*, **276**, 593-596.
- Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iversen, S.D., & David, A.S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, **10**, 2619-2623.
- Calvert, G.A., Campbell, R., & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Current Biology*, **10**, 649-657.
- Calvert, G.A. (2001). Crossmodal processing in the human brain : insights from functional neuroimaging studies. *Cerebral Cortex*, **11**, 1110-1123.
- Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of multisensory integration sites by application of electrophysiological criteria to the BOLD response. *Neuroimage*, **14**, 427-438.
- Campanella, S., Hanoteau, C., Dépy, D., Rossion, B., Bruyer, R., & Crommelinck, M. (2000). Right N170 modulation in a face discrimination task : an account for categorical perception of familiar faces. *Psychophysiology*, **37**, 796-806.
- Campanella, S., Joassin, F., Rossion, B., De Volder, A., Bruyer, R., & Crommelinck, M. (2001). Association of the distinct

- visual representations of faces and names : a PET activation study. *Neuroimage*, **14**, 873-882.
- Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guérit, J.M. (2002). Categorical Perception of Happiness and Fear Facial Expressions : an ERP Study. *Journal of Cognitive Neuroscience*, **14**(2), 1-18.
- Caramazza, A., & Shelton, J. (1998). Domain-specific knowledge in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, **10**, 1-34.
- Carmel, D., & Bentin, S. (2002). Account of the N170 face-effect : a reply to Rossion, Curran and Gauthier. *Cognition*, **85**, 197-202.
- Carney, R., & Temple, C.M. (1993). Prosopagnosia ? A possible category-specific anomia for faces. *Cognitive Neuropsychology*, **10**(2), 185-195.
- Chao, L.L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, **12**(4), 478-484.
- Cipolotti, L., McNeil, J.E., & Warrington, E.K. (1993). Spared written naming of proper nouns: a case report. *Memory*, **1**(4), 289-311.
- Clark, C.R., Egan, G.F., McFarlane, A.C., Morris, P., Weber, D., Sonkkilla, C., Marcina, J., & Tochon-Danguy, H.J. (2000). Updating working memory for words: a PET activation study. *Human Brain Mapping*, **9**(1), 42-54.
- Clarke, S., Lindemann, A., Maeder, P., Borruat, F.X., & Assal, G. (1997). Face recognition and postero-inferior hemispheric lesions. *Neuropsychologia*, **35**(12), 1555-1563.

- Cohen, N.J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal System and Declarative (Relational) Memory : Summarizing the Data from Functional Neuroimaging Studies. *Hippocampus*, **9**, 83-98.
- Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.
- Coull, J.T. (1998). Neural Correlates of Attention and Arousal : Insights from Electrophysiology, Functional Neuroimaging and Psychopharmacology. *Progress in Neurobiology*, **55**, 343-361.
- Damasio, A.R., Damasio, H., & Van Hoesen, G.W. (1982). Prosopagnosia : anatomical basis and behavioral mechanisms. *Neurology*, **32**, 331-341.
- Damasio, A.R. (1989). Time-Locked Multiregional Retroactivation : a System-level Proposal for Neural Substrates of Recall and Recognition. *Cognition*, **33**, 25 – 62.
- Damasio, A.R., Damasio, H., Tranel, D., & Brandt, J.P. (1990). The Neural Regionalization of Knowledge Access : Preliminary Evidence in *The Brain*. Cold Spring Harbor Symposium Series.
- Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D., & Damasio, A.R. (1996). A neural basis for lexical retrieval. *Nature*, **180**, 499-505.
- Debruille, J.B., Guillem, F., & Renault, B. (1998). ERPs and chronometry of face recognition : following-up Seeck et al. and George et al. *Neuroreport*, **9**(15), 3349-3353.

- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ears and by eyes. *Cognition & Emotion*, **14**, 289-311.
- de Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio-visual integration in schizophrenia. *Schizophr. Res.*, **59**(2-3), 211-218.
- de Gelder, B., Vroomen, J., de Jong, S.J., Masthof, E.D., Trompenaars, F.J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophr. Res.*, **72**(2-3), 195-203.
- de Haan, E.H.F., Young, A.W., & Newcombe, F. (1987). Face recognition without awareness. *Cogn. Neuropsychol.*, **4**, 385-415.
- Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport*, **6**, 2153-2157.
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B., & Rivière, D. (2001). Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming. *Nature Neuroscience*, **4**(7) : 752-758.
- De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Aperceptive and associative forms of prosopagnosia. *Cortex*, **27**, 213-221.
- De Renzi, E., Peroni, D., Carlesimo, G.A., Silveri, M.C., & Fazio, F. (1994). Prosopagnosia Can Be Dissociated with Damage Confined to the Right Hemisphere – an MRI and PET Study and a Review of the Literature. *Neuropsychologia*, **32**(8), 893-902.

- De Renzi, E., & di Pellegrino, G. (1998). Prosopagnosia and alexia without object agnosia. *Cortex*, **34**, 404-415.
- Di Russo, F., & Spinelli, D. (1999). Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vision Research*, **39**, 2975-2985.
- Driver, J. (1996). Enhancement of selective listening by illusory mis-location of speech sounds due to lip-reading. *Nature*, **381**, 66-68.
- Driver, J., & Spence, C. (2000). Multisensory perception : beyond modularity and convergence. *Current Biology*, **10**(20), 731-735.
- 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**(3), 278-289.
- Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, **17** (2), 249-261.
- Duhamel, J.R., Colby, C.L., & Goldberg, M.E. (1991). In : Brain and Space (Paillard, J, Ed.). pp : 223-236. New York : Oxford University Press.
- Edwards, J., Jackson, H.J., & Pattison, P.E. (2002). Emotion recognition via facial expression and affective prosody in schizophrenia : a methodological review. *Clin. Psych. Rev.*, **22**(6), 789-832.
- Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye detector ? *Neuroreport*, **9**, 2945-2948.

- Eimer, M. (2000a). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, **111**, 694-705.
- Eimer, M. (2000b). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Cognitive Brain Research*, **10**, 145-158.
- Ellis, H.D., Jones, D.M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology*, **88**, 143–156.
- Engel, A.K., König, P., Kreiter, A.K., & Singer, W. (1991). Inter-hemispheric synchronisation of oscillatory neuronal responses in cat visual cortex. *Science*, **252**, 1177-1179.
- Eslinger, P.J., Easton, A., Grattan, L.M., & Van Hoesen, G.W. (1996). Distinctive forms of partial retrograde amnesia after asymmetric temporal lobe lesions: possible role of the occipitotemporal gyri in memory. *Cerebral Cortex*, **6**, 530-539.
- Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., Reiterer, S., Grodd, W., & Wildburger, D. (2006). Impact of voice on emotional judgment of faces : an event-related fMRI study. *Human Brain Mapping*, article online in advance of print.
- Farah, M.J. (1990). Visual Agnosia. Disorders of Object Recognition and what they tell us about Normal Vision. Cambridge : a Bradford Book, MIT Press.
- Farah, M.J., Wilson, K.D., Drain, M., & Tanaka, J.N. (1998). What is “special” about face perception? *Psychological Review*, **105**(3), 482-498.

- Fecteau, S., Armony, J.L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex ? An fMRI study. *Neuroimage*, **23**(3), 840-848.
- Fecteau, S., Armony, J.L., Joanette, Y., & Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *Journal of Neurophysiology*, **94**(3), 2251-2254.
- Feindberg, T.E., Schindler, R.J., & Ochoa, E. (1994). Associative visual agnosia and alexia without prosopagnosia. *Cortex*, **30**, 395-412.
- Fletcher, P.C., Shallice, T., Frith, C.D., Frackowiak, R.S., & Dolan, R.J. (1996). Brain activity during memory retrieval: the influence of imagery and semantic cueing. *Brain*, **119**, 1587-1596.
- Flude, B.M., Ellis, A.W., & Kay, J. (1989). Face processing and name retrieval in an anomic aphasic: names are stored separately from semantic information about familiar people. *Brain and Cognition*, **11**(1), 60-72.
- Fodor, J. (1983). The modularity of mind. Cambridge : MIT Press.
- Fort, A., Delpuech, C., Pernier, J., & Giard, M.H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, **12**(10), 1031-9.
- Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., & Schroeder, C.E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, **10**(1-2), 77-83.

- Frens, M. A., Van, O. A., & Van-der, W. R. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept. Psychophys.*, **57**, 802-816.
- 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.
- Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform « face area » increases with expertise in recognizing novel objects. *Nature Neuroscience*, **2**, 568-573.
- Gauthier, I., Behrmann, M., & Tarr, M.J. (1999). Can face recognition be dissociated from object recognition ? *Journal of Cognitive Neuroscience*, **11**, 349-370.
- Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, **3**, 191-197.
- Gauthier, I., Tarr, M.J., Moylan, 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. *Journal of Cognitive Neuroscience*, **12**, 495-504.
- Gentileschi, V., Sperber, S., & Spinnler, H. (1999). Progressive defective recognition of familiar people. *Neurocase*, **5**, 407-424.
- 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.
- Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., & Logothetis, N.K. (2005). Multi-sensory integration of dynamic faces and

- voices in rhesus monkey auditory cortex. *J. Neurosci.*, **25**(20), 5004–5012.
- Giard, M.H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, **11**(5), 473-490.
- Giovanello, K.S., Alexander, M., & Verfaellie, M. (2003). Differential impairment of person-specific knowledge in a patient with semantic dementia. *Neurocase*, **9**, 15-26.
- Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex*, **14**(3), 247–255.
- Gomez, C.M., Clark, V.P., Luck, S.J., Fan, S., & Hillyard, S.A. (1994). Sources of Attention Sensitive Visual Event-Related Potentials. *Brain Topography*, **7**, 41-51.
- Gorno-Tempini, M.L., Price, C.G., Josephs, O., Vandenberghe, R., Cappa, S.F., Kapur N., & Frackowiack, R.S.J. (1998). The neural system sustaining face and proper-name processing. *Brain*, **121**, 2103-2118.
- Graboswki, T.J., Damasio, H., & Damasio, A.R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage*, **7**(3), 232-243.
- Graboswki, T.J., Damasio, H., Tranel, D., Ponto, L.L., Hichwa, R.D., & Damasio, A.R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Human Brain Mapping*, **13**(4), 199-212.

- Gray, C.M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences of USA*, **86**, 1698-1702.
- Graziano, M.S., & Gross, C.G. (1998). Spatial maps for the control of movement. *Current Opinion in Neurobiology*, **8**, 195-201.
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, **13**, 159-166.
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, **7**, 555-562.
- Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., & Pantev, C. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neuroscience Letters*, **348**, 13-16.
- Guo, C., Voss, J.L., & Paller, K.A. (2005). Electrophysiological correlates of forming memories for faces, names and face-name associations. *Cognitive Brain Research*, **22**, 153-164.
- Hadjikani, N., & Roland, P.E. (1998). Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron tomographic study. *The Journal of Neuroscience*, **18**(3), 1072-1084.
- Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B., Marinkovic, K., & Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping*, **7**, 29-37.

- Hatwell, Y. (1994). Transferts intermodaux et intégration intermodale. In Richelle, M., Requin, J. & Robert, M. (Eds). Traité de Psychologie Expérimentale (tome 1). Paris : Presses Universitaires de France, pp. 543-584.
- 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. *Journal of Neuroscience*, **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**(1), 189-199.
- Haxby, J.V., Hoffman, E.A., & Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, **4**(6), 223-233.
- Heinze, H.J., Mangun, G.R., Burchert, W., Hinrichs, H., Scholtz, M., Münte, T.F., Göss, A., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M.S., & Hillyard, S.A. (1994). Combined Spatial and Temporal Imaging of Brain Activity During Selective Attention in Humans. *Nature*, **372**, 543-546.
- Heisz, J.J., Watter, S., & Shedden, J.M. (2006). Progressive N170 habituation to unattended repeated faces. *Vision Research*, **46**, 47-56.
- Henson, R.N., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., & Rugg, M.D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral Cortex*, **13**, 793-805.

- Herholz, K., Ehlen, P., Kessler, J., Strotmann, T., Kalbe, E., & Markowitsch, H.J. (2001). Learning face-name associations and the effect of age and performance : a PET activation study. *Neuropsychologia*, **39**, 643-650.
- Hillis, A. E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, **114**, 2081-2094.
- Hittmair-Delazer, M., Denes, G., Semenza, C., & Mantovan, C. (1994). Anomia for people's names. *Neuropsychologia*, **32**(4), 465-476.
- Howard, I.P., & Templeton, W.B. (1966). Human Spatial Orientation. London: Wiley.
- 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. (2002). Inversion and Contrast Polarity Reversal Affect both Encoding and Recognition Processes of Unfamiliar Faces : a Repetition Study Using ERPs. *Neuroimage*, **15**, 353-372.
- Jackson, O3rd, & Schacter, D.L. (2004). Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *Neuroimage*, **22** (1), 456-462.
- Jay, M.F., & Sparks, D.L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, **309**, 345-347.
- Jeannerod, M. (2002). Le cerveau intime. Paris : Odile Jacob.
- Jeffreys, D.A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, **78**, 193-202.

- Jeffreys, D.A. (1996). Evoked Potentials Studies of Face and Object Processing. *Visual Cognition*, **3**, 1-38.
- Joassin, F. (2000). Approche anatomo-fonctionnelle de la récupération d'associations entre visages et noms propres au moyen de la tomographie par émission de positrons. Université catholique de Louvain: mémoire de licence en Sciences Psychologiques, non publié.
- Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., & Campanella, S. (2004). When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience Letters*, **369**, 132-137.
- Joassin, F., Maurage, P., Campanella, S., & Bruyer, R. Is associative priming a valid method to differentiate the serial and parallel models of face identification ? *Visual Cognition*, sous presse.
- Johnson, R., Jr. (1988). The amplitude of the P300 component of the event-related potential: review and synthesis. *Advances in Psychophysiology*, **3**, 69-137.
- Johnson, J.S., & Olhausen, B.A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, **5**, 299-312.
- Jolicoeur, P., Gluck, M., & Kosslyn, S.M. (1984). Pictures and names : making the connection. *Cognitive Psychology*, **16**, 243-275.
- Joseph, J.E. (2001). Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, **1**(2), 119-136.

- Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes : the effect of reference electrode site. *Clinical Neurophysiology*, **116**(11), 2613-2631.
- Kaas, J.H., & Hackett, T.A. (1999). 'What' and 'where' processing in auditory cortex. *Nature Neuroscience*, **2**(12), 1045-1047.
- Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The Fusiform Face Area : a Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, **9**, 462-475.
- Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport*, **10**, 183-187.
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, **3**, 759-763.
- Kellenbach, M.L., Hovius, M., & Patterson, K. (2005). A PET study of visual and semantic knowledge about objects. *Cortex*, **41**(2), 121-132.
- Kim, J.J., Andreasen, N.C., O'Leary, D.S., Wiser, A.K., Boles Ponto, L.L., Watkins, G.L., & Hichwa, R.D. (1999). Direct Comparison of the Neural Substrates of Recognition Memory for Words and Faces. *Brain*, **122**, 1069-1083.
- Kirwan, C.B., & Stark, C.E. (2004). Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*, **14**(7), 919-930.
- Lachaux, J.P., Rodriguez, E., Martinerie, J., & Varela, F.J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, **8**, 194-208.

- Lambon Ralph, M.A., McClelland, J.L., Patterson, K., Galton, C.J., & Hodges, J.R. (2001). No right to speak ? The relationship between object naming and semantic impairment: neuropsychological evidence and an computational model. *Journal of Cognitive Neuroscience*, **13**(3), 341-356.
- Lancelot, C., Ahad, P., Noulhiane, M., Hasboun, D., Baulac, M., & Samson, S. (2005). Loss of memory for auditory-spatial associations following unilateral medial temporal-lobe damage. *Neuropsychologia*, **43**(13), 1975-1982.
- Leube, D.T., Erb, M., Grodd, W., Bartels, M., & Kircher, T.T.J. (2001). Activation of right fronto-temporal cortex characterizes the 'living' category in semantic processing. *Cognitive Brain Research*, **12**, 425-430.
- 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. *Journal of Neuroscience*, **20**(2), 878-886.
- Levy, D.A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport*, **12**(12), 2653-2657.
- Levy, D.A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology*, **40** (2), 291-305.
- Lewkowicz, D.J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception Performance*, **22**(5), 1094-1106.

- Lewkowicz, D.J. (2000a). The development of intersensory temporal perception: an epigenetic systems/limitations review. *Psychological Bulletin*, **126**(2), 281-308.
- Lewkowicz, D.J. (2000b). Infant's perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development*, **71**(5), 1241-1257.
- Lewkowicz, D.J. (2002). Heterogeneity and heterochrony in the development of intersensory perception. *Cognitive Brain Research*, **14**, 41-63.
- Luchelli F., & De Renzi, E. (1992). Proper name anomia. *Cortex*, **28**, 221-230.
- Lyons, F., Hanley, J.R., & Kay, J. (2002). Anomia for common names and geographical names with preserved retrieval of names of people: a semantic memory disorder. *Cortex*, **38**, 431-444.
- Macaluso, E., Frith, C.D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, **289**, 1206-1208.
- Macaluso, E., & Driver, J. (2001). Spatial attention and cross-modal interactions between vision and touch. *Neuropsychologia*, **39**, 1304-1316.
- Macaluso, E., Driver, J., & Frith, C.D. (2003). Multimodal spatial representations engaged in human parietal cortex during both saccadic and manual spatial orienting. *Current Biology*, **13**(12), 990-999.
- Maguire, E.A., Frith, C.D., & Cipolotti, L. (2001). Distinct neural systems for the encoding and recognition of topography and faces. *Neuroimage*, **13**(4), 743-750.

- Mattson, A.J., Levin, H.S., & Grafman, J. (2000). A case of prosopagnosia following moderate closed head injury with left hemisphere focal lesion. *Cortex*, **36**(1), 125-137.
- McCarthy, R. A., & Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. *Nature*, **334**, 428-430.
- McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II. Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, **9**, 431-444.
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.
- McNeill, J.E., & Warrington, E.K. (1993). Prosopagnosia : a face specific disorder. *The Quarterly Journal of Experimental Psychology*, **46A**(1), 1-10.
- Meredith, M.A., & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, **221**, 389-391.
- Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in cat superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, **7**, 3215-3229.
- Miceli, G., Capasso, R., Daniele, A., Esposito, T., Magarelli, M., & Tomaiuolo, F. (2000). Selective deficit for people's names following left temporal damage: an impairment of domain-specific conceptual knowledge. *Cognitive Neuropsychology*, **17**(6), 489-516.
- Mitchell, J.P., Heatherton, T.F., & Macrae, C.N. (2002). Distinct neural systems subserve person and object knowledge.

- Proceedings of the National Academy of Sciences*, **99**(23), 15238-15243.
- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans : a high-density electrical mapping study. *Cognitive Brain Research*, **14**(1), 115-128.
- Morris, R.G., Ahmed, S., Syed, G.M.S., & Toone, B.K. (1993). Neural Correlates of Planning Ability: Frontal Lobe Activation during the Tower of London Test. *Neuropsychologia*, **31**, 1367-1378.
- Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition ? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, **9**(5), 555-604.
- Müller, H.M., & Kutas, M. (1996). What's in a name ? Electrophysiological differences between spoken nouns, proper names and one's own name. *Neuroreport*, **8**, 221-225.
- Münste, T.F., Brack, M., Grootheer, O., Wieringa, B.N., Matzke, M., & Johannes, S. (1998). Brain potentials reveal the timing of face identity and expression judgments. *Neurosci. Res.*, **30**, 25-34.
- Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., Toma, K., Nakamura, K., Hanakawa, T., Konishi, J., Fukuyama, H., & Shibasaki, H. (1999). Transient Neural Activity in the Medial Superior Frontal Gyrus and Precuneus Time-Locked with Attention Shift between Object Features. *Neuroimage*, **10**, 193-199.

- Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia*, **39**, 1047-1054.
- Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Brain Research Cognitive Brain Research*, **12**(2), 225-231.
- Neuner, F., & Schweinberger, S.R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn.*, **44**(3), 342-366.
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, **25**(3), 653-660.
- Niznikiewicz, M., Donnino, R., McCarley, R.W., Nestor, P.G., Losifescu, D.V., O'Donnell, B., Levitt, J., & Shenton, M.E. (2000). Abnormal angular gyrus asymmetry in schizophrenia. *Am J Psychiatry*, **157**(3), 428-437.
- Nolde, S.F., Johnson, M.K., & Raye, C.L. (1998). The Role of the Prefrontal Cortex during Tests of Episodic Memory. *Trends in Cognitives Sciences*, **2**(10), 399 - 406.
- Olhausen, B., Anderson, C., & Van Essen, D. (1995). A multiscale dynamic routing circuit ofr forming size- and position-invariant object representations. *Journal of Computational Neuroscience*, **2**, 45-62.
- Olson, I.R., Gatenby, J.C., & Gore, J.C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cogn. Brain Res.*, **14**, 129-138.

- Owen, A.M., Milner, B., Petrides, M., & Evan, A.C. (1996). Memory for object features versus memory for object location: a positron-emission tomography study of encoding and retrieval processes. *Proc. Natl. Acad. Sci.*, **93**, 9212-9217.
- Paller, K.A., Gonsalves, B., Grabowecky, M., Bozic, V.S., & Yamada, S. (2000). Electrophysiological Correlates of Recollecting Faces of Known and Unknown Individuals. *Neuroimage*, **11**(2), 98-110.
- Paller, K.A., Ranganath, C., Gonsalves, B., LaBar, K., Parrish, T.B., Gitelman, D.R., Mesulam, M.M., & Reber, P.J. (2003). Neural correlates of person recognition. *Learning & Memory*, **10**, 253-260.
- Pallis, C.A. (1955). Impaired identification of faces and places with agnosia for colors. *Journal of Neurology, Neurosurgery and Psychiatry*, **18**, 218-224.
- Papagno, C., & Capitani, E. (1998). Proper name anomia: a case with sparing of the first-letter knowledge. *Neuropsychologia*, **36**(7), 669-679.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, **117**(6), 1283-1301.
- Petrides, M., Alivisatos, B., & Frey, S. (2002). Differential activation of the human orbital, mid-ventrolateral, and mid-dorsolateral prefrontal cortex during the processing of visual stimuli. *Proc. Natl. Acad. Sci.*, **99**(8), 5649-5654.
- Piaget, J. (1952). The Origins of Intelligence in Children. New York : International Universities Press.

- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., & Gabrieli, J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, **10**(1), 15-35.
- Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, **41**, 103-146.
- Posamentier, M.T. (2002). Developmental prosopagnosia : a case study. PhD dissertation, The University of Texas at Dallas.
- Posamentier, M.T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychology Review*, **13**(3), 113-143.
- Pourtois, G., & de Gelder, B. (2002). Semantic factors influence multisensory pairing: a transcranial magnetic stimulation study. *Neuroreport*, **13**(12), 1567-1573.
- Pourtois, G., Schwartz, S., Seghier, M.L., Lazeyras, F., & Vuilleumier, P. (2005). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity : an event-related fMRI study. *Neuroimage*, **24**, 1214-1224.
- Price, J.C., Moore, C.J., Humphreys, G.W., Frackowiak, R.S.J., & Friston, K.J. (1996). The Neural Pathways Sustaining Object and Colour Naming. *Proc R Soc Lond B Biol Sci*, **263**(1376), 1501-1507.
- Price, J.C., Moore, C.J., Humphreys, G.W., & Wise, R.J. (1997). Segregating Semantic from Phonological Procesess during Reading. *Journal of Cognitive Neuroscience*, **9**(6), 727-733.
- Proverbio, A.M., Lilli, S., Semenza, C., & Zani, A. (2001). ERP indexes of functional differences in brain activation during

- proper and common names retrieval. *Neuropsychologia*, **39**(8), 815-827.
- Puce., A., Allison, T., Gore, J.C., & McCarthy, G. (1995). Face-Sensitive Regions in Human Extrastriate Cortex Studied by Functional MRI. *Journal of Neurophysiology*, **74**, 1192-1199.
- Puce , A., Allison, T., Asgari, M., Gore, J.C., & McCarthy, G. (1996). Differential Sensitivity of Human Visual Cortex to Faces, Letter-Strings and Textures : a Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, **16**, 5205-5215.
- Puce, A., Syngieniotis, A., Thompson, J.C., Abbott, D.F., Wheaton, K.J., & Castiello, U. (2003). The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage*, **19**(3), 861-869.
- Raij, T., Uutela, K., & Hari, R. (2001). Audiovisual integration of letters in the human brain. *Neuron*, **28**, 617-625.
- Rämä, P., & Courtney, S.M. (2005). Functional topography of working memory for face or voice identity. *Neuroimage*, **24**, 224-234.
- Reinkemeier, M., Markowitsch, H.J., Rauch, M., & Kessler, J. (1997). Differential impairments in recalling people's names: a cas study in search of neuroanatomical correlates. *Neuropsychologia*, **35**(5), 677-684.
- Reisberg, D. (1987). In Dodd, B., & Campbell, R. (Eds). Hearing by eye: the psychology of lip-reading. Hillsdale NJ : Lawrence Erlbaum.
- Rhodes, G., Byatt, G., Michie, P.T. & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation ? *J. Cogn. Neurosci.*, **16**(2), 189-203.

- Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways : evidence from Balint's syndrome. *Journal of Cognitive Neuroscience*, **9**(3), 295-317.
- Roskies, A.L. (1999). Introduction : the binding problem. *Neuron*, **24**(1), 7-9.
- Rossion, B., Campanella, S., Delinte, A., Debatisse, D., Liard, L., Dubois, S., Bruyer, R., Crommelinck, M., & Guérit, J.M. (1999). Task modulation of brain activity related to familiar and unfamiliar face processing : an ERP study. *Clinical Neurophysiology*, **110**, 449-462.
- 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. *Biological Psychology*, **50**(3), 173-189.
- Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, **11**(1), 69-74.
- 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.
- Rothstein, P., Henson, R.N., Treves, A., Driver, J., & Dolan, R.J. (2005). Morphing marylin into maggie dissociates physical and identity face representation in the brain. *Nature Neuroscience*, **8**(1), 107-113.

- Rousselet, G.A., Macé, M.J., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component? *Journal of Vision*, **4**, 13-21.
- Rugg, M.D., and Coles, M. (1995). *Electrophysiology of Mind*. Oxford Univ. Press, Oxford.
- Rugg, M.D., Doyle, M.C., & Wells, T. (1995). Word and non-Word Repetition within-and across-Modality : an Event-Related Potential Study. *Journal of Cognitive Neuroscience*, **7**, 209-227.
- Rugg, M.D., Fletcher, P.C., Chua, P.M.L., & Dolan, R.J. (1999). The Role of the Prefrontal Cortex in Recognition Memory and Memory for Sources : an fMRI Study. *Neuroimage*, **10**, 520 – 529.
- Rumiati, R., Humphreys, G.W., Riddoch, M.J., & Bateman, A. (1994). Visual object agnosia without alexia or prosopagnosia. *Visual Cognition*, **1**, 181-225.
- Saetti, M.C., Marangolo, P., De Renzi, E., Rinaldi, M.C., & Lattanzi, E. (1999). The nature of disorder underlying the inability to retrieve proper names. *Cortex*, **35**(5), 675-685.
- Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces : holistic and part-based processes. *Journal of Cognitive Neuroscience*, **13**, 937-951.
- Scanlan, L.C., & Johnston, R.A. (1997). I recognize your face but I can't remember your name : a grown-up explanation ? *Quarterly Journal of Experimental Psychology*, **50**(1), 183-198.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime user's guide. Pittsburgh: Psychology Software Tools Inc.
- Schweinberger, S.R., Herholz, A., & Stief, V. (1997). Auditory

- long-term memory: repetition priming of voice recognition. *Q. J. Exp. Psych.*, **50**(A), 498–517.
- Schweinberger, S.R., Burton, A.M., & Kelly, S.W. (2001). Priming the access to names of famous faces. *British Journal of Psychology*, **92**, 303-317.
- Schweinberger, S.R., Pickering, E.C., Jentzsch, I., Burton, M., & Kaufmann, J.M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Research Cognitive Brain Research*, **14**, 398-409.
- Seeck, M., Michel, C.M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., & Schomer, D.L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport*, **8**(12), 2749-2754.
- Sekiyama, K & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, **21**, 427-444.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, **47**, 277–287.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.
- Semenza, C., & Zettin, M. (1988). Generating proper names: a case of selective inability. *Cognitive Neuropsychology*, **5**(6), 711-721.
- Semenza, C., & Sgaramella, T.M. (1993). Production of proper names: a clinical case study of the effects of phonemic cueing. *Memory*, **1**(4), 265-279.

- Semenza, C., Mondini, S., & Zettin, M. (1995). The anatomical basis of proper name processing. A critical Review. *Neurocase*, **1**, 183-188.
- Senkowski, D., & Herrmann, C.S. (2002). Effect of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology*, **113**, 1742-1753.
- Sergent, J., & Signoret, J.L. (1992a). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex*, **2**(5), 375-388.
- Sergent, J., Ohta, S., & McDonald, B. (1992b). Functional Neuroanatomy of Face and Object Processing. *Brain*, **115**, 15-36.
- Sergent, J., McDonald, B., & Zuck, E. (1994). Structural and functional organisation of knowledge about faces and proper names: a positron emission tomography study. In C. Umiltà & M. Moscovitch (Eds). Attention and performance XV. Conscious and non-conscious information processing. (pp. 203-228). Hillsdale NJ: Lawrence Erlbaum Associates, Inc.
- Shadlen, M.N., & Movshon, J.A. (1999). Synchrony unbound : a critical evaluation of the temporal binding hypothesis. *Neuron*, **24**(1), 67-77.
- Shafritz, K.M., Gore, J.C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proc. Natl. Acad. Sci. USA*, **99**(16), 10917-10922.
- Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001). The neural correlates of person familiarity: a functional magnetic resonance imaging study with clinical implications. *Brain*, **124**,

804–815.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, **12**(17), 3849-52.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cogn. Brain Res.*, **14**, 147–152.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, **11**, 505–509.

Simons, J.S., Graham, K.S., Owen, A.M., Patterson, K., & Hodges, J.R. (2001). Perceptual and semantic components of memory for objects and faces: a PET study. *Journal of Cognitive Neuroscience*, **13**(4), 430-443.

Small, M. (1988). Visual Evoked Potentials in a Patient with Prosopagnosia. *Electroencephalogr. Clin. Neurophysiol.*, **71**(1), 10-16.

Sommer, W., Komoss, E., & Schweinberger, S.R. (1997). Differential localization of brain systems subserving memory for names and faces in normal subjects with event-related potentials. *Electroencephalography and Clinical Neurophysiology*, **102**, 192-199.

Spelke, E.S. (1976). Perceiving bimodally specified events in infancy. *Developmental Psychology*, **15**, 626-636.

Spence, C.J., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, **22**(4), 1005-1030.

Sperling, R., Bates, J.F., Cocchiarella, A.J., Schacter, D.L., Rosen, B.R., & Albert, M. (2001). Encoding novel face-name

- associations : a functional MRI study. *Human Brain Mapping*, **14**, 129-139.
- Sperling, R., Chua, E., Coccharella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D., & Albert, M. (2003). Putting names to faces : successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage*, **20**(2), 1400-1410.
- Steeves, J.K.E., Culham, J.C., Duchaine, B.C., Pratesi, C.C., Valyear, K.F., Schindler, I., Humphrey, G.K., Milner, A.D., & Goodale, M.A. (2006). The fusiform face area is not sufficient for face recognition : evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, **44**, 594-609.
- Stein, B.E., & Meredith, M.A. (1993). Merging of the senses. Cambridge, MA : MIT press.
- Stein, B.E., Meredith, M.A., & Wallace, M.T. (1993). The visually responsive neurons and beyond : multisensory integration in cat and monkey. *Progress in Brain Research*, **95**, 79-90.
- Stein, B.E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, **123**(1-2), 124-135.
- Stein, B.E., Jiang, W., Wallace, M.T., & Stanford, T.R. (2001). Nonvisual influences on visual-information processing in the superior colliculus. *Prog Brain Res.*, **134**, 143-56.
- Streit, M., Wulver, W., Brinkmeyer, J., Ihl, R., & Gaebel, W. (2000). Electrophysiological correlates of emotional and structural face processing in humans. *Neurosci. Lett.*, **278**, 13-16.

- Takahashi, N., Kawamura, M., Hirayama, K., Shiota, J., & Isono, O. (1995). Prosopagnosia : a Clinical and Anatomical Study of Four Patients. *Cortex*, **31**, 317-329.
- Tallon-Baudry, C., Bertrand, O., Wienbruch, C., Ross, B., & Pantev, C. (1997a). Combined EEG and MEG recordings of visual 40 Hz responses to illusory triangles in human. *Neuroreport*, **8**(5), 1103-1107.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997b). Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *Journal of Neuroscience*, **17**(2), 722-734.
- Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *Journal of Neuroscience*, **18**(11), 4244-4254.
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, **3**(4), 151-162.
- Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol. A*, **46**(2), 225-245.
- Tanaka, J.W., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *Neuroreport*, **10**, 829-835.
- Tanaka, J.W. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology: General*, **130**(3), 534-543.
- Tanaka, J.W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, **12**, 43-47.

- Tarr, M.J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, **3**(8), 764-769.
- Teder-Sälejärvi, W.A., Münte, T.F., Sperlich, F.J., & Hillyard, S.A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cognitive Brain Research*, **8**(3), 327-343.
- Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., & Hillyard, S.A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potentials (ERP) recordings. *Cognitive Brain Research*, **14**, 106-14.
- Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., & Farah, M.J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci.*, **94**, 14792-14797.
- Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, **17**(1/2/3), 257-279.
- Tsukiura, T., Fujii, T., Fukatsu, R., Otsuki, T., Okuda, J., Umetsu, A., Suzuki, K., Tabuchi, M., Yanagawa, I., Nagasaka, T., Kawashima, R., Fukuda, H., Takahashi, S., & Yamadori, A. (2002). Neural basis of the retrieval of people's names: evidence from brain-damaged patients and fMRI. *Journal of Cognitive Neuroscience*, **14**, 922-937.
- Tsukiura, T., Namiki, M., Fujii, T., & Iijima, T. (2003). Time-dependent neural activations related to recognition of people's names in emotional and neutral face-name

- associative learning : an fMRI study. *Neuroimage*, **20**, 784-794.
- Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2005). Dissociable roles of the bilateral anterior temporal lobe in face-name associations : an event-related fMRI study. *Neuroimage*, Epub ahead of print.
- Turk, D.J., Rosenblum, A.C., Gazzaniga, M.S., & Macrae, C.N. (2005). Seeing John Malkovich: the neural substrates of person categorization. *Neuroimage*, **24**(4), 1147-1153.
- Valentine, T., Brédart, S., Lawson, R., & Ward, G. (1991). What's in a name ? Access to information from people's names. *European Journal of Cognitive Psychology*, **3**, 147-176.
- Valenza, N., Murray, M.M., Ptak, R., & Vuilleumier, P. (2004). The space of senses: impaired crossmodal interactions in a patient with Balint syndrome after bilateral parietal damage. *Neuropsychologia* **42**(13), 1737-1748.
- Van Lancker, D.R., & Canter, G.J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cognition*, **1**, 185-195.
- Van Lancker, D.R., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, **25**, 829-834.
- Van Lancker, D.R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neurophysiology*, **11**(5), 665-674.
- Varela, F.J. (1995). Resonant cell assemblies : a new approach to cognitive functions and neuronal synchrony. *Biological Research*, **28** (1), 81-95.

- Von der Malsburg, C. (1981). The Correlation Theory of Brain Function. *Internal Report*, **81** – 2. Göttingen: Max-Plant-Institute for Biophysical Chemistry. Reprinted (1994) in Models of Neural Networks II. Domany, E., Van Hemmen, J.L., and Schulten, K., Eds. Berlin: Springer.
- Von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A.L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, **17**, 48-55.
- Von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A.L. (2005). Interaction of face and voice areas during speaker recognition. *J. Cogn. Neurosci.*, **17**(3), 367–376.
- Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International journal of psychophysiology*, **50**, 117–124.
- Wallace, M.T., Meredith, M.A., & Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, **91**, 484-488.
- Wallace, M.T., & Stein, B.E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *Journal of Neurophysiology*, **83**, 3578-3582.
- Warrington, E.K., & Shallice, T. (1984). Category-specific semantic impairments. *Brain*, **107**, 829-859.
- Wolfe, J.M., & Cave, K.R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, **24**(1), 11-17.
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, **15**(8), 1234-1242.

- Yasuda, K., & Ono, Y. (1998). Comprehension of famous personal and geographical names in global aphasic subjects. *Brain and Language*, **61**, 274-287.
- Young, A.W., McWeeny, K.H., Hay, D.C., & Ellis, A.W. (1986). Access to identity-specific semantic codes from familiar faces. *Quarterly Journal of Experimental Psychology*, **38**(2), 271-295.

Travaux annexes et collaborations



D. Goossens

L'encyclopédie des bébés (l'intégrale), Editions Fluide Glacial (2004)

Article publié dans *Neuroimage*, 14 (2001), 873-882.

**Association of the distinct visual
representations of faces and names : A PET
activation study**

S. Campanella^{1,3}, F. Joassin¹, B. Rossion^{1, 3,4}, A. De Volder²,
R. Bruyer¹, M. Crommelinck³

¹ Unité de Neuropsychologie Cognitive (NECO), Université Catholique
de Louvain, Louvain-la-Neuve, Belgium

² Cyclotron, Unité TOPO, Université Catholique de Louvain, Louvain-la-
Neuve, Belgium

³ Laboratoire de Neurophysiologie (NEFY), Université Catholique de
Louvain, Brussels, Belgium

⁴ Department of Cognitive and Linguistic Sciences, Brown University,
USA

Corresponding author :

Campanella Salvatore
Université Catholique de Louvain
Faculté de Psychologie - Unité NECO
Place du Cardinal Mercier, 10
B- 1348 Louvain-la-Neuve
Tel: +32 10 473808
Fax: +32 10 473774
e-mail: campanella@neco.ucl.ac.be

Abstract

A PET study of 7 normal individuals was carried out to investigate the neural populations involved in the retrieval of the visual representation of a face when presented with an associated name, and conversely. Face-name associations were studied by means of 4 experimental matching conditions, including the retrieval of previously learned (1) name-name (NN), (2) face-face (FF), (3) name-face (NF) and (4) face-name (FN) associations, as well as a resting scan with eyes closed. Before PET images acquisition, subjects were presented with 24 unknown face-name associations to encode in 12 male/female couples. During PET scanning, their task was to decide whether the presented pair was a previously learned association.

The right fusiform gyrus was strongly activated in FF condition as compared to NN and Rest conditions. However, no specific activations were found for NN condition relative to FF condition. A network of three areas distributed in the left hemisphere, both active in (NF-FF) and (FN-NN) comparisons, was interpreted as the locus of the integration of visual faces and names representations. These three regions were localized in the inferior frontal gyrus (BA 45), the medial frontal gyrus (BA 6) and the supramarginal gyrus of the inferior parietal lobe (BA 40). An interactive model accounting for these results, with BA 40 seen as an amodal binding region, is proposed.

Key Words : PET, faces, names, association

Article sous presse dans *Visual Cognition*.

**Is associative priming a valid method to
differentiate the serial and parallel models of
face identification ?**

F. Joassin, P. Maurage, S. Campanella, and R. Bruyer

Unité de Neurosciences Cognitives (NESC), Université catholique de
Louvain, Louvain-la-Neuve, Belgium

Corresponding author :

F. Joassin, Ph. D student
Université catholique de Louvain
Unité de Neurosciences Cognitives (PSP / NESC)
Place du Cardinal Mercier, 10
B 1348 Louvain-la-Neuve
Belgium
Tél. : +32(10) 47.90.62
Fax : +32(10) 47.37.74
e-mail : frederic.joassin@psp.ucl.ac.be

Abstract

One current controversy in face identification is whether names are accessed after or in parallel to semantic information. In prior research, Schweinberger et al. (British Journal of Psychology, 2001, 92, 303-317) have shown that phonological decisions to the names of famous faces were facilitated by name but not by semantic primes, while semantic decisions were facilitated by semantic but not by name primes. They proposed a parallel rather than a serial model of face identification. We replicated these experiments by (1) adding specific semantic primes in order to examine the effect of the uniqueness of the relation between semantic or phonological information and a face, and (2) adding a neutral baseline to differentiate facilitation from inhibition effects when primes and targets were related or unrelated. Our study revealed that uniqueness plays an important role in associative priming, with the specific primes (whether lexical or semantic) producing the greatest effects. However, it was also revealed that the same primes generated different effects according to the difficulty of the task. Our results highlight the importance, in associative priming, of the specificity of the information and of the relation between the primes and the task, and raise the question of the validity of such a method to test cognitive models of face identification.

Article publié dans *Neuroscience Letters*, 367 (2004), 14-18.

Human gender differences in an emotional visual oddball task: an ERP study.

S. Campanella, M. Rossignol, S. Mejias, F. Joassin, P. Maurage, D. Debatisse, R. Bruyer, M. Crommelinck & J.M. Guérit

Pictures from the Ekman and Friesen series were used in an event-related potentials study to define the timing of occurrence of gender differences in the processing of positive (happy) and negative (fear) facial expressions. Ten male and 10 female volunteers were confronted with a visual oddball design, in which they had to detect, as quickly as possible, deviant happy or fearful faces amongst a train of standard stimuli (neutral faces). Behavioral results suggest that men and women detected fearful faces more quickly than happy ones. The main result is that the N2b component, functionally considered as an attentional orienting mechanism, was delayed in men for happy stimuli as compared with fearful ones. Gender differences observed in the processing of emotional stimuli could then originate at the attentional level of the information processing system.

Article publié dans *Visual Cognition*, 10 (2003), 129-156.

Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique.

S. Campanella, C. Hanoteau, X. Seron, F. Joassin, & R. Bruyer

Categorization is a fundamental property of the human brain. We used an image-morphing procedure to investigate the categorical perception of unfamiliar and familiar facial identities. Two delayed matching tasks are reported here. First, independently of face familiarity, subjects discriminate more easily two different morphed faces belonging to two different identities than two different morphed faces belonging to the same one, even if the physical distance within the pairs is identical. Second, this holds on only when faces are upright. The results are interpreted in terms of how representations of faces are encoded and stored in long-term memory. The inability of the face--space metaphor, developed by Valentine (1991), is discussed. An alternative explanation, based on the morphing characteristics and the concept of typicality, is proposed.

