



HAL
open science

Représentation des individus par le macaque Rhésus : approche neurophysiologique et comportementale

Julia Sliwa

► **To cite this version:**

Julia Sliwa. Représentation des individus par le macaque Rhésus : approche neurophysiologique et comportementale. Neurosciences. Université Claude Bernard - Lyon I, 2012. Français. NNT : 2012LYO10011 . tel-00979701

HAL Id: tel-00979701

<https://theses.hal.science/tel-00979701>

Submitted on 16 Apr 2014

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

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

UNIVERSITE DE LYON
Ecole Doctorale Neurosciences et Cognition

Thèse présentée par :

Julia SLIWA

Pour l'obtention du DIPLOME DE DOCTORAT (arrêté du 7 août 2006)

REPRESENTATION DES INDIVIDUS PAR LE MACAQUE RHESUS

Approche neurophysiologique et comportementale

Soutenue publiquement le 17 février 2012

Directeurs de thèse : Jean-René DUHAMEL et Sylvia WIRTH

JURY :	Pascal BELIN	Rapporteur
	Michèle FABRE-THORPE	Examineur
	Rémi GERVAIS	Président du jury
	Olivier PASCALIS	Invité
	Sidney WIENER	Rapporteur

REMERCIEMENTS

Je tiens tout d'abord à remercier vivement les rapporteurs, Pascal BELIN et Sidney WIENER, pour avoir accepté de lire et d'analyser le manuscrit. Merci également à Olivier PASCALIS qui a participé à l'élaboration de la première étude, ainsi qu'à Michèle FABRE-THORPE, et Rémi GERVAIS pour m'avoir fait l'honneur d'examiner et de présider ma soutenance de thèse.

Bien sûr je remercie chaleureusement mes directeurs de thèse, Sylvia WIRTH et Jean-René DUHAMEL. Merci de m'avoir initiée à l'art de l'électrophysiologie chez le macaque, de m'avoir proposé un projet super excitant à l'interface de différentes disciplines, et surtout de m'avoir fait confiance tout au long du chemin.

Plusieurs personnes ont participé à la réalisation de ce travail. Tout d'abord un grand merci à Jean-Luc CHARIEAU et Fabrice HERANT pour leur présence et leur précieuse aide quotidienne durant les périodes d'expérimentation ; à Aurélie PLANTE pour avoir fait du spike sorting jusque dans ses rêves ; à Pierre BARADUC et Philippe VINDRAS pour avoir aiguillé mes choix d'analyses ; et à Elise METEREAU pour les belles images de cerveau qui illustrent la thèse. J'en profite pour remercier les membres de mon comité de suivi de thèse, Martine MEUNIER et Bruno BONTEMPI. Merci également à l'unité Inserm U846 et au Centre de Primatologie de Rousset, à Danielle IBAROLLA du CERMEP, et à Sylvain MAURIN, Johan PACQUIT, Hassen SLIMANI et Christèle VALLERIE. Je remercie le CNRS, la DGA et la FRM d'avoir financé ces projets.

Un grand merci à tous les membres de l'équipe pour leur entourage scientifique et humain : à Suliann BEN HAMED qui m'a encadré au début de ma thèse, Guilhem IBOS, Pierre MOREL, Eric KOUN, Claire WARDAK, João AZZI, Jérôme MUNUERA, Serge PINEDE, Elaine ASTRAND, Olivier GUIPPONI, Sébastien BALLESTA, Nazied HUGUET. Enfin je remercie les autres chercheurs du Centre de Neurosciences Cognitive et plus particulièrement Angela SIRIGU, Michel DESMURGET, Léon TREMBLAY, Véronique SGAMBATO-FAURE, Bénédicte BALLANGER, Philippe BOULINGUEZ, Jean-Claude DREHER, Giorgio CORRICELLI et Giovanna VANNI-MERCIER.

Pendant ces années, j'ai aussi beaucoup appris en dehors du laboratoire, auprès des membres des journal-clubs « Cognition Animale » et « Oscillations Lyon », et grâce à Jean-Baptiste VAN DER HENST, qui m'a permis de d'enseigner à l'Université Catholique de Lyon. J'en profite pour remercier également mes anciens directeurs de stages, Stéphane SUPPLISSON, Thierry BAL et Zuzanna PIWKOWSKA, Michael GOLDBERG et Anna IPATA, et d'étude Frédéric MARION-POLL rencontrés avant ma thèse et qui m'ont progressivement orienté vers l'électrophysiologie.

Cette thèse doit beaucoup à la bonne ambiance qui a régné à l'Institut des Sciences Cognitives. Je remercie donc Audrey, Augustin, Aurélie, Elaine, Elise, Elissar, Eric, Flavia, Guilhem et Laura et Vittorio, Guillaume B, Guillaume S, Guillaume J, Irene, Jérôme M, Jérôme P, Jérôme R, Julie, Juliette,

Justine, Ludovic et l'Ayers, Marie, Mariia, Mathilde, Maude, Nadège, Nicola, Olivier, Philippe, Pierre M, Pierre W, Rawan, Rémi, Romain, Ligneul, Romuald, Sara, Sébastien, Suzanne, Tiffany, Valeria, Valérian, Xavier. Enfin mes colocataires Frédéric, Erwan, Sabrina et plus sporadiquement Hélène, ainsi que mes amis méritent de grands mercis.

And last but not least, je remercie particulièrement les personnes qui comptent le plus, ma sœur Suzanna et mes parents, pour leur présence régénérante. Et Jérémie qui, en plus de tout ce qu'il représente intimement à mes yeux, a endossé un rôle supplémentaire durant ma thèse, celui du coach sportif et scientifique. Je suis heureuse d'avoir partagé cette aventure avec toi !

RESUME en français

Nous possédons la faculté de reconnaître individuellement des centaines d'individus. Ceci nous permet d'évoluer dans une société complexe dont l'organisation est en partie forgée par les relations interindividuelles. La reconnaissance individuelle peut être réalisée par l'identification de divers éléments distincts, comme le visage ou la voix, qui forment chez l'Homme une seule représentation conceptuelle de l'identité de la personne. Nous avons démontré que les singes rhésus, comme les humains, reconnaissent individuellement leurs congénères familiers, mais également les individus humains connus. Ceci montre que la reconnaissance fine est une compétence partagée par un éventail d'espèces de primates pouvant servir de fondement à la vie en réseaux sociaux sophistiqués, et également que le cerveau s'adapte de façon flexible pour reconnaître les individus d'autres espèces lorsque ceux-ci ont une importance socio-écologique. Par la suite, au niveau neuronal, ce projet a mis en lumière que les connaissances sociales concernant autrui sont représentées par les neurones hippocampiques ainsi que par les neurones inferotemporaux. Ainsi nous avons observé l'existence de neurones sélectifs aux visages non seulement dans le cortex inferotemporal, comme ceci a été décrit précédemment, mais également dans l'hippocampe. La comparaison des propriétés de ces neurones au sein de ces deux structures, suggère que les deux régions joueraient des rôles complémentaires au cours de la reconnaissance individuelle. Enfin, parce que l'hippocampe est une structure qui a évolué à des degrés divers chez différents mammifères pour soutenir la mémoire autobiographique et les représentations spatiales, la caractérisation des différents types de neurones et de leur connectivité a fourni un cadre commun pour comparer les fonctions de l'hippocampe à travers les espèces.

TITRE en anglais

Rhesus monkeys' behavioral and neuronal responses to voices and faces of known individuals

RESUME en anglais

Humans can individually recognize some hundreds of persons and therefore operate within a rich and complex society. Individual recognition can be achieved by identifying distinct elements such as the face or voice as belonging to one individual. In humans, those different cues are linked into one conceptual representation of individual identity. I demonstrated that rhesus monkeys, like humans, recognize familiar peers but also familiar humans individually and that they match their voice to their corresponding memorized face. Thus it shows that fine individual recognition is a skill shared across a range of primate species, which may serve as the basis of a sophisticated social network. It also suggests that animals' brains flexibly adapt to recognize individuals of other species when socio-ecologically relevant. Following at the neuronal level, this project put in light that social knowledge about other individuals is represented by hippocampal neurons as well as by inferotemporal neurons. For instance I observed the existence of face preferring neurons not only in the inferotemporal cortex as previously described but also in the hippocampus. Comparison of their properties across both structures, suggests that they could play complementary roles in recognition of individuals. Finally, because the hippocampus is a structure that evolved in various degrees to support autobiographical memory and spatial information in different mammals, I characterized the different subtypes of neurons and their network connectivity in the monkey hippocampus to provide a common anatomical framework to discuss hippocampal functions across species.

DISCIPLINE

Neurosciences cognitives

MOTS-CLES

neurophysiologie, électrophysiologie, comportement, primates, Macaca mulatta, hippocampe, cortex inferotemporal, reconnaissance individuelle, voix, visages, intégration multisensorielle, activité neuronale, neurones pyramidaux, interneurones, caractérisation cellulaire

**Centre de Neurosciences Cognitive, UMR5229
CNRS – Université Lyon 1
67 Boulevard Pinel
69 675 Bron cedex**

Table des matières

ABREVIATIONS	7
INTRODUCTION	9
A. REPRESENTATION DES INDIVIDUS CHEZ L'HOMME	11
A.1. LE CONCEPT DE PERSONNE	11
A.2. RECONNAISSANCE INDIVIDUELLE : UN PROCESSUS TRIPLE	12
A.3. SPECIFICITE DU TRAITEMENT PERCEPTIF DES VISAGES ET PERCEPTION DE LA SINGULARITE	14
A.4. TRAITEMENT PERCEPTIF DE LA VOIX	17
A.5. LES DIVERSES ASSOCIATIONS ENTRE VOIX ET VISAGE	19
A.6. RECONNAISSANCE INDIVIDUELLE A PARTIR DU VISAGE OU DE LA VOIX	23
A.7. LIMITES ET QUESTIONS RELATIVES AU MODELE SERIEL	24
B. REPRESENTATION DES INDIVIDUS CHEZ LE MACAQUE RHESUS	27
B.1. LES MACAQUES RHESUS S'ORGANISENT EN SOCIETE COMPLEXES	27
B.2. PERCEPTION, INDIVIDUALISATION ET FAMILIARITE DES VISAGES	29
B.3. PERCEPTION ET INDIVIDUALISATION DES VOIX	30
B.4. ASSOCIATIONS DYNAMIQUES ENTRE LES VOIX ET LES VISAGES	32
B.5. LA POSSIBILITE D'UN CONCEPT DE PERSONNE	34
B.6. CONCLUSION	37
C. BASES NEURALES DE L'IDENTIFICATION DES VISAGES	39
C.1. UN SYSTEME DEDIE A LA PERCEPTION DE CE STIMULUS SOCIAL	39
C.2. ENCODAGE DES VISAGES LE LONG DE LA VOIE VENTRALE	47
C.3. RECONNAISSANCE DU VISAGE	54
C.4. CONCLUSION	58
D. BASES NEURALES DE L'IDENTIFICATION DES VOIX	61
D.1. UN SYSTEME DEDIE A LA PERCEPTION DE CE STIMULUS SOCIAL	61
D.2. ENCODAGE DES VOIX LE LONG DE LA VOIE VENTRALE	68
D.3. RECONNAISSANCE DE LA VOIX	71
D.4. CONCLUSION	73
E. RECONNAISSANCE INDIVIDUELLE ET EMERGENCE DU CONCEPT DE PERSONNE	77
E.1. UNE ZONE D'INTEGRATION MULTISENSORIELLE DE L'IDENTITE ?	77
E.2. LES CANDIDATS AUX PERSONS' IDENTITY NODES (PINs)	82
E.3. L'HIPPOCAMPE, UNE STRUCTURE ASSOCIATIVE IMPLIQUEE DANS LA MEMOIRE AUTOBIOGRAPHIQUE	85
E.4. CONCLUSION	93

F. PROBLEMATIQUES	95
<u>EXPERIENCES ET RESULTATS</u>	<u>99</u>
A. RECONNAISSANCE INDIVIDUELLE CROSS-MODALE PAR LES MACAQUES RHESUS	101
A.1. ARTICLE N°1: SPONTANEOUS VOICE-FACE IDENTITY MATCHING BY RHESUS MONKEYS FOR FAMILIAR CONSPECIFICS AND HUMANS.	103
A.2. ARTICLE N°1: SUPPORTING INFORMATION	109
A.3. ANALYSES SUPPLEMENTAIRES NON PUBLIEES	112
B. REPONSES NEURONALES AUX VISAGES : COMPARAISON DE L'HIPPOCAMPE ET DE L'AIRE TEMPORALE CHEZ LE SINGE	115
B.1. ARTICLE N°2: COMPARING FACE-CELLS PROPERTIES IN THE MONKEY HIPPOCAMPUS AND TE	117
C. DISSECTION DES MICROCIRCUITS DANS L'HIPPOCAMPE DU SINGE: LIEN ENTRE LES REPONSES DES CELLULES ET LEUR SIGNATURE ELECTROPHYSIOLOGIQUE <i>IN VIVO</i>	149
C.1. ARTICLE N°3 : DISSECTION OF MICROCIRCUITRY IN MONKEY HIPPOCAMPUS: LINKING CELLS' RESPONSE PROPERTIES TO THEIR ELECTROPHYSIOLOGICAL FEATURES <i>IN VIVO</i>	151
<u>DISCUSSION GENERALE</u>	<u>171</u>
A.1. ETENDUE DU CONCEPT DE PERSONNE CHEZ LE MACAQUE RHESUS ET D'AUTRES ESPECES ANIMALES	171
A.2. REPRESENTATION DES INDIVIDUS DANS LE LOBE TEMPORAL	174
A.3. CONTINUITES ET DISCONTINUITES EVOLUTIVES DANS LE FONCTIONNEMENT DE L'HIPPOCAMPE	179
A.4. CONCLUSION ET RETOMBES POTENTIELLES DANS LE DOMAINE DE LA SANTE	182
<u>ANNEXES</u>	<u>185</u>
ANNEXE A. PARADIGMES DES « VISAGES COMPOSITES » ET DU « RENVERSEMENT DES VISAGES »	187
ANNEXE B. UNE PROPENSION INNEE A DETECTER LES VISAGES ?	189
ANNEXE C. DIFFERENTS MODELES DE REPRESENTATION NEURALE DU CONCEPT	191
ANNEXE D. CARTES COGNITIVES ET MODELES TOP-DOWN DE RECONNAISSANCE	195
ANNEXE E. BASES NEURONALES DES ASSOCIATIONS VOIX-VISAGE AMODALES	197
<u>BIBLIOGRAPHIE</u>	<u>203</u>
<u>TRAVAUX SCIENTIFIQUES</u>	<u>231</u>

Abréviations

A1	cortex auditif primaire	OFA	occipital face area
AF	anterior fundus face patch	PET	tomographie par émission de positron
aIT	anterior inferotemporal cortex	PIN	person identity node
aITv	anterior inferotemporal ventral cortex	PL	posterior lateral face patch
AL	anterior lateral auditory belt	R	rostral auditory core
AL	anterior lateral face patch	RM	rostromedial auditory belt
AM	anterior medial face patch	RP	rostral parabelt
BA	Broadmann area	RT	rostrotemporal auditory core
BOLD	blood oxygen level-dependant	RTL	lateral rostrotemporal auditory belt
CA 1-4	champs ammoniens	RTM	medial rostrotemporal auditory belt
CBV	contrast agent-dependant cerebral blood volume	SE-EPI	spin-echo echo-planar
CL	caudolateral auditory belt	STG	superior temporal gyrus
CM	caudomedial auditory belt	STS	superior temporal sulcus
CP	caudal parabelt	TE	anterior IT
EEG	électroencéphalographie	TEO	posterior IT
ERP	potentiel évoqué, event-related potential	TF	parahippocampal face patch
FFA	fusiform face area	TPJ	temporo-parietal junction
IRMf	imagerie par résonance magnétique fonctionnelle	Ts1-3	temporal auditory association areas
IT	inferior temporal cortex	TVAA	temporal voice area anterior
MF	middle fundus face patch	TVAm	temporal voice area middle
ML	medial lateral auditory belt	TVAp	temporal voice area posterior
ML	middle lateral face patch	TVAs	temporal voice areas
MT	middle temporal visual area	V2-4	visual association areas
		vV4	ventral V4 face patch

Introduction

Nous possédons la capacité de reconnaître individuellement des centaines d'autres individus. Cette faculté nous permet d'évoluer dans une société complexe dont l'organisation est en partie forgée par les relations interindividuelles. La reconnaissance des autres est protéiforme, permettant d'identifier quelqu'un par les traits de son visage, sa voix, son odeur ou encore sa démarche. Reconnaître un individu signifie que la perception d'un de ces indices amène à se remémorer les informations le concernant, comme son nom ou un épisode vécu ensemble. Par exemple au téléphone, le son d'une voix connue fait écho intérieurement au prénom et au visage visualisé mentalement pendant la discussion. Ainsi les informations sensorielles (voix, visage, etc.) et les informations biographiques (nom, profession, personnalité, souvenirs communs, etc.) convergent en une représentation cognitive, ou un concept, de l'identité de l'Autre.

La richesse de ces représentations amène à s'interroger sur la spécificité humaine à créer des concepts d'individus. L'Homme est-il le seul à utiliser ces concepts, ou partage-t-il cette faculté avec d'autres espèces ? Par exemple le singe rhésus (*Macaca mulatta*) est une espèce de primate qui vit en larges colonies et présente des comportements sociaux complexes, tout comme l'Homme. Au vu des nombreuses relations interindividuelles qu'il entretient avec ses congénères, on peut se demander s'il utilise une représentation des autres similaire à celle que nous utilisons. Ceci renseignerait sur l'évolution de cette faculté cognitive au sein de la famille des primates.

Par ailleurs le terme « représentation » sous-entend que nous possédons un code nerveux *représentant* l'individu, qui est une réalité environnementale extérieure à nous. La question qui se pose alors est de savoir quelle est la nature de ce code. Quelles activités neuronales permettent d'associer les différents indices de l'identité en une représentation commune ? Des aires corticales impliquées dans l'analyse des informations unisensorielles sur les voix et les visages sont identifiées dans le lobe temporal chez l'humain et le singe rhésus. L'étude conjointe des activités des neurones de ces régions pendant que l'on perçoit un individu connu permettrait d'approcher une explication.

Enfin, la reconnaissance d'une personne s'accompagne du rappel des informations biographiques la concernant, comme son nom ou un épisode que l'on vient de vivre ensemble, à partir de la perception d'un indice spécifique, comme sa voix ou son visage. Une aire cérébrale associative, pourrait-elle établir un lien entre les différentes aires sensorielles unimodales et sémantiques, au cours de l'identification ? L'hippocampe est une structure impliquée dans la mémoire autobiographique. L'étude de son activité lors de la perception de voix ou visages connus permettrait de tester son rôle dans la reconnaissance des individus.

Nous allons commencer par présenter au cours de cette introduction une description générale de la représentation des individus chez l'Homme et l'animal. Nous nous intéresserons particulièrement à la perception des voix et des visages. Puis nous exposerons en détail les réseaux cérébraux impliqués dans la perception de ces stimuli sociaux et les différentes interactions connues entre ces deux systèmes. Nous nous interrogerons sur l'émergence du concept d'individu à partir de ces réseaux. Enfin, nous décrirons pourquoi l'hippocampe, de part son activité de carte cognitive, est un bon candidat pour l'association des différents composants du concept de « personne » lors de la reconnaissance.

À partir de cette revue bibliographique, nous dégagerons les problématiques plus précises du travail de thèse. Les résultats expérimentaux obtenus dans deux projets distincts seront présentés dans le reste du document. Premièrement, en testant une cohorte de six macaques rhésus dans une tâche de priming sensoriel cross-modal, j'ai observé que ces primates sont capables d'associer spontanément le visage à la voix des individus qu'ils connaissent (Sliwa et al., 2011). Cette faculté à reconnaître de façon individuelle et multimodale des personnes s'applique à leurs congénères mais également aux humains lorsque ces singes sont élevés en leur présence. Ainsi l'individualisation et la représentation cognitive des individus est une aptitude que l'Homme partage avec d'autres espèces de primates. Deuxièmement, en menant une étude électrophysiologique engageant deux macaques rhésus, j'ai mis en lumière que les connaissances sociales concernant autrui sont représentées par les neurones hippocampiques ainsi que par les neurones infero-temporaux. Par exemple j'ai observé l'existence de neurones répondant préférentiellement aux visages non seulement dans le cortex infero-temporal comme décrit dans des études antérieures mais également dans l'hippocampe. La comparaison de leurs propriétés à travers les deux structures, suggère qu'ils pourraient jouer des rôles complémentaires lors de la reconnaissance individuelle. Enfin, parce que l'hippocampe est une structure qui a évolué à des degrés divers chez différents mammifères pour soutenir la mémoire autobiographique et les représentations spatiales, la caractérisation des différents types de neurones et de leur connectivité a fourni un cadre commun pour comparer les fonctions de l'hippocampe à travers les espèces.

Pour conclure, la dernière partie fait le bilan des résultats expérimentaux et les compare à ceux de la littérature.

A. Représentation des individus chez l'Homme

Au cours de cette partie, nous allons d'abord définir la notion d'individu, pour ensuite présenter plus particulièrement la façon dont nous percevons les voix et les visages et enfin exposer comment nous les associons.

A.1. Le concept de Personne

Les termes d'*individu* et de *personne* se ressemblent et se répondent dans leurs définitions données par le dictionnaire de l'Académie française présentées ci-dessous. Principalement les deux définitions font appel à la *singularité* d'un membre de l'espèce par rapport aux autres. Par contre, si le terme d'individu s'applique à diverses espèces animales ou végétales, le terme de personne semble spécifiquement dédié à l'Homme, comme le suggère l'apposition des termes de *conscience, morale, raison et volonté* à sa définition.

INDIVIDU Tout être concret formant une unité distincte et identifiable, et qui ne peut être divisé sans être détruit. **1.** Chacun des êtres vivants dont se compose une société, un groupe, une collectivité.

PERSONNE 1. Individu en tant qu'être physique singulier. Considéré selon ses caractéristiques physiques et morales. **2.** Être singulier universel, conscient de lui-même, doué de raison et de volonté.

Comment individualise-t-on les autres ? Comment perçoit-on leur singularité ? Pour éprouver la singularité d'un individu, l'Académie française propose de le diviser puis d'observer si cette opération le détruit ou non. Moins attentatoire à son intégrité physique, l'identification d'une personne, peut-être réalisée à partir de l'observation de ses caractéristiques physiques et morales. Il est donc intéressant de s'interroger sur la nature des caractères propres, qui constituent l'originalité d'une personne et la distingue des autres.

Dans le domaine visuel, c'est principalement le visage, la silhouette mais aussi la démarche et la posture, ou plus généralement les façons de se mouvoir qui sont utilisés pour l'identification (Cutting and Kozlowski, 1977; Loula et al., 2005; Westhoff and Troje, 2007). Dans le domaine

auditif, il s'agit de la voix, qui est perçue à travers la langue parlée mais également à travers divers bruits comme les éternuements, les rires ou encore la toux. Enfin il a été montré que les informations tactiles, telles celles fournies par la texture de la peau, (Kaitz, 1992), ou encore les informations olfactives peuvent être utilisées pour identifier les personnes ou discriminer deux individus inconnus (Wallace, 1977; Porter, 1998; Olsson et al., 2006), mais de façon largement inconsciente (Lundstrom et al., 2009). D'autre part les informations sémantiques (comme le nom, la profession, les liens familiaux, les goûts, les aptitudes et divers autres éléments biographiques) et épisodiques (comme les souvenirs partagés) sont associées aux caractéristiques physiques et permettent également l'identification. Enfin les caractéristiques morales d'un individu (sa *personnalité*) sont dérivées de l'historique de nos interactions. La personnalité concerne les traits de caractère (gentil, sérieux, de confiance, précautionneux, etc.), les intentions, les comportements sociaux, les états mentaux et les relations avec nous-mêmes et les autres (Winter et al., 1985; Greenwald and Banaji, 1995; Bargh et al., 1996; Todorov and Uleman, 2002; Saxe et al., 2004; Zaki and Ochsner, 2009). La permanence et l'unité des caractères propres physiques, biographiques et comportementaux permettent de reconnaître une personne, de prévoir son comportement futur et d'interagir avec elle adéquatement.

Dans la suite nous nous intéresserons à la reconnaissance des individus à partir du visage et de la voix. Notons que les visages, comme les voix, sont des stimuli particuliers car ils ne renseignent pas uniquement sur l'identité mais aussi sur des caractéristiques plus générales des individus comme l'âge, le genre, l'origine ethnique, géographique (accents), l'attractivité, le statut social, le lien familial, l'état de santé, etc. Par ailleurs ils jouent un rôle de premier plan dans la communication symbolique et affective, étant les media privilégiés - mais non les seuls - du langage et des émotions. L'ensemble de ces fonctions font de la voix et du visage, deux des stimuli les plus utiles et les plus fondamentaux de notre environnement social. Cette importance sociale peut être mise en parallèle de l'aptitude spécifique qui s'est développée pour leur perception par rapport au traitement d'autres éléments visuels ou auditifs de l'environnement, que nous décrirons par la suite.

A.2. Reconnaissance individuelle : un processus triple

L'individualisation de personnes non familières peut être réalisée à partir des seules caractéristiques physiques. Par exemple on discrimine facilement un visage A d'un visage B, ou une voix A d'une voix B. Ceci est réalisé par la perception des caractéristiques physiques propres (**Figure 1**). Par contre, la reconnaissance d'un individu familier (« C'est A ! », « C'est B ! ») suppose un niveau cognitif supplémentaire au cours duquel on fait appel à l'ensemble des informations concernant cet individu, non perçues à l'instant présent mais gardées en mémoire. L'ensemble de ces informations comprend les informations sensorielles, sémantiques, épisodiques et émotionnelles décrites ci-dessus. Il constitue la représentation cognitive - le concept - de chaque individu. Par exemple, entendre la voix de A fera écho intérieurement notamment au prénom et au visage de A. C'est-à-dire que l'indice perceptif déclenchera un rappel

plus large de l'ensemble des caractères propres de l'individu A. Ainsi, contrairement à la *discrimination* qui se joue sur le terrain perceptif, la *reconnaissance individuelle*, recrute en plus des connaissances conceptuelles.

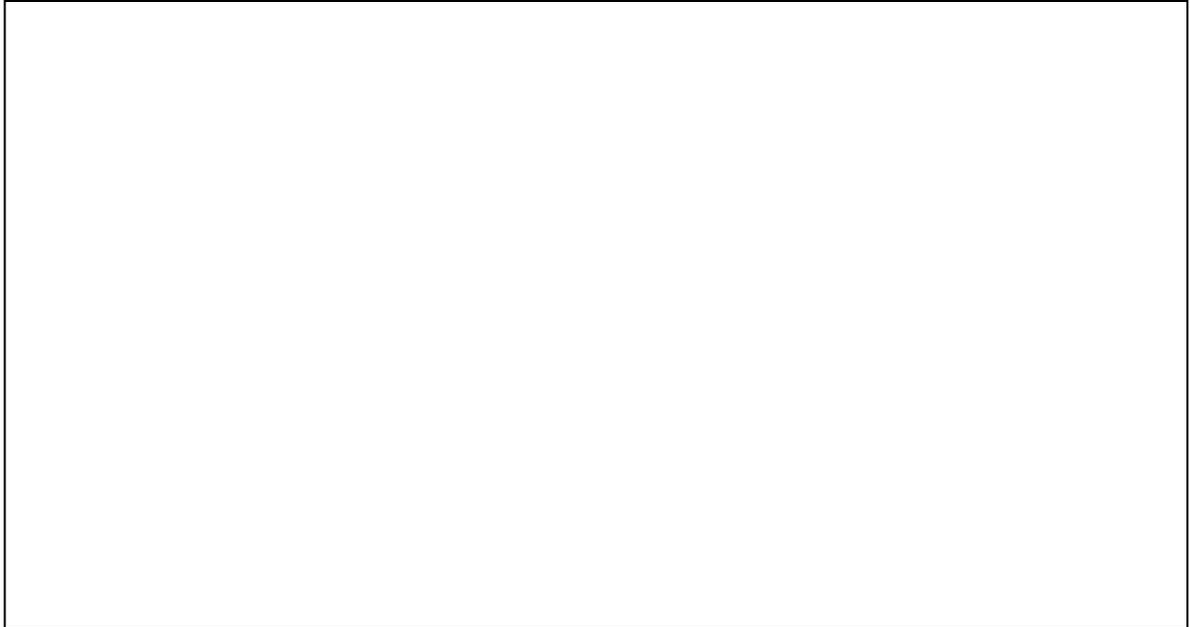


Figure 1 : Modèle cognitif proposé par Bruce et Young (1986) pour l'analyse des visages (gauche). Ce modèle a été augmenté pour prendre en compte l'accès à l'identité à partir d'autres indices : la voix et le nom (droite). D'après Bruce and Young, 1986; Burton et al., 1990; Ellis et al., 1997a; Neuner and Schweinberger, 2000.

BRUCE et YOUNG (1986) ont proposé un modèle sériel du traitement de l'information faciale, dans lequel différents processus se succèdent pour permettre finalement la reconnaissance de la personne et le rappel de son nom (**Figure 1**). Le visage serait d'abord analysé perceptivement afin d'extraire sa configuration d'ensemble (*Structural encoding*) : c'est à cette étape que différents visages seraient distingués les uns des autres. Le modèle prédit ensuite une analyse séparée de l'identité (*Face recognition units*), de celle des émotions et des signaux de communication. La représentation de l'identité du visage serait à ce stade comparée à l'ensemble de visages connus mémorisés et reconnue comme familière ou non. Dans l'étape suivante (*Person identity node*) l'ensemble des caractères propres physiques est assemblé. Enfin les deux dernières étapes postulent l'accès aux informations sémantiques puis au nom. La reconnaissance individuelle aurait lieu dans ces trois dernières étapes. Ce modèle a été adopté par la suite et augmenté pour expliquer les étapes menant à la reconnaissance de l'individu à partir de la voix (Ellis et al., 1997a) et du nom (Burton et al., 1990) (**Figure 1**). La finalité de leurs traitements étant la même, l'analyse des visages, voix et noms pourrait se rejoindre et partager certaines étapes du modèle. Dans un premier temps nous décrirons les processus de discrimination des visages et des voix puis comment ces stimuli jouent un rôle d'indice pour le rappel global de l'individu.

A.3. Spécificité du traitement perceptif des visages et perception de la singularité

La première et non des moindres caractéristique du traitement perceptif des visages est son extrême rapidité : la détection d'un visage nécessite moins de 100 ms (Braeutigam et al., 2001; Johnson, 2005). Des expériences menées à la lisière de l'aveuglement attentionnel, ont montré que des présentations d'images de visages de l'ordre de 200 ms suffisent à l'identification du genre et de l'identité (Reddy et al., 2004; Reddy et al., 2006), ce qui est extrêmement faible au vu de la grande complexité visuelle des visages mais plus encore de la très grande homogénéité entre les différents exemplaires de visages. L'homogénéité provient du fait que les visages sont analysés au niveau subordonné et fin de l'identité, contrairement aux objets qui sont analysés au niveau générique de la catégorie (Tanaka, 2001). La rapidité et la précision avec laquelle les visages sont analysés tiennent aux processus perceptifs particuliers qui leur sont dédiés et qui diffèrent de ceux utilisés pour les autres objets visuels. Dans la suite nous décrivons plus particulièrement le traitement des informations relatives à l'identité, mais notons que des processus similaires permettent l'analyse des émotions faciales.

A.3.1. Perception holistique de la singularité d'un visage

C'est du détail des traits (les composants du visage : la couleur et la forme des yeux, du nez, de la bouche), et de leurs relations (la configuration du visage : les distances, les symétries entre les traits) que résultent l'originalité d'un visage par rapport à un autre (Valentine, 1991). Alors que pour un objet complexe, les détails sont analysés séquentiellement avant qu'une structure d'ensemble ne soit définie, un visage est tout de suite analysé de façon globale – *holistique* – et en parallèle pour les différents traits et leurs relations. Cette aptitude spécifique explique la rapidité d'analyse des visages. En effet alors que dans le cas d'un objet les yeux font des allers-retours constants entre les différents détails, augmentant le temps de traitement, dans le cas d'un visage les saccades oculaires sont de faible amplitude et restent concentrées au centre du visage, captant ainsi par une vision d'ensemble l'originalité à la fois des composants et de leur configuration (Orban de Xivry et al., 2008, **Figure 16**). Deux paradigmes expérimentaux ont historiquement mis en évidence ce processus perceptif spécifique aux visages: le paradigme des visages composites et celui du renversement des visages (**Annexe A**).

Beaucoup moins documenté que l'identification à partir de l'apparence physique statique, un autre type d'identification faciale peut être réalisé à partir des attributs dynamiques du visage : par exemple la façon de bouger la tête, les yeux, les sourcils, la façon de sourire et en règle générale les différentes mimiques faciales idiosyncratiques apportent de nombreuses informations sur l'identité de la personne (O'Toole et al., 2002; Lander and Davies, 2007). Cependant lorsque les informations statiques et dynamiques sont toutes deux accessibles, les personnes se fondent principalement sur les informations statiques pour la reconnaissance, ces informations étant plus fiables et plus rapidement identifiées que les informations dynamiques.

Comment les visages sont comparés les uns aux autres ? Les études psychologiques ont montré que les représentations mentales des visages pourraient s'inscrire dans un *espace des visages* multidimensionnel (**Figure 2**, Valentine, 1991). Dans cet espace chaque visage est représenté par un point plus ou moins éloigné d'un *visage prototypique* central. Les différentes dimensions représentent les variations configurales et componentielles possibles le long desquelles les identités faciales évoluent. Ainsi des visages similaires seraient représentés par deux points proches. Théoriquement des caricatures de visages s'inscriraient aux extrémités des axes, loin du visage moyen prototypique. De plus à l'opposé de chaque visage, centré sur le prototype, ce trouverait un *anti-visage*, ayant des caractéristiques faciales opposées.

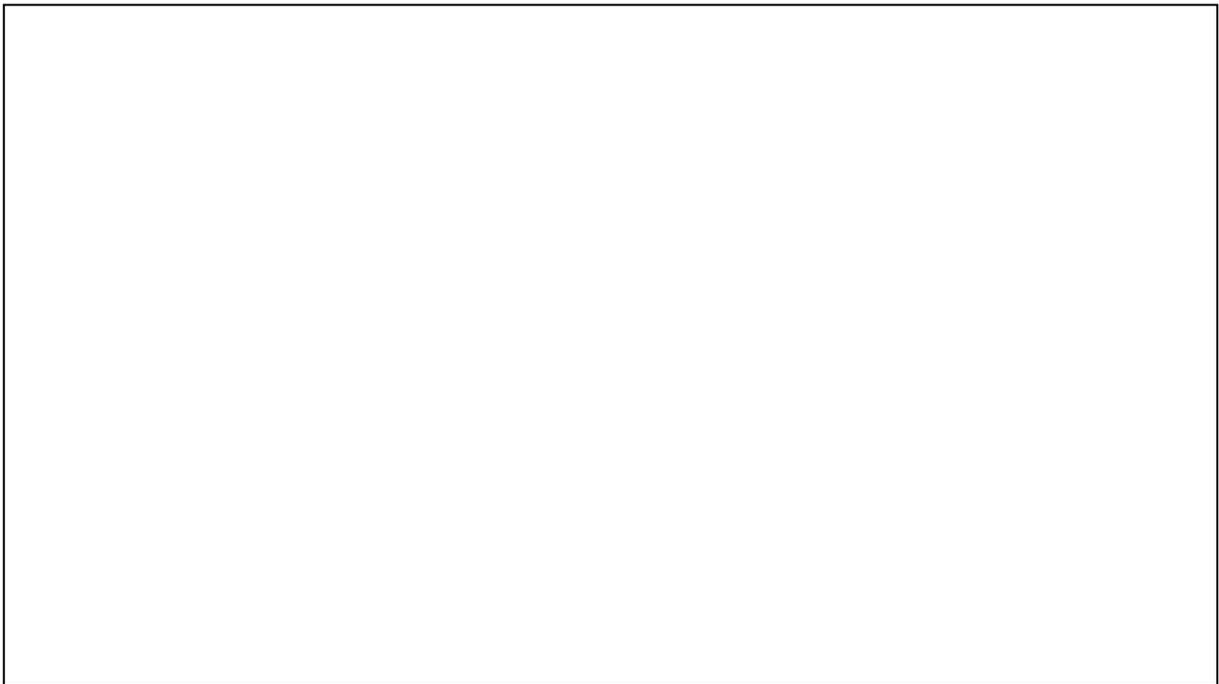


Figure 2 : « Espace des visages » multidimensionnel. Dans cet espace chaque visage est représenté par un point plus ou moins éloigné d'un visage prototypique central. Les différentes dimensions représentent les variations configurales et componentielles possibles le long desquelles les identités faciales évoluent. Théoriquement des caricatures de visages s'inscriraient aux extrémités des axes, loin du visage moyen prototypique. Des visages similaires seraient représentés par deux points proches. A l'opposé de chaque visage, centré sur le prototype, ce trouverait un « anti-visage », ayant des caractéristiques faciales opposées. D'après Tsao et al., 2006.

Ce cadre théorique permet de rendre compte des capacités de discrimination utilisées chaque jour envers un grand nombre de visages inconnus. En effet chaque nouveau visage rencontré serait mentalement disposé sur cet espace et comparé aux visages connus. Une telle représentation mentale a été démontrée par l'utilisation d'un protocole fondé sur une illusion perceptive. Dans ces études, les sujets regardaient un visage pendant plusieurs minutes ; suite à cette longue période d'exposition, leur perception d'autres visages était évaluée (Webster and MacLin, 1999; Leopold et al., 2001; Leopold et al., 2005; Afraz and Cavanagh, 2008). Les auteurs ont observé que le second visage était perçu de façon distordue. Ainsi suite à l'habituat

visage caricatural, le visage présenté en second semblait caricaturé dans l'autre sens. De la même façon l'habitué à une identité faciale, biaisait la perception de l'identité du second visage vers l'*anti-visage* du premier visage. Cet effet a été observé en manipulant l'identité mais également l'âge, le genre et l'ethnicité d'un visage (Webster et al., 2004; Schweinberger et al., 2010). De façon intéressante le point de symétrie entre les visages perçus en premier et les visages perçus de façon distordue en second était unique et correspondait au visage prototypique.

Notons que le visage prototypique est déterminé par l'expérience propre de chacun, et s'organise autour du pool de visages personnellement familiers. Ce modèle de représentation mentale suggère que le réseau pour l'identification des visages est soumis à un apprentissage et change au cours de la vie. Ainsi ce cadre permettrait de prendre en compte l'effet comportemental *Other-race effect*, selon lequel la discrimination des visages éloignés physiquement du visage prototypique propre est plus difficile à réaliser.

A.3.2. Progression de l'expertise pour sa propre espèce et identification

A la naissance, dès les premières minutes, et pendant le premier mois de la vie, le nouveau-né préfère observer un visage plutôt qu'un visage inversé ou un objet (Goren et al., 1975; Johnson et al., 1991; Mondloch et al., 1999; De Schonen, 2009) suggérant une aptitude innée à la détection des visages. Par ailleurs les nouveau-nés ne présentent pas de préférence parmi les visages de singes et les visages humains. Les propriétés du visage qui attirent le nouveau-né sont donc présentes chez ces deux espèces ; et les développementalistes se sont très tôt attelés à déterminer les propriétés élémentaires du visage qui guident sa détection (**Annexe B**). Il faut attendre l'âge de deux ou trois mois pour que la préférence visuelle aille vers le faciès humain plutôt que le faciès primate général (Kelly et al., 2005). De façon concomitante jusqu'à l'âge de 6 mois les enfants discriminent aussi bien les visages de singes entre eux que les visages humains entre eux (Pascalis et al., 2002). Par contre dès l'âge de 9 mois, les enfants n'arrivent plus à discriminer deux visages de singes, alors que leur discrimination des visages humains s'améliore. L'absence de préférence pour sa propre espèce à la naissance (*broad tuning*) se canalise (*perceptual narrowing*) de façon drastique pour les individus du cercle socio-écologique, jusqu'à affaiblir fortement l'aptitude à discriminer les individus de groupes ethniques différents de celui dans lequel on évolue (*Other race effect*), comme prédit par le modèle de *l'espace des visages*. L'importance de l'expérience et de l'apprentissage dans l'évolution de la représentation des visages a été mise en évidence en observant que la discrimination des visages de singes peut être maintenue pendant un certain temps si l'enfant continue à être en contact avec des visages de singe (Pascalis et al., 2005).

A.4. Traitement perceptif de la voix

Les voix ont longuement été analysées en tant qu'outil de la parole. Par conséquent l'étude des voix comme moyen de communication non-linguistique est plus récente que celle des visages. Cependant un parallèle a été créé entre ces deux stimuli sociaux, renforçant des investigations complémentaires. Ainsi d'une part le rôle des visages dans la compréhension de la parole a été étudié, par exemple par l'analyse de leur dynamique. D'autre part des investigations sur les informations non-linguistiques contenues dans la voix telles l'identité, le genre, l'âge, l'attractivité ou les émotions, ont été menées, en considérant la voix comme un *visage sonore* (Belin et al., 2004).

A.4.1. Perception de la singularité de la voix

L'originalité d'une voix est perçue par divers indices temporels et fréquentiels contenus dans la vocalisation. La vocalisation suit un modèle *source-filtre* en deux temps (Fant, 1960). La source est constituée des sons créés au niveau du larynx et peut avoir deux origines distinctes (**Figure 3a et 3d**). Premièrement, lorsque les cordes vocales vibrent, elles émettent un son dont la *fréquence fondamentale* f_0 détermine la hauteur de la voix. La fréquence fondamentale diffère d'un individu à l'autre, par exemple celle des hommes est en moyenne plus basse que celle des femmes. Elle varie également au cours de l'élocution, donnant lieu à la musicalité de la prosodie caractéristique de chaque individu. Ces caractéristiques fréquentielles fournissent des indices sur l'identité du locuteur. Deuxièmement, un son peut être produit même lorsque les cordes vocales sont closes. L'air expulsé est alors soumis à un écoulement turbulent résultant en un son bruité à large bande de fréquence, comme lors de chuchotement ou de la prononciation de certaines consonnes (p, t, k, s, ch, f).

Le filtre acoustique du tractus vocal agit sur ces différents sons émis par la source. Le tractus vocal est constitué de plusieurs cavités et tissus supra-laryngés, qui s'étendent de la glotte jusqu'aux lèvres et aux narines, et dont l'anatomie est spécifique à chaque individu. Notamment la taille du tractus vocal est corrélée positivement avec la taille de l'individu. La configuration du tractus vocal est dynamique, changeant en fonction des mouvements d'articulation effectués. Ainsi le souffle pulmonaire résonne de différentes façons, en fonction des différentes configurations prises par le canal vocal et par la langue. Chaque configuration est caractérisée par un ensemble de fréquences de résonance (**Figure 3b**). L'énergie du son provenant de la source est alors fidèlement transmise à ces fréquences-là et largement filtrée aux autres, créant des pics de grande amplitude dans l'enveloppe spectrale appelés les *formants* (**Figure 3e**, Fant, 1960). Les formants jouent un rôle central dans le langage humain puisqu'ils sont les paramètres acoustiques les plus importants pour distinguer les différentes voyelles et une partie des consonnes entre elles (Fitch, 2000). La descente du larynx chez l'Homme a augmenté la taille du tractus vocal (**Figure 3**), permettant à la langue des mouvements à la fois verticaux et horizontaux qui sont à la base d'une variété plus grande de formants prononcés possibles (par exemple ceux utilisés pour les sons /i/ et /u/). Les formants remplissent également un second rôle que l'on partage cette-fois

ci avec les autres mammifères : ils constituent un indicateur de la taille et de l'identité de l'émetteur (Bachorowski and Owren, 1999; Ives et al., 2005; Baumann and Belin, 2010). Par exemple les individus de grande taille possèdent un tractus vocal plus long créant des formants plus bas en fréquence et plus espacés les uns des autres. Par conséquent l'identité peut être perçue grâce à la voix parlée mais également grâce à d'autres types de sons soumis au modèle source-filtre comme un rire, un bâillement, un éternuement, etc.

Les voix sont des stimuli complexes, caractérisés par une unique combinaison de composants physiques qui correspondent à la configuration de l'appareil vocal de chacun. C'est grâce à cette combinaison unique et à son traitement efficace par l'auditeur, que la discrimination et la reconnaissance de l'individu sont rendues possibles (Schweinberger et al., 1997b). Cependant contrairement aux visages l'identification vocale n'est pas instantanée, et nécessite un échantillonnage plus long sur plusieurs centaines de millisecondes, permettant le calcul de signaux auditifs moyens concernant le ton et les formants. Ainsi les temps de réactions pour identifier une voix sont de 1 s contre 853 ms pour un visage (Joassin et al., 2004).



Figure 3 : Représentation schématique du tractus vocal (rouge) chez l'orang-outan, le chimpanzé et l'humain (gauche). Chez l'Homme le tractus vocal est augmenté par la descente du larynx (jaune). En dessous sont présentés les spectrogrammes et les oscillogrammes de sons produits par ces trois espèces. Les formants sont bien visibles sur les trois spectrogrammes. Illustration du modèle source-filtre (droite) : (a) composition spectrale de la source laryngée, (b) fonction de transfert du tractus vocal, (c) effet radiatif des lèvres. En dessous sont représentés les résultats spectraux à chaque étape de la production du son : (e) renforcement sélectif des harmoniques créant des formants, (f) redressement de la pente du spectre. D'après Fitch, 2000; Belin, 2006; Ghio and Pinto, 2007.

De la même façon que les représentations des visages s'inscrivent dans un espace multidimensionnel des visages, on pourrait s'attendre à ce que les représentations des voix s'inscrivent dans un *espace des voix*. En utilisant le même protocole d'illusion perceptive que celui décrit précédemment, plusieurs études ont mis en évidence une perception distordue des voix suivant l'adaptation à une première voix. Ainsi l'habituation à une identité vocale, biaisait la perception de l'identité de la seconde voix vers l'*anti-voix* de la première voix. Cet effet a été observé en manipulant l'identité mais également l'âge et le genre (Schweinberger et al., 2008;

Zäske et al., 2010; Zäske and Schweinberger, 2011). A cause de limitations techniques ne permettant pas de morpher les voix dans plus d'une dimension, ces premières études n'ont pas mis en évidence l'existence d'une voix prototypique. Celle-ci a donc été démontrée récemment en morphant les voix dans plus d'une dimension et en observant que les sujets adaptaient leur jugement par rapport à une voix prototypique (Latinus and Belin, 2011).

A.4.2. Une spécificité pour les voix, similaire à celle pour les visages ?

De façon similaire aux visages, les études comportementales ont montré que l'attention des nouveau-nés s'oriente plutôt vers les voix humaines que vers d'autres stimuli sonores non-sociaux (Ockleford et al., 1988; Ecklund-Flores and Turkewitz, 1996; Vouloumanos and Werker, 2007). Cependant contrairement aux visages, les nouveau-nés ont été exposés pendant la grossesse à des voix et particulièrement à celle de leur mère (Kisilevsky et al., 2003). Il n'est donc pas possible de conclure quant au caractère inné ou acquis de cette préférence.

L'aptitude à distinguer certaines voix est tout de même présente très tôt au cours du développement. En effet les nouveau-nés préfère la voix de leur mère à celle d'autres individus (DeCasper and Fifer, 1980). Mais ils distinguent également la voix de leur père de celle d'hommes inconnus et sont capables de discriminer entre deux voix mâles non-familiales (DeCasper and Prescott, 1984). Il n'a pas été étudié si la discrimination des voix est également soumise à l'*other race effect*, c'est-à-dire s'il est plus aisé de discriminer l'identité vocale de deux personnes de son environnement socioculturel que d'un environnement plus éloigné. Cependant il a été montré que les nouveau-nés sont capables de discriminer les prononciations de différentes syllabes et sons présents dans des langues étrangères et non dans la langue maternelle et que cette capacité se perd par la suite (Werker and Tees, 1984; Kuhl et al., 1992; Best et al., 1995; Cheour et al., 1998; Kuhl et al., 2003).

A.5. Les diverses associations entre voix et visage

La plupart des signaux faciaux et vocaux sont redondants. En effet l'identité, l'âge, le genre, l'espèce, l'attractivité, la carrure, l'origine sociale ou ethnique, mais également les émotions ou la parole peuvent être perçus à travers chacun des deux canaux sensoriels (**Figure 4**, Campanella and Belin, 2007). Etant donnée l'importance de ces signaux pour la communication interindividuelle, leur redondance n'est pas étonnante et devrait augmenter la fiabilité avec laquelle on juge et perçoit les autres.

A.5.1. Associations arbitraires et associations amodales

Les informations provenant de la voix et du visage sont congruentes quand il s'agit d'émotions et de parole, cependant ce n'est pas toujours le cas avec les autres associations. Par exemple les associations liées à la spécificité de l'individu (l'âge, le genre, etc.) sont des stéréotypes appris au cours du développement (Bahrick et al., 2005). Ces stéréotypes sont suffisamment fiables pour permettre de prédire le genre (Bachorowski and Owren, 1999) et l'âge d'une personne seulement en entendant sa voix mais ne le sont pas assez par exemple pour prédire sa taille (Collins and Missing, 2003; González, 2004; Rendall et al., 2005; Bruckert et al., 2006).

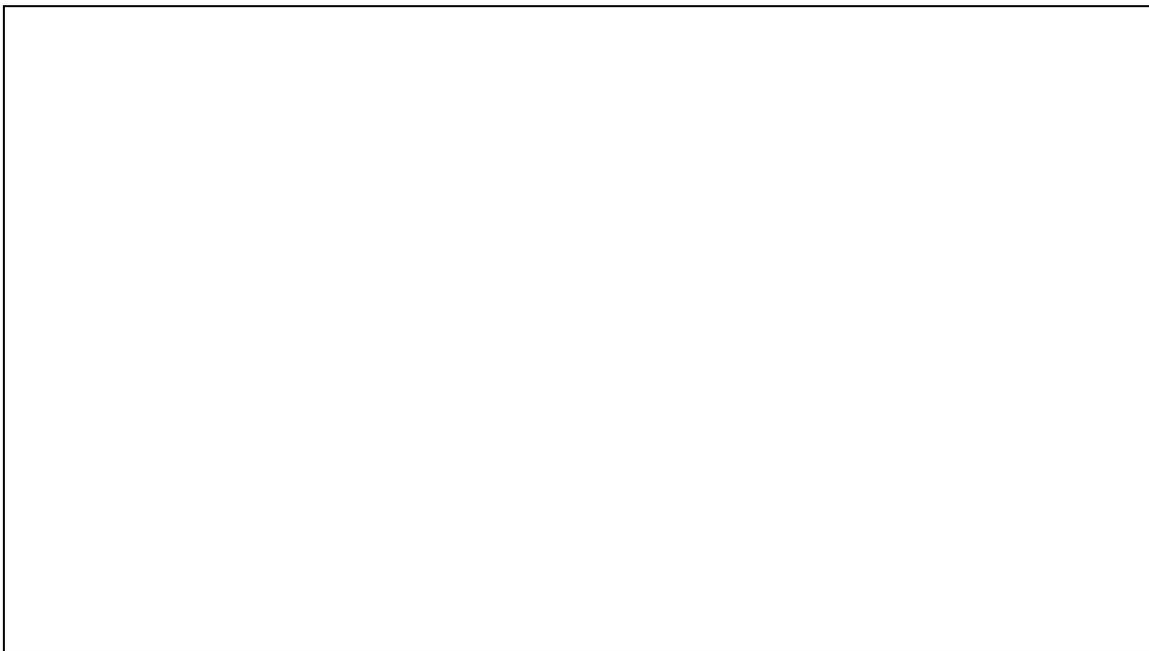


Figure 4 : Diverses associations entre les voix et les visages. Le modèle reprend les étapes du modèle de Bruce et Young, les duplique pour la modalité auditive et met en valeur les ponts qui existent à différents endroits entre les deux lignées. Notamment les auteurs présentent les associations entre les voix et les visages pour la compréhension de la parole, i.e. la communication symbolique, pour la compréhension des émotions, i.e. la communication affective, et la reconnaissance des individus. D'après Belin et al., 2004.

Une distinction a donc été proposée entre les associations apprises d'une part et les associations non apprises d'autre part. Les associations apprises sont également appelées *associations arbitraires*, car les attributs visuels (par exemple, la configuration faciale, la couleur de la peau, la pilosité faciale) ne peuvent être déduits des attributs auditifs (par exemple le ton et le timbre de la voix) autrement que par un apprentissage. Cet apprentissage est réalisé passivement grâce à la coterporalité et la colocalisation de leur survenue. Par opposition les associations liées à la parole ou aux émotions sont détectées sans apprentissage très tôt par les jeunes enfants (Bahrick et al., 1998). Dans ces cas, la voix et le visage sont dynamiques et partagent en plus de

leur colocalisation et leur cotemporalité, un rythme, une intensité, une durée, et un tempo communs (Chandrasekaran et al., 2009). Lorsque ces dynamiques visuelles et auditives concordent, la voix et le visage sont perçus comme étant associés (Bahrack and Lickliter, 2000, 2002; Lander et al., 2007), même lorsqu'il s'agit d'un leurre comme dans le cas du ventriloquisme. L'association entre la voix et le visage est alors réalisée à un niveau perceptif qualifié d'*amodal* car il ne dépend pas d'une modalité particulière. De même que la perception unimodale des visages ou celle de la langue, la perception des associations amodales voix-visage se spécifie pendant le développement. Elle est présente très tôt chez le nouveau-né et ceci de façon non-spécifique (*broad perception*) (Bahrack et al., 1998) puis se spécialise et s'intensifie pour les individus de sa propre espèce (*perceptual narrowing*) entre 6 et 10 mois (Lewkowicz and Ghazanfar, 2006; Lewkowicz et al., 2008; Lewkowicz and Ghazanfar, 2009).

A.5.2. Association de l'identité faciale et vocale

L'association entre la voix et le visage au niveau de l'identité est apprise. Cependant elle ne suit pas une règle générale (comme dans le cas du genre ou de la taille), mais représente une association unique, ce qui la rend fiable. L'utilisation des deux modalités de façon complémentaire pour percevoir les autres a été mise en évidence expérimentalement par un protocole de priming cross-sensoriel. Ainsi entendre la voix d'un individu augmente la rapidité avec laquelle on juge la familiarité pour le visage de cet individu, et *vice versa* (Ellis et al., 1997a; Schweinberger et al., 1997a).

L'association d'identité n'est pas présente à la naissance comme les associations amodales, car elle doit être apprise. Cependant cet apprentissage est réalisé très tôt. Ainsi dans une expérience, des enfants de 2, 4 et 6 mois ont été familiarisés avec plusieurs associations uniques de voix et de visages individuels. Lors de la phase de test, les enfants de 4 et 6 mois, mais non ceux de 2 mois, ont été capables de reconnaître des associations erronées entre ces voix et visages familiers. De plus c'est seulement à partir de 6 mois, que les enfants ont démontré une aptitude à mémoriser la relation unique liant l'identité d'une voix et celle d'un visage, et à indiquer quel visage appartenait à quelle voix (Bahrack et al., 2005). Ainsi l'association de l'identité voix-visage serait constituée de deux phénomènes conjoints : d'une part l'association voix-visage comme une unité à part entière et d'autre part la relation flexible entre l'association et les stimuli unimodaux la constituant, permettant de retrouver l'information d'une modalité à partir de la seconde.

Enfin il est possible de comprendre que l'identité d'une voix correspond à l'identité d'un visage même si ceux-ci sont inconnus, en se fondant cette fois-ci non pas sur les caractéristiques idiosyncratiques du visage statique (les traits du visages et leur configuration) mais sur les caractéristiques idiosyncratiques dynamiques (les mouvements et mimiques du visage, la ponctuation du discours par des hochements de tête, les roulements des yeux ou haussement des sourcils, etc.) qui suivent le même tempo que le discours (Kamachi et al., 2003; Lachs and Pisoni, 2004; Lander et al., 2007). Cependant en conditions non expérimentales de choix forcé, ces caractéristiques sont principalement utilisées pour la compréhension du discours et non pour

l'identification, comme ceci a été démontré chez des patients prosopagnosiques qui perçoivent et utilisent ces informations pour faciliter leur compréhension du discours mais ne peuvent les utiliser pour faciliter l'identification des voix (von Kriegstein et al., 2008).

A.5.3. Les associations amodales renforcent l'apprentissage des associations arbitraires

Les stimuli cohérents présentent une forte saillance et attirent l'attention dès le plus jeune âge (Bahrick and Lickliter, 2000; Flom and Bahrick, 2010). Dès lors les associations amodales jouent un rôle important au cours du développement car elles dirigent l'attention vers les associations arbitraires et aident leur mémorisation. D'ailleurs la facilitation de l'apprentissage de l'association voix-visage au niveau de l'identité par une exposition préalable au discours audio-visuel s'observe également à l'âge adulte (Sheffert and Olson, 2004). Le fait que l'apprentissage des associations voix-visage soit principalement multisensoriel donne lieu à des attentes cross-modales fortes et automatisées de nos systèmes sensoriels. Ainsi tous les éléments dynamiques peuvent ne pas concorder entre la voix et le visage, par le biais d'expérimentations. Des illusions tenaces sont alors mises à jour, comme celle de l'effet McGurk dans le domaine de la parole (McGurk and Macdonald, 1976; Vatakis et al., 2008). Dans le domaine de l'identité, des effets d'interférence sont également observés montrant qu'il est difficile d'identifier une voix lorsqu'un visage incongruent dynamique est présenté simultanément (Joassin et al., 2004). Au contraire la reconnaissance de l'identité de la voix est plus rapide si le visage est présenté simultanément, qu'il soit dynamique (comme dans une condition naturelle) ou statique (Joassin et al., 2004; Schweinberger et al., 2007; Joassin et al., 2011).

A.5.4. Facilitation de l'identification unimodale par l'apprentissage préalable de l'association multimodale

La reconnaissance de l'identité vocale ou faciale est facilitée lorsque la voix et le visage ont été perçus et appris ensemble, c'est-à-dire lorsqu'ils ont été associés de façon multisensorielle au préalable. Cependant si la facilitation est quantitative dans le cas de la reconnaissance des visages, elle est qualitative pour la reconnaissance des voix. En effet l'apprentissage puis le rappel d'identités vocales isolées apprises dans un contexte unimodal est très pauvre (von Kriegstein and Giraud, 2006; Joassin et al., 2011), alors qu'il est fiable même s'il est plus lent pour les identités faciales apprises dans un contexte unimodal (Paller et al., 2000; Joassin et al., 2011). Les deux phénomènes ne sont donc pas équivalents. Par exemple dans des situations pratiques de psychologie légale, les témoins oculaires sont plus fiables que des témoins auditifs (Olsson et al., 1998; Mullennix et al., 2009). Une partie de cet effet peut s'expliquer par le temps d'analyse plus long nécessaire à l'identification des voix par rapport aux visages, qui augmente l'incertitude quant à l'identification (Joassin et al., 2004). Cependant une part plus importante de cette

différence peut être imputée au fait que l'apprentissage de l'identité vocale en conditions normales repose fermement sur l'identité multisensorielle, perçue antérieurement à travers le discours audio-visuel (Sheffert and Olson, 2004; von Kriegstein and Giraud, 2006). Dans le cas où l'identité faciale ne peut être perçue pendant le discours audio-visuel, par exemple chez les personnes prosopagnosiques, l'identification vocale est mauvaise alors qu'on pourrait s'attendre au contraire à ce qu'elle soit meilleure pour compenser l'incapacité des personnes prosopagnosiques à identifier les visages (von Kriegstein et al., 2008). Par ailleurs chez des sujets sains, lorsque la composante visuelle est absente de l'apprentissage, des phénomènes de compensation se mettent en place : les personnes simulent mentalement des visages dynamiques et forment des prédictions mentales sur la taille, le genre ou l'apparence physique des locuteurs (von Kriegstein and Giraud, 2006). La vision serait donc le sens dominant pour permettre l'identification.

A.6. Reconnaissance individuelle à partir du visage ou de la voix

Les voix et les visages sont des portes d'accès vers le rappel complet des informations sémantiques (le nom, les informations biographiques, les autres indices perceptifs idiosyncratiques) et épisodiques (les souvenirs autobiographiques partagés) concernant une personne. Cependant cette voie n'est pas toujours ouverte. Ainsi un effet fréquemment expérimenté par chacun est l'impression de reconnaître une personne sans pour autant réussir à la replacer dans un contexte ou à se rappeler de son nom. Cet effet appelé aussi *Butcher on the bus* quand il s'agit d'individus, ou plus généralement *Feeling of Knowing*, résulte du fait que l'impression de familiarité pour un individu peut être préservée sans reconnaissance individuelle de celui-ci (Yovel and Paller, 2004; Gruppuso et al., 2007). Une même séquence a été proposée par BRUCE et YOUNG : dans leur modèle, la familiarité est analysée au niveau des *Face Recognition Units* en même temps que l'identité et précède la reconnaissance individuelle. Au contraire c'est seulement au niveau du *Person Identity Node* que l'ensemble des caractères propres de l'individu serait rappelé et mis en parallèle des connaissances sémantiques et épisodiques concernant la personne.

Sur la base des exemples précédents, TULVING (1985) a développé le paradigme *Remember/ Know*, d'auto-évaluation du degré de reconnaissance. Une réponse *Remember* signifie une reconnaissance accompagnée du rappel de plusieurs éléments liés à l'item présenté. Une réponse *Know* signifie une familiarité avec l'item sans rappel d'éléments contextuels ou associés. L'application de ce paradigme à des tâches de reconnaissance individuelle a mis en évidence que les voix et les visages n'ont pas la même propension à rappeler les informations sémantiques et épisodiques liées à un individu. En effet les voix seraient des indices plus faibles que les visages pour accéder aux connaissances épisodiques et sémantiques concernant une personne (Hanley et al., 1998; Hanley and Turner, 2000; Damjanovic and Hanley, 2007). Cependant, ceci serait dû à une plus grande difficulté à identifier les voix apprises seules et non à accéder aux informations supplémentaires une fois que la voix est reconnue. Ainsi dans le modèle de BRUCE et YOUNG,

cette difficulté à identifier les voix se traduirait à l'accès aux *Voice recognition units* plutôt qu'à l'accès au *Person identity node*.

Une question qui se pose concerne l'ordre de réactivation des différents éléments du concept lié à l'individu. La réactivation de la voix et du nom à partir du visage précède-t-elle le rappel des connaissances sémantiques et épisodiques ? Au contraire, le rappel du nom viendrait-il seulement à la fin de la réactivation des connaissances sémantiques et épisodiques ? Peut-on se rappeler de certains éléments mais non d'autres ?

A.7. Limites et questions relatives au modèle sériel

BRUCE et YOUNG furent les premiers à proposer un modèle très simple et didactique du traitement des visages, de leur perception à la reconnaissance individuelle. Cependant de nombreuses études de cas psychiatriques et neurologiques, ainsi que les avancées en neurosciences, ont remis en cause son côté sériel et plaidé pour sa complexification (Gobbini and Haxby, 2007; Lucchelli and Spinnler, 2008, **Figure 5**). Citons deux exemples. Le syndrome de CAPGRAS est un trouble psychiatrique engendrant un délire spécifique d'identification des personnes. Les patients qui en sont atteints perdent le sentiment de familiarité envers les personnes connues tout en continuant à les identifier correctement, ayant l'illusion d'être en présence d'un sosie de l'individu et non de l'individu lui-même (Ellis and Young, 1990; Ellis et al., 1997b; Ellis and Lewis, 2001). Ce cas remet donc en cause le fait que l'analyse de la familiarité du visage précéderait le rappel des informations sémantiques liées à la personne. De façon similaire, les personnes *prosopagnosiques*, bien qu'elles ne puissent pas accomplir la première étape d'encodage structural des visages proposée par BRUCE et YOUNG, reconnaissent de façon non consciente qu'elles sont en présence de visages familiers, comme ceci peut être révélées par des tests électrodermaux (Bauer, 1984 ; Tranel and Damasio, 1985). Ainsi, il a été proposé que parallèlement à la voie consciente d'analyse des visages proposée par BRUCE et YOUNG, une seconde voie inconsciente effectuerait également une analyse en parallèle et permettrait notamment l'intégration de la familiarité. Les cas de prosopagnosie et de syndrome de CAPGRAS seraient la démonstration par double dissociation de ce processus (Burton et al., 1991 ; Schweinberger and Burton, 2003).

Une seconde question se pose à propos de ce modèle, concernant les ramifications auxquelles il a donné lieu suite à l'ajout d'autres modalités (voix, corps) et données sémantiques (noms, initiales). En effet, on peut se demander si les informations des différentes modalités se rejoignent seulement au niveau du *Person Identity Node* ou si des ponts existent tout au long de leur acheminement. Ces deux hypothèses ne sont pas mutuellement exclusives. Les expériences de BHRICK et coll. (2005) menées chez l'enfant sur le développement de l'apprentissage des associations voix-visages concernant l'identité, corroborent l'existence conjointe de ces deux phénomènes : d'une part, les stimuli sont associés sans mémorisation longue dès 4 mois, et d'autre part, l'association peut être mémorisée permettant par la suite de retrouver l'information

d'une modalité à partir de la seconde à partir de 6 mois (Bahrick et al., 2005). Ces hypothèses supposeraient de plus la possibilité de feedback entre les différents niveaux du modèle.

Enfin, nous avons vu qu'une identité dynamique du visage peut être perçue, à côté d'une identité statique du visage. Ces deux traitements pourraient nécessiter des *Face structural encoding unit* et *Face recognition unit* dédoublés (O'Toole et al., 2002).

Comment les autres espèces de primates perçoivent-elles les voix et les visages ? Ont-ils la même importance sociale que chez l'humain ? Leur analyse est-elle sérielle ? Qu'elles parties du modèle psychologique de la perception des voix et des visages peut être transposée chez le macaque ?

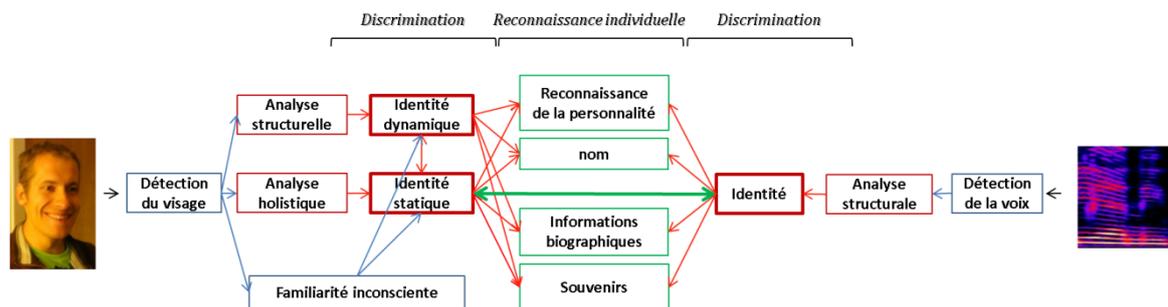


Figure 5 : Représentation schématique de différentes adaptations pouvant être apportées au modèle de Bruce et Young (1986) et Ellis et coll. (1987). La ligne verte indique l'association directe entre les voix et les visages. La reconnaissance individuelle comprend divers éléments comme le rappel de la personnalité de la personne, de son nom, des informations biographiques et des souvenirs communs. D'après O'Toole et al., 2002; Schweinberger and Burton, 2003.

B. Représentation des individus chez le macaque Rhésus

Au cours de cette partie, nous présenterons rapidement l'espèce *Macaca mulatta*, puis nous décrirons comment les objections à une pensée animale ont été dépassées par l'éthologie cognitive permettant l'étude de la perception de stimuli sociaux comme les visages et les voix par ces singes.

B.1. Les macaques Rhésus s'organisent en société complexes

Les singes rhésus (*Macaca mulatta*) font partie de l'une des vingt-deux espèces de macaques. Ce sont des singes de l'Ancien Monde (famille des *Cercopithecidae*, parvordre des Catarrhiniens), ne vivant qu'en Afrique ou en Asie, par opposition aux singes du Nouveau Monde (parvordre des Platyrrhiniens) vivant en Amérique. Au sein des Catarrhiniens, les macaques se distinguent notamment des Grands singes, dont font parties l'Homme et les chimpanzés (**Figure 6**).

L'histoire évolutive des macaques est un cas d'école de radiation adaptative : processus par lequel les organismes colonisent de nouveaux environnements, s'adaptent aux conditions locales et se diversifient en différentes espèces (Richard et al., 1989). Les macaques se différencièrent des autres *Cercopithecidae* à la même époque où les *Hominidae* se distinguèrent des autres Grands singes (**Figure 6**, Stewart and Disotell, 1998). Les ancêtres des macaques migrèrent du nord de l'Afrique vers l'Europe puis vers l'Asie (**Figure 6**). Les seuls descendants des macaques d'Afrique du Nord sont les macaques de Barbarie habitant sur le rocher de Gibraltar. En se développant en Asie, les macaques colonisèrent toute la partie sud du continent jusqu'en Chine et au Japon, et se différencièrent alors en espèces distinctes (Fooden, 1982; Zhang and Shi, 1993; Hayasaka et al., 1996 ; Morales and Melnick, 1998). Les macaques rhésus ont été particulièrement habiles à la colonisation de larges espaces géographiques (Afghanistan, Inde, Thaïlande, Chine, Pakistan, Bhoutan, Birmanie, Népal, Bangladesh, Laos, Vietnam). Ils habitent dans quasiment tous les types d'habitats, des forêts tropicales aux régions désertiques en passant par des montagnes culminants à plus de 4000 m (Richard et al., 1989). Les macaques rhésus sont aujourd'hui la seconde espèce de primates la plus étendue, après l'Homme.



Figure 6 : Hypothèse sur l'évolution des primates catarhiniens. Les chiffres encadrés représentent les estimations des dates de divergences. Les noms en italiques représentent les espèces fossiles. Les espèces se trouvant en Afrique sont en rouges et celles d'Asie en noir. La dispersion intercontinentale est indiquée par les flèches. Notons que le plus ancien ancêtre commun entre les macaques rhésus et l'Homme daterait de 25 millions d'années. Illustration d'après Stewart and Disotell, 1998.

Les groupes de rhésus peuvent compter jusqu'à 180 individus, comprenant quatre fois plus de femelles que de mâles. La hiérarchie sociale est matrilineaire, c'est-à-dire que le rang de chacun dépend de son lien de parenté avec la femelle dominante. Les groupes sont ainsi composés de plusieurs lignées de femelles apparentées, les *matrilines*, et de un ou quelques mâles non apparentés. Les rhésus, comme d'autres singes de l'Ancien monde, entretiennent des relations sociales complexes entre eux (Cheney et al., 1986; Sterck et al., 1997; Barrett et al., 2007; Silk, 2007). Ils entretiennent des interactions compétitives et d'autres coopératives entre et à l'intérieur des matriline. La plupart de leurs comportements sont népotiques, c'est à dire biaisés en faveur des individus apparentés, comme on peut l'observer lors d'épouillages préférentiels ou lors de coopérations dans les combats (Widdig et al., 2001). Cependant ils créent également des alliances avec des individus non apparentés, se réconcilient après les combats et peuvent rediriger les agressions vers d'autres membres du groupe (Kaplan, 1977 ; 1978; De Waal and Yoshihara, 1983; de Waal and Luttrell, 1988 ; Judge and De Waal, 1997). Ainsi à l'intérieur du groupe l'élévation à un rang hiérarchique donné, le succès dans les compétitions et l'accès aux ressources, dépendent tous du recrutement de supports sociaux. Ces deux habiletés : faire et défaire des alliances pour servir son intérêt propre, un comportement *machiavélique*, mais également maintenir des groupes

stables, culminent bien entendu chez les humains mais sont déjà présentes chez les macaques rhésus (Byrne, 1997; Gavrilets et al., 2008). Certains considèrent d'ailleurs que le succès adaptatif des macaques rhésus comme de l'humain proviendrait de ces aptitudes spécifiquement dédiées à la vie en groupe (Maestripieri, 2007).

Les comportements sociaux chez les macaques rhésus dépendent ainsi finement de l'identité et des liens de parentés de leurs congénères. Ceci implique fortement que les rhésus soient capables à la fois de reconnaissance individuelle et reconnaissance des liens de parenté. Les rhésus perçoivent les sons et les indices visuels de façon similaire aux humains (Heffner, 2004; Waitt and Buchanan-Smith, 2006), ces deux sens peuvent donc leur fournir des informations importantes sur leurs congénères.

B.2. Perception, individualisation et familiarité des visages

Les macaques rhésus passent la plupart de leur temps à observer leurs congénères et se montrent beaucoup plus motivés pour regarder des films mettant en scène leurs congénères plutôt que des objets (Anderson, 1998). Particulièrement les visages jouent un rôle pivot dans les interactions sociales chez les primates non-humains, étant un média privilégié de leurs communications sociales (Parr and Maestripieri, 2005). Ainsi l'habileté à discriminer les individus sur la base de leur visage a été observée dans plusieurs espèces, notamment chez les chimpanzés (*Pan troglodytes*, Parr et al., 2000), les macaques fascicularis (*Macaca fascicularis*, Dasser, 1987) et les macaques rhésus (*Macaca mulatta*, Parr et al., 2000).



Figure 7: Différents protocoles utilisés pour tester la discrimination des visages au niveau individuel : choix actif (gauche), préférence visuelle pour la nouveauté (milieu, les lignes rouges indiquent les mouvements des yeux) et test de l'adaptation (droite). Dans ce dernier, les singes s'adaptent à une image puis présente un regain d'intérêt pour la seconde image si celle-ci est perçue comme différente. Au niveau 'basic' le changement de catégorie est toujours perçu comme différent, mais au niveau subordonné seul le changement d'identité du visage de sa propre espèce élicite un regain d'intérêt. D'après Parr et al., 2000; Dahl et al., 2007; Gothard et al., 2009.

Chez les macaques rhésus la mise en évidence de l'individualisation de visages de singes ainsi que d'humains s'est basée sur deux types de protocoles, l'un de choix actif et l'autre de

préférence visuelle (**Figure 7**). Dans le premier cas les singes ont appris à appairer une photo de visage avec une autre photo du même visage présenté d'un point de vue différent, plutôt qu'avec le visage d'un autre individu (Parr et al., 2000). Puis il a été montré qu'ils sont capables de transférer cette règle abstraite à de nouveaux visages de singes non-rencontrés auparavant, démontrant ainsi la capacité des macaques rhésus à discriminer les visages au niveau individuel. Dans le second cas, les singes pouvaient observer deux photos de visages d'un même individu, suite auxquelles deux autres photos étaient présentées (**Figure 7**). Dans la nouvelle paire, l'une des photos correspondait à l'individu présenté précédemment mais d'un point de vue différent alors que la seconde photo présentait un nouvel individu. Les singes ont montré une préférence pour le visage du nouvel individu, suggérant à nouveau leur capacité de discrimination des visages (Gothard et al., 2004; Gothard et al., 2009). Ces deux types d'expériences démontrent que les singes discriminent les visages au niveau individuel, qui est le plus fin niveau de discrimination, et non seulement à des niveaux plus génériques tels l'espèce, le genre, la taille, etc. (Tanaka, 2001). Ils réservent ce type d'analyse aux visages d'individus dont ils sont familiers, comme les individus de leur espèce (**Figure 7**, Pascalis and Bachevalier, 1998; Dufour et al., 2006; Dahl et al., 2007) et dans certains cas pour les humains (Gothard et al., 2009).

De fait les systèmes de perceptions des visages présentent de fortes similitudes entre les humains et les primates non-humains. D'une part les jeunes macaques présentent également une préférence visuelle pour les visages par rapport aux objets à la naissance, montrant que ces stimuli bénéficient d'un traitement perceptif particulier et inné (Kuwahata et al., 2004 ; Sugita, 2008). D'autre part ils perçoivent les visages de façon holistiques comme démontré par des tests d'inversion (Tomonaga, 1994; Guo et al., 2003; Gothard et al., 2004; Adachi et al., 2009a; Dahl et al., 2009b; Dahl et al., 2010b) et de visages composites (Dahl et al., 2007). La discrimination des individus se fonderait donc sur une perception globale du visage (ses composants et sa configuration) et non seulement sur le détail de certains traits.

Enfin les singes utilisent d'autres informations perçues dans un visage, notamment la familiarité de l'individu (Pokorny and de Waal, 2009 ; Mahajan et al., 2011 ; Schell et al., 2011), mais également la hiérarchie, le genre, l'état sexuel, etc. (Waitt et al., 2003; Deaner et al., 2005; Waitt et al., 2006; Gerald et al., 2009). L'ensemble de ces informations pourrait leur permettrait de déployer des comportements complexes.

B.3. Perception et individualisation des voix

Les macaques peuvent identifier leurs congénères visuellement. Cependant dans leur habitat naturel ces indices visuels sont souvent inaccessibles, à cause d'une végétation dense, des longues distances sur lesquelles le groupe se déploie ou encore de la luminosité. On pourrait donc s'attendre à ce que les singes puissent identifier les autres grâce à d'autres sens, par exemple grâce aux indices auditifs. Contrairement aux signaux visuels, les signaux auditifs peuvent être transmis sur de longues distances et sans interférences. Effectivement les vocalisations constituent l'un des

éléments les plus importants de l'environnement auditif des singes. Elles apportent des informations sur l'émetteur lui-même et potentiellement des informations sémantiques.

En ce qui concerne les informations sémantiques, leur signification est largement débattue. D'une part les tenants d'une communication référentielle proposent que les vocalises de singes transmettent des messages représentationnels et codés (Gouzoules et al., 1984; Hauser and Marler, 1993; Hauser, 1998; Ghazanfar and Hauser, 1999; Seyfarth and Cheney, 2003b). Elles seraient alors des précurseurs proches du langage humain à un niveau de complexité près. D'autre part leurs contradicteurs considèrent que les vocalises permettent d'orienter l'attention et les états émotionnels du récepteur, ainsi que certains comportements qui auraient au préalable été couplés de façon pavlovienne à ces types de vocalises, mais n'auraient pas de contenu référentiel (Deacon, 1989; Owren et al., 2005). Les vocalises des macaques seraient le pendant des sons humains émis sans signification référentielle comme les rires, mais ayant un impact émotionnel. Dans cette optique les vocalises différencieraient du langage qualitativement et non seulement quantitativement.



Figure 8 : Spectrogrammes de deux vocalises 'coo' de deux macaques rhésus différents. Ces vocalises sont produites dans différents contextes, comme durant les mouvements de groupe, la présentation de nourriture ou l'exploration spatiale isolée. Il existe une variation considérable entre les 'coo' de différents individus mais une variation intra-individuelle moindre. Notamment on peut noter que les fréquences des formants sont plus rapprochées chez le singe 1 que chez l'individu 2. D'après Ghazanfar and Santos, 2004.

Par ailleurs, les vocalises apportent potentiellement un large pan d'informations sur l'émetteur à propos de sa taille, de ses liens de parenté, de son appartenance au groupe et enfin de son identité (Rendall et al., 1996; Fitch, 1997). En ce qui concerne l'identité qui nous intéresse particulièrement ici, sa discrimination a été démontrée par une expérience d'habituation à des appels de contact, *coo*, joués en playback (Rendall et al., 1996; Hauser, 2007). En effet lorsque

plusieurs vocalises ‘coo’ d’un même individu sont présentées à la suite, les macaques rhésus diminuent leur intérêt pour le micro, mais présentent un regain d’intérêt lorsque l’identité de l’émetteur change. Cette expérience démontre d’une part que les macaques rhésus perçoivent le changement d’identité du locuteur et d’autre part qu’ils peuvent abstraire l’identité de la voix à partir de plusieurs exemplaires différents de vocalises. Une expérience subséquente a montré que la discrimination entre les deux singes locuteurs se base principalement sur la fréquence des formants, que les singes identifient spontanément (**Figure 8**, Fitch and Fritz, 2006). En effet chez les singes les formants sont principalement des marqueurs de l’âge et donc de la taille de l’animal : plus l’animal est grand, plus son tractus vocal est long et plus ses formants sont situés à basse fréquence. Les formants sont très marqués dans les sons de type ‘coo’ et ‘threat’ et peu dans les cris et les ‘grunts’. L’utilisation de ‘coo’ dans l’expérience d’habituation était donc une condition nécessaire à la mise en évidence de l’aptitude à identifier la voix. En effet des modélisations à partir de tests acoustiques ont montré que l’identité du vocalisateur pouvait être dérivée à partir de ‘coo’ et de ‘threat’ mais non de cris ou de ‘grunts’ (Fitch, 1997; Rendall et al., 1998; Owren and Rendall, 2003 mais voir Gouzoules et al., 1984, pour les cris).

Par ailleurs la discrimination des individus grâce à la voix est présente à un jeune âge chez les macaques, au moins en ce qui concerne la mère. En effet il a été observé que les jeunes macaques distinguent ses vocalises, de celles de femelles inconnues à l’âge de 10 semaines (Fischer, 2004). L’ensemble de ces études montre que les voix sont de bons indicateurs de l’identité, notamment à travers les formants qui les constituent. Les macaques pourraient donc utiliser les variations dans la structure acoustique des différentes vocalisations pour déterminer qui appelle et où il se situe, guidant des comportements élaborés.

B.4. Associations dynamiques entre les voix et les visages

De même que les jeunes enfants humains préfèrent regarder des stimuli visuels et auditifs synchronisés, les singes préfèrent spontanément regarder des stimuli visuels correspondant aux stimuli auditifs qu’ils perçoivent. Par exemple lorsqu’ils entendent deux voix de singes et ont le choix entre regarder un film présentant deux singes faisant des mouvements de la bouche synchronisés aux sons ou un film présentant trois singes faisant des mouvements de la bouche synchronisés aux sons, ils préféreront regarder dans la direction des deux singes (Jordan et al., 2005). De façon similaire lorsqu’ils perçoivent un son ‘coo’, long et tonal, ils orienteront leur regard préférentiellement vers le visage de singe synchronisé mimant un ‘coo’, c’est-à-dire avec une petite ouverture de bouche et des lèvres décollées plutôt que vers le visage désynchronisé mimant un ‘threat’, c’est-à-dire avec une plus grande ouverture de bouche et les lèvres non décollées. Ils orienteront leur regard inversement s’ils entendent un ‘threat’, un son plus court, pulsatile et bruité (Ghazanfar and Logothetis, 2003).

Chez l’Homme, nous avons vu que l’identité pouvait être perçue visuellement et auditivement et qu’à partir d’un indice dans l’une des modalités on pouvait retrouver l’indice de l’autre modalité, après avoir mémorisé l’association arbitraire existant entre un visage et sa voix. Qu’en

est-il chez les macaques rhésus ? Les macaques sont capables de discriminer les visages entre eux au niveau individuel, et les voix entre elles au niveau individuel également. Sont-ils capables d'associer ces deux éléments ? Un premier élément de réponse a été fourni par une étude dans la lignée des deux précédentes jouant cette fois-ci sur la taille et l'âge des individus observés par le singe (Ghazanfar et al., 2007). L'étude met en scène d'une part des films muets de singes inconnus grands et adultes ou juvéniles et petits émettant un même type de vocalise et met en scène d'autre part des sons de 'coo' artificiellement resynthétisés pour en changer la structure des formants. Pour que la vocalise ait l'air de provenir d'un individu juvénile et petit, elle était filtrée avec des formants présents à des fréquences plus élevée et plus rapprochés entre elles (**Figure 9 bas**), et inversement pour des vocalises mimant un individu grand et adulte (**Figure 9 haut**). Les chercheurs ont observé que les singes préfèrent regarder les stimuli concordants, c'est-à-dire le visage du jeune singe, lorsque la voix mimait un jeune singe et inversement.

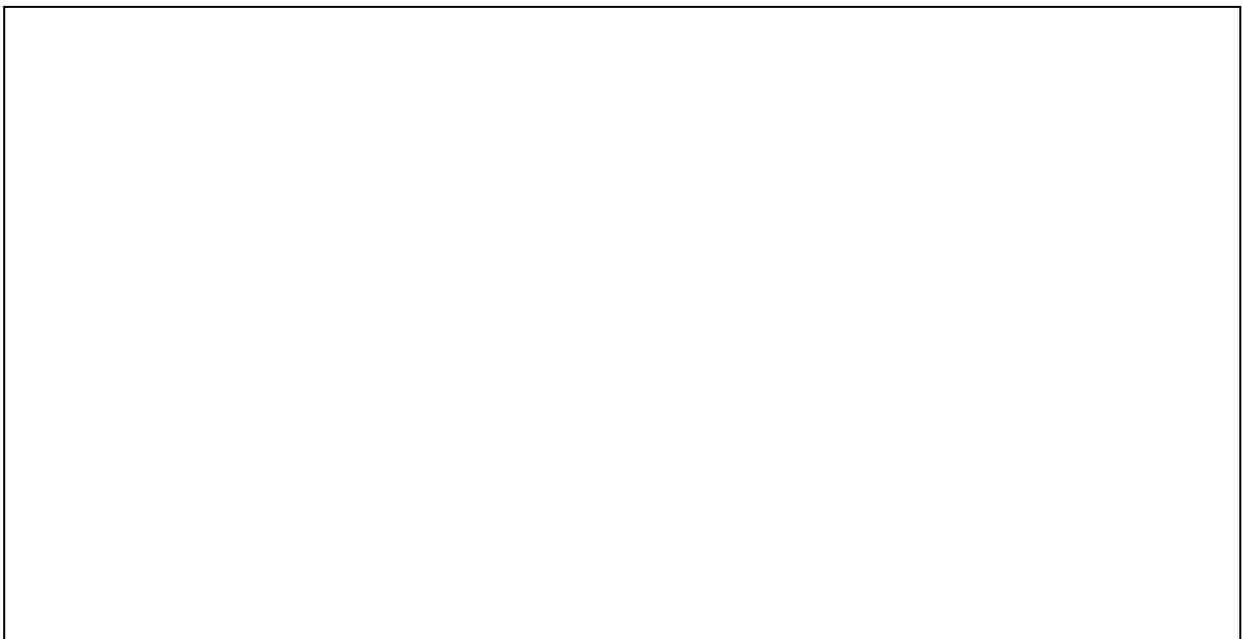


Figure 9: Protocole d'appariement cross-modal de l'âge de l'individu. Stimuli auditifs et visuels caractérisant un individu adulte (haut) et un juvénile (bas). Les vidéos (c) proviennent d'individus inconnus. Les sons (a) ont été resynthétisés à partir d'une même vocalises 'coo'. En effet on peut observer que l'allure des oscillogrammes des deux vocalises est la même, cependant leur spectrogrammes présentes des différences majeures dans la distribution des formants. Leurs fréquences sont plus basses et plus éloignées les unes des autres pour un individu adulte et inversement pour un juvénile (a et b). D'après Ghazanfar et al., 2007.

Cette expérience apporte la démonstration directe que les formants présents dans la structure vocale apportent des informations au locuteur sur les caractéristiques visuelles de l'émetteur. Par ailleurs ceci laisse entrevoir la possibilité que les singes utilisent ces informations au niveau de l'individu, et non à un niveau plus générique comme l'âge démontré ici, pour mémoriser les associations voix-visages des différents individus qu'ils connaissent. Cependant, comme ceci a été proposé par les auteurs de l'article, la question de savoir si les formants permettent de renseigner sur la taille de l'individu à l'intérieur d'une classe d'âge reste entière puisque des résultats

contradictaires ou ambigus ont été obtenus chez l'Homme à ce sujet (Collins and Missing, 2003; González, 2004; Rendall et al., 2005; Bruckert et al., 2006). D'autre part, comme ceci a été présenté précédemment, chez les enfants humains, l'association entre les voix et les visages au niveau individuel se fait en deux temps (Bahrick et al., 2005). D'abord ils acquièrent la capacité à percevoir qu'une fausse association voix-visage n'est pas familière puis seulement ils se montrent capables d'associer un indice, par exemple la voix, à un autre, par exemple le visage, en se référant à une représentation mentale de l'individu gardée en mémoire.

Ainsi la reconnaissance individuelle semble nécessiter une représentation conceptuelle des autres. Nous verrons dans la suite que l'étude des représentations et concepts chez les animaux bien qu'apparue avec les théories évolutionnistes darwiniennes, n'est devenue effective que récemment.

B.5. La possibilité d'un concept de Personne

B.5.1. De l'étude du comportement à celle de la cognition animale

Dès le XIX^{ème} siècle, les théories évolutives de DARWIN (1871) recherchent, la continuité entre l'animal et l'humain non seulement dans le domaine de la morphologie et de la physiologie, mais aussi dans celui des « facultés mentales ». Pourtant ce n'est que récemment que cette continuité du mental est redevenue une question centrale en éthologie, grâce aux méthodes des sciences cognitives appliquées à l'animal (Cézilly, 2004; Kreutzer and Vauclair, 2004).

A partir des années 1950, quatre écoles proposent d'étudier les comportements animaux de façon observable mais aucune ne prend en compte de possibles états mentaux. Ce n'est qu'à partir des années 1970 que la nouvelle discipline d'*éthologie cognitive* explore les représentations cognitives des animaux. Elle postule que les animaux élaborent leurs conduites à partir d'états mentaux et de représentations de l'environnement (Griffin, 1978). Les *représentations* seraient des constructions mentales faisant appel notamment à la mémoire et aux émotions, qui peuvent être approchées en étudiant les catégorisations faites par l'animal. Du fait de cet intérêt, les approches de l'éthologie cognitives ont souvent été taxées d'*anthropomorphiques*, c'est à dire postulant a priori des capacités conceptuelles animales calquées sur la pensée humaine (Bekoff and Allen, 1997; Barrett et al., 2007). En réaction des contours exigeants pour la discipline ont été tracés, en précisant par exemple que les représentations mentales n'auraient pas pour corrélat l'existence d'une conscience animale.

La cognition est abordée par l'éthologie cognitive en tant que stratégie adaptative pour optimiser les chances de survie et de reproduction et présente une forte variation interindividuelle. La *psychologie comparée* par contre se défait de la dimension interindividuelle et explore les mécanismes mentaux moyens d'une espèce pour les comparer à ceux d'autres espèces, notamment pour établir des convergences évolutives de stratégies cognitives. Dans un premier temps la psychologie cognitive a souvent comparé les représentations cognitives de différentes espèces animales à celles de l'humain, dans le but de comprendre les précurseurs du

langage ou de l'intelligence humaine. Cette approche *anthropocentrée*, à l'inverse de l'anthropomorphisme, juge les capacités des primates par rapport à un standard humain et perçoit souvent en elles une fraction seulement de ce standard (Barrett et al., 2007). Elle ne prend donc pas en compte le fait que chaque espèce ait évolué pour s'adapter à son environnement et ait développé des particularités originales (citons par exemple l'écholocation chez la chauve-souris) ou ressemblant à celles présentes chez d'autres espèces par convergence évolutive (comme les différents types d'ailes). L'approche contemporaine en éthologie cognitive, prend donc en compte les deux limites formées par l'anthropocentrisme et l'anthropomorphisme (Barrett et al., 2007; de Waal and Ferrari, 2010). Elle s'attache à étudier les représentations des espèces non-humaines soit en laboratoire, soit dans l'environnement naturel en considérant les comportements en tant qu'adaptations riches à des niches écologiques.

B.5.2. Une pensée conceptuelle animale

Les représentations mentales des primates non-humains ont été étudiées à l'aide de trois types de protocoles principalement : la catégorisation perceptive, l'utilisation de règles relationnelles et enfin les tests cross-modaux (Fabre-Thorpe, 2003; Fagot et al., 2004).

La catégorisation perceptive repose sur l'abstraction d'attributs pertinents pour regrouper les stimuli connus entre eux et appréhender les stimuli nouveaux. Elle a été étudiée chez différents animaux, du pigeon aux singes, avec des stimuli créés en laboratoire ou des stimuli naturels (« arbre », « poisson », « aliment », « homme », « singe », « animal »). Ainsi différentes études ont mis en évidence que les primates non-humains catégorisent les stimuli perceptifs dans leur environnement naturel (Schrier and Brady, 1987; Vonk and Macdonald, 2002; Fize et al., 2011), sont capables d'apprendre de nouvelles catégories dans des protocoles en laboratoire (Roberts and Mazmanian, 1988; Vauclair and Fagot, 1996), et utilisent potentiellement la catégorisation perceptive dans leurs interactions avec l'environnement et les autres (Seyfarth et al., 1980; Seyfarth and Cheney, 2003b, a). De plus en étudiant la catégorisation visuelle dans des conditions particulières de présentation des images (sans couleurs, aux hautes fréquences filtrées, avec des présentations de 30s seulement), FABRE-THORPE et coll. (1998, 2001) ont démontré de façon intéressante que les catégorisations perceptives peuvent être réalisées par l'humain et le singe sans accès conscient au contenu des images (Fabre-Thorpe et al., 1998; Delorme et al., 2000; Thorpe et al., 2001; Mace et al., 2010). Un deuxième type de représentations mentales concerne l'utilisation de règles relationnelles par les animaux. En effet il a été mis en évidence que les primates non-humains perçoivent différents concepts relationnels (dedans/dehors, dessus/dessous, la symétrie, la numérosité et enfin l'identité relationnelle) (Brannon and Terrace, 1998; Dépy et al., 1999; Bovet and Vauclair, 2001; Wasserman et al., 2001; Livingstone et al., 2010; Truppa et al., 2010), et peuvent utiliser les mêmes règles abstraites dans des situations très différentes comme l'exploration de l'environnement ou lors d'interactions sociales tel le partage de nourriture.

Ainsi les singes catégorisent les stimuli visuels ou auditifs de leur environnement en abstrayant les attributs perceptifs communs entre les stimuli. De plus ils conçoivent et peut-être utilisent des règles relationnelles entre des stimuli similaires environnementaux ou sociaux. Cependant ces deux types de pensées abstraites ne nous permettent pas de conclure à la possibilité de l'existence d'un concept de Personne chez le singe. En effet le concept d'individu suppose en premier lieu et principalement l'association de différents attributs (voix, visage, odeur...) sans similarité physique entre eux car perçus par des canaux sensoriels différents, d'autre part un sentiment de familiarité avec la personne et finalement l'apposition de connaissances sur les interactions sociales passées. Ainsi l'utilisation de protocoles cross-modaux est une méthode de choix pour étudier la partie principale du concept de Personne. Plusieurs études ont utilisé des protocoles cross-modaux chez le singe afin de tester l'existence de certains concepts. En effet ce protocole diffère des précédents car dans celui-ci il ne s'agit pas d'extraire des composants perceptifs communs entre les différents exemplaires de stimuli mais de tester l'existence d'associations abstraites entre des éléments jugés équivalents fonctionnellement et gardés en mémoire. Ainsi dans les années 1970, des chercheurs se sont intéressés aux équivalences cross-modales entre l'identité visuelle et l'identité tactile d'objets perçus pour la première fois pendant une expérience. Dans les premières études, les singes faisaient l'expérience des objets par la modalité tactile dans le noir mais ne semblaient pas transférer cet apprentissage à leur présentation visuelle ultérieure (Ettlinger and Blakemore, 1967; Milner and Ettlinger, 1970; Milner, 1973). De nouvelles études ont utilisé des objets à la fois palpables et mangeables faits avec de la nourriture et présentés dans le noir ; certains de ces objets ayant un goût désagréable par l'ajout de quinine. Les chercheurs ont observé que les singes choisissaient les objets savoureux lorsqu'ils étaient ensuite présentés uniquement visuellement, démontrant que les singes réactivaient l'apprentissage tactile mais dans la modalité visuelle, ayant réalisé un transfert de connaissance entre ces deux modalités (Covey and Weiskrantz, 1975; Weiskrantz and Covey, 1975; Elliott, 1977; Jarvis and Ettlinger, 1977). Par la suite d'autres études se sont demandé si les singes utilisent spontanément les informations sociales de façon cross-modale et équivalente. MARTIN-MALIVEL et FAGOT (2001) ont appris à deux babouins à indiquer à l'aide d'un joystick si une vocalise jouée par un haut-parleur provenait d'un humain ou d'un babouin. A la suite de cet apprentissage, avant chaque présentation de vocalise, une image de babouin ou d'humain était présentée sur un écran. Lorsque l'image était congruente avec le son, alors les réponses de l'un des babouins étaient plus rapides que lorsque le son apparaissait seul, suggérant une facilitation cross-modale. Les auteurs en ont conclu que les babouins sont capables de former le concept d'« espèce » (Martin-Malivel and Fagot, 2001). Des études similaires ont par la suite été menées chez les macaques japonais jeunes, et ont mis en évidence que lorsqu'ils entendent une vocalise de singe ils regarderont plus longtemps un visage de singe mais non d'humain s'affichant ensuite sur un écran (Adachi et al., 2006). De façon intéressante l'inverse était vrai (un plus long temps de regard sur un écran présentant un humain si celui-ci est précédé d'une vocalise humaine) uniquement chez les macaques japonais ayant été élevée à la fois en présence d'humains et de singes mais non chez ceux élevés uniquement en présence d'autres macaques japonais, indiquant que les concepts de « macaque japonais » et d'« humain » étaient mis en place chez les animaux les ayant expérimenté (Adachi et al., 2009b).

Ainsi il est possible de mettre en évidence des représentations mentales conceptuelles sociales à l'aide de protocoles cross-modaux. Il serait donc intéressant d'utiliser cette technique pour savoir si les macaques rhésus sont capables d'associer les différents attributs de leurs congénères mais également des humains au niveau individuel et d'utiliser les informations des différentes modalités de façon équivalente. Ceci constituerait un premier pas dans la compréhension du concept de Personne chez le macaque rhésus. Un tel essai a été réalisé avec les chimpanzés et s'est révélé fructueux (Bauer and Philip, 1983; Kojima et al., 2003; Izumi and Kojima, 2004; Martinez and Matsuzawa, 2009). Cependant de nombreuses différences cognitives existent entre les deux espèces ne permettant pas de conclure à cette habileté chez le macaque rhésus.

B.6. Conclusion

Les macaques rhésus perçoivent et discriminent les identités faciales et vocales. Cependant il n'a pas été mis en évidence s'ils possèdent la faculté de les associer et de retrouver l'information d'une modalité à partir de l'autre, c'est-à-dire la faculté de reconnaître individuellement leurs congénères (**Figure 10**). Cette mise en évidence permettrait d'approcher l'existence d'un concept de Personne chez le singe.

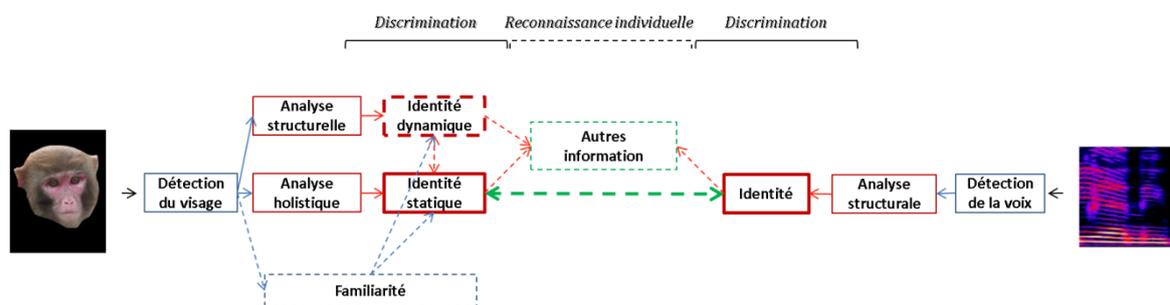


Figure 10 : Schématisation du processus possible de reconnaissance individuelle chez le macaque rhésus. Les lignes en pointillé indique les processus n'ayant pas été testé expérimentalement.

En parallèle et indépendamment des études sur le comportement animal, les avancées en neurosciences ont depuis longtemps assimilé l'appareil mental animal à un système de traitement d'informations basé sur des représentations. Dans cette perspective, les représentations prennent tout leur sens en tant que correspondance entre certains aspects de l'environnement et un code nerveux spécifique, permettant d'adapter le comportement de l'animal (Roitblat and von Fersen, 1992). Cependant dans le domaine des neurosciences les animaux, notamment les rats et les singes rhésus, ont principalement été étudiés en tant que modèles simplifiés du cerveau humain. Pour prendre en compte les adaptations propres à chaque espèce, de nouvelles disciplines ont émergé, la *neuroéthologie* ou *neuroécologie*, tentant de réaliser la synthèse entre étude du système nerveux et spécificité de l'espèce étudiée (Hampton et al., 2002; Platt and Ghazanfar, 2010). Par

ailleurs l'utilisation de stimuli écologiques s'est généralisée. Notamment l'utilisation de visages et de voix comme stimuli visuels et auditifs a mis en évidence les traitements neuronaux spécifiques dont bénéficient leurs traitements.

C. Bases neurales de l'identification des visages

Chez l'humain comme chez le singe, la perception des visages est réalisée par un système neural distribué comprenant des régions multiples et bilatérales. La mise en évidence d'un système spécifiquement dédié à l'analyse des visages est venue en premier lieu de données cliniques de patients lésés présentant des troubles de reconnaissance visuels circonscrits aux visages, la *prosopagnosie*. Les données d'imagerie par résonance magnétique fonctionnelles (IRMf), de tomographie par émission de positron (PET), d'électroencéphalographie (EEG) chez l'humain, ainsi que d'enregistrements de neurones unitaires et d'imagerie optique chez le singe, ont aidé à définir les différentes aires constituant ce système et à dégager leurs spécificités.

C.1. Un système dédié à la perception de ce stimulus social

C.1.1. Mise en évidence par des lésions neurologiques d'aires spécialisées

Les données cliniques ont décrits plusieurs types de syndromes d'*agnosie*, c'est-à-dire d'incapacité à identifier et nommer différentes catégories d'objets bien que les systèmes sensoriels soient intacts. Les cas cliniques ont mis en évidence une distinction entre le système de perception des objets et celui des visages, par double-dissociation. D'un part dans le cas de la *prosopagnosie*, les personnes sont incapables d'identifier un visage, bien qu'elles ne nient pas sa présence et peuvent en décrire les détails (Bruyer et al., 1983; Etcoff, 1984; Tranel and Damasio, 1988; Tranel et al., 1988; Young et al., 1993). Notons que, seule l'identification mais non la détection, des visages est détériorée dans les cas de prosopagnosie. En effet la détection des visages dépendrait d'un réseau d'aires sous-corticales actives précocement (Johnson, 2005). Dans le cas inverse d'agnosie relative aux objets, les personnes perçoivent et identifient les visages mais non les objets. Par exemple la dissociation du patient agnosique CK peut être illustrée par sa perception intacte du visage constitué de légumes dans les tableaux d'Arcimboldo, mais non des légumes constituant le visage (Moscovitch et al., 1997). La double dissociation apportée par l'observation de patients agnosique et prosopagnosiques démontre que la perception des visages

ne constitue pas une étape ultime du traitement perceptif des objets, comme il a longtemps été cru, mais qu'elle empreinte une voie dédiée parallèle.

Différents types de prosopagnosie peuvent être différenciés. Ils peuvent être classés en deux groupes. La *prosopagnosie aperceptive* est un déficit dans le traitement perceptif du visage alors que la *prosopagnosie associative* concerne les déficits dans le traitement mnésique du visage (Damasio et al., 1990; Schweinberger and Burton, 2003; Calder and Young, 2005). La prosopagnosie associative inclue elle-même différents types de déficits : la *prosopamnésie antérograde* ou l'incapacité à apprendre de nouveaux visages (Hanley et al., 1990; Tippett et al., 2000); la *prosopagnomie* ou l'incapacité à retrouver le nom des personnes à partir de la vue de leur visage (Carney et al., 1993); enfin la *propamnésie rétrograde* ou l'incapacité à reconnaître les visages familiers (Ellis et al., 1989; Hanley et al., 1989; Evans et al., 1995). L'ensemble de ces déficits suggère que différentes étapes de perception et reconnaissance de visages pourraient être codées par des modules différents.

Les lésions des patients agnosiques et prosopagnosiques concernent des aires les lobes temporaux indiquant le rôle majeur de cette partie du cerveau dans la reconnaissance et l'identification des stimuli visuels. Les lésions sont bilatérales. Cependant dans certains cas moins fréquents, une lésion unilatérale droite suffit à produire un syndrome prosopagnosique indiquant son rôle prépondérant dans la perception des visages (Landis et al., 1986).

C.1.2. Les deux voies corticales d'analyse des stimuli visuels

La spécificité du lobe temporal dans le traitement perceptif de l'identité des objets et des individus a été intégrée dans un modèle postulant deux voies d'analyse par le cortex (**Figure 11**).

Ce modèle a été développé pour expliquer l'analyse des stimuli visuels, sur la base de données lésionnelles obtenues chez l'humain et neurophysiologiques obtenues chez le singe. En effet au-delà de l'aire visuelle primaire, la trentaine d'aires visuelles corticales décrites par FELLEMAN et VAN ESSEN s'organisent selon deux voies anatomiques : une voie occipito-pariétale ou voie dorsale et une voie occipito-temporale ou voie ventrale (FELLEMAN AND VAN ESSEN, 1991). UNGERLEIDER et MISHKIN ont proposé une première interprétation fonctionnelle de cette ségrégation, en distinguant la voie du « quoi » de la voie du « où » (Mishkin et al., 1983; Haxby et al., 1991; Ungerleider, 1995). La première serait impliquée dans les mécanismes liés à la perception de l'identité de l'objet ou du visage alors que la seconde participerait à sa localisation spatiale. GOODALE et MILNER ont proposé une autre interprétation fonctionnelle : les deux voies seraient impliquées dans le traitement des dimensions structurelles et spatiales des stimuli sensoriels, mais c'est la finalité de l'action (perceptive ou motrice) qui déterminerait la voie sollicitée (Goodale and Milner, 1992).

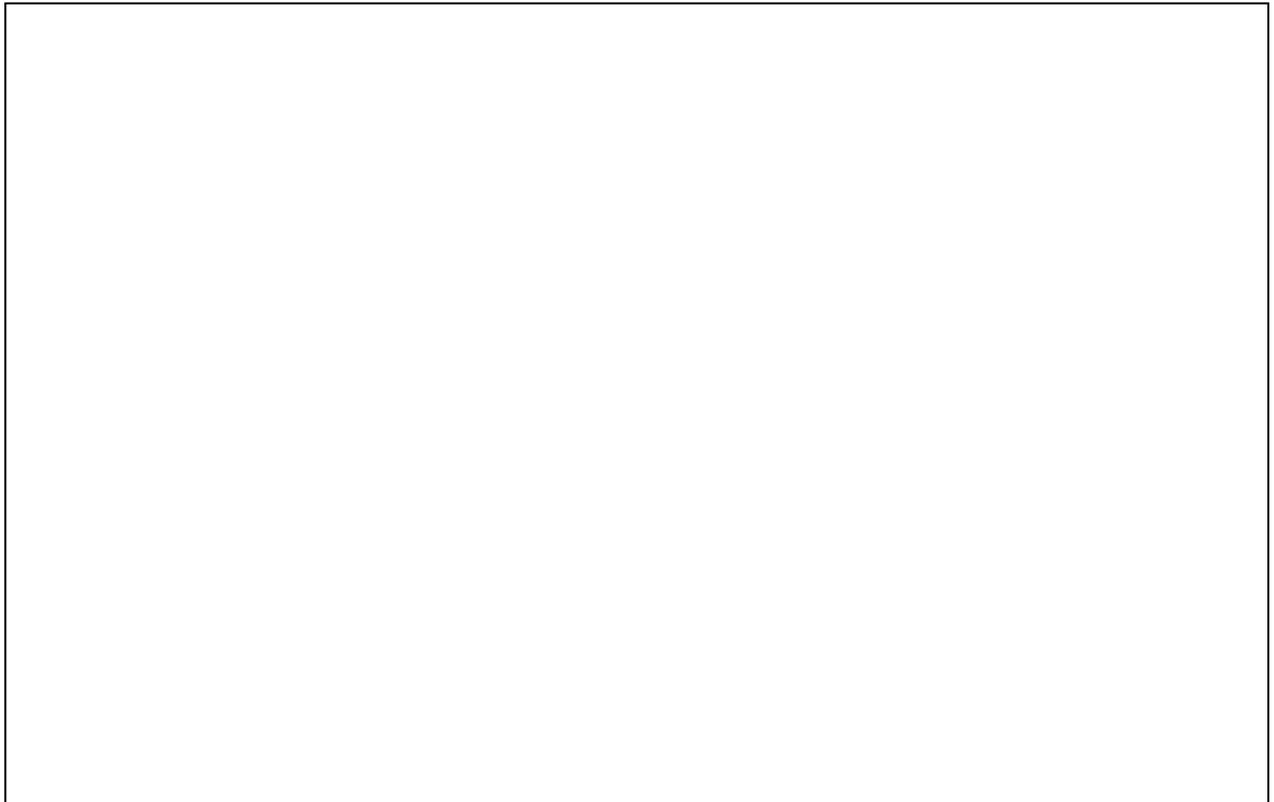


Figure 11: Organisation hiérarchique des aires visuelles chez le singe (gauche) et ségrégation en deux voies (droite). À droite, les flèches représentent le sens de projection des informations. Les lignes pointillées indiquent des connexions « feedback » (.....) les lignes pleines des connexions « feed-forward » (—). Le schéma montre une divergence entre les voies temporale et pariétale au niveau du sulcus temporal supérieur et désigne cette région comme un site potentiel d'interaction entre les deux voies. D'après Felleman and Van Essen, 1991; Distler et al., 1993; Ungerleider, 1995.

C.1.3. Un système distribué pour la perception des visages

Durant la dernière décennie, les études en IRMf et PET chez l'humain ont révélé des aires faisant partie de la voie ventrale présentant une forte sélectivité pour les visages par rapport aux objets et paysages (Kanwisher et al., 1997), mais également pour les corps humains par rapport aux objets (Cohen et al., 2000; Downing et al., 2001), les paysages par rapport aux objets et visages (Epstein and Kanwisher, 1998) ou encore la forme visuelle des mots par rapport aux objets (Petersen et al., 1990; Cohen et al., 2000; Cohen and Dehaene, 2004). Ces zones correspondaient aux zones lésées chez les patients agnosiques et prosopagnosiques.

Notamment il a été montré à plusieurs reprises que la perception des visages implique une région du gyrus fusiforme latéral, nommée *aire fusiforme des visages FFA* (**Figure 12**) (Sergent et al., 1992; Sergent and Signoret, 1992; Haxby et al., 1994; Puce et al., 1995; Kanwisher et al., 1997; McCarthy et al., 1997). Pour tester si cette zone s'active spécifiquement pour le traitement perceptif des visages, MCCARTHY et coll. (1997) ont tenu le raisonnement suivant : si on fait

L'hypothèse que les visages sont traités séparément par le système visuel, alors si un visage est perçu pendant que le système de perception visuelle des objets est déjà occupé au traitement d'un certain nombre d'objets, une zone corticale additionnelle devrait être activée (McCarthy et al., 1997). Au contraire si les visages sont perçus grâce à un système général d'analyse des objets alors une telle activation additionnelle ne devrait pas être visible. Pour tester cette hypothèse MCCARTHY et coll. (1997) ont mesuré par IRMf l'activation évoquée par des visages, par rapport à des fleurs, présentés au milieu d'un montage continuellement changeant d'objets ou d'objets brouillés. Les résultats obtenus concordent avec leurs hypothèses et démontrent l'implication d'une aire sélective pour la perception des visages, l'aire FFA. Cette étude prend de plus en compte et contrôle les effets attentionnels générés par la vue d'un visage. Par ailleurs elle n'utilise pas de tâche de discrimination active, qui peut générer une inégalité inhérente au traitement particulier des visages par rapport aux objets. En effet la distinction entre deux visages s'opère à l'intérieur de la catégorie des visages au niveau fin de l'identité, au contraire la distinction entre deux objets s'opère à un niveau supérieur entre des catégories d'objets (Tanaka, 2001). Le traitement des visages et celui des objets ne sont donc pas équivalents l'un à l'autre. Pour circonvier à ce problème KANWISHER et coll. (1997) ont demandé aux sujets de distinguer différents visages les uns des autres et ont soustrait à cette activité celle obtenue lorsque les sujets distinguaient différentes mains humaines entre elles (Kanwisher et al., 1997). Comme dans l'étude précédente, les résultats ont montré une activation plus forte de la FFA pour les visages. Son activité ne serait donc pas fonction du niveau de discrimination opéré mais bien de la nature des visages eux-mêmes.



Figure 12 : Régions du cerveau humain (gauche) répondant sélectivement aux stimuli indiqués. En rose sont représentées les trois régions sélectives aux visages faisant partie du système central : FFA, OFA et STS. Modèle du système distribué pour la perception des visages (droite). Le modèle est divisé en un système central et un système étendu. D'après Haxby et al., 2000; Kanwisher, 2006.

Notons cependant qu'une objection à cette étude reste non résolue, concernant le type de traitement induit par la discrimination des visages. En effet comme nous l'avons vu précédemment (§A.3.1), le traitement des visages présente une originalité par rapport au traitement d'autres objets car il est réalisé de façon holistique, ce qui confère notre expertise et notre rapidité pour la

discrimination des visages. Or dans certain cas particuliers, le développement d'une expertise autre que celle pour les visages peut-être observé, par exemple pour les chiens, dans le cas des experts en chiens, ou pour les oiseaux ou les voitures, ou encore, plus généralement au niveau de la population, pour la lecture. Le traitement des objets d'expertise est alors réalisé au niveau très fin de l'identité et de façon holistique, de façon similaire aux visages. Dans ces cas, les objets d'expertises peuvent également préférentiellement activer la FFA (Gauthier et al., 1999; Gauthier and Logothetis, 2000; Gauthier et al., 2000a) ou son homologue controlatéral (Dehaene et al., 2010).

Par la suite d'autres aires sont venues s'ajouter à la liste des zones spécifiquement dédiées à la perception des visages, faisant parties du lobe temporal mais également du lobe occipital, du système limbique et du cortex préfrontal (Ishai et al., 2005). HAXBY et coll. (2000) ont proposé que le réseau dédié à la perception des visages soit distribué et qu'un système central puisse être distingué d'un système étendu (Haxby et al., 2000). Le système central comprend une aire du gyrus occipital inférieur, *l'aire occipitale des visages OFA*, *l'aire fusiforme des visages FFA* ainsi qu'une aire du *sulcus temporal supérieur STS* (**Figure 12**). Plus récemment deux aires sélectives aux visages supplémentaires ont été observées dans le lobe temporal antérieurement à la FFA (Kriegeskorte et al., 2007; Tsao et al., 2008a).

Le système étendu inclus des aires qui s'activent plus fortement mais non exclusivement à la perception de visages (**Figure 12**): l'hippocampe, l'amygdale, l'insula et le système limbique, le gyrus frontal inférieur, et des régions du circuit de la récompense comme le cortex orbitofrontal et le nucleus accumbens (Ishai et al., 2002; Ishai et al., 2004; Ishai et al., 2005; Ishai, 2007). Bien que l'ensemble de ces aires s'active bilatéralement, l'activité en est plus forte dans l'hémisphère droit, se manifestant par des clusters plus larges, plus fortement activés et de façon constante à travers les sujets testés.

C.1.4. Des systèmes homologues chez l'humain et le primate non-humain ?

GROSS et coll. (1969, 1972) sont les premiers à avoir donné une description qualitative des neurones visuels du cortex temporal inférieur chez le singe (Gross et al., 1969; Gross et al., 1972). Ils ont observé leurs larges champs récepteurs et leurs propriétés sélectives pour des stimuli complexes, dont des stimuli biologiques comme des mains ou des visages. La découverte de ces neurones confirmait l'hypothèse de KONORSKI sur l'existence de cellules hautement sélectives, ou *unités gnostiques*, notamment pour les visages (Konorski, 1967). Par la suite de nombreuses études ont fait état de cellules sélectives aux visages regroupées en clusters discontinus dans le cortex temporal inférieur, le sulcus temporal supérieur et le lobe temporal antérieur (**Figure 13 haut gauche**) (Bruce et al., 1981; Perrett et al., 1982; Desimone et al., 1984; Perrett et al., 1984; Yamane et al., 1988; Hasselmo et al., 1989; Heywood et al., 1992; Perrett et al., 1992; Young and Yamane, 1992; Kobatake and Tanaka, 1994; Nakamura et al., 1994; Sugase et al., 1999; Eifuku et al., 2004; Foldiak et al., 2004; Tsao et al., 2006). Il a été proposé que les neurones sélectifs aux visages occupent des colonnes corticales distribuées jouxtant des colonnes

sélectives à d'autres objets complexes (Haxby et al., 2001; Tanaka, 2003). Cette vue est contradictoire avec l'observation en IRMf chez l'humain de régions corticales sélectives aux visages bien plus larges qu'une colonne corticale de singe. Ce désaccord a été en partie levé grâce à l'utilisation d'IRMf chez le singe et d'enregistrements électrophysiologiques guidés par les résultats d'imagerie dans les zones sélectives aux visages (Tsao et al., 2006; Bell et al., 2011). Notons que des neurones sélectifs aux visages ont également été enregistrés dans le cortex préfrontal (O Scalaidhe et al., 1997, 1999; Romanski and Diehl, 2011), orbitofrontal (Thorpe et al., 1983; Rolls et al., 2006) et dans l'amygdale (Nakamura et al., 1992; Kuraoka and Nakamura, 2006; Gothard et al., 2007).

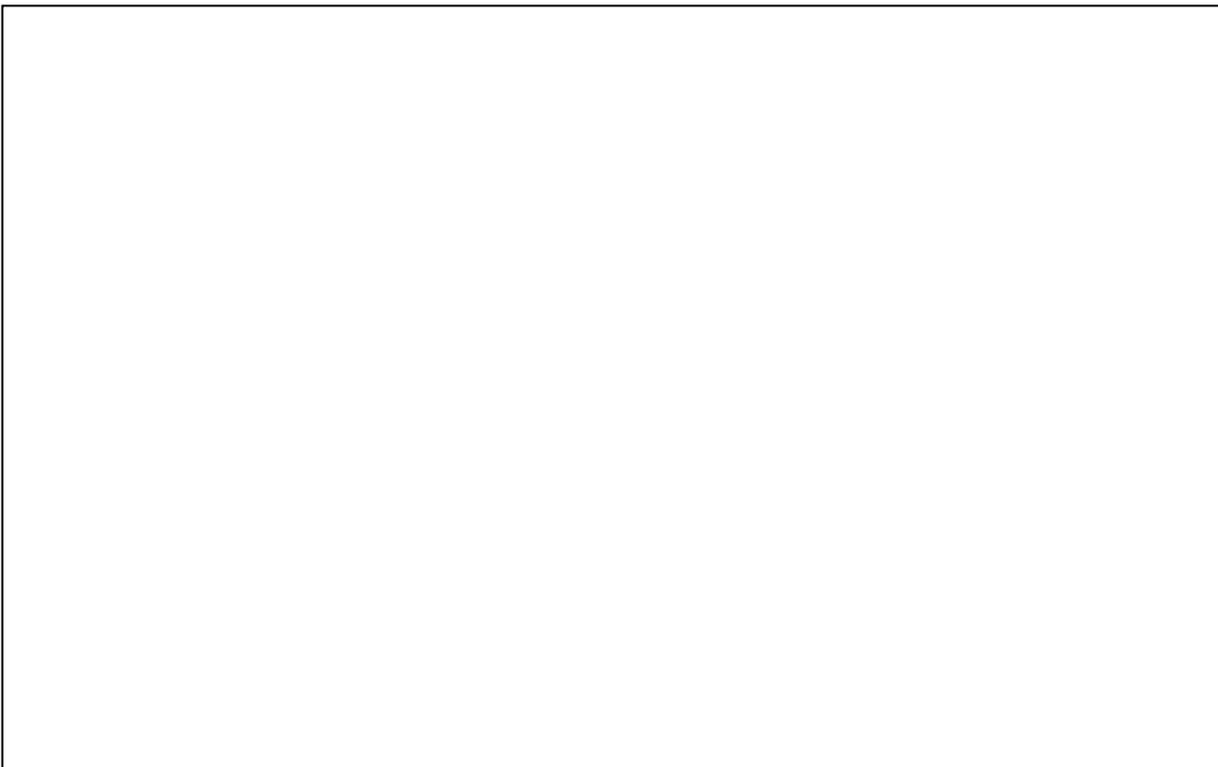


Figure 13 : Meta-analyse présentant les localisations des enregistrements de neurones sélectifs aux visages chez le macaque (haut gauche). Le point rose représente les enregistrements réalisés par Tsao et coll. (2006) guidés par IRMf. Activités des neurones enregistrés à cet endroit (haut droite) pour différents stimuli visuels. La figure de cerveau (bas gauche) représente les six aires sélectives aux visages mises en évidence par TSAO et coll. (2006) et les changements de signal BOLD pour les différents stimuli visuels dans ces régions (bas droite). D'après Tsao et al., 2006; Tsao and Livingstone, 2008; Tsao et al., 2008a.

Les images acquises par résonance magnétique fonctionnelle chez le singe, ont révélés des aires cérébrales sélectives aux visages (**Figures 13 bas et 14**, Logothetis et al., 1999; Tsao et al., 2003; Pinsk et al., 2005; Hoffman et al., 2007b; Hadj-Bouziane et al., 2008; Tsao et al., 2008b; Ku et al., 2011). Les aires sélectives aux visages, nommées *face patches*, sont situées principalement dans le lobe temporal, mais également dans les lobes occipitaux et frontaux et dans l'amygdale, montrant un réseau d'aires distribuées. Dans le lobe temporal, un pattern stéréotypé de six aires est observé chez le macaque, reproductible entre les individus (**Figure 13 bas**, Tsao et al., 2008a).

Les aires s'organisent selon un axe antéropostérieur principalement le long du sulcus temporal supérieur : une aire est postérieure PL, deux sont médianes ML et MF et trois antérieures AL, AF et AM. De nouvelles régions ventrales et difficiles à imager ont été découvertes par la suite grâce à l'utilisation d'une antenne volumétrique et d'imagerie de haut champ spin-echo echo-planar (SE-EPI) dont le signal n'est pas dégradé par la susceptibilité aux artefacts dus au canal de l'oreille (**Figure 14**, Ku et al., 2011).

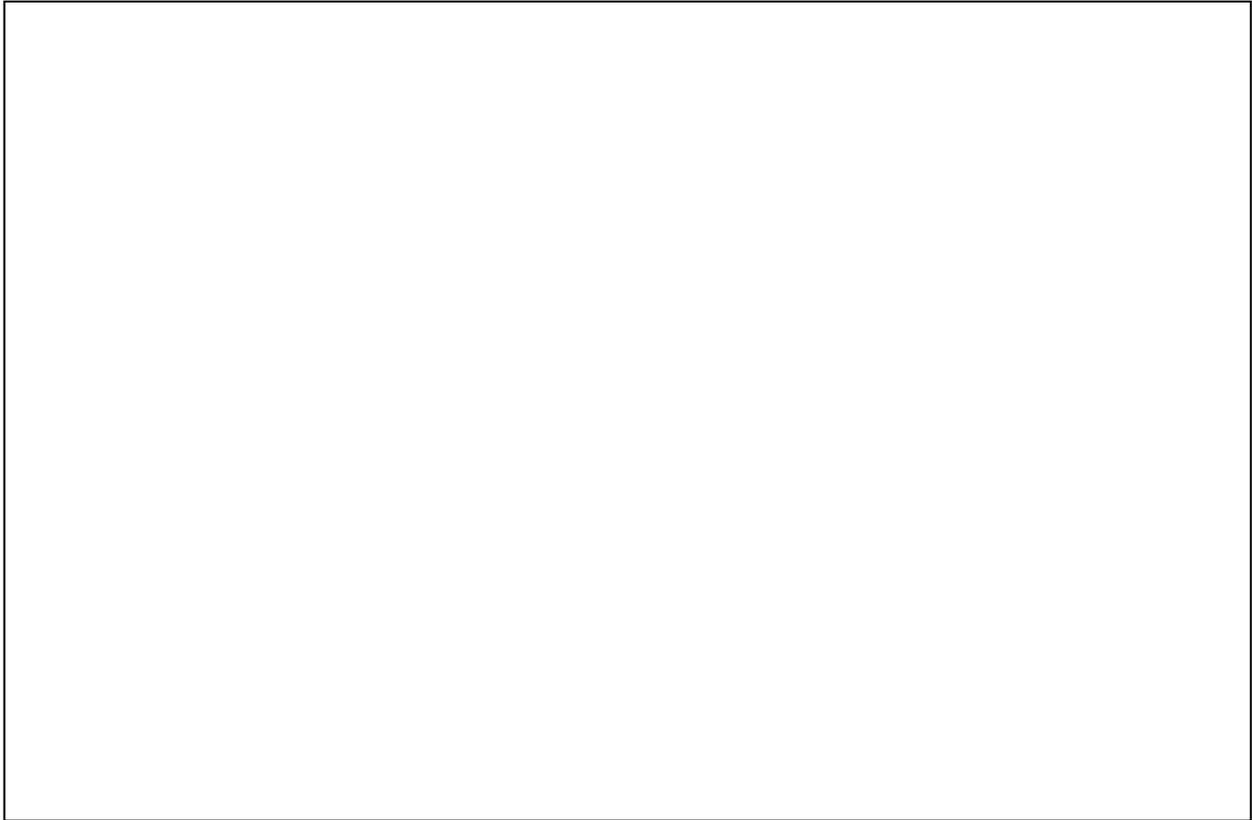


Figure 14 : Méta-analyse des régions cérébrales chez le macaque sélectivement activées à la vue d'un visage, présentées sur une vue de profil et une vue ventrale de cerveau (gauche). Sur cette figure sont superposées les régions définies par différentes études menées en IRMf (Logothetis et al., 1999; Pinsk et al., 2005; Tsao et al., 2008a; Tsao et al., 2008b; Bell et al., 2009; Ku et al., 2011). Localisation des activations ventrales et hippocampiques (droite). D'après Ku et al., 2011.

Des enregistrements neuronaux effectués dans les aires actives en IRMf, ont révélé un nombre remarquable de cellules spécifiquement sélectives aux visages, de l'ordre de 97% au centre de ces aires (**Figure 13 haut droite**, Tsao et al., 2006) et en moyenne entre 41 et 69% sur l'ensemble de l'aire activée (Bell et al., 2011). Ces nombres dépassent largement le pourcentage de cellules sélectives aux visages enregistrées dans le lobe temporal par les méthodes classiques, qui est de l'ordre de 15 à 35% (Perrett et al., 1982; Desimone et al., 1984; Eifuku et al., 2004). De façon intéressante, une étude comparative a guidé des enregistrements neuronaux à l'intérieur des *face patches*, avec les pourcentages de neurones sélectifs présentés ci-dessus, et à l'extérieur des *face patches* (Bell et al., 2011). Les auteurs ont mis en évidence que, dans le second cas, les neurones

présentent une sélectivité aux visages dans 20% des cas, c'est-à-dire dans les mêmes proportions que celles observées dans les études classiques. Ainsi les visages pourraient être encodés par un système hiérarchique incorporant à la fois des éléments modulaires (les *face patches*) et des éléments distribués (les neurones en dehors des *face patches* organisés en colonnes corticales).

Des études comparatives en IRMf ont été menées afin de définir des correspondances entre les activations chez le singe et l'humain (Tsao et al., 2008a; Bell et al., 2009; Pinsk et al., 2009). L'aire activée la plus large, ML *middle face patch*, est située dans le sillon du sulcus temporal supérieur. Par sa taille et son intense activation elle pourrait être un bon candidat pour être l'homologue simien de l'aire FFA humaine, cependant sa localisation est proportionnellement plus dorsale que celle de FFA. En effet chez le macaque la plupart des *face patches* imagés par TSAO et coll. (2006) est localisée à l'intérieur ou proche du sulcus temporal supérieur alors que chez l'humain, toutes les régions, à l'exception de STS, sont localisées sur la face ventrale du lobe temporal, à l'intérieur ou en dessous du sulcus collatéral. Ainsi les régions plus ventrales TF et vV4 découvertes récemment chez le macaque par KU et coll. (2011) pourraient constituer de meilleurs homologues aux aires FFA et OFA humaines respectivement, que les aires définies par TSAO et coll. (2006) (Tsao et al., 2008a; Ku et al., 2011). Comme l'aire vV4 est la seule aire occipitale sélective aux visages observée chez le singe, elle pourrait constituer un homologue d'OFA. Par conséquent les différentes aires imagées dans le sulcus temporal supérieur chez le singe pourraient être collectivement un homologue de STS. Les aires plus antérieures correspondraient aux deux aires du lobe temporal antérieur également sélectives aux visages chez l'humain. Enfin les activations hippocampiques et des lobes frontaux sont similaires chez les deux espèces.

Ces études comparatives ont également mis à jour qu'un plus grand nombre de régions sélectives aux visages chez le singe que chez l'humain. Par exemple l'étude de TSAO et coll. (2006) rapportent cinq aires chez l'humain incluant FFA, OFA, STS ainsi que deux régions du lobe temporal antérieur alors qu'elle dénombre jusqu'à six aires chez le singe. Si à cela s'ajoute les nouvelles aires ventrales découvertes par KU et coll. (2011), notamment TF et vV4, ceci porte à huit le nombre d'aires sélectives aux visages chez le macaque. Des différences méthodologiques existent entre les études chez l'humain et le singe qui pourraient partiellement expliquer cette différence (Op de Beeck et al., 2008). En effet chez l'humain, les études en IRMf sont basées sur un signal intrinsèque, le contraste BOLD (blood oxygen level-dependant) alors que chez le singe le signal est souvent optimisé par l'utilisation d'agents de contrastes (CBV, contrast agent-dependant cerebral blood volume) qui améliorent la résolution des images. Par conséquent, chez l'humain la résolution spatiale de la plupart des études est supérieure ou égale à 3 mm³ par voxel alors que celle-ci est meilleure chez le singe, de 1.25 mm³. Une meilleure résolution spatiale pourrait amener à distinguer deux régions là où une technique à moins bonne résolution n'en verrait qu'une. En effet ce type de comparaison a été entrepris chez l'humain grâce à l'IRMf haute résolution, c'est à dire utilisant des voxels isotropiques de 1 mm³ (Schwarzlose et al., 2005; Grill-Spector et al., 2006). L'augmentation de la résolution spatiale montre que les régions

sélectives aux visages précédemment distinguées par IRMf classique chez l'humain peuvent être des amas de régions de plus petite taille sélectives aux visages.

C.2. Encodage des visages le long de la voie ventrale

C.2.1. Dynamique d'activation du réseau

Des études alliant l'IRMf à des micro-stimulations électriques ont permis de mettre en évidence la connectivité entre les différentes aires sélectives aux visages chez le singe. Elles ont montré que la micro-stimulation d'une aire sélective aux visages résulte en l'activation exclusive des autres aires sélectives aux visages, dans les aires temporales (Moeller et al., 2008) et les aires frontales (Tsao et al., 2008b). Une stimulation d'une aire ne faisant pas partie de ce réseau a par opposition peu de chance d'activer une aire sélective aux visages. La stimulation des aires temporales influe également sur l'activation de l'amygdale et du pulvinar, deux régions sous-corticales impliquées dans la détection des visages (Johnson, 2005). Ainsi le réseau distribué est fortement interconnecté et pourrait constituer un système unifié pour le traitement perceptif des visages. Ces observations appellent deux questions : l'activation de ce réseau est-elle hiérarchique lors de la perception d'un visage ? Y a-t-il des différences fonctionnelles entre les différentes aires sélectives aux visages ?



Figure 15 : Dynamique d'activation des différentes aires cérébrales à la vue d'un visage connu (ici Johnny Depp). Schéma de la dynamique d'activation de différentes aires cérébrales pendant la reconnaissance d'un visage (gauche). Les rectangles indiquent les principaux pics des composantes dans chaque région. Les traits indiquent les débuts des composantes. Les rectangles bleus foncés indiquent les pics durant lesquels un effet de reconnaissance est observé. Les images de cerveau montrent l'emplacement des aires (milieu) dans lesquelles les potentiels évoqués étaient sélectifs aux visages (droite). D'après Barbeau et al., 2008.

La position anatomique d'OFA suggère qu'elle pourrait être une voie d'entrée vers les autres aires et notamment vers les deux aires du système central, FFA et STS. Une étude de connectivité fonctionnelle menée par dynamic causal modeling (DCM), a mis en évidence une organisation hiérarchique du système central avec une influence d'OFA sur FFA et STS (Fairhall and Ishai, 2007) et un lien majeur entre FFA et le système étendu de perception des visages notamment avec l'amygdale, le cortex orbitofrontal et le cortex préfrontal inférieur.

Une autre étude menée par électrophysiologie chez des patients épileptiques pharmaco-résistants, a suivi le décours d'activation des différentes aires du réseau dans huit régions cérébrales en comparant les potentiels évoqués par la vue du visage d'une personnalité célèbre (Barbeau et al., 2008). Un premier traitement en parallèle a été observé dans les aires du gyrus fusiforme à 110 ms et 170 ms (**Figure 15**). La composante à 170 ms est typique de la perception des visages, car le pic en est plus fort pour les visages que pour les autres stimuli visuels (Bentin et al., 1996; Sagiv and Bentin, 2001, mais voir Rousselet, 2004; Thierry et al., 2007). Ensuite un second traitement en parallèle a été observé à 240 ms dans les régions corticales plus antérieures incluant les cortex périorhinal, parahippocampique et le cortex temporal antérieur. Finalement l'activation la plus tardive a été observée dans l'hippocampe à 480 ms. Une comparaison entre les activations élicitées par des visages connus et inconnus a mis en évidence que les composantes de certaines régions sont modulées par la reconnaissance des individus et non seulement par le traitement perceptif des visages, par exemple dans l'hippocampe, le lobe temporal antérieur, le cortex périorhinal et le cortex parahippocampique.

Ces deux études pointent vers une organisation hiérarchique du système de perception des visages et supposent donc des fonctions différentes aux différentes régions sélectives aux visages.

C.2.2. Distinction entre le traitement statique et dynamique du visage

Le modèle de HAXBY et coll. (2000) pose une distinction majeure entre le traitement de l'identité d'un visage et celui des informations dynamiques qu'il contient, en proposant deux routes séparées dès le *système central*. En ce sens ce modèle suit la dissociation proposée par le modèle psychologique de BRUCE et YOUNG qui distingue le traitement de l'identité des divers traitements dynamiques (émotions, discours, attention du regard) (§ **A.2.**, **Figure 1**). L'aire OFA serait impliquée dans l'analyse des composants du visage, l'aire ventrale FFA dans la discrimination des visages et l'aire plus dorsale STS participerait à la reconnaissance de signaux sociaux comme la perception de la direction du regard, des mouvements des lèvres liés à la parole et des expressions faciales.

L'aire la plus dorsale, STS, est plus activée en IRMf par des visages dynamiques que par des visages statiques (Ghazanfar et al., 2010; Pitcher et al., 2011). STS est également activée par des visages statiques mais présentant un mouvement sous-entendu, par exemple par des visages dont le regard est tourné vers une certaine direction, des visages présentant des expressions faciales émotionnelles, ou des visages effectuant des mouvements des lèvres (Hoffman and Haxby,

2000). Ces observations ont été confirmées par des enregistrements électrophysiologiques de neurones chez le singe. En effet PERRETT et coll. ont montré que différentes cellules de STS sont sélectives pour différentes directions du regard ou différentes orientations de la tête (Perrett et al., 1982; Perrett et al., 1985; Perrett et al., 1991; Hietanen et al., 1992; Perrett et al., 1992), ainsi que pour différentes expressions du visage (Perrett et al., 1984). Une autre étude électrophysiologique a observé que les neurones codant la nature des expressions sont plus nombreux dans le STS alors que les neurones sélectifs à l'identité sont plus nombreux dans la convexité du gyrus temporal inférieur (Hasselmo et al., 1989), confirmant la spécificité du codage faciale dans STS par rapport à celui effectué par les autres aires temporales. Comme la perception de la direction du regard est à la base du mécanisme social d'attention partagée (*joint attention*), il a été proposé que STS joue un rôle majeur dans un réseau dédié à la cognition sociale (Allison et al., 2000; Adolphs, 2003 ; 2009). Ce réseau inclurait également des régions du système étendu pour la perception des visages, participant à des jugements sociaux. En effet l'amygdale et l'insula participent à la reconnaissance des émotions faciales particulièrement de la peur du dégoût et de la colère (Breiter et al., 1996; Morris et al., 1996; Phillips et al., 1997; Ishai et al., 2004), alors que l'attractivité d'un visage évoquerait une plus forte activation du cortex orbitofrontal et du circuit de la récompense dont l'amygdale et le nucleus accumbens (Nakamura et al., 1998; Aharon et al., 2001; O'Doherty et al., 2003; Kranz and Ishai, 2006; Ishai, 2007; Kim et al., 2007).

Enfin il a été proposé que l'identité faciale dynamique (§ A.3.1) contenues dans les expressions faciales, les mouvements idiosyncratiques des lèvres pendant le discours et le regard, soit également analysée par STS qui constituerait une voie alternative ou complémentaire pour la reconnaissance individuelle (O'Toole et al., 2002). Cette voie pourrait être utile lorsque les conditions visuelles ne sont pas optimales (faible illumination, résolution basse, reconnaissance à distance). Elle pourrait également apporter des informations complémentaires aux aires OFA et FFA ventrales et les deux pourraient communiquer en ce sens.

Par opposition à STS, les deux aires ventrales OFA et FFA seraient impliquées dans la perception de l'identité. L'étude de cas de patients prosopagnosiques cérébro-lésés a mis en évidence que OFA et FFA sont les deux régions nécessaires au traitement holistique des composants et de la configuration des visages. Par exemple l'étude des saccades oculaires de la patiente prosopagnosique P.S. présentant une lésion d'OFA gauche, mais une FFA droite intacte, montre qu'elle regarde plus longuement les différents détails du visage, et particulièrement la bouche, au lieu de poser son attention au centre du visage pour le percevoir dans sa globalité comme le fait un sujet sain (**Figure 16**, Rossion et al., 2003; Orban de Xivry et al., 2008). P.S. fait des allers-retours entre les différents détails pour les analyser un à un. En effet elle utilise les particularités de ces détails, et plus particulièrement de la bouche, pour reconnaître les individus connus. De façon similaire le patient prosopagnosique L.R. ayant également une lésion droite de OFA présente un temps de fixation plus long sur la bouche des individus plutôt que sur le haut du nez, indiquant un traitement non-holistique du visage (Bukach et al., 2006; Bukach et al., 2008). Cependant bien qu'il ne perçoive pas la configuration spatiale de la partie haute du visage incluant les yeux, L.R. perçoit la configuration spatiale de la partie basse du visage située au tour

de la bouche. Les études de cas cliniques de prosopagnosie due à une lésion de FFA montrent quant à elles un déficit global dans la perception des configurations spatiales des visages dans les parties hautes et basses du visage (De Gelder and Rouw, 2000; Barton et al., 2002). De façon intéressante, l'aire FFA reste active chez la patiente P.S. comme mis en évidence par IRMf. L'ensemble de ces études cliniques indique qu'OFA et FFA sont toutes deux nécessaires à la perception holistique des visages. Ces observations concordent avec les études en IRMf présentées précédemment qui démontrent le rôle de l'aire FFA pour la perception holistique des visages (Gauthier et al., 2000b; Grill-Spector et al., 2004) et éventuellement d'autres objets d'expertise (Gauthier et al., 1999; Gauthier and Logothetis, 2000; Gauthier et al., 2000a; Dehaene et al., 2010).

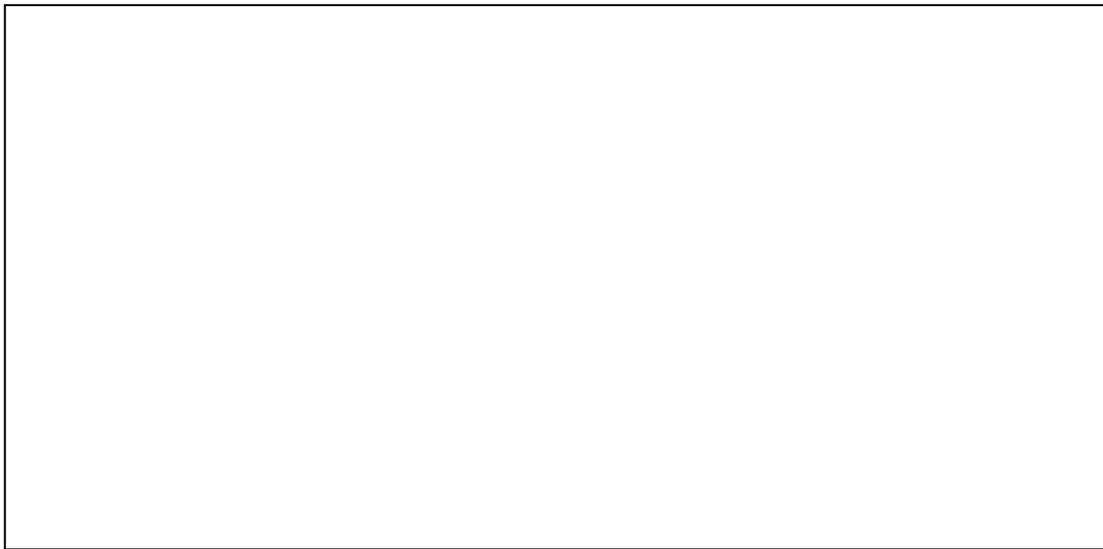


Figure 16 : Représentation du temps de regard alloué par la patiente prosopagnosique P.S. (droite) et un témoin contrôle (gauche) aux différentes parties du visage lors d'une tâche de reconnaissance. La couleur de chaque région correspond au temps de fixation moyen. Le point noir central représente le point de fixation présenté juste avant l'apparition d'un visage à l'écran. Chaque petit point noir représente une fixation des yeux. Notons le grand nombre de fixation effectué par P.S. sur la bouche et au centre des détails alors que les fixations du contrôle restent proches du point central. D'après Orban de Xivry et al., 2008.

C.2.3. Discrimination et identification des visages

La discrimination et l'identification des visages nécessitent au moins deux processus : la catégorisation des visages selon leurs caractéristiques physiques et la perception de l'invariance de l'identité d'un visage qui résisterait à des changements physiques. Existe-t-il une distinction de ce type de traitement entre OFA et FFA ?

Une étude en IRMf montre qu'OFA serait impliquée dans la mesure des propriétés physiques faciales et FFA dans la classification des visages (Rotshtein et al., 2005). En utilisant des morphing de visages entre Marilyn Monroe et Margaret Thatcher, les auteurs ont mis en évidence

que la courbe d'adaptation du signal provenant de l'aire OFA suit la courbe de morphing physique, alors que celle de FFA suit celle du changement d'identité perçu (**Figure 17**). OFA serait sensible aux changements physiques alors que FFA serait sensible aux changements d'identité consciemment perçu. Ainsi FFA réalise une identification de la personne qui résiste à des changements physiques moindres, ici caractérisés par un certain degré de morphing. FFA et OFA sont donc toutes deux impliquées dans la catégorisation des visages mais à des niveaux d'abstraction différents. L'invariance de FFA à des petits changements physiques, a également été montrée par deux autres études faisant varier la taille des visages (Andrews and Ewbank, 2004), et leur filtrage à différentes bandes de fréquences (Eger et al., 2004). Cependant cette invariance ne résiste pas à la perception d'un même visage présenté de points de vue différents (Andrews and Ewbank, 2004).

En conclusion ces trois résultats confirment le rôle d'OFA pour la mesure des propriétés physiques des visages et montrent un début de codage invariant de l'identité des visages par l'aire FFA.

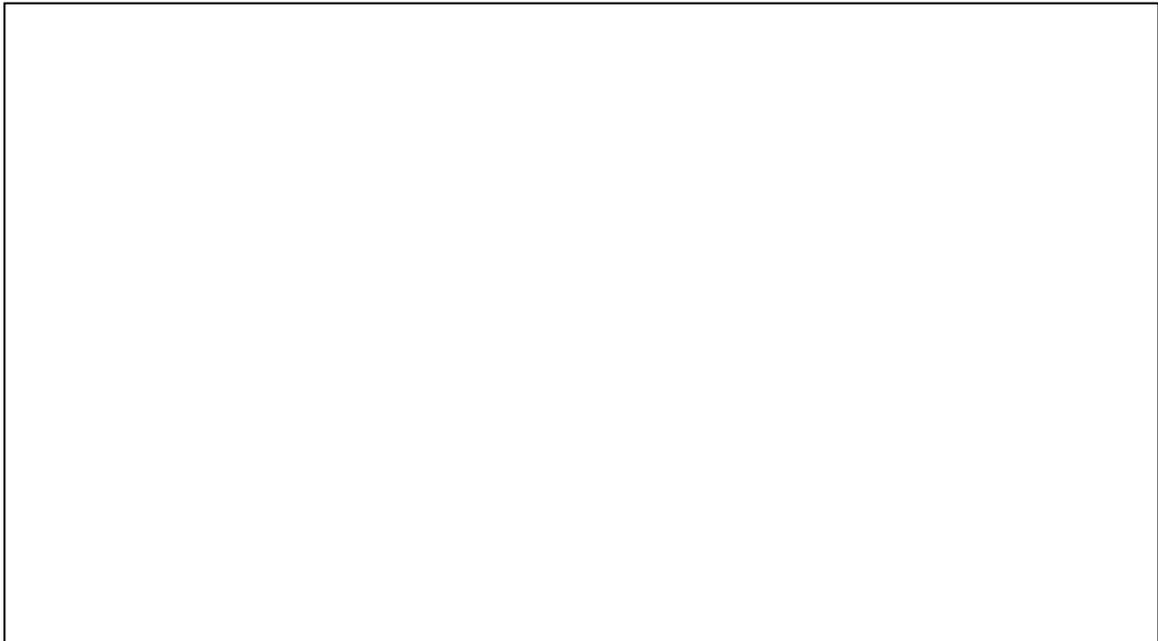


Figure 17 : Continuum du morphing entre Marilyn Monroe et Margaret Thatcher. Les sujets de l'étude devaient identifier en choix forcé si l'image correspondait à l'une ou l'autre des personnes. Le graphique de gauche représente la proportion de réponses identifiant Marilyn Monroe en fonction du continuum. Les images en dessous montrent les trois types de comparaisons effectués dans l'étude IRMf : 1) même identité perçue mais propriétés physiques différentes, 2) même identité et même propriétés physiques, 3) identités perçues différentes et propriétés physiques différentes. Les figures de droites montrent l'activation des aires OFA droite et gauche dans le cas de changements physiques (haut et milieu) et l'activation de FFA pour les changements d'identité (bas). D'après Rotshtein et al., 2005.

C.2.4. Catégorisation des visages

Les études psychologiques ont montré que les représentations des visages s'inscrivent dans un *espace des visages* multidimensionnel, qui pourrait servir de trame à la façon dont les visages sont encodés au niveau neuronal (§ A.3.1). Selon ce modèle les neurones ne seraient pas hautement sélectifs pour un visage individuel mais répondraient de façon de plus en plus soutenue le long d'un continuum des visages à partir du visage prototypique moyen jusqu'au visage extrême – ou caricatural (**Figure 18 gauche**). Cette hypothèse a été testée par l'enregistrement des activités neuronales chez le singe (**Figure 18 droite**). Ainsi LEOPOLD et coll. (2006) ont montré à l'aide de morphing de visages que les neurones du lobe temporal antérieur organisent leur activité autour d'un visage prototypique moyen et présentent un maximum d'activité pour les visages les plus éloignés de ce prototype (Leopold et al., 2006). Les neurones du lobe temporal antérieur pourraient donc représenter une carte de l'espace des visages. De façon similaire les neurones du middle face patch, présentent des préférences le long de continuums concernant les composantes (par exemple la longueur des cheveux ou la forme des sourcils) et les configurations faciales (par exemple la distance entre les yeux), avec un maximum d'activité à un extrême et un minimum de réponse à l'autre extrême d'une même dimension (Freiwald et al., 2009). Notons que ces deux études ont observé un codage par carte dans des zones différentes du lobe temporal, le middle face patch et le lobe temporal antérieur, laissant supposer que la catégorisation des visages pourrait être progressive dans le lobe temporal, et que la représentation sous forme de cartes pourrait être répandue dans le lobe temporal.



Figure 18 : Les deux modèles proposés pour la représentation des visages par les neurones et leur test. A) Dans le modèle d'exemplaire unique (gauche haut), chaque visage active sélectivement un neurone. Dans le modèle axial (gauche bas), l'espace des visages est organisé de façon paramétrique, telle que l'identité des différents visages présents le long d'un axe soit représentée par les activités relatives de seulement deux cellules. B) Exemple de l'activité de deux neurones du lobe temporal antérieur du singe, à la présentation de différents visages morphés organisés selon un continuum (droite). À gauche se situe le visage prototypique (0%), à droite les visages caricaturaux (160 %) et entre les deux les visages normaux (100%). Sous chaque visage sont représentés les rasters et les histogrammes des taux de décharge. Le trait noir indique les 400 ms de présentation du visage. D'après Leopold et al., 2006 et Tsao et al., 2006.

La sélectivité progressive plutôt qu'exclusive des activités neuronales met en évidence qu'une petite population de neurones suffit à coder tous les types de visages (**Figure 18 gauche**). En effet dans un cas le nombre de neurones sélectifs augmenterait proportionnellement au nombre de visages connus, alors que dans l'autre les neurones ne coderaient que pour les visages extrêmes. En plus de l'optimisation des ressources neuronales qu'il génère, l'encodage sous forme de carte est plus puissant pour la réalisation de comparaisons de visages entre eux et par conséquent pour l'identification.

C.2.5. Invariance des visages

L'identification d'un visage nécessite qu'il soit discriminé des autres visages quelle que soit l'orientation dans laquelle il est présenté. Nous avons vu que l'aire STS code spécifiquement les différentes orientations d'un visage indépendamment de son identité. A l'opposé l'aire FFA présente un début d'invariance à la taille du visage, mais non à son orientation. Une comparaison de la sélectivité pour l'identité et pour l'orientation du visage a été réalisée entre les neurones des six *face patches* du sillon temporal (**Figure 19**). Les chercheurs ont montré que les régions les plus postérieures sont les plus sélectives à l'orientation du visage indépendamment de l'identité, inversement les régions les plus antérieures sont plus sélectives à l'identité qu'à l'orientation. Les aires du milieu présentent un continuum de sélectivité pour les deux aspects des visages, avec par exemple une même activité pour les orientations en miroir (Freiwald and Tsao, 2010).



Figure 19 : Sélectivité neuronale pour l'identité et l'orientation du visage dans le réseau dédié au traitement perceptif des visages. Schématisation idéalisée de trains de potentiel d'action élicités par la vue des différentes photos de visages. Les neurones des régions postérieures sont sélectifs à l'orientation mais non à l'identité, les neurones des régions intermédiaires présentent plus de sélectivité à l'identité et présente une invariance pour les orientations en miroir. Les neurones des régions antérieures seraient fortement sélectifs aux identités indépendamment de l'orientation du visage. D'après Connor, 2010 et Freiwald and Tsao, 2010.

Des neurones présentant une même activité pour les deux vues en miroir avaient déjà été observés précédemment dans une région proche de AL (De Souza et al., 2005). De même la mise en évidence de neurones codant l'identité d'un visage de façon invariante a été répliquée par la suite dans les aires les plus antérieures du lobe temporal (Eifuku et al., 2010; Eifuku et al., 2011). Cependant dans cette dernière étude les auteurs suggèrent que l'invariance ne s'observe que pour

les visages d'individus humains personnellement connus par les singes et non pour les visages d'individus inconnus, contrairement à ce qui est observé dans l'étude de FREIWALD et coll. (2010) dans laquelle tous les visages présentés sont inconnus. Ces études longitudinales mettent en évidence un codage hiérarchique de l'invariance des visages entre le STS et le lobe temporal antérieur. L'identité invariante des individus semble donc codée de la façon la plus abstraite au niveau du lobe temporal antérieur ventral.

C.3. Reconnaissance du visage

Lorsque nous percevons le visage d'une personne connue, en plus de la discriminer des autres individus, nous reconnaissons son identité. La reconnaissance nécessite au moins cinq étapes (§ A.6) : 1) une comparaison au pool de visages connus stocké en mémoire, 2) l'activation d'un sentiment de familiarité, et la réactivation des 3) connaissances sémantiques, 4) épisodiques et 5) morales associées à l'individu.

C.3.1. Implication du lobe temporal antérieur : Comparaison avec un pattern stocké en mémoire

Sur la base d'études de neuroimagerie, il a d'abord été proposé que le cortex temporal antérieur (aIT) et FFA jouent des rôles similaires et complémentaires dans l'identification des visages (Sergent et al., 1992; Kriegeskorte et al., 2007; Nestor et al., 2011). Ceci conforte les données neurophysiologiques obtenues chez le singe. D'une part celles-ci décrivent des neurones codant l'identité sous la forme d'un « espace des visages » (Leopold et al., 2006). D'autre part l'étude des temps de latences comportementaux et neuronaux, a mis en évidence une corrélation entre les activités neuronales de aIT et l'identification des visages – absente dans les aires postérieures de STS (Eifuku et al., 2004). Ainsi FFA et aIT arbitrent toutes deux le passage d'une représentation physique à une représentation conceptuelle de l'identité.

Cependant contrairement à FFA, l'aire aIT présente la particularité d'être active dans des tâches de jugement de la familiarité des visages, et est en règle générale plus activée par les visages familiers qu'inconnus (**Figure 15**, § C.2.1) (Nakamura et al., 2000; Sugiura et al., 2001; Ishai et al., 2005; Barbeau et al., 2008). Ainsi l'identification des visages dans aIT pourrait être fonction de leur familiarité. C'est ce qui a été montré par ROTSHTEIN et coll. (2005), dans l'étude morphant le visage de Marilyn Monroe en celui de Margaret Thatcher (**Figures 17 et 20 gauche**). Les auteurs ont observé une relation positive entre l'ampleur de l'activité dans aIT due au changement d'identité perçu et le niveau de familiarité avec les deux personnages (Rotshtein et al., 2005). Ces résultats corroborent plusieurs études en IRMf et PET chez l'humain montrant une augmentation de signal dans aIT lors de l'identification de visages familiers (Gorno-Tempini et al., 1998; Leveroni et al., 2000; Gorno-Tempini and Price, 2001; Sugiura et al., 2001). Ceci concorde également avec les enregistrements électrophysiologiques mettant en évidence une

activité neuronale différente pour les individus familiers par rapports aux autres (Eifuku et al., 2011). Enfin ces résultats sont en accord avec les études de patients ayant une lésion du lobe temporal antérieur. Les lésions créent un type particulier de prosopagnosie associative, *la prosopamnésie rétrograde*, rendant les patients incapables de reconnaître une personne bien qu'ils puissent percevoir et discriminer les visages inconnus correctement (Ellis et al., 1989; Hanley et al., 1989; Evans et al., 1995) ; ces déficits étant plus souvent associés à une résection du lobe temporal droit plutôt que gauche.



Figure 20: Régions cérébrales présentant une corrélation positive entre l'expérience personnelle de familiarité avec les visages de Marilyn Monroe et Margaret Thatcher et l'effet BOLD du changement d'identité, lorsque les changements physiques étaient manipulés : pôle temporal antérieur gauche (gauche) et hippocampe antérieur droit (droite). Et graphiques présentant l'effet du changement d'identité physique en IRMf en fonction de la familiarité des sujets avec les deux personnes. D'après Rotshtein et al., 2005.

L'ensemble ces études démontre le rôle important joué par le lobe temporal antérieur dans la reconnaissance des visages familiers. Ceci suggère par ailleurs que l'espace des visages mis en évidence par électrophysiologie s'organise à partir des visages connus, comme prédit par les études psychologiques sur le *Other race effect*. De façon intéressante une activité similaire à celle d'aIT a été observée dans l'hippocampe dans l'étude présentée ci-dessus (**Figure 20 droite**).

C.3.2. Implication de l'hippocampe et des régions parahippocampiques et périrhinales

Une aire sélective aux visages a été mise en évidence plus récemment dans l'hippocampe par IRMf chez l'humain (Ishai et al., 2005) et le singe (Ku et al., 2011). De même que pour aIT, l'activité de l'hippocampe lors de l'identification est corrélée positivement avec la familiarité des sujets pour les visages présentés (**Figure 20 droite**, Rotshtein et al., 2005). Cet effet concorde avec les résultats d'études précédentes montrant que l'hippocampe s'active préférentiellement à la vue de visages connus plutôt qu'inconnus (**Figure 15**, § C.2.1, Leveroni et al., 2000; Haist et al., 2001; Bernard et al., 2004; Denkova et al., 2006; Ishai and Yago, 2006; Barbeau et al., 2008). De plus cette activation est spécifique d'un rappel des informations associées préalablement au visage (voix, nom, occupation) et non seulement de la familiarité avec le visage (Paller et al., 2003). Enfin sa lésion peut également résulter en une prosopamnésie rétrograde (Cipolotti et al., 2001).

Au niveau neuronal, des enregistrements réalisés chez des patients épileptiques pharmacorésistants ont mis en évidence des neurones sélectifs aux visages au coté de neurones sélectifs à d'autres catégories d'images (Fried et al., 1997; Kreiman et al., 2000; Fried et al., 2002; Kraskov et al., 2007). Dans ces études 18% des neurones répondaient sélectivement à une et parfois à deux catégories. Des études subséquentes ont révélées l'existence d'une petite population de neurones très épars et peu nombreux mais incroyablement invariants à l'identité de personnalités connues ou de personne connues de l'entourage des patients (Quiroga et al., 2005; Quian Quiroga et al., 2009). Par exemple des cellules étaient actives pour de multiples vues d'un même individu ainsi qu'à son nom écrit ou prononcé par l'expérimentateur (**Figure 21**). Il a été mis en évidence que ces neurones sont modulés par la familiarité avec les stimuli (Viskontas et al., 2009). En effet ils s'activent préférentiellement pour les visages d'individus personnellement connus, par rapport aux visages de célébrités ou de personnes inconnues.

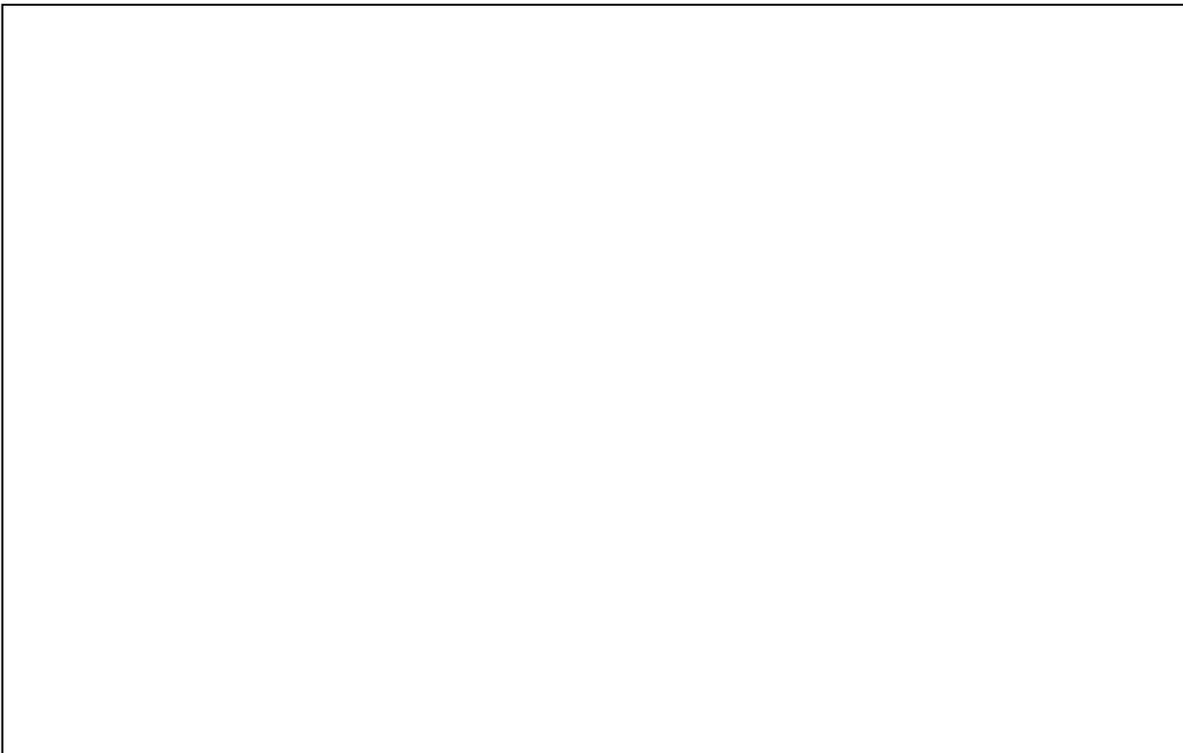


Figure 21 : Exemple d'un neurone présentant une activité sélective, invariante et multimodale pour les stimuli se référant à l'animatrice de télévision américaine Oprah Winfrey. Le neurone décharge à la vue de son visage (stimuli 40, 39, 11), de son nom écrit (56) et de son nom prononcé (73). Ce neurone ne décharge pas ou peu pour les autres stimuli visuels ou auditifs. D'après Quian Quiroga et al., 2009.

Cependant des différences existent entre les activités d'aIT et de l'hippocampe relativement aux visages. D'une part la dynamique des deux aires est différente : l'étude des potentiels évoqués par des visages de célébrités chez l'humain montre que l'activation de l'hippocampe présente un pic tardif à 480 ms alors que le lobe temporal antérieur présente deux pics d'activation à 240 et 360 ms (**Figure 15**, § C.2.1, Barbeau et al., 2008) ; une comparaison des latences d'activation

neuronales chez le singe et l'humain montre une activité plus tardives dans l'hippocampe par rapport à TE et par rapport au cortex parahippocampique respectivement (Mormann et al., 2008). D'autre part, l'hippocampe est également engagé dans l'apprentissage de nouveaux visages comme mis en évidence par imagerie (Haxby et al., 1996) et par des lésions à la base de troubles *prosopamnésiques antérogrades* (Tippett et al., 2000; Cipolotti et al., 2001). Enfin l'hippocampe est aussi impliqué durant l'imagerie mentale de visages de célébrités (Ishai et al., 2002) et est actif seulement lorsque les stimuli sont perçus consciemment (Quiroga et al., 2008b).

Ainsi l'aire aIT et l'hippocampe semblent jouer des rôles majeurs dans la reconnaissance de l'identité d'un visage mais avoir des contributions distinctes. La connaissance de leur rôle respectif pour le rappel en *mémoire sémantique* et *épisode* (Westmacott et al., 2001; Westmacott et al., 2004) suggère que leurs contributions pourraient concerner les accès aux connaissances sémantiques et épisodiques associées à une personne familière (Denkova et al., 2006). Un type particulier de connaissance sémantique concerne la connaissance de la *personnalité* de l'individu.

C.3.3. Implication des régions dévolues au traitement de la théorie de l'esprit et des émotions

La vue d'individus personnellement connus (parents, enfants, amis, partenaire, collègues) réactive spontanément nos connaissances sur leur personnalité et leurs attitudes. Une activation de régions supplémentaires à celles activées par des visages appris pendant une expérience et par celles activées par des visages de célébrités a été mise en évidence par IRMf.

Les visages personnellement connus évoquent des réponses plus importantes que les visages célèbres, dans le cortex paracingulaire, le sulcus temporal postérieur et le cortex cingulaire postérieur/precuneus (Ida Gobbin et al., 2004; Leibenluft et al., 2004; Taylor et al., 2009). D'autres études ont associé ces zones à la représentation des attributs personnels et aux états mentaux d'autrui – ou *Théorie de l'esprit* (Frith and Frith, 1999; Mitchell et al., 2002; Gallagher and Frith, 2003; Saxe et al., 2004; Moll et al., 2005; Zaki and Ochsner, 2009). Notamment ces aires sont actives dans des tâches qui requièrent l'interprétation et la prédiction du comportement d'autrui.

Par conséquent l'activation de ces aires à la vue d'un visage personnellement connu pourrait être due à une réactivation spontanée des connaissances sur la personnalité, les attitudes et les intentions des individus connus personnellement. Par ailleurs l'activité d'une aire particulière semble croître avec la *familiarité* des visages et atteindre son optimum pour les visages personnellement connus.

C.3.4. Plusieurs corrélats neuronaux du sentiment de familiarité

Certaines études se sont attelées à dissocier la composante de familiarité associée à un visage de celle de la réactivation d'autres connaissances associées à cette personne. Différentes méthodes ont été utilisées (Gobbini and Haxby, 2007).

GOBBINI et coll. (2006) ont par exemple familiarisé les sujets de leur expérience avec des photos de visages sans donner d'indication relatives aux individus présents sur les photos. Les visages étaient alors dénués de connaissances associées. Les auteurs ont mis en évidence une activation croissante du precuneus avec l'acquisition de la familiarité (Gobbini and Haxby, 2006). Ce groupe avait précédemment mis en évidence une gradation dans l'activité du precuneus pour les personnes personnellement familière par rapport aux célébrités (Ida Gobbini et al., 2004; Leibenluft et al., 2004). D'autres études ont également montré que l'activation precuneus est corrélée à la familiarité (Kosaka et al., 2003), suggérant que le precuneus joue un rôle important dans l'acquisition de la familiarité avec les visages.

Par ailleurs une activation plus faible de l'amygdale à la vue de visage familiers a été mise en évidence par ces mêmes études (Ida Gobbini et al., 2004; Gobbini and Haxby, 2006) et par d'autres (Bartels and Zeki, 2000; Schwartz et al., 2003; Wright et al., 2003), montrant que l'amygdale serait sensible aux événements inattendus ou inconnus ayant une importance biologique potentielle, comme les visages nouveaux. La diminution de l'activité de l'amygdale pourrait refléter une baisse du niveau de vigilance accompagnant la vue d'une personne familière, par rapport à une personne inconnue. Cependant le vue du visage de ses propres enfants provoquait une augmentation inattendue du niveau d'activation de l'amygdale, accompagné d'une augmentation d'activité dans l'insula (Leibenluft et al., 2004). L'activité de l'insula avait déjà été observée précédemment (Bartels and Zeki, 2004) et interprétée comme le reflet de l'empathie (Singer et al., 2004). Les auteurs suggèrent que ces augmentations d'activité à la vue de ses propres enfants reflètent la protection vigilante ainsi que l'intense attachement spécifique à la relation maternelle.

D'autres études ont utilisée des effets de dissociation entre la familiarité et la reconnaissance. Nous avons vu précédemment (§ A.6) qu'une double dissociation entre la reconnaissance et la familiarité (Burton et al., 1991 ; Schweinberger and Burton, 2003) peut être mise en évidence par l'étude de personnes prosopagnosiques (Bauer, 1984; Tranel and Damasio, 1985) et de personnes atteintes du syndrome de CAPGRAS (Ellis and Young, 1990; Ellis et al., 1997b; Ellis and Lewis, 2001). En effet les personnes prosopagnosiques ayant une lésion bilatérale occipito-temporale reconnaissent de façon non consciente qu'elles sont en présence de visages familiers, comme ceci peut être révélées par des tests électrodermaux. Ces études indiquent qu'une réponse émotionnelle pourrait de plus véhiculer la familiarité par des voies sous-corticales.

C.4. Conclusion

La perception et la reconnaissance des visages engagent de nombreuses aires cérébrales corticales et sous-corticales distribuées à travers le cerveau. Ces traitements suivent un continuum le long du lobe temporal des aires primaires vers des aires associatives. Durant le

trajet le passage d'une représentation physique à une représentation conceptuelle de l'identité est progressivement réalisé. Cependant des allers-retours entre la perception et la reconnaissance sont constamment effectués, comme mis en évidence par les études de connectivités (Moeller et al., 2008; Nestor et al., 2011). Ainsi notre perception des visages se nourrit de notre connaissance des différents types de visages rencontrés au cours de notre vie. Ceci est drastiquement démontré par les études psychologiques sur le *Other race effect* (Vizioli et al., 2010). En effet *l'espace des visages* qui permet à la fois la discrimination des individus inconnus et l'identification des individus connus, est basé sur notre propre expérience avec les visages construite au cours de la vie. Ceci signifie que l'ensemble des circuits neuronaux ayant participé à l'analyse et au traitement de l'information sensorielle se modifie à la suite de l'expérience sensorielle, permettant notamment une optimisation du traitement des nouvelles informations. Ainsi la plasticité des régions perceptives contribue autant à la reconnaissance individuelle que l'activité d'aires associatives. Le rôle d'aires associatives, et plus spécifiquement de l'hippocampe, serait à la fois de faciliter cette plasticité cérébrale et de réactiver d'autres régions plus lointaines, par exemple auditives, pour un rappel complet d'un épisode.

Les avancées en neurosciences ont mis en évidence des composantes ignorées par le modèle de BRUCE et YOUNG. D'une part la mise en évidence de régions cérébrales associées aux connaissances de la personnalité d'autrui montre que la reconnaissance d'un individu familier implique plus que la simple représentation de l'apparence de cette personne. D'autre part l'étude de la familiarité a mis en évidence différentes composantes indépendantes les unes des autres et de la reconnaissance. Ceci a amené à revisiter les modèles de perception des visages proposé par BRUCE et YOUNG et par HAXBY et al (**Figure 22**, Johnson, 2005; Gobbini and Haxby, 2007). Enfin les études de connectivité ont démontré que les différentes étapes du modèle ne seraient pas sérielles mais bénéficieraient des projections en feedback des niveaux ultérieurs.

Ainsi le PIN proposés par BRUCE et YOUNG pourrait être redéfinis en différents sous-processus : 1) l'association entre le visage et d'autres indicateurs perceptifs de l'identité comme la voix, 2) la familiarité avec le visage, 3) le rappel du nom à partir du visage, 4) le rappel des autres informations sémantiques biographiques, 5) le rappel des informations épisodiques, 6) le rappel de la personnalité. Des questions restent en suspens quant aux aires cérébrales impliquées dans ces différents processus. Les études sur la reconnaissance des voix et des associations entre les voix et les visages pourraient clarifier ces questions.

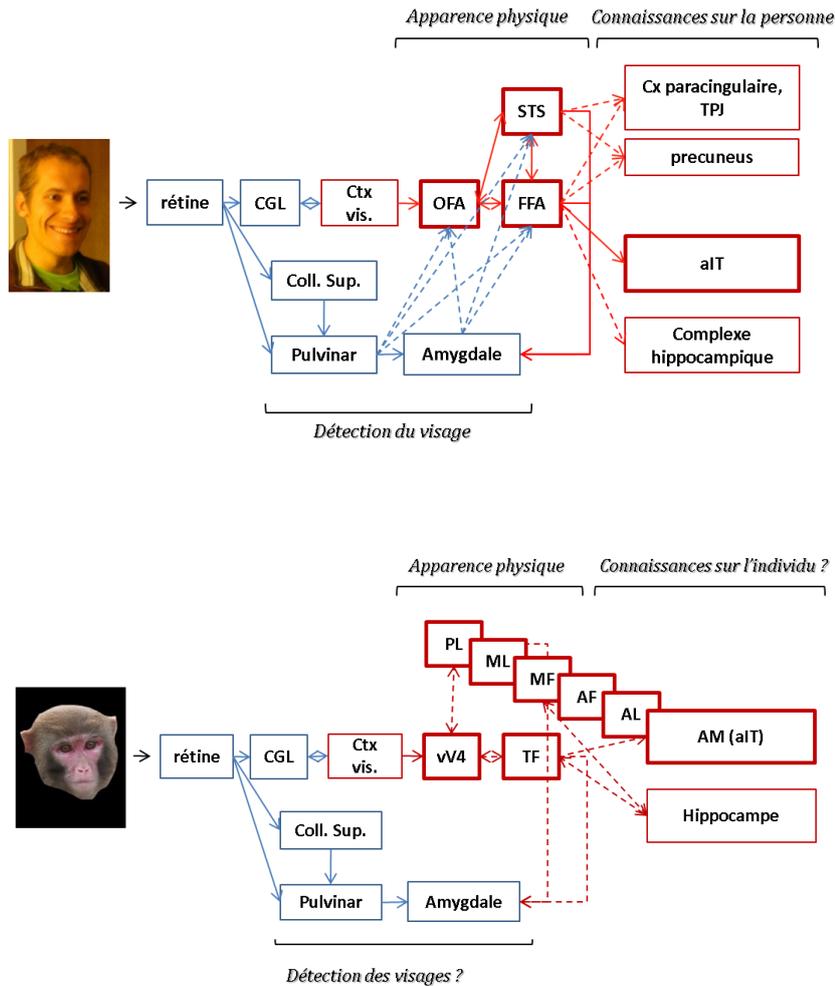


Figure 22 : Schémas prenant en compte des deux nouveaux modèles proposés pour l'analyse des visages (Johnson, 2005; Gobbini and Haxby, 2007). Le modèle des deux routes de Johnson et coll. (2005) suppose une voie sous-corticale d'analyse et de détection des visages (bleu) qui se projette sur le système central (OFA-FFA-STS) et interagit avec lui. Dans la nouvelle version du modèle distribué de Gobbini et Haxby, les aires du système central liées à l'analyse de l'apparence visuelle (OFA-FFA-STS) sont distinguées des structures impliquées dans la familiarité (Amygdale et precuneus), le rappel des informations biographiques (aIT), épisodiques (complexe hippocampique) et des traits de personnalité (Cx paracingulaire, TPJ). Schéma similaire chez le singe. D'après Johnson, 2005 et Gobbini and Haxby, 2007.

D. Bases neurales de l'identification des voix

L'étude des bases neurales de la perception des voix s'est d'abord intéressée aux bases neurales du langage mais a bénéficié par la suite des mêmes avancées techniques et des données sur la perception des visages pour expliquer le réseau impliqué dans la perception des voix. Dans cette partie nous allons présenter les bases neurales engagées dans la perception des voix ainsi que les liens qui existent entre ce système et celui pour la perception des visages.

D.1. Un système dédié à la perception de ce stimulus social

D.1.1. Mise en évidence par des lésions neurologiques d'aires spécialisées

Les données cliniques ont décrit un syndrome d'agnosie auditive spécifique aux voix, nommé *phonagnosie* (Van Lancker and Canter, 1982; Vanlancker et al., 1989; Neuner and Schweinberger, 2000). Il s'agirait ici d'une incapacité à identifier différents individus d'après leur voix malgré des systèmes sensoriels intacts. Ce syndrome serait le pendant de la prosopagnosie dans le domaine auditif. A l'opposé du syndrome de phonagnosie, des cas cliniques ont fait états de déficits circonscrits à la musique (Peretz et al., 1994) ou aux bruits d'objets (Van Lancker and Canter, 1982), laissant intacte l'identification des voix. Cette double dissociation concernant la perception des voix par rapport aux autres stimuli auditifs indique que des aires spécialisées pour les voix sont distinctes d'un système pour la perception des sons en général. Par ailleurs, différents cas de phonagnosie peuvent être distingués. Certains types produisent des déficits pour la reconnaissance des voix familières (Hailstone et al., 2010) alors que d'autres empêchent la discrimination des voix inconnues (Van Lancker and Kreiman, 1987; Van Lancker et al., 1988). Bien qu'ils n'aient pas été nommés ainsi, ils pourraient être qualifiés respectivement de *phonagnosies associative* et *aperceptive*. Les cas de phonagnosies associatives documentés ne concernent que des cas de *phonamésie rétrograde* (Neuner and Schweinberger, 2000; Hailstone et al., 2010).

Les lésions des patients agnosiques et phonagnosiques concernent les lobes temporaux et non l'aire de Broca, frontale antérieure, ou l'aire de Wernicke temporo-pariétale, comme dans les cas d'aphasies (Lancker et al., 1989). Les lésions sont donc situées dans le même lobe que celles des

patients prosopagnosiques, indiquant un rôle de ce lobe pour le traitement de l'identité perçue à travers diverses modalités.

D.1.2. Les deux voies corticales d'analyse des stimuli auditifs

Au niveau cortical, le traitement de l'information auditive suit des routes parallèles à celles de l'information visuelle, en étant dissocié en deux voies anatomiques : une voie occipito-pariétale et une voie occipito-temporale (**Figure 23 haut**) (Romanski et al., 1999). Sur la base de données neurophysiologiques RAUSCHECKER et coll. ont proposé la même interprétation fonctionnelle pour ces deux voies que celle proposée par MISHKIN et UNGERLEIDER pour le traitement visuel ; ils distinguèrent la voie du « quoi » de la voie du « où ». Une seconde interprétation fonctionnelle a été proposée par la suite en tenant compte du rôle prépondérant de la voie dorsale pour le langage chez l'Homme, en distinguant cette fois une voie du « quoi » et une voie du « comment » (Belin and Zatorre, 2000; Rauschecker and Scott, 2009).

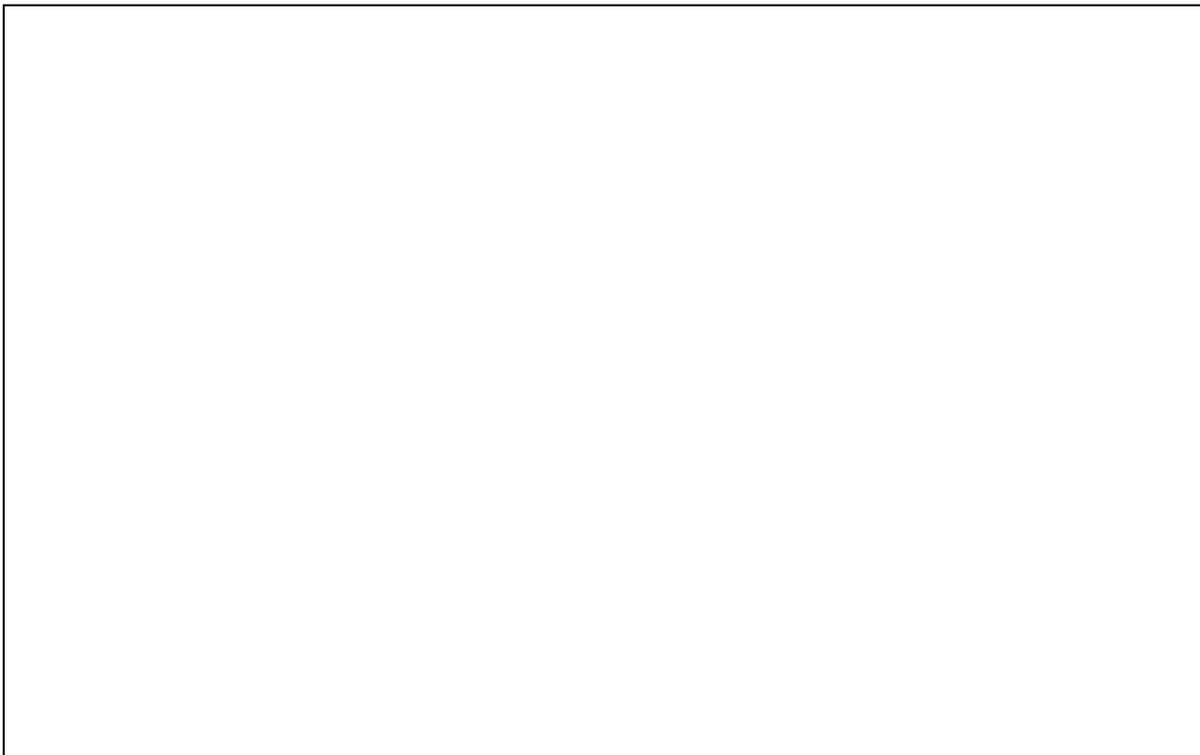


Figure 23 : Projections du cortex auditif vers le lobe temporal et le lobe frontal (haut gauche) et division en deux voies de traitement de l'information (haut droite). Le cortex auditif peut être divisé en trois régions (bas) : un noyau – *core* comprenant les aires A1 R et RT en violet –, une première région périphérique – *belt* comprenant les aires CM, CL, ML, AL, RTL, RTM, MM en jaune –, une seconde région périphérique – *parabelt* comprenant les aires RP et CP en orange. D'après Kaas and Hackett, 1999, 2000; Romanski and Averbeck, 2009.

Les signaux initiaux provenant des deux cochlées sont acheminés vers un réseau complexe de voies et de noyaux dans le tronc cérébral et le thalamus. L'information spectrale et temporelle extraite par ces structures est alors utilisée pour déterminer l'identité et la localisation de la source du son. Bien que la perception de l'identité puisse être réalisée à partir des informations provenant d'une seule oreille, celle de la localisation nécessite les informations provenant des deux oreilles. C'est l'une des raisons pour lesquelles un traitement différent est opéré pour l'identité et la localisation dès le tronc cérébral, et est poursuivi ensuite dans le cortex. La transformation non-isomorphe de l'information acoustique en information perceptive est réalisée dans ces structures, en amont du cortex auditif primaire (Griffiths and Warren, 2004; Wang, 2007).

Le *cortex auditif* peut être divisé en trois régions principales, chacune constituée de plusieurs aires (**Figure 23 bas**) (Galaburda and Pandya, 1983; Kaas et al., 1999; Kaas and Hackett, 2000; Petkov et al., 2006; Recanzone and Sutter, 2008). L'organisation comprend un noyau et deux régions périphériques : *core – belt – parabelt*. Chaque régions se distingue des autres sur la base de son apparence histologique, de ses connexions thalamiques et corticales et des propriétés fonctionnelles.

Le noyau comprend trois aires corticales : le cortex auditif primaire A1, le champ rostral R et le champ rostrotemporal RT (Hackett et al., 2001). A1 et R ont une organisation tonotopique (Formisano et al., 2003; Petkov et al., 2006) et R encode la tonalité du son (Bendor and Wang, 2005). La première zone périphérique contient des aires dont les neurones préfèrent les sons bruités aux sons purs : CL, CM, RM, RTM, RTL, AL, ML (Rauschecker et al., 1995; Tian et al., 2001; Rauschecker and Tian, 2004; Tian and Rauschecker, 2004; Petkov et al., 2006; Kusmirek and Rauschecker, 2009). Enfin latéralement se situe la seconde zone périphérique, qui est subdivisée en régions rostrale RP et caudale CP. Cette zone, le parabelt, traite les caractéristiques acoustiques complexes.

Le traitement des stimuli auditifs s'étend au-delà du cortex auditif *via* des connexions du parabelt, et dans une moindre mesure du belt, avec des régions spécifiques du cortex temporal, préfrontal et pariétal (**Figure 23 haut gauche**). Notons par exemple les connexions avec les aires antérieures du lobe temporal : Ts1, Ts2, Ts3 et proisocortex du pôle temporal (Galaburda and Pandya, 1983).

La distinction en deux voies de traitement des stimuli auditifs commence dès le belt (Tian et al., 2001) et se poursuit en dehors du cortex auditif. Ce traitement dual a également été mis en évidence par IRMF et par l'étude de potentiels évoqués chez l'humain (Alain et al., 2001; Maeder et al., 2001). Notamment la reconnaissance de différents sons environnementaux active une large partie du lobe temporal (Maeder et al., 2001).

D.1.3. Un système distribué pour la perception des voix

BELIN et coll. (2000, 2002) ont mis en évidence par IRMF des aires du lobe temporal présentant une forte sélectivité pour les voix par rapports à d'autres bruits de l'environnement,

nommées *TVA – Temporal voice areas* (Belin et al., 2000; Belin et al., 2002). Ces aires sont situées bilatéralement le long du sulcus temporal supérieur, celles de l'hémisphère droit présentant des activités plus soutenues (**Figure 24**). Les auteurs ont par ailleurs démontré que ces zones s'activent spécifiquement pour le traitement perceptif des voix, et non à des caractéristiques acoustiques partagées avec d'autres sons. En effet ils ont explicitement recherché une aire dont l'activité diffère de celles de bruits blancs ayant la même enveloppe que des voix, de celle de voix brouillées, et de celle d'autres sons humains non vocaux, comme des applaudissements ou des claquements de doigts.

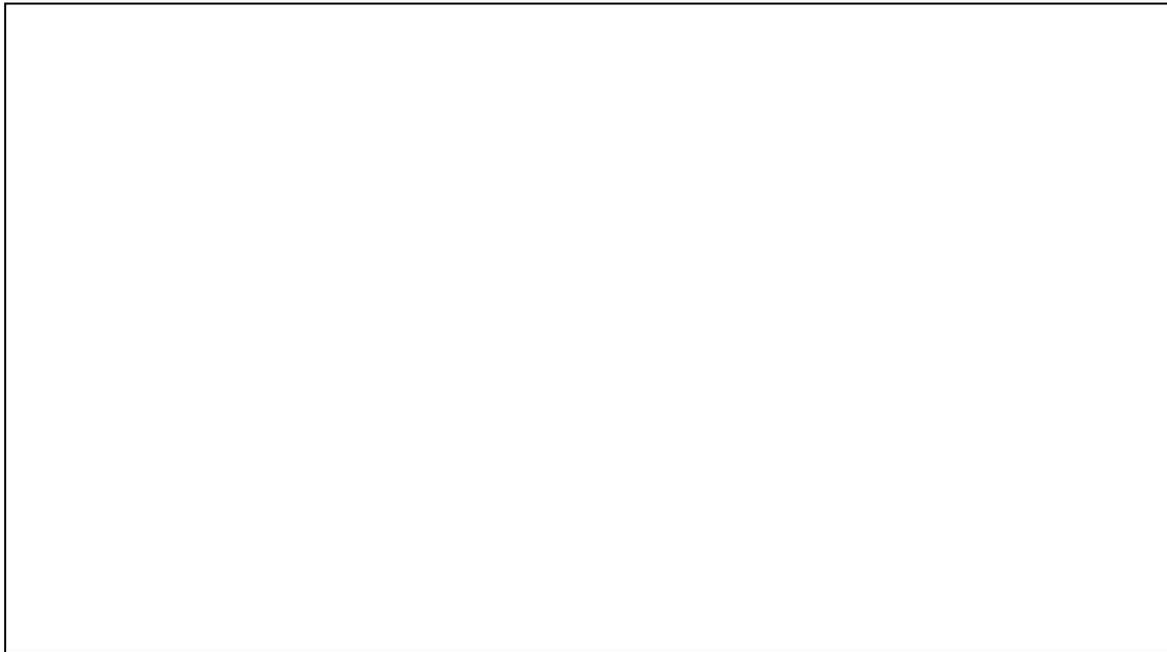


Figure 24 : Aires des voix, Temporal voice areas TVAs. Spectrogrammes et oscillogrammes de stimuli auditifs non-vocaux et vocaux, présentés dans l'étude par IRMf (haut gauche). Vues sagittales gauche et droite et transversale des aires plus activées par les voix que par l'ensemble des stimuli contrôles (bas gauche). Elles sont situées le long du sillon temporal supérieur de façon bilatérale. Vues transversale et sagittale droite de l'aire plus antérieure présentant une adaptation à différents sons d'un même locuteur (droite). D'après Belin et al., 2000 et Belin and Zatorre, 2003.

Les voix sont analysées au niveau subordonné et fin de l'identité, contrairement aux objets auditifs qui diffèrent au niveau générique de la catégorie. La distinction entre deux voix s'opère donc à un niveau inférieur à celui des catégories d'objets et ces deux traitements pourraient donc ne pas être équivalents l'un à l'autre. Pour circonvier à ce problème BELIN et coll. (2000) ont fait écouter aux sujets différents sons de cloches, analysés à un niveau subordonné similaire à celui des voix, et ont soustrait à cette activité celle obtenue lorsque les sujets écoutaient des voix humaines. Comme dans les expériences précédentes, les résultats ont montré une activation des mêmes aires. Leur activité ne serait donc pas fonction du niveau de discrimination opéré mais bien de la nature des voix elles-mêmes. Enfin une deuxième étude a mis en évidence que ces aires

s'activent plus fortement et de façon plus étendue pour des voix humaines que pour des vocalises animales (Fecteau et al., 2004). Ces études confirment et renforcent par leur fine caractérisation, d'autres observations d'activations temporelles bilatérales pour des vocalises non-linguistiques (DéMonet et al., 1992; Binder et al., 2000). L'ensemble de ces études met en évidence que les TVAs pourraient être le pendant des aires dédiées à la perception des visages et correspondre aux zones lésées chez les personnes phonagnosiques (Belin et al., 2004). Leur spécialisation pour l'analyse des voix se définirait au cours du développement entre 4 et 7 mois (Grossmann et al., 2010).

De façon importante, les auteurs ont montré que ces aires sont actives pour la voix même lorsqu'elle ne véhicule pas de langage, en utilisant des sons vocaux comme des rires, des toux, des soupirs. Les aires de perception des composants propres au signal vocal seraient donc distinctes de celles dédiées à la compréhension du langage latéralisées à gauche (Vouloumanos et al., 2001). Une expérience a explicitement testé cette distinction hémisphérique par IRMf en présentant aux sujets des voix prononçant des phrases et en leur demandant soit d'identifier le locuteur soit de reconnaître le contenu verbal. Alors que la stimulation perceptive restait constante, le fait d'orienter l'attention vers des analyses distinctes activait différemment les deux hémisphères : une activation droite était observée pour l'analyse de l'identité du locuteur et une activation gauche pour celle du contenu de la voix (von Kriegstein et al., 2003; von Kriegstein and Giraud, 2004). Il a été proposé qu'une telle distinction se fonde sur un échantillonnage auditif asymétrique des deux hémisphères, avec une fenêtre temporelle large pour les voix et fine pour le langage.

L'activation sélective d'aires temporelles droites a été confirmée par la suite en utilisant d'autres types de protocoles (Nakamura et al., 2001; Andics et al., 2010). Trois aires semblent se distinguer parmi les TVAs : une aire du sulcus temporal postérieur, une aire du sulcus temporal médio antérieur et une aire du sulcus temporal antérieur (von Kriegstein and Giraud, 2004).

Enfin des régions situées en dehors du lobe temporal pourraient également présenter une activité plus forte pour les voix, par exemple dans le cortex préfrontal (Zatorre et al., 1992; Fecteau et al., 2005; Andics et al., 2010) et l'amygdale (Fecteau et al., 2007). Ainsi une aire a été découverte dans le cortex préfrontal gauche, BA47, distincte de l'aire de Broca (Fecteau et al., 2005). De façon similaire il a été montré que l'amygdale s'active pour les vocalises humaines émotionnelles non-linguistiques (Sander and Scheich, 2001; Fecteau et al., 2007). La présence de plusieurs aires dans le lobe temporal et en dehors laisse supposer que le réseau dédié à la perception des voix puisse être distribué.

D.1.4. Des systèmes homologues chez l'humain et le primate non-humain ?

Les premières études décrivant des neurones auditifs sélectifs aux vocalisations ont été menées chez des espèces de singes du Nouveau Monde (parvordre des Platyrrhiniens) : le singe écureuil (genre *Saimiri*) et le marmouset (genre *Callithrix*). Dès les années 1970, les chercheurs se sont attelés à chercher des neurones sélectifs aux vocalises, appelés *call detectors*, et en ont trouvé quelques exemplaires non clairement définis (Wang, 2000). Les enregistrements des neurones

auditifs du marmouset ont ensuite démontré l'existence de deux classes de neurones, l'une répondant de façon sélective aux types de vocalises ou aux émetteurs, l'autre répondant à un large panel de sons incluant des vocalises et d'autres bruits (Wang et al., 1995; Wang and Kadia, 2001). Le premier groupe étant proportionnellement plus important dans les aires auditives périphériques que dans A1. De façon intéressante ces études ont mis en évidence que certains neurones sélectifs aux vocalises sont également excités par les propres vocalises de l'animal alors que d'autres sont inhibés par ce feedback tant que celui-ci concorde avec le son créé espéré (Eliades and Wang, 2008).

Les premiers enregistrements de neurones auditifs répondants aux vocalises chez le macaque, ont été menés en dehors des régions du noyau du cortex auditif. Ainsi des neurones répondant aux vocalises de congénères ont été mis en évidence dans les aires AL, ML, CL du latéral belt (Rauschecker et al., 1995; Tian et al., 2001; Russ et al., 2008). 50% des neurones de ces aires s'activaient préférentiellement pour un ou plusieurs des sept exemplaires de vocalises utilisés. Cependant ces études n'ont pas comparé la sélectivité pour les vocalises à celles pour d'autres sons complexes, rendant la conclusion de la mise en évidence de réels *call-detectors* difficile. De plus une étude subséquente a montré que les neurones de ces aires, ainsi que des aires A1, R et CM, s'activent également pour les vocalises jouées à l'envers, remettant en cause cette sélectivité (Recanzone, 2008). Ainsi l'existence d'aires sélectives aux vocalises n'a pas été clairement mise en évidence par électrophysiologie dans le lobe temporal. Par contre des neurones auditifs sélectifs aux vocalises ont été observés dans l'insula (Remedios et al., 2009), le cortex préfrontal (Romanski and Goldman-Rakic, 2002; Cohen et al., 2004; Gifford et al., 2005; Cohen et al., 2006; Cohen et al., 2007; Russ et al., 2008) et le cortex orbitofrontal (Rolls et al., 2006). Cette vue est contradictoire avec l'observation en IRMf chez l'humain de régions corticales sélectives aux voix. Cependant ce désaccord a été en partie levé grâce à l'utilisation d'IRMf chez le singe et d'enregistrements électrophysiologiques guidés par les résultats d'imagerie dans les zones sélectives aux voix (Perrodin et al., 2011).

Effectivement l'utilisation d'IRMf chez le singe a révélé l'existence d'aires cérébrales sélectives aux vocalises des congénères dans le lobe temporal (Petkov et al., 2008). Dans cette étude les vocalises étaient contrastées par d'autres sons complexes environnementaux, des vocalises d'autres animaux et des vocalises brouillées. Une première aire est située dans les régions Ts1 et Ts2 du lobe temporal, antérieurement au cortex auditif (**Figure 25**). Une seconde est située dans la partie postérieure de la région A1 du cortex auditif. La première région est également active lorsque le singe est anesthésié. Par ailleurs une troisième région était située dans le cortex pariétal postérieur. L'activation du cortex pariétal postérieur a également été mise en évidence par PET-scan, dans une étude contrastant des vocalises avec des sons non-biologiques ayant des propriétés acoustiques similaires (fréquence, taux, amplitude et durée) (Gil-da-Costa et al., 2006). Des activations bien plus étendues et moins spécifiques avaient été mises en évidence précédemment par PET-scan incluant plusieurs régions visuelles (V2, V3, V4, TEO, TE, MT). Par ailleurs en contrastant les cris d'alerte aux cris d'appel et aux sons non-biologiques, les

auteurs avaient observé l'activation du cortex préfrontal ventro-médian, de l'amygdale et de l'hippocampe (**Figure 25**) (Gil-da-Costa et al., 2004).

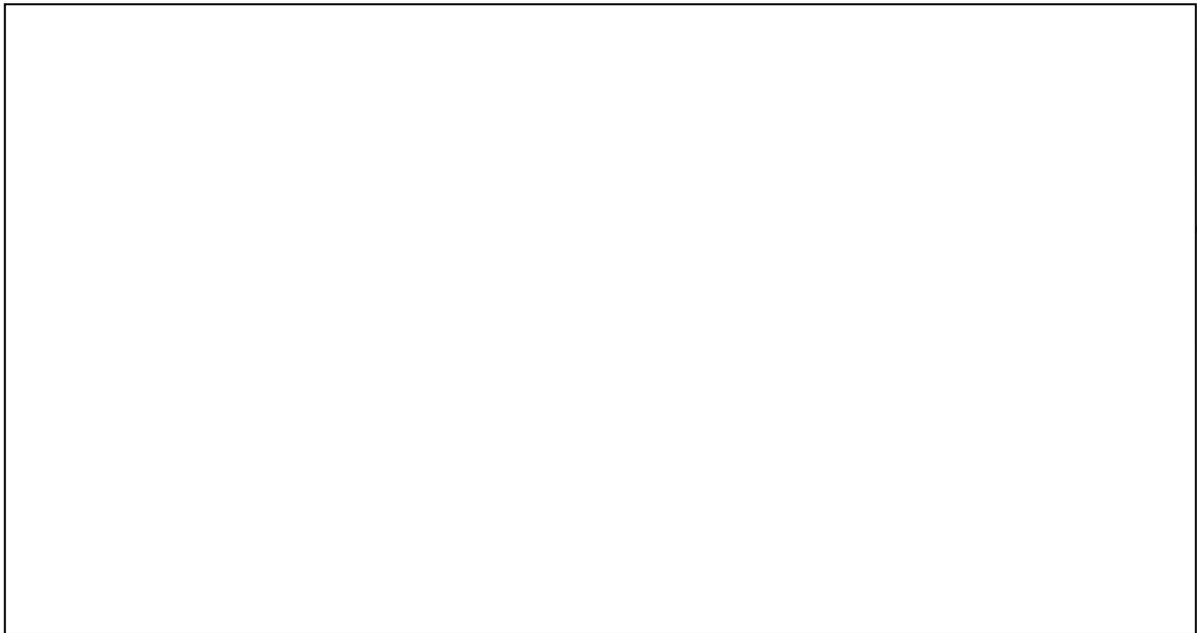


Figure 25 : Aires sélectives aux vocalises mises en évidence chez deux macaques rhésus. Le cortex auditif et les régions du lobe temporal ont été identifiés fonctionnellement ou anatomiquement chez chaque singe et sont représentés par les contours noirs sur des coupes transversales obliques suivant le plan d'inclinaison indiqué sur la figure basse (haut gauche). Les régions sélectives aux voix sont superposées sur ces cartes. Le code couleur indique que les voxels orange à rouges présentent une préférence significative pour les voix. Représentation en rouge de l'aire présentant une adaptation pour différentes vocalises d'un même singe sur un cerveau en 3D (haut droite). Représentation des aires visuelles et limbiques (TE, amygdale et hippocampe) mises en évidence par PET (bas gauche). Représentation schématique de la localisation de l'aire des voix chez le singe et de sa localisation prédite à partir des données d'IRMf humaine (bas droite). D'après Gil-da-Costa et al., 2004; Ghazanfar, 2008; Petkov et al., 2008.

Des enregistrements neuronaux effectués dans l'aire la plus antérieure activée en IRMf, ont révélé des populations de neurones s'activant préférentiellement et de façon maximale pour les vocalises de singes par rapport aux sons d'objets complexes ou aux vocalises d'autres animaux (Perrodin et al., 2011), mettant en évidence l'existence de *call detectors* dans le lobe temporal. Cependant seulement 25% des neurones auditifs de cette aire présentaient une activation deux fois plus forte pour les vocalises de congénères par rapport aux autres sons, ce qui est peu par rapport aux forts pourcentages de cellules répondant sélectivement aux visages dans les face-patches, mettant en évidence des différences dans le codage des deux types de stimuli.

Ainsi les aires du lobe temporal mises en évidence par PETKOV et coll. (2008) pourraient constituer des homologues des TVAs humaines. Par ailleurs la mise en évidence de l'activité spécifique des aires du cortex primaire en réponse aux voix serait concordante avec les études électrophysiologiques. Enfin les activations frontales semblent cohérentes avec les activations

chez l'Homme et les enregistrements électrophysiologiques chez le singe. Cependant il reste à éclaircir le fait que les zones sélectives aux voix chez l'humain et le macaque ne soient pas des homologues anatomiques (**Figure 25**). En effet l'aire simienne est plus postérieure que l'aire homologue humaine (Ghazanfar, 2008).

D.2. Encodage des voix le long de la voie ventrale

D.2.1. Distinction entre le traitement des composants de la voix et l'identification

Une distinction principale a été proposée entre les traitements réalisés par les parties postérieure et médiane des TVAs d'une part et ceux réalisés par la partie antérieure. En effet seule la partie antérieure droite a montré une adaptation à différents sons provenant d'un même locuteur, indiquant son rôle dans l'identification abstraite des voix (**Figure 24 droite**) (Belin and Zatorre, 2003). Par conséquent il a été proposé que la partie antérieure soit dédiée à l'identification et que par opposition la partie médiane soit dédiée au traitement structurel des composants acoustiques de la voix.

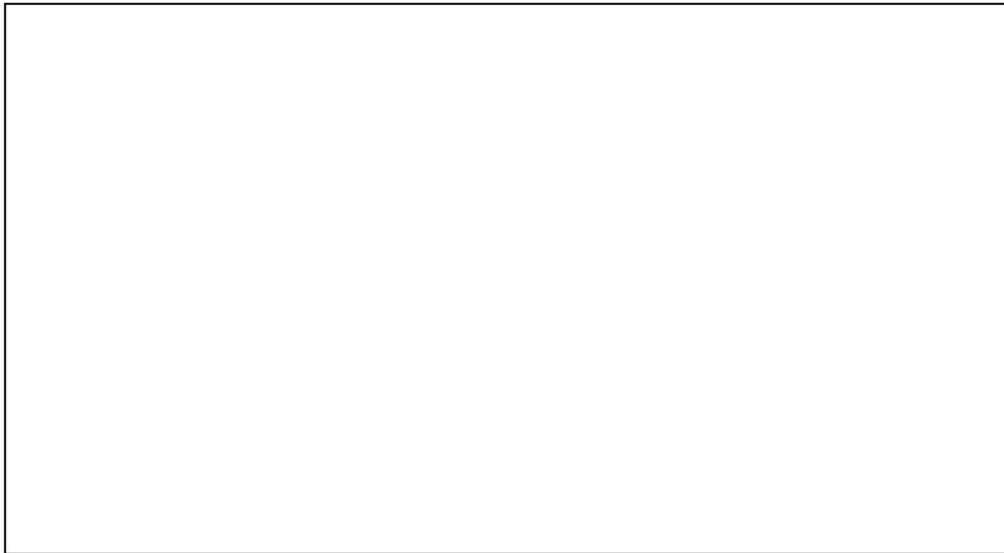


Figure 26 : Cerveau 3D montrant la répartition de l'analyse des voix par le système distribué dédié. Parmi les aires sélectives aux voix sont distinguées : les régions impliquées dans la perception des changements acoustiques (vert), les régions impliquées dans la perception du changement d'identité (jaune), les régions impliquées dans le maintien en mémoire à court terme des changements acoustiques (rose). Une séparation fonctionnelle entre les aires antérieures et postérieure du lobe antérieur se dessine. Par ailleurs une région du cortex préfrontal ventromédian semble également impliquée dans la perception des changements acoustiques et une aire du cortex orbitofrontal semble impliquée dans le maintien en mémoire à court terme des changements acoustiques. D'après (Andics et al., 2010).

Une étude similaire à celle menée par ROTSHEIN et coll. (2005) sur les visages (Rotshtein et al., 2005), mais menée cette fois-ci sur les voix a démontré que l'on peut distinguer les zones sensibles aux changements acoustiques des composants de la voix, des zones sensibles aux changements perçus de l'identité de la voix (Andics et al., 2010). Dans cette étude, les auteurs ont appris aux sujets à distinguer et à se familiariser avec une nouvelle voix. Ils leur ont ensuite présenté des voix issues d'un continuum de morphing de la voix apprise et leur ont demandé d'indiquer s'il s'agissait de la voix apprise ou non. Au sein du continuum, les voix proches de la voix apprise étaient perçues comme identiques alors qu'elles différaient acoustiquement, au contraire les voix éloignées étaient perçues comme différentes de la voix apprise. Cette expérience a mis en évidence que l'aire médiane/postérieure de STS suit la courbe de changement acoustique alors que l'aire temporale antérieure suit le changement perçu consciemment d'identité (**Figure 26**). Ainsi les deux aires du lobe temporal sont impliquées dans la catégorisation des voix mais à des niveaux d'abstraction différents.

D.2.2. Analyse des composants de la voix

La partie postérieure des TVAs traque les changements acoustiques. Des études se sont attelées à rechercher la nature des composants acoustiques dont les changements sont détectés par ces régions. VON KRIEGSTEIN et coll. (2010) ont réussi à séparer les deux éléments principaux constitutifs de l'identité de la voix, le ton de la voix créé à la source du son, au niveau des cordes vocales, et sa modulation par le filtre du tractus vocal, c'est-à-dire les formants (von Kriegstein et al., 2010). Ils ont mis en évidence que la perception des formants active la partie postérieure de STS/STG bilatéralement alors que la perception de la tonalité du son active une aire immédiatement adjacente au cortex auditif primaire. La localisation de l'aire spécifique au traitement des formants est en accord avec leur rôle à la fois dans la compréhension du langage chez l'humain et pour l'accès à l'identité et la taille du locuteur (von Kriegstein et al., 2006b; von Kriegstein et al., 2007). La localisation de l'aire spécifique à la tonalité est également très proche de celle mise en évidence chez le singe pour des tonalités non-vocales (Bendor and Wang, 2005).

D.2.3. Identification des voix

Plusieurs études corroborent la spécificité de l'aire temporale antérieure droite dans l'identification, comme mis en évidence par une étude d'adaptation du signal BOLD à l'identité du locuteur (Belin and Zatorre, 2003). Par exemple VON KRIEGSTEIN et coll. (2003) ont montré par IRMF que lorsque les sujets devaient activement discriminer l'identité du locuteur et non son discours, le lobe temporal antérieur droit était plus fortement activé (von Kriegstein et al., 2003). De même le rôle du lobe temporal antérieur dans l'identification des voix avait été mis en évidence initialement par PET-scan. (Imaizumi et al., 1997). Enfin la principale aire sélective aux voix mise en évidence chez le singe le plus antérieurement dans le lobe temporal, est également

sensible à l'identité vocale singe émetteur (Petkov et al., 2008). Ceci a été mis en évidence en utilisant trois conditions de stimuli présentés successivement : dans la première condition la même vocalise d'un individu était répétée, dans une deuxième condition des vocalises d'un même individu étaient présentées, enfin dans une troisième condition la même vocalise réalisée par des individus différents était présentée. Les auteurs ont traqué une adaptation du signal BOLD à l'identité indépendamment du type de vocalise utilisée (condition 2) et mis en évidence que la région antérieure joue un rôle important dans l'identification des voix chez le singe également (**Figure 25**). La discrimination et l'identification nécessitent une catégorisation des différentes voix. Bien qu'au niveau psychologique il ait été mis en évidence que les voix sont représentées dans un *espace des voix* multidimensionnel (Latinus and Belin, 2011), les corrélats neuronaux laissent supposer un codage des voix épars et sélectif à l'identité des voix et des types de vocalises (Perrodin et al., 2011).

D.2.4. Vers un traitement invariant de la voix

Le codage de l'invariance du locuteur se fait-il de façon hiérarchique ? Des études de décodage des signaux enregistrés en EEG et en IRMf, laisse penser que le codage abstrait de l'identité et de la signification des vocalises pourrait émerger de l'activité du réseau par réaligement d'activités oscillatoires (Formisano et al., 2008; Bonte et al., 2009). Cependant la distinction observée par IRMf entre les processus réalisés par les parties postérieure et antérieure de STS laisse supposer qu'un codage invariant de l'identité pourrait être observé à l'échelle neuronale dans le lobe temporal supérieur antérieur. En effet deux types opposés de codages invariants des vocalises ont été montrés : d'une part un codage abstrait des « significations sémantiques » des vocalises a été proposé au niveau du cortex préfrontal du singe (Gifford et al., 2005; Cohen et al., 2006; Cohen et al., 2007; Russ et al., 2007; Russ et al., 2008), d'autre part l'invariance à l'identité du locuteur a été proposée dans le cortex temporal supérieur antérieur (Ts1) (Perrodin et al., 2011).

La mise en évidence d'un traitement hiérarchique des vocalisations entre le cortex primaire et les aires temporeales antérieures pourraient expliquer la progression de l'invariance le long du lobe temporal supérieur. En enregistrant dans les aires A1, RT et l'aire Ts1 (donc proche de l'aire mise en évidence par PETKOV et coll. (2008) dans le traitement de l'identité chez le singe), KIKUCHI et coll. (2010) ont montré une évolution dans la façon dont sont représentées les vocalises par l'activité neuronale : d'une représentation fidèle aux composantes acoustiques modulés par des attributs fréquentiels et temporels spécifiques celle-ci se transforme en une représentation abstraite (Kikuchi et al., 2010). Les auteurs observent des changements dans la nature de l'activité des neurones, la sélectivité aux stimuli et la latence des réponses évoquées par les vocalises (**Figure 27 gauche**). Ainsi une invariance de l'activité neuronale aux variations intra-individuelles la voix pourrait émerger au niveau neuronal et populationnel (**Figure 27 droite**).



Figure 27 : Activités neuronales le long du cortex temporal. Les neurones A, B, C ont été enregistré dans les aires A₁, RT et Ts₁ respectivement (carte de gauche). Les activités moyennes élicitées par les vocalises ainsi que les spectrogrammes des vocalises sont présentés (milieu). L'activation pour chaque vocalise (rouge) ou non-vocalise (bleu) par rapport à la ligne de base de l'activité du neurone est représentée sous forme d'histogramme (milieu). On observe que le neurone A présente une faible latence d'activation, son activité est liée aux caractéristiques temporelles et fréquentielles du son, sa sélectivité est très faible. Les neurones B et C présentent des latences et des sélectivités croissantes et une indépendance par rapports aux caractéristiques fréquentielles des vocalises. Représentation schématique de la façon dont pourrait être codée une vocalise harmonique par un réseau neuronal hiérarchique (droite). La vocalise est représentée par son spectrogramme et son oscillogramme. Le réseau neuronal comprend des neurones spécialisés pour des composants acoustiques simples convergeant vers un neurone codant une représentation complexe et invariante. D'après Rauschecker and Scott, 2009 et Kikuchi et al., 2010.

D.3. Reconnaissance de la voix

Lorsque nous entendons la voix d'une personne connue, en plus de la discriminer des autres individus, nous reconnaissons son identité et nous réactivons l'ensemble du concept associé à cette personne. Par rapport à la perception de voix inconnues, la reconnaissance de voix familières active plus fortement un set de régions supplémentaires.

Dans une expérience menée par PET-scan des sujets devaient réaliser une tâche de discrimination de voix d'individus familiers (collègues, amis et propre voix) et d'individus non familiers (Nakamura et al., 2001). Les régions plus activées par la perception des voix familières mises en évidence par cette expérience étaient : le cortex antérieur temporal droit, le cortex entorhinal, le cortex frontal gauche ainsi que le precuneus. Par ailleurs les auteurs montrent que les activités du cortex frontal gauche et du cortex antérieur temporal droit sont corrélées positivement avec le nombre d'identifications correctes des voix familières, démontrant que le cortex temporal antérieur droit joue un rôle majeur non seulement dans la discrimination des voix mais également dans la reconnaissance individuelle à partir des voix.

Une étude subséquente par IRMf a démontré des activations plus importantes pour les voix personnellement connues par rapport aux voix inconnues dans des régions similaires : le précuneus, le cortex parahippocampique et le cortex frontal (von Kriegstein and Giraud, 2004). De façon intéressante cette étude met également en évidence l'activité augmentée dans deux régions visuelles : le gyrus fusiforme et le cortex occipital. Par opposition les voix inconnues activaient plus fortement les régions du STS postérieur, liées à l'analyse acoustique de la voix, et l'amygdale, suggérant une augmentation du niveau de vigilance comme dans le cas des visages inconnus. Bien que dans cette seconde étude le cortex temporel supérieure antérieure TVAa ne présente pas d'augmentation d'activité pour les voix familières, l'analyse des connectivités fonctionnelles a mis en évidence qu'il interagit fonctionnellement avec le complexe hippocampique droit et le cortex frontal médian supérieur lors de la reconnaissance de locuteurs familiers (**Figure 28**).

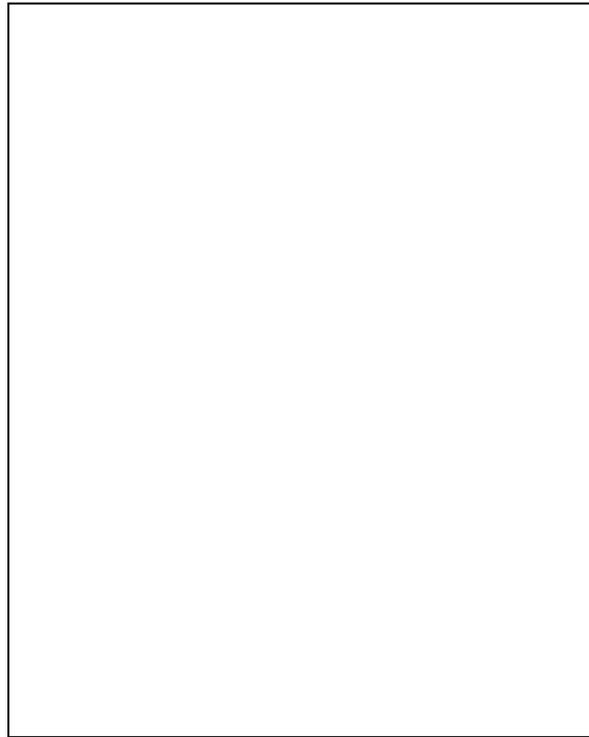


Figure 28 : Connectivité fonctionnelle des trois TVAs (A: STS antérieure, M: STS médiane/antérieure et P: STS postérieure) pendant la reconnaissance des voix par rapport au contenu linguistique (vert), et dans les conditions où le locuteur était personnellement connu (rouge) ou non (bleu). D'après von Kriegstein and Giraud, 2004.

Ainsi alors que la discrimination de voix inconnues requière une analyse plus importante des composantes acoustiques impliquant les régions postérieures du STS, la perception de voix familières active plus fortement TVAa et un ensemble de régions similaires à celles mises en évidence pour la reconnaissance de visages familiers et impliquées plus largement dans le rappel autobiographique, le sentiment de familiarité ou encore la théorie de l'esprit. Enfin l'activation de régions visuelles lorsque des voix connues sont perçues montre une interaction entre ces deux

système et appelle de nombreuses questions sur la signification de cette activation : serait-elle liée au rappel du visage correspondant, à une imagerie mentale, à une préparation du réseau en vue du rappel de l'identité ultérieure ?

D.4. Conclusion

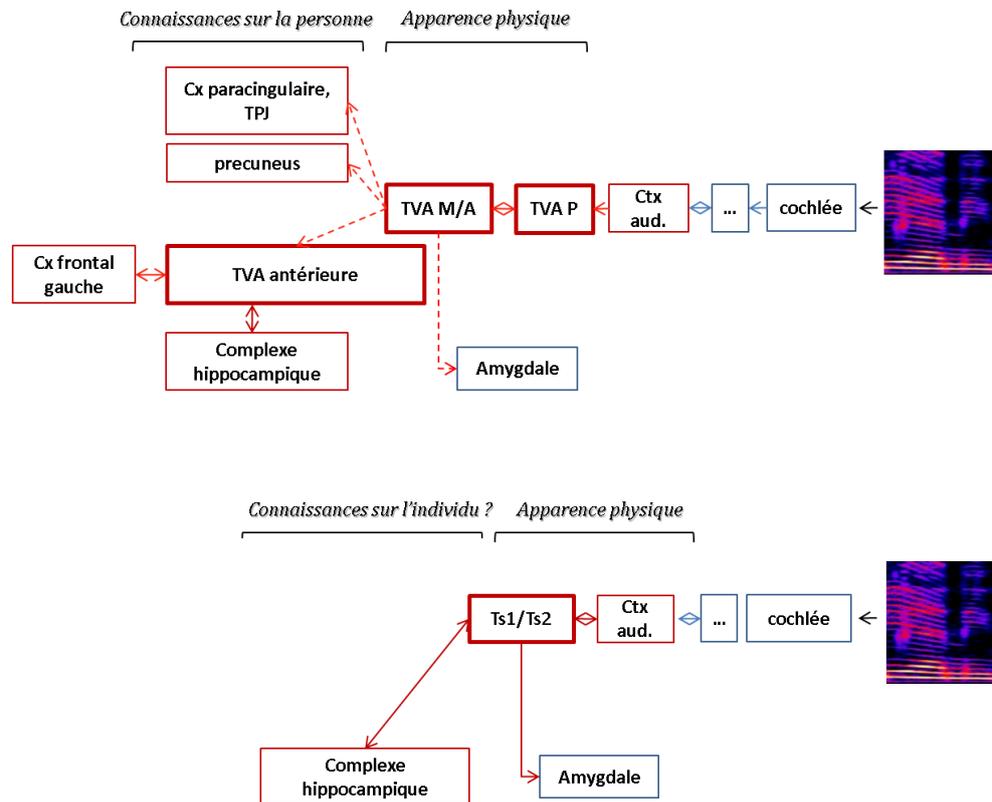


Figure 29 : Schémas de l'identification des voix chez l'Homme et le singe, mettant l'accent sur les similitudes avec le système d'identification des visages. Le rôle de la TVA antérieure a été distingué de celui des deux autres TVAs postérieures et médianes, et présentés comme étant autant un analogue possible de all que de FFA pour l'identification des voix familières. Les aires TVA P et TVA M/A seraient des équivalents du système central, liées à l'analyse de l'apparence visuelle. Les structures impliquées dans la familiarité (Amygdale et precuneus), le rappel des informations biographiques (TVA antérieure), épisodiques (complexe hippocampique) et des traits de personnalité (cx paracingulaire, TPJ) ainsi qu'une aire possible du rappel du nom (Cortex frontal gauche) sont indiquées.

La perception et la reconnaissance des voix engagent de nombreuses aires cérébrales corticales et sous-corticales distribuées à travers le cerveau. Ces traitements suivent un continuum le long du lobe temporal, de façon similaire à celle décrite pour l'analyse des visages (Figure 29). Dans un premier temps, les deux voies sont cependant ségréguées : l'analyse des voix engageant des régions situées plus postérieurement au niveau du lobe temporal que celle dédiées à l'analyse des visages. Durant le trajet le passage d'une représentation physique à une représentation conceptuelle de l'identité est également progressivement réalisé.

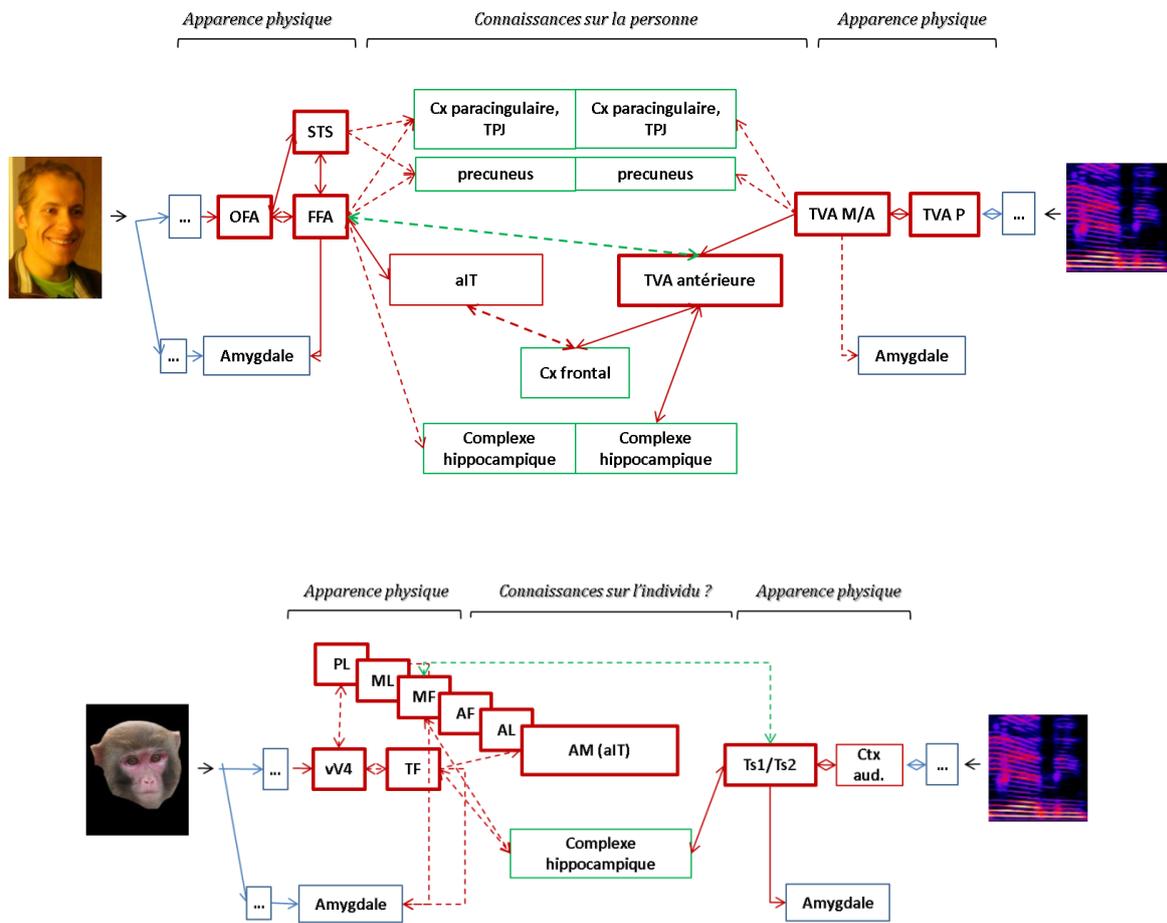


Figure 30: Schémas des possibles intégrations (vert) entre l'identité des voix et des visages, au vu des aires activées par les deux systèmes et des hypothèses sur les activations directes entre les deux systèmes.

Bien que des inconnues demeurent en ce qui concerne la perception et la reconnaissance des voix au niveau neuronal, de nombreux parallèles entre les deux systèmes apparaissent clairement. De plus plusieurs zones similaires sont engagées dans les deux processus (**Figure 30**). En effet le cortex paracingulaire, la jonction temporo-pariétale, le precuneus, le cortex temporel antérieur, le cortex frontal et enfin le complexe hippocampique semblent impliqués dans la reconnaissance individuelle qu'elle soit initiée par la vue d'un visage ou par l'entente d'une voix personnellement connus. Cependant ces aires sont très larges et les apports de chaque stimulus pourraient être distincts en leur sein. Par ailleurs les études chez l'humain et le singe précédemment citées ont mis en évidence que les vocalises peuvent activer des aires du gyrus fusiforme et du cortex occipital proches de celles impliquées dans la perception des visages, suggérant que des liens directs entre les deux systèmes pourraient également émerger par leur activation synchrone. Ainsi des études supplémentaires manipulant conjointement ou dans un même protocole ces deux stimuli pourraient éclaircir l'apport de ces différentes zones à la reconnaissance individuelle.

E. Reconnaissance individuelle et émergence du concept de Personne

Une combinaison unique des différents attributs (visage, voix, silhouette, personnalité, etc.) de l'individu doit être représentée par le cerveau, alors que chacun est traité par une ou plusieurs modalités distinctes. L'ensemble de ces attributs a été encodé en mémoire en un concept unique, grâce à la colocalisation et la cotemporalité de leurs survenues répétitives. Une fois encodé nous pouvons ré-expérimenter ce concept. Toutefois tous les attributs de l'individu peuvent ne pas être présents en même temps (par exemple au téléphone ou sur des photos) et sa reconnaissance dépend alors de la complétion d'un pattern incomplet avec les connaissances antérieures. Notre expertise pour certains de ces attributs, principalement les voix et les visages, est tellement importante qu'à eux seuls ils peuvent permettre le rappel du concept complet.

E.1. Une zone d'intégration multisensorielle de l'identité ?

Les deux systèmes de traitement perceptif des voix et des visages avancent en parallèle dans le cortex temporal. Bien que ségréguées anatomiquement, les voies d'analyse des voix et des visages interagissent pendant la reconnaissance individuelle. En conséquence les modèles psychologiques de reconnaissance de l'identité incluent la possibilité de connections réciproques entre les deux voies. Le modèle influant de BRUCE et YOUNG, adapté et augmenté par ELLIS et coll. (§ **A.2.**, **Figure 1**), prévoit l'existence d'une zone de convergence supramodale, la *Person Identity Node*, liant les deux systèmes. Inversement d'autres modèles suggèrent que les voies spécifiques unimodales pourraient être connectées fonctionnellement directement et de façon réciproque (**Figure 32 droite**). Au niveau neuronal plusieurs modèles ont été proposés pour expliquer comment les populations cellulaires distantes codent une représentation conceptuelle unique et explicite. Deux possibilités principales se distinguent : l'existence de *neurones gnostiques* situés au sommet de hiérarchies pyramidales à complexité croissante et l'existence d'*assemblées neuronales* distribuées et connectées anatomiquement ou fonctionnellement (**Annexe C**). Dans le cas d'un codage par unités gnostiques, la réactivation de l'ensemble des attributs du concept est directement réalisée dès que l'unité gnostique est activée. Dans le cas d'un codage distribué deux possibilités non mutuellement exclusives émergent pour le rappel de l'ensemble des composants de l'individu à partir d'un seul indice (**Annexe D**) : soit la perception de l'un des attributs permet

à elle seule de réactiver l'ensemble de l'assemblée neuronale, soit la réactivation globale est réalisée par une zone de *convergence-divergence* supplémentaire, appelée aussi *concentrateur* (ou *hub*). Dans ce cas la perception de l'attribut (par exemple le visage) activerait le hub, qui à son tour permettra la réactivation de l'ensemble du réseau en *feedback* ou *top-down*. Pour la reconnaissance des individus une telle zone serait le *Person identity node* défini psychologiquement. Ainsi deux questions complémentaires se posent. Les aires unimodales interagissent-elles lors de la reconnaissance individuelle ? Le *Person identity node* est-il nécessaire pour le rappel de l'identité de la voix à partir de l'identité du visage et vice-versa ?

E.1.1. Activation cross-modale des aires unimodales

Après avoir observé une activité cross-modale dans la région *visuelle* du gyrus fusiforme lors de la perception de *voix* d'individus personnellement connus (§ D.3.), VON KRIEGSTEIN et coll. (2005) se sont attelés à caractériser précisément cette aire pour comprendre s'il s'agirait de l'aire FFA impliquée dans la perception des visages. Ainsi dans une étude ultérieure ils ont de nouveau utilisé un protocole d'identification de voix familières contrasté par l'identification de voix non-familières et l'identification de contenu linguistique (von Kriegstein et al., 2005). Comme dans l'étude précédente, plusieurs aires impliquées dans l'identification de voix familières ont été observées : les cortex temporaux antérieurs, le complexe hippocampique, TPJ, le precuneus et le cortex fusiforme. Par la suite les auteurs ont caractérisé l'aire FFA séparément, en recherchant spécifiquement une aire plus activée par la vue passive de visages que d'objets, et ont observé que cette aire se superpose effectivement à la région du gyrus fusiforme activée par les voix familières (**Figure 31**). Ainsi la colocalisation de ces activations visuelles et auditives dans FFA met en évidence le recrutement cross-modal de FFA lors de l'identification de voix familières. Cette activation cross-modale rappelle les activations d'aires visuelles, dont TE, observées par GIL-DA-COSTA et coll. (2004) par PET-scan, lorsque les macaques rhésus entendaient des vocalises 'coo' ou des cris d'alerte (**Figure 25**, Gil-da-Costa et al., 2004).

Quelles sont les conditions nécessaires et suffisantes pour une activation de la FFA par des voix familières ? Dans une seconde étude, VON KRIEGSTEIN et GIRAUD ont fait apprendre aux participants diverses associations voix-visages ou voix-noms. Après avoir appris ces associations les sujets devaient à nouveau identifier les voix de façon unimodale. L'activation de la FFA pendant l'identification des voix a été observée uniquement pour les voix ayant été apprises en association avec des visages (von Kriegstein and Giraud, 2006). Ainsi la présence d'un visage pendant l'apprentissage de la voix est une condition nécessaire à l'activation ultérieure cross-modale de la FFA. Ces mêmes expériences ont été menées chez des personnes prosopagnosiques congénitales (von Kriegstein et al., 2006a; von Kriegstein et al., 2008). Bien que ne reconnaissant pas les visages, ces personnes présentent une activité plus importante dans l'aire FFA droite et les régions temporales antérieures à la vue de visages qu'à la vue d'objets (**Figure 31**). Bien entendu ces activations sont nettement inférieures à celles des sujets contrôles. Lors du rappel de voix

familiales (von Kriegstein et al., 2006a) ou de voix préalablement associées à des visages (von Kriegstein et al., 2008) une activation de FFA est observée. Ces études démontrent que le visage présent pendant l'apprentissage ne doit pas forcément être reconnu pour induire une réponse cross-modale de FFA. Ainsi ces deux résultats mettent en évidence que la présence d'un visage, même s'il n'est pas reconnu, pendant l'apprentissage de la voix est une condition nécessaire et suffisante à l'activation cross-modale subséquente de FFA.



Figure 31 : Colocalisation de FFA lors de la vue de visages ou de l'écoute de voix familières. Activation cross-modale en réponse aux voix familières vs non-familiales (rouge), activation visuelle en réponse aux visages vs objets (bleu). A droite sont représentées les activations de FFA par les voix (rouge) et les visages (violet) chez une personne prosopagnosique. D'après von Kriegstein et al., 2005 et von Kriegstein et al., 2006a.

A l'inverse une activation cross-modale des aires auditives par des visages a-t-elle été observée ? Plusieurs études ont mis en évidence une augmentation d'activité dans les cortex auditifs à la vue de visages dynamiques, par exemple la lecture sur les lèvres sans voix audible active le cortex auditif (Calvert et al., 1997; Calvert and Campbell, 2003). Cependant ces activations ne sont pas spécifiques aux visages familiers et ne sont pas liées à l'identification de la personne, mais à la compréhension de la communication (**Annexe E**). Deux études symétriques à celle de KRIEGSTEIN et coll. ont analysé les régions engagées dans la reconnaissance de personnes nouvellement apprises à partir des visages (Paller et al., 2000; Paller et al., 2003). L'apprentissage consistait en une présentation seule de visage ou d'une présentation conjointe de photos de visages et d'une voix annonçant un nom et une occupation. La reconnaissance des visages appris avec d'autres informations engageait la région hippocampique gauche et l'insula gauche. Elle impliquait également le gyrus temporal supérieur et médian (BA 21/22), impliqué dans l'analyse de la parole. Cependant l'activation était présente du côté gauche et non du côté droit. Ainsi l'existence d'activations cross-modales visage-vers-voix pour l'identité analogue à celles décrites ci-dessus reste à démontrer. Cependant l'association voix-visage n'est pas symétrique pour la reconnaissance individuelle (§ **A.5.4.** et § **A.6.**), les visages étant la modalité dominante pour l'apprentissage et le rappel de l'identité. Ainsi les voix bénéficient plus fortement de l'apprentissage multisensoriel pour le rappel ultérieur que ne le font les visages. Une activation

cross-modale symétrique de celle mis en évidence par KRIEGSTEIN et coll. pourrait être plus faible ou ne pas exister.

E.1.2. Interactions des aires unimodales FFA et TVA antérieure pendant la perception des voix

Les auteurs se sont ensuite demandé si l'activation cross-modale de FFA provenait d'une activation directe par TVA ou si elle provenait d'une zone relai par un mécanisme top-down. En analysant la connectivité fonctionnelle entre FFA et l'aire TVA antérieure lors de l'identification de voix d'individus personnellement connus, les auteurs ont révélé un couplage direct entre les aires TVA et FFA (von Kriegstein et al., 2005). Ce couplage était également présent après un rapide apprentissage d'associations entre les voix et les visages, indiquant qu'une connaissance personnelle des identités vocales n'est pas nécessaire à la mise en place de ce couplage (von Kriegstein and Giraud, 2006). Enfin ce couplage était également présent chez les personnes prosopagnosiques, bien qu'elles ne reconnaissent pas les visages (von Kriegstein et al., 2006a; von Kriegstein et al., 2008). Par contre dans aucun des cas un couplage direct n'était observé lorsque les voix étaient inconnues, ou n'avaient pas été au préalable associées à des visages ou encore avait été associées avec d'autres éléments comme le nom ou l'occupation. Enfin l'étude de la connexion fonctionnelle de FFA avec les autres régions activées par l'identification de voix connues ou associées à des visages (IPJ, precuneus, cortex temporal inférieur, cortex préfrontal) mit en évidence que FFA n'était pas connectées à ces aires pendant l'identification des voix. Les auteurs en ont conclu que le couplage entre TVA et FFA observé dans l'ensemble des ces études ne proviendrait pas d'une influence *top-down* d'une aire supplémentaire mais serait *direct*. Récemment cette prédiction a été confirmée par l'étude de la connectivité anatomique entre FFA et les trois régions auditives (a, m et pTVA), révélant des connexions directes entre FFA et les régions auditives antérieure et médiane ainsi que dans une moindre mesure postérieure (**Figure 32**).

Notons cependant qu'il aurait été intéressant d'étudier le couplage de ces deux aires avec l'amygdale ou d'autres régions sous-corticales véhiculant un sentiment de familiarité pour les visages, préservé notamment chez les personnes prosopagnosiques (de Gelder et al., 2003). En effet la connexion entre les deux aires unimodales est établie chez les personnes prosopagnosiques uniquement pour les individus familiers ou dont la voix et le visage ont été associés par apprentissage, ce qui laisse entrevoir que l'aire sous-tendant la représentation de la familiarité des individus puisse moduler à la fois TVA et FFA. Par ailleurs les régions du complexe hippocampique et du cortex temporal antérieur sont également impliquées dans la reconnaissance individuelle, comme mis en évidence dans ces études, cependant leur lien avec FFA n'a pas été testé ici. Il a été testé et s'est révélé positif pour l'hippocampe dans une autre étude présenté dans la suite (§ E.1.4., Joassin et al., 2011).



Figure 32 : Connexions anatomiques entre les régions spécifiques aux voix (pSTS vert, mSTS rouge, aSTS bleu) et aux visages (vFFA jaune) mises en évidence par tracking de fibres (gauche). Les connexions sont représentées de la même couleur que les régions de départ des fibres. Les connexions principales sont entre aSTS et FFA et mSTS et FFA. Les deux modèles de reconnaissance des individus (droite) : l'intégration a lieu au niveau supramodal du PIN ou une intégration a également lieu précocement par connexions réciproques entre les régions unimodales. D'après (Blank et al., 2011).

E.1.3. Corrélats fonctionnels et psychologiques du couplage précoce

L'aire TVA antérieure semble engagée dans deux types d'interactions : l'une avec les régions impliquées dans le rappel de l'identité de la personne et l'autre directement avec FFA. Le couplage direct avec FFA réalisé en présence de stimuli familiers pourrait constituer une étape préliminaire au Person identity node. Quels sont les corrélats fonctionnels et psychologiques d'un tel couplage précoce ?

Plus précisément le couplage suffit-il à rappeler en mémoire le visage correspondant à la voix entendue ? Le couplage serait-il alors associé à une imagerie mentale du visage familier ? En effet des études antérieures avaient mis en évidence que FFA était active lors d'imagerie mentale des visages sans support perceptif (Ishai et al., 2000; Ishai et al., 2002; Mechelli et al., 2004) ; il est donc intéressant de se demander si une telle réactivation mentale peut survenir par le couplage entre FFA à TVA. Un premier élément de réponse est apporté par les résultats obtenus chez les personnes prosopagniques. En effet dans leur cas, l'activation de FFA par les voix et son couplage direct à TVA ont bien lieu alors que ces personnes sont incapables d'identifier leurs amis par leur visage, de décrire un visage de mémoire ou d'imaginer un visage. Ceci suggère que le couplage entre TVA et FFA est dissocié de l'habileté à évoquer mentalement un visage. Un autre élément de réponse est apporté par l'étude où les associations voix-visages étaient apprises. Avant l'apprentissage les sujets étaient testés sur leur discrimination de voix inconnues. Bien que les voix aient été inconnues, les sujets ont déclaré dans un questionnaire avoir formé

automatiquement des prédictions mentales sur la taille, le genre et l'apparence physique des locuteurs activant une imagerie visuelle. FFA était alors fonctionnellement connectée à des régions des cortex pariétaux et préfrontaux précédemment mises en évidence dans l'évocation explicite d'images mentales, mais non à TVA. Le cortex préfrontal serait nécessaire à l'initiation de l'imagerie visuelle mentale et à l'activation des aires occipitales-temporales telle FFA, et le cortex pariétal contribuerait au maintien de l'imagerie visuelle mentale. Cependant ces deux aires n'étaient plus activées et connectées à FFA après l'apprentissage de l'association voix-visage, contrairement à TVA qui avait alors établie une liaison avec FFA. Ceci indique que l'établissement de la liaison entre TVA et FFA n'est pas suffisant et n'est pas nécessaire à l'imagerie mentale ou au rappel du visage à partir de la voix.

Le couplage permet-il d'optimiser l'accès aux étapes ultérieures de reconnaissance individuelle, similairement à une stimulation multimodale (§ A.5.) ? Effectivement les auteurs de ces études ont mis en évidence que l'accroissement de la connectivité entre les deux modules unisensoriels était accompagné d'une facilitation à identifier la voix du locuteur. Ainsi le fait que non seulement TVA mais également FFA (par l'intermédiaire de TVA) se projettent en feedforward sur des régions ultérieures du traitement de l'identité pourrait augmenter la vitesse et la précision de la reconnaissance individuelle. Au niveau comportemental l'apprentissage conjoint d'une voix avec le visage associé augmente considérablement la quantité d'information disponible pour la reconnaissance individuelle. Comme la reconnaissance individuelle à partir des voix est entachée d'une grande incertitude inhérente au signal auditif (§ A.6.), elle bénéficie amplement des signaux visuels gardés en mémoire pour l'identification. Par ailleurs cette optimisation semble spécifique à l'association avec un visage, qui véhicule des signaux fiables et optimisés pour l'identification. En effet la facilitation de la reconnaissance individuelle était réduite dans le cas des sujets ayant appris des associations voix-noms et inexistante chez les sujets prosopagnosiques, qui présentaient des difficultés à identifier les voix connues pourtant associées par expérience avec divers attributs individuels autres que les visages.

En conclusion le couplage direct entre TVA et FFA pourrait constituer une étape préliminaire au *Person identity node*. En effet le couplage n'aurait pas pour conséquence le rappel direct du visage à partir de la voix mais plutôt une préparation du réseau en vue du rappel de l'identité et de la reconnaissance individuelle. VON KRIEGSTEIN et al proposent qu'il soit un signal de familiarité facilitant le passage à une étape ultérieure. Ainsi la question de l'existence d'un *Person Identity Node* ultérieur reste entière. L'interaction fonctionnelle de TVA avec plusieurs aires impliquées dans le rappel de l'identité (TPJ, precuneus, cortex temporal inférieur, cortex préfrontal, complexe hippocampique), indique que plusieurs candidats au Person Identity Node existent.

E.2. Les candidats aux Persons' Identity Nodes (PINs)

Des zones similaires émergent lors de la phase finale de reconnaissance à partir des voix et des visages (**Figure 30**), cependant les expériences présentées précédemment n'incluaient pas

dans un même protocole la reconnaissance des voix et des visages rendant une réponse difficile quant à l'existence, l'unicité et la localisation du PINs. Une zone de convergence peut être définie de deux façons (Calvert, 2001; Stein et al., 2010). Une première définition suppose que la zone de convergence réalise une *intégration multisensorielle* de l'information, caractérisée par une superadditivité des activités auditives et visuelles [$AV > A+V$]. Dans ce cas, la présence des deux modalités simultanément apporte un gain d'information supérieur à celui apporté par la présence de l'une ou l'autre des modalités ou encore à la somme des activités unimodales. Ce type d'intégration caractérise principalement les associations amodales, par exemple celles ayant cours entre les voix et visages pendant la communication (**Annexe C**). Une seconde définition suppose que la zone de convergence présente une *invariance cross-modale* à l'information, caractérisée par une égalité des activités auditives et visuelles [$A=V$]. Dans ce cas chaque modalité est retranscrite de façon abstraite et représente donc la même information. Ce type d'invariance caractériserait principalement les associations arbitraires. Quatre études en imagerie ont testé dans un même protocole l'activité induite par les identités vocales et faciales. Une première étude a analysé les corrélats neuraux de la mémoire de travail pendant le rappel d'identités faciales ou vocales inconnues (Rämä and Courtney, 2005). Cette étude a montré l'existence d'aires proches mais dissociées du lobe ventral préfrontal impliquées dans la mémoire de travail à partir des deux modalités, n'apportant pas d'indice sur la localisation du PIN.

E.2.1. Le precuneus : aire cross-modale de la familiarité

Une seconde étude a mené une analyse de conjonction afin de mettre en évidence les régions cérébrales présentant des augmentations d'activité similaires pour les voix et les visages d'individus personnellement connus (parents, enfants, amis, partenaire, collègues) par rapport à ceux d'individus inconnus (Shah et al., 2001). Dans cette étude était testée la familiarité (*Know*) plutôt que la reconnaissance individuelle (*Remember*) car l'identification explicite n'était pas exigée des participants. Les auteurs ont mis en évidence l'implication du cortex cingulaire postérieur et du cortex retrosplénial dans la familiarité avec les voix et les visages. Cette zone serait donc engagée de façon cross-modale dans la représentation invariante de la familiarité des individus. L'activation du cortex cingulaire/precuneus et d'aires limbiques proches (cortex pariétal inférieur, TPJ) a été observée précédemment dans des études sur le sentiment de familiarité évoqué par la perception de voix (Nakamura et al., 2001; von Kriegstein and Giraud, 2004) et de visages (Gobbini and Haxby, 2007) familiers séparément. L'étude confirme donc que les régions précédemment mises en évidence sont colocalisées au sein du precuneus. Une voie dorsale d'analyse de la familiarité incluant le cortex pariétal inférieur et le gyrus cingulaire pourrait donc être engagée quelque soit la modalité de présentation des stimuli.

E.2.2. L'hippocampe intègre les voix et les visages

Dans une troisième étude la reconnaissance individuelle (*Remember*) était cette fois-ci testée après l'apprentissage d'association voix-visage-nom (Joassin et al., 2011). Suite à l'apprentissage, la voix, le visage ou les deux étaient présentés à l'écran et le sujet avait pour tâche de les dénommer. Une analyse de la superadditivité des activités auditives et visuelles [AV>A+V] démontra l'engagement de deux aires unimodales (le gyrus fusiforme et le sulcus temporal supérieur) et de l'hippocampe droit dans l'intégration multisensorielle des voix et des visages. L'étude confirme donc que les régions du complexe hippocampique précédemment observées pendant la reconnaissance individuelle à partir de visages (Leveroni et al., 2000; Haist et al., 2001; Paller et al., 2003; Bernard et al., 2004; Rotshtein et al., 2005; Denkova et al., 2006; Ishai and Yago, 2006) ou des voix (Nakamura et al., 2001; von Kriegstein and Giraud, 2004) pourraient être colocalisées au sein de l'hippocampe. La mise en évidence de neurones ayant une activité invariante au visage et au nom des individus connus abonde également dans ce sens (Quian Quiroga et al., 2009), bien que les auteurs n'aient pas testé l'invariance à la voix. L'analyse des connectivités fonctionnelles confirma l'observation de KRIEGSTEIN et coll. d'une connexion directe entre FFA et TVA et révéla que FFA et STS sont également connectés bidirectionnellement à l'hippocampe. Le renforcement de la connectivité de façon bidirectionnelle entre l'hippocampe et les aires unimodales lui confère un rôle dans l'intégration des voix et visages en une représentation multimodale pour le rappel d'autres informations.

Une étude similaire s'est intéressée à l'engagement de l'hippocampe dans l'encodage et le rappel d'associations voix-visage (Holdstock et al., 2010). Les auteurs ont montré que l'hippocampe est impliqué dans l'encodage de cette association mais, contrairement à l'étude précédente, non dans son rappel. Cependant des différences existent dans le protocole des deux études qui pourraient expliquer ce désaccord. En effet dans cette deuxième étude les sujets doivent indiquer si l'association voix-visage est correcte ou fautive (match/non-match), alors que dans l'étude précédente les sujets doivent associer la voix ou le visage ou les deux avec le nom gardé en mémoire (match-to-sample). Or ces deux types de reconnaissance cross-modale d'identité sont appris à des âges différents par l'enfant (§ A.5.2) et pourraient être dissociés sur le plan psychologique et aussi sur le plan neuronal. L'hippocampe peut être nécessaire pour retrouver un élément à partir de l'autre et non à tester si une association est juste ou non.

E.2.3. Le pôle temporal droit : un accès aux informations biographiques

Enfin le dernier candidat au PIN est le pôle temporal droit, s'étendant jusqu'à aIT et au gyrus frontal inférieur. Bien que celui-ci n'ait pas été mis en évidence par des études en imagerie testant l'invariance cross-modale ou l'intégration multisensorielle, des modèles influents basés sur des études lésionnelles ont proposé qu'il soit une zone de convergence où les composants de la représentation d'un concept seraient liés (Damasio, 1989; Martin and Chao, 2001; Patterson et al., 2007; Meyer et al., 2010). Ces modèles sont cependant basés sur des études de personnes lésées testées uniquement dans la modalité visuelle. Par exemple DAMASIO et coll. (2004), en testant une cohorte de 186 sujets lésés et en réalisant des expériences par PET chez des sujets

contrôles, ont mis en évidence que le rappel d'informations biographiques à partir de la vue d'un visage connu est détériorée lors de lésions du pôle temporal droit alors que la dénomination des visages implique le pôle temporal gauche et le gyrus frontal inférieur gauche (Damasio et al., 2004). Au niveau cellulaire, une étude a également mis en évidence que les neurones de aITv peuvent représenter sélectivement à la fois une identité faciale présentée sous différents points de vues et une image abstraite qui lui a été précédemment associée, pointant un rôle de aITv dans le rappel biographique (Eifuku et al., 2010). Une étude de lésion du pôle temporal droit due à une démence sémantique abonde dans le sens de ce modèle en mettant en évidence un type particulier d'agnosie, *l'agnosie des personnes*, qui se caractérise par une incapacité à se rappeler les informations relatives à une personne à partir de son visage, puis de sa voix et de son nom (Gainotti et al., 2003). Ce type d'agnosie avait déjà été observé sans que la localité de la lésion ne soit caractérisée (Ellis et al., 1989; Neuner and Schweinberger, 2000). Cependant les dégénérescences progressives du pôle temporal droit peuvent également donner lieu à des formes pures de prosopagnosie ou phonagnosie associative questionnant la délimitation des zones lésées causant les syndromes d'agnosie des personnes (Hailstone et al., 2010). En effet la lésion pourrait être suffisamment étendue pour englober à la fois aIT et aTVA située postérieurement et dans ce cas le pôle temporal ne constituerait pas une zone de convergence.

Pour conclure, les aires des systèmes centraux dédiées à l'analyse de l'apparence physique des voix (aTVA) et des visages (FFA) sont directement connectées entre elles. Par ailleurs d'autres aires semblent associer les voix et les visages connus à d'autres niveaux. D'une part le sentiment de familiarité est perçu de façon cross-modale et impliquerait le cortex cingulaire/precuneus et les aires limbiques proches (cortex pariétal inférieur, TPJ) ; cette familiarité pourrait être entendue au sens large incluant la reconnaissance de la personnalité des individus, comme proposé par GOBBINI et coll. (2006) D'autre part la reconnaissance individuelle implique deux autres régions du lobe temporal : l'hippocampe qui intégrerait les informations multisensorielles d'identité et le système créé par les pôles temporaux antérieurs droits et gauche et le gyrus frontal inférieur. La contribution de ces deux dernières zones pourrait être fonction de leur rôle plus général dans la mémoire épisodique et sémantique. Nous allons présenter l'hippocampe et son rôle dans la mémoire épisodique afin de comprendre comment les informations sociales comme les voix et les visages pourraient être représentées dans cette aire au niveau neuronal.

E.3. L'hippocampe, une structure associative impliquée dans la mémoire autobiographique

E.3.1. Une structure archaïque complexe au carrefour du lobe temporal

L'hippocampe est situé à l'extrémité du cortex temporal médian, à l'endroit où celui-ci s'enroule sur lui-même. C'est une structure plus ancienne évolutivement que le néocortex, appelée *archécortex*, présente et semblable chez tous les mammifères. Elle est composée d'une

seule couche dense de neurones pyramidaux, contre six dans le neocortex, dont les particularités morphologiques permettent de distinguer plusieurs *champs ammoniens* (CA1, CA2, CA3, CA4). A ceux-ci s'ajoutent une couche centrale de neurones excitateurs granulaires appelée *gyrus denté* et le *subiculum*, une structure à la jonction entre le néocortex et l'hippocampe (Ramón y Cajal, 1899-1904). L'hippocampe est un circuit tri-synaptique en forme de boucle, fortement récurrent, dont la morphologie laisse entrevoir qu'il puisse maintenir et associer les informations en son sein. Les informations sont transmises du cortex entorhinal vers le gyrus denté par les *fibres perforantes*, puis du gyrus denté vers CA3 par les *fibres moussues*, enfin de CA3 vers CA1 par les *collatérales de Schaffer*. La récurrence du circuit provient d'une part du fait que l'information des fibres perforantes arrive simultanément à différents endroits de la boucle (au gyrus denté mais également à CA3, CA1 et au subiculum) et d'autre part car l'information est maintenue et intégrée à plusieurs reprises par des collatérales récurrentes dans CA3 principalement et dans CA1 (**Figure 33**).

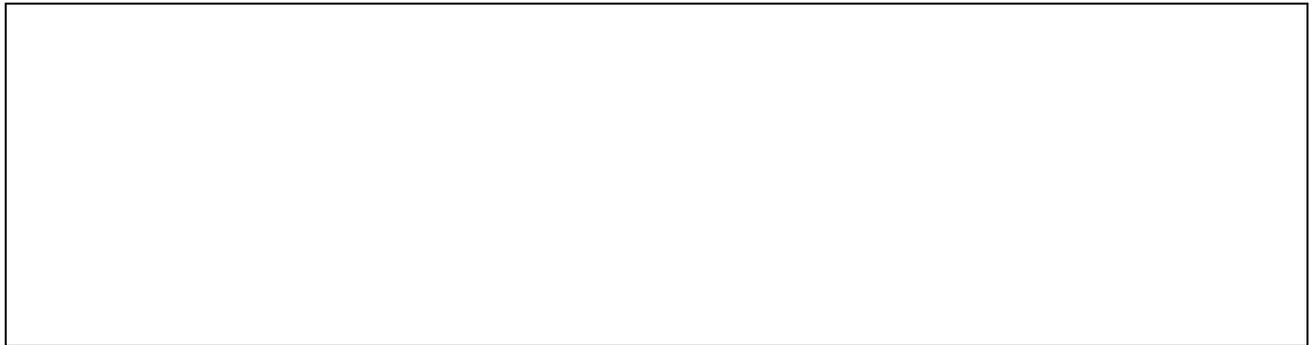


Figure 33 : Histologie de l'hippocampe (gauche) et schématisation des différents champs constituant l'hippocampe et du circuit trisynaptique (milieu et droite). D'après Ramón y Cajal, 1899-1904 et Nakazawa et al., 2002.

L'hippocampe a une position centrale dans le cerveau car il est connecté à la plupart des cortex associatifs et sensoriels ainsi qu'aux structures sous-corticales (amygdale, ganglions de la base, septum, **Figure 34**). Les connections sont soit directes via la *fimbria-fornix*, qui connecte également les hippocampes droits et gauches entre eux (Insausti and Muñoz, 2001), soit indirectes *via* la région parahippocampique, constituée des cortex entorhinal, périrhinal et parahippocampique (Burwell et al., 1995; Suzuki and Amaral, 2003). Notons que dans certaines littératures le pôle temporal est également considéré comme une région parahippocampique. L'hippocampe est notamment anatomiquement connecté à STS et TE directement (Rockland and Van Hoesen, 1999; Yukie, 2000; Zhong and Rockland, 2004; Zhong et al., 2005) et via le cortex entorhinal (Mohedano-Moriano et al., 2008), ce qui nous intéresse particulièrement ici. Cette organisation anatomique suggère que l'hippocampe contribue à la mémorisation en altérant la nature, la persistance et l'organisation des représentations stockées par les cortex cérébraux lors de l'encodage et en les réactivant lors du rappel. Par exemple l'hippocampe jouerait un rôle central pour favoriser la plasticité cérébrale des régions sensorielles dédiées à la perception des visages ou des voix à l'origine de la nature empirique des « espace des visages et des voix ».

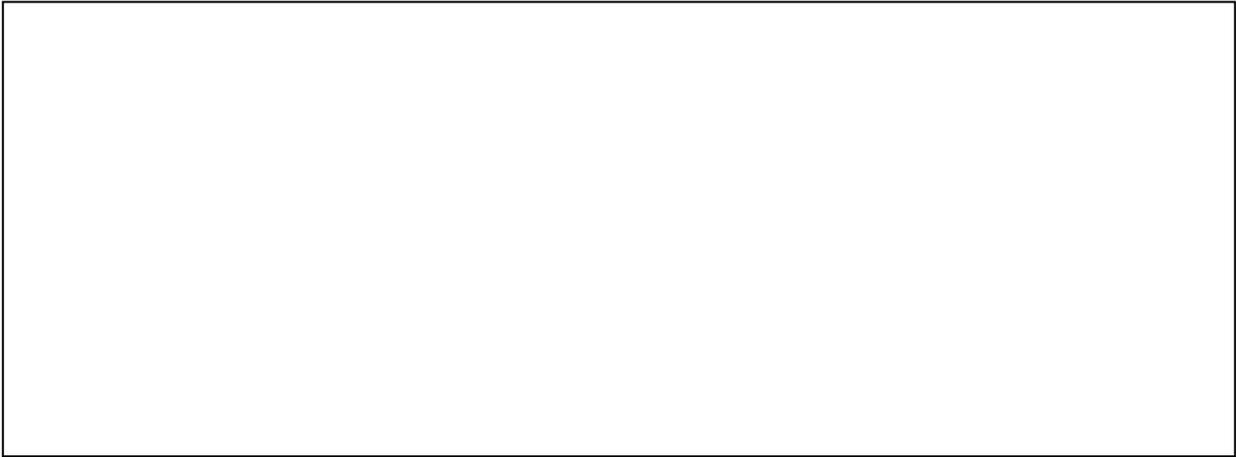


Figure 34 : Morphologie de l'hippocampe (gauche) et son observation par IRM anatomique (milieu). Schématisation de ses connexions avec le reste du lobe temporal (droite). D'après Hendelman, 2000 et Mayes et al., 2007.

E.3.2. Rôles de l'hippocampe dans l'encodage et le rappel en mémoire épisodique

Depuis les travaux psychologiques de TULVING (1972), il est considéré que différentes formes de mémoire peuvent être distinguées ; cependant la contribution des différentes structures anatomiques à chaque type de mémoire commence à faire consensus seulement récemment (Tulving, 2002). L'hippocampe a été associé à la mémoire dès les travaux de MILNER sur l'amnésie développée par le patient H.M. suite à l'ablation bilatérale de ses formations hippocampiques pour des raisons médicales (Scoville and Milner, 1957). Des travaux ultérieurs menés chez H.M. par CORKIN et chez d'autres patients ayant des lésions de l'hippocampe notamment par VARGHA-KHADEM ont affiné ces résultats et délimité les formes de mémoire dans lesquelles l'hippocampe est impliqué (Corkin et al., 1997; Vargha-Khadem et al., 1997; Corkin, 2002; O'Kane et al., 2004). Ils ont mis en évidence que l'hippocampe est engagé dans la mémoire déclarative mais non dans l'acquisition incidente de compétences et de programmes moteurs comme les habiletés motrices ou les habitudes, i.e. dans la mémoire procédurale. Ils ont également observé que l'hippocampe n'est pas nécessaire pour garder à l'esprit des données pendant un temps court de quelques minutes, ou mémoire de travail. Enfin ils ont observé que deux formes de mémoires déclaratives seraient discernables : la première concernant les souvenirs des expériences personnelles, ou mémoire épisodique, et la seconde les connaissances factuelles accumulées tout au long de la vie, ou mémoire sémantique. La détérioration de l'*encodage* en mémoire épisodique et sémantique a été mise en évidence en étudiant les patients cérébro-lésés. Au contraire, le *rappel* des connaissances sémantiques serait largement conservé après une lésion hippocampique indiquant qu'il est indépendant de l'hippocampe. Cependant, la question du *rappel* des souvenirs épisodiques est difficile à investiguer par observation des

patients cérébro-lésés et reste en suspend. En effet chez le sujet sain, certains souvenirs autobiographiques sont oubliés, d'autres sont sémantisés, ressemblant à des photos figées dont les scènes ne peuvent être revécues avec la même vivacité, et enfin certains souvenirs resteraient vivaces et pourraient être qualifiés d'épisodiques. Dès lors il est difficile de savoir si les souvenirs oubliés et ceux gardés en mémoire par les patients cérébro-lésés sont pathologiques ou non. Ainsi la question du *rappel* des souvenirs autobiographiques a été principalement étudiée par neuroimagerie PET et IRMf, mettant effectivement en évidence une implication de l'hippocampe, mais également d'autres aires comme le cortex préfrontal médian, le cortex retrosplénial, le gyrus parahippocampique, la jonction temporo-pariétale et le pôle temporal (Maguire, 2001; Viard et al., 2007; Viard et al., 2010). L'utilisation d'enregistrements cellulaires chez des patients épileptiques, a montré l'extraordinaire sélectivité de certains neurones hippocampiques pour des séquences d'événements à la fois lorsqu'elles étaient perçues et encodées et lorsqu'elles étaient rappelées librement, mettant en évidence un codage des souvenirs épisodiques au niveau neuronal (Gelbard-Sagiv et al., 2008).

Qu'est-ce qui distingue un souvenir épisodique d'une connaissance sémantique ? Il a été proposé que les souvenirs épisodiques soient constitués de trois éléments récapitulés en « Quoi-Quand-Où » et qu'on puisse les rappeler consciemment, par expérience *autonoétique* (Clayton and Dickinson, 1998; Tulving, 2002). Au contraire les connaissances sémantiques ne seraient constituées que du « Quoi ». Ainsi si l'expérience autonoétique semble réservée aux humains, la mémoire épisodique aurait des racines plus anciennes partagées par un grand nombre d'espèces fondée sur les éléments « Quoi-Quand-Où » (Clayton and Russell, 2009), et notamment par d'autres primates (Schwartz and Evans, 2001; Schwartz et al., 2002; Hampton et al., 2004; Hampton and Schwartz, 2004; Schwartz et al., 2004; Hampton et al., 2005; Basile and Hampton, 2011). Les études par imagerie ont mis en évidence plusieurs aires engagées dans la mémoire épisodique, les études ultérieures se sont donc attelées à définir exactement la nature des informations nécessitant une implication de l'hippocampe pour être correctement encodées et rappelées, contrairement à celles ayant besoin d'une implication des cortex parahippocampiques. Au moins cinq théories ont émergé, sur la base de nouveaux tests psychologiques plus précis réalisés chez des patients cérébro-lésés mais également chez des modèles animaux, rats ou singes ayant des lésions cytotoxiques bien définies de l'hippocampe ou des cortex parahippocampiques.

- 1) La théorie *Multiple-Trace* postule que l'encodage des données sémantiques ou épisodiques nécessite l'implication des mêmes structures du lobe temporal médian. Par contre alors que le rappel des souvenirs épisodiques reste dépendant de l'hippocampe, celui des connaissances sémantiques en devient progressivement indépendant (Nadel and Moscovitch, 1997; Moscovitch et al., 2005; Moscovitch et al., 2006; Winocur et al., 2007).
- 2) La théorie de la *consolidation*, postule au contraire que les deux types de rappel deviennent indépendants de l'hippocampe (Maviel et al., 2004; Frankland and Bontempi, 2005; Squire et al., 2007; Tse et al., 2007).

- 3) La théorie *Dual-Process* se fonde sur le paradigme *Remember/Know* (§ A.6) et propose que les cortex parahippocampiques soient impliqués dans le sentiment de familiarité (Know) alors que l'hippocampe serait impliqué dans le rappel contextualisé (Remember) (Eldridge et al., 2000; Brown and Aggleton, 2001; Yonelinas et al., 2002; Daselaar et al., 2006; Eichenbaum et al., 2007; Wiltgen et al., 2010).
- 4) La théorie *relationnelle* postule que l'hippocampe est engagé dans l'association flexible des informations présentes dans les modules corticaux (Mayes et al., 2007; Konkel and Cohen, 2009), permettant ainsi l'encodage, le rappel et l'imagination de nouvelles situations à partir des différentes associations entre les éléments mais également des éléments eux-mêmes (Hassabis and Maguire, 2007; Schacter et al., 2007).
- 5) Enfin un cas spécifique de la théorie relationnelle est la théorie de la *carte cognitive*, initialement proposée pour rendre compte de l'existence des cellules de lieux observées chez le rat, qui postule que le rôle premier de l'hippocampe est de construire des représentations allocentriques de l'environnement pour permettre une navigation flexible (O'Keefe, 1976; O'Keefe, 1999). Chez l'humain ces processus spatiaux auraient évolué pour assister la mémorisation des contextes spatio-temporels des souvenirs épisodiques (Bird and Burgess, 2008).

D'après quatre de ces théories (**1, 3, 4, 5**) l'hippocampe serait engagé dans la reconnaissance individuelle. Pour deux d'entre elles l'hippocampe permettrait de rappeler l'association entre une voix et un visage connu car ils appartiennent à des domaines sensoriels différents et l'un étant un contexte de l'autre (**3, 4**). Pour les deux autres théories, la voix et le visage d'individus connus évoqueraient le rappel d'épisodes vécus avec cette personne ayant un contexte spatio-temporel, engageant l'hippocampe (**1, 5**). Notre travail n'a pas pour but de favoriser l'une de ces théories, mais plutôt d'observer comment l'hippocampe représente des stimuli sociaux comme les voix et les visages d'individus connus ou inconnus.

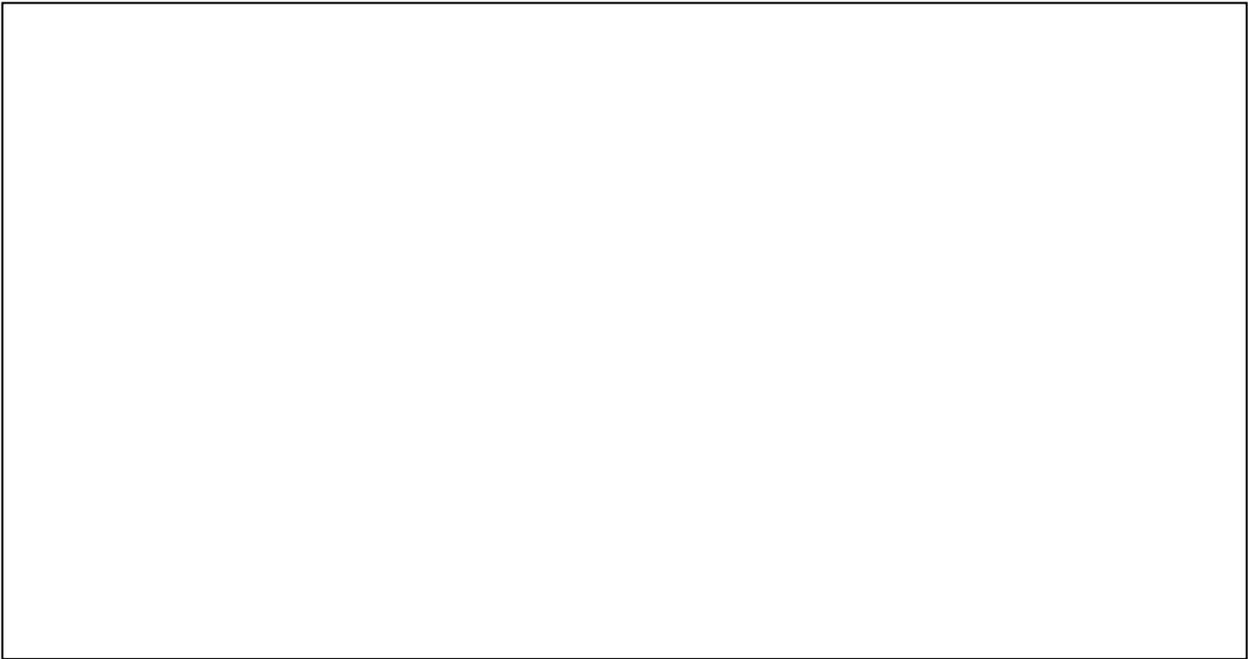


Figure 35: Illustration de la théorie Dual Process. A) L'apprentissage et le rappel d'items seuls non contextualisé est réalisé par le cortex périrhinal. Le rappel se fait alors sur la base de la familiarité ou de la connaissance factuelle (*Know*) de ce concept. B) Les associations d'items d'un même domaine sont encodées par les deux structures. La reconnaissance ultérieure de la familiarité de l'association implique uniquement le cortex périrhinal alors que le rappel flexible des différents composants les uns à partir des autres nécessite l'hippocampe. C) Les associations d'items de domaines différents nécessite l'hippocampe à la fois pour l'encodage et le rappel. D'après Mayes et al., 2007.

E.3.3. Associations dans le temps, dans l'espace et entre les domaines

L'hippocampe participe à la mémoire épisodique en reliant entre eux les attributs spatiaux, temporeux et sémantiques d'un épisode. Il pourrait également selon certaines théories mais non toutes permettre le rappel flexible de ces associations. Comment les différentes associations ou relations sont représentées par les neurones hippocampiques ?

Les relations les plus étudiées sont les relations spatiales chez le rat. Ces études ont mis en évidence que l'hippocampe représente des cartes spatiales de l'environnement où chaque neurone, appelé *cellule de lieu*, code un lieu de l'espace (O'Keefe, 1976). Ces représentations permettent un comportement flexible car elles sont codées de façon *allocentrique*, c'est-à-dire indépendamment de la position ou de l'endroit où se situe l'animal. Par la suite une organisation similaire à celle des cellules de lieu a été observée chez l'humain (Ekstrom et al., 2003) et le singe (Rolls et al., 1989; Ono et al., 1993b; Ono et al., 1993a; O'Mara et al., 1994; Eifuku et al., 1995). De façon intéressante il a été mis en évidence que l'activité des cellules de lieux peut être modulée par des objets, des comportements ou des changements sensoriels apparaissant dans le lieu qu'elles codent (Wiener et al., 1989; Tamura et al., 1992; Wiebe and Staubli, 1999, 2001). De

façon similaire les cellules codant un lieu cette fois-ci visuel, modulent également leur activité en fonction du stimulus présenté dans ce lieu (Tamura et al., 1992; Rolls et al., 2005; Rutishauser et al., 2008). Ainsi des données factuelles ou sémantiques seraient associées à la carte spatiale. Cependant il est important de préciser que ces cellules ne sont pas des cellules « grand-mères » stockant la représentation d'un objet précis à un endroit précis, au contraire l'activité des cellules formant les cartes spatiales est entièrement liée à une tâche donnée réalisée à un moment donné. En effet dès que l'animal change d'environnement les mêmes neurones se spécialisent pour les lieux du nouvel environnement (*remapping*) (Wills et al., 2005). Ainsi les neurones catégorisent les endroits et créent des associations spatiales et spatio-sémantiques éphémères.

Le rappel des séquences temporelles implique l'hippocampe chez l'humain (Lehn et al., 2009; Staresina and Davachi, 2009; Paz et al., 2010). Au niveau neuronal, les cellules hippocampiques du rat (Wiener and Korshunov, 1995; Manns et al., 2007) et du lapin (McEchron and Disterhoft, 1997, 1999) peuvent suivre les séquences temporelles d'événements discontinus. Il a été mis en évidence récemment que l'association entre la séquence temporelle et les stimuli sensoriels serait codée au niveau neuronal par l'action conjointe de l'hippocampe, du cortex périrhinal et de TE (Naya and Suzuki, 2011), l'hippocampe fournissant des signaux temporels incrémentiels de la présentation d'un stimulus au stimulus suivant, le cortex périrhinal signalant la conjonction entre les items et leur ordre temporel relatif et TE signalant la nature des stimuli visuels.

Le dernier type d'associations concerne les associations non-spatiales et non-temporelles entre différents éléments. Leur acquisition peut être distinguée de leur rappel. Ainsi WIRTH et coll. (2003) ont montré que les neurones hippocampiques signalent l'apprentissage de nouvelles associations (par exemple un stimulus visuel associé à un mouvement) en changeant leur sélectivité ; ce changement ayant lieu avant, pendant ou après l'apprentissage (Wirth et al., 2003). Par ailleurs il a été montré que cet encodage est réalisé avec succès lorsque les potentiels d'action du neurone sont coordonnés au rythme thêta (3-8 Hz) ou au rythme gamma (30-100 Hz) des oscillations du réseau local (Jutras et al., 2009; Jutras and Buffalo, 2010a; Rutishauser et al., 2010). Pendant le rappel lorsque des stimuli connus sont perçus, l'hippocampe signale à la fois leur reconnaissance et classe les stimuli les uns par rapport aux autres. Les signaux de reconnaissance peuvent être des augmentations d'activité pour ces stimuli (Rutishauser et al., 2008; Jutras and Buffalo, 2010b) mais voir (Xiang and Brown, 1998)), des sélectivités affinées (Yanike et al., 2004) ou un recrutement plus nombreux de neurones actifs (Viskontas et al., 2009). Le classement des différents stimuli peut être représenté sous forme de carte cognitive non-spatiale.

E.3.4. Cartographie de l'espace cognitif par les neurones de l'hippocampe

La théorie de la *carte cognitive* postule que les neurones hippocampiques réalisent une carte spatiale allocentrique de l'environnement (O'Keefe, 1976; O'Keefe, 1999). Cependant l'exploration de l'activité des neurones unitaires a mis en évidence qu'un certain pourcentage de neurones (plus faible chez le rat que le pourcentage de cellules de lieux) est engagé dans la représentation de stimuli non spatiaux. Ainsi des neurones signalent la nature des stimuli visuels

(Tamura et al., 1992; Korshunov et al., 1996), tactiles (Young et al., 1994), auditifs (Sakurai, 1994, 1996, 2002), olfactifs (Eichenbaum et al., 1987; Wiener et al., 1989; Wiebe and Staubli, 1999; Wood et al., 1999; Wiebe and Staubli, 2001) ou comportementaux (Wood et al., 2000). Cette vision a été combattue par les tenants de la théorie de la carte cognitive indiquant que ces activités pourraient être attribuées, soit à des neurones spatiaux modulant leur réponse à des événements non-spatiaux, soit à une augmentation de l'éveil attentionnel (O'Keefe, 1999). Cependant plusieurs études ont montré que les trois catégories de neurones hippocampiques peuvent être dissociées : des neurones spatiaux purs, des neurones spatiaux modulés par la nature des stimuli et enfin des neurones sélectifs à la nature des stimuli indépendamment de leur localisation (Wiebe and Staubli, 1999; Wood et al., 1999; Wiebe and Staubli, 2001). Comment ces neurones s'inscrivent-ils dans la carte cognitive spatiale ? D'une part il a été proposé que les neurones spatiaux et non-spatiaux appartiennent à des classes de neurones différentes (Christian and Deadwyler, 1986) mais l'étude des neurones pyramidaux et des interneurones a mis en évidence que, dans les deux classes de neurones, les différents types d'informations peuvent être distingués (Wiebe and Staubli, 1999, 2001). D'autre part il a été proposé que l'hippocampe soit engagé dans la représentation d'une carte relationnelle globale dont la carte spatiale serait un sous-exemplaire (Konkel and Cohen, 2009). Ainsi les différents types de relations (spatiale mais également temporelle ou entre différents domaines sémantiques) seraient représentées par les cellules hippocampiques ; les cellules de lieux pourraient alors être considérées comme un sous-type de cellules relationnelles, les autres relations étant représentées par des cellules non-spatiales.

En utilisant un grand nombre de stimuli visuels, des études chez le singe et chez des patients épileptiques humains ont mis en évidence que les neurones hippocampiques catégorisent l'espace cognitif des stimuli, comme ils catégorisent l'espace spatial lorsque le rat se déplace dans un labyrinthe. Ces études ont mis en évidence la représentation de catégories telles « animaux », « humains », « objets » (Kreiman et al., 2000; Hampson et al., 2004), ou de combinaisons telles le genre et l'expression des visages (Fried et al., 1997; Fried et al., 2002) par les neurones hippocampiques. Des catégories encore plus précises ont été observées, codées par un petit nombre de neurones (Quiroga et al., 2005; Mormann et al., 2008; Quiroga et al., 2008a; Quiroga et al., 2008b; Quiroga et al., 2009; Viskontas et al., 2009; Ison et al., 2011). Ces cellules étaient par exemple actives pour de multiples vues d'un même individu ainsi qu'à son nom écrit ou prononcé par l'expérimentateur (**Figure 21**). Ces cellules éparses dans l'hippocampe et hautement sélectives ressemblent à des unités gnostiques représentant les concepts de façon explicite. Notons cependant qu'il ne s'agit pas dans ce cas de neurones « grand-mères » pour plusieurs raisons. La première étant qu'il y aurait plus d'un neurone codant chaque concept (Quiroga et al., 2008b). En effet un calcul probabiliste indique que parmi le milliard de neurones constituant le lobe temporal médian, quelques deux millions d'entre eux représenteraient un même concept d'individu (Waydo et al., 2006). La deuxième raison se réfère aux propriétés de *remapping* des cellules hippocampiques. En effet l'activité de ces cellules serait éphémère et liée à la tâche expérimentale. Ainsi du point de vue de la théorie relationnelle, ces neurones ultra-sélectifs, invariants et épars seraient les nœuds d'une carte multidimensionnelle, dans laquelle chaque axe représente une dimension différente de l'espace cognitif.

En résumé l'ensemble des associations pourrait être représenté sous forme de carte cognitive relationnelle par l'hippocampe. Cette représentation spéciale permettrait un encodage et une réactivation flexible des différents éléments, de leurs associations et de leurs relations. L'hippocampe ne stockerait pas l'ensemble des liens entre les items d'un épisode, au contraire l'organisation fonctionnelle sous forme de carte suggère qu'il contribue à la mémorisation en altérant la nature, la persistance et l'organisation des représentations stockées par les cortex cérébraux et en réactivant ces représentations de façon flexible lors du rappel. Ainsi l'hippocampe agirait comme un lien entre les différentes régions corticales, qu'il réactiverait de façon répétée pendant l'encodage, notamment pendant le sommeil, pour consolider un épisode en mémoire en modifiant les réseaux neuronaux de ces régions. Il pourrait également permettre la réactivation vivace de cet épisode au cours du rappel, en réactivant l'ensemble des régions corticales impliquée lorsque l'épisode fut vécu. Comment l'hippocampe communique avec les autres cortex ? L'hippocampe est une structure fortement oscillante. L'existence de rythmes créés par l'hippocampe ou modulant son activité, comme les rythmes thêta et gamma mais également les sharp-wave ripples ou les rythmes du sommeil lent, suggère qu'ils ont un rôle de choix dans la communication entre l'hippocampe et les autres cortex pendant l'encodage (Sirota et al., 2003; Jones and Wilson, 2005; Sirota et al., 2008; Girardeau et al., 2009; Peyrache et al., 2009; Benchenane et al., 2010; Jutras and Buffalo, 2010a) mais également pendant le rappel (Guderian and Düzel, 2005; Düzel et al., 2010; Steinvorh et al., 2010).

E.4. Conclusion

Les voix et les visages permettent à eux seuls le rappel complet du concept d'individu. La reconnaissance individuelle engage alors l'activation cross-modale des aires unisensorielles FFA et TVAA *via* leur connexion anatomique directe ainsi qu'au moins trois aires multisensorielles. Ces aires sont le precuneus impliqué dans le sentiment de familiarité, possiblement les pôles temporaux impliqués dans le rappel biographique, ainsi que l'hippocampe, impliqué dans l'encodage et le rappel flexible d'associations notamment dans le contexte de la mémoire épisodique (**Figure 36**). Le Person identity node défini par BRUCE et YOUNG (1986) pourrait alors ne pas être unique mais au contraire regrouper au moins trois aires complémentaires. L'hippocampe, dans le cas de la reconnaissance individuelle, serait l'aire phare pour la complétion du pattern incomplet avec les connaissances antérieures. En effet BHRICK et coll. (2005) ont déterminé la nature polymorphe de l'association voix-visage : une première connaissance concerne l'association figée entre les deux éléments dont la reconnaissance pourrait engager le lien directe entre les aires unimodales ou encore le cortex périrhinal ; la deuxième connaissance est flexible permettant le rappel de l'identité de la voix à partir de celle du visage et *vice versa*, et pourrait engager l'hippocampe. Ainsi le rôle de l'hippocampe reste à définir : permet-il le rappel de l'identité de la voix à partir de celle du visage ? Ou permet-il le rappel épisodique subséquent à partir de ces indices puissants ?

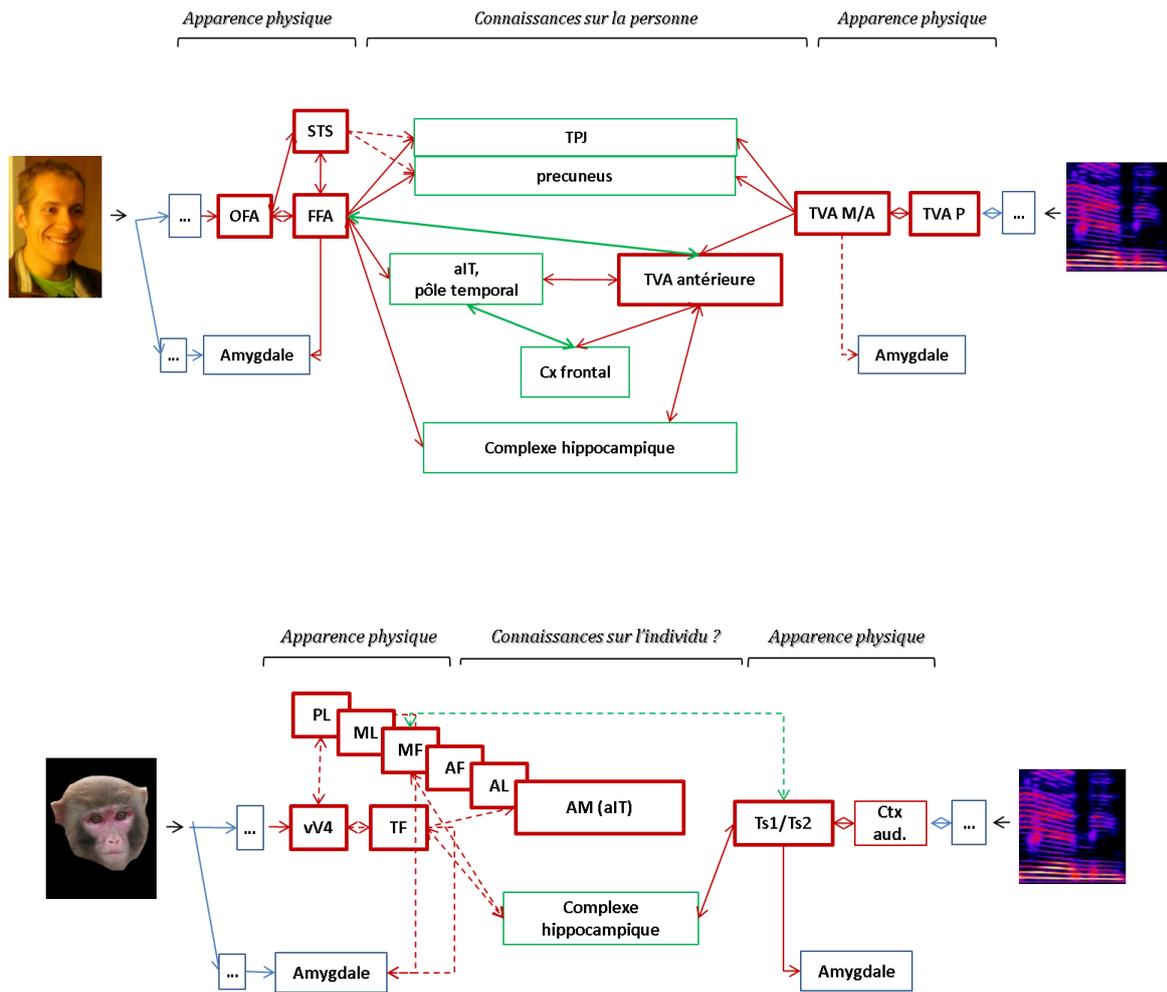
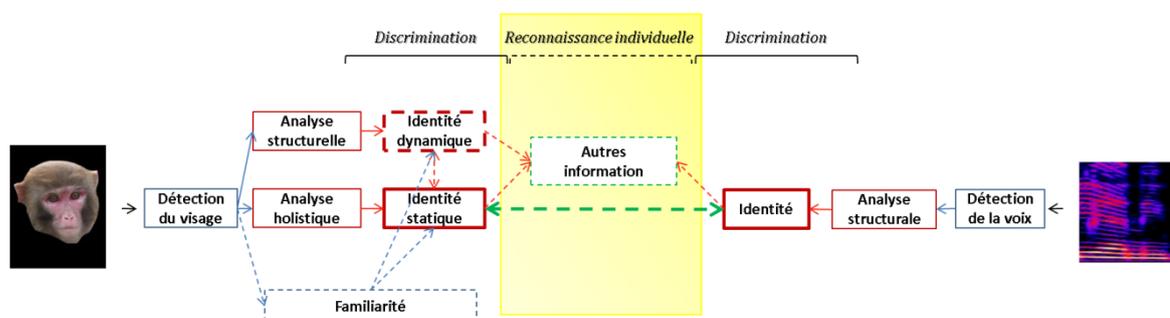


Figure 36 : Schémas des aires intégratives (vert) entre l'identité des voix et des visages. Notons un lien direct entre FFA et TVAa. Et trois régions multisensorielles: une région dorsale comprenant le precuneus et la jonction temporo-pariétale véhiculant le sentiment de familiarité, le complexe hippocampique impliqué dans la mémoire épisodique, et enfin le pôle temporal antérieur et gyrus frontal inférieur possiblement impliqué dans le rappel du nom et des informations biographiques.

F. Problématiques

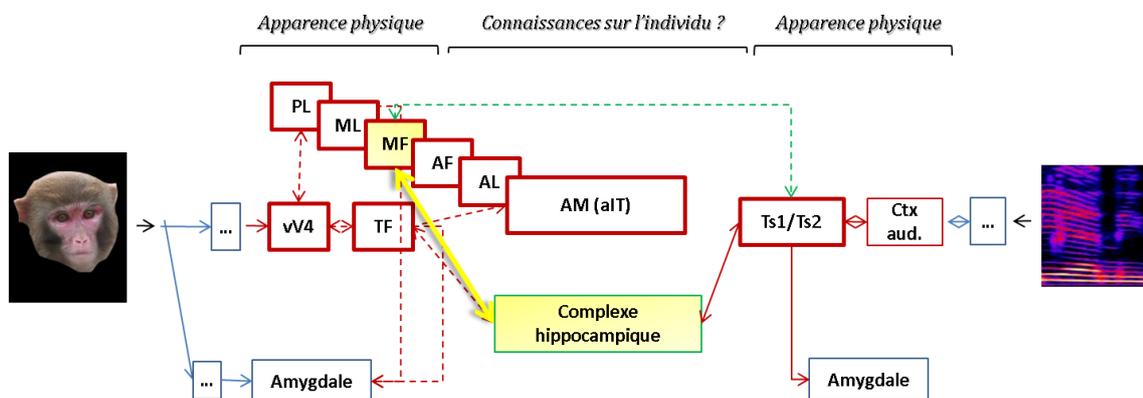
Question 1: Les macaques rhésus reconnaissent-ils individuellement leurs congénères et les humains familiers ?

Les macaques rhésus perçoivent et discriminent les identités faciales et vocales. Cependant il n'a pas été démontré qu'ils possèdent la faculté de les associer et de retrouver l'information d'une modalité à partir de l'autre, c'est-à-dire la faculté de reconnaître individuellement leurs congénères. Or cette mise en évidence permettrait d'approcher l'existence d'un concept de Personne chez le singe. Dans l'étude décrite dans le **chapitre A (§ Expériences et Résultats)**, nous avons donc développé une tâche de reconnaissance spontanée nous permettant de tester l'existence d'une représentation cross-modale des individus. Six macaques rhésus entendaient une voix connue, puis à la fin de l'enregistrement audio pouvaient librement regarder deux visages affichés à l'écran. L'un des visages correspondait à la voix jouée précédemment, alors que le second, également connu du singe, ne lui correspondait pas. Nous avons fait l'hypothèse que les singes regarderont plus longtemps le visage correspondant à la voix. Par ailleurs nous avons également voulu comparer la reconnaissance individuelle de congénères à celle d'humains familiers, car les singes testés ont été élevés à la fois en présence d'humains et de singes. Ainsi dans le même protocole la moitié des essais était constituée de voix et de visages humains et l'autre de vocalises et de visages simiens. Dans son ensemble, cette étude démontre que les singes possèdent une représentation multimodale des individus incluant des congénères mais également des individus humains familiers.



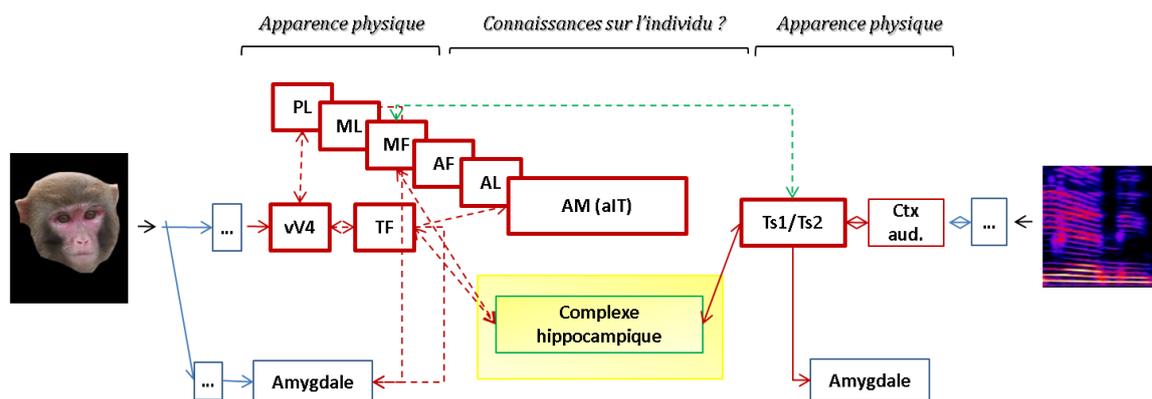
Question 2 : Quelles sont les contributions respectives de TE et de l'hippocampe à la représentation des stimuli sociaux au niveau neuronal ?

Les individus sont représentés cognitivement par une multitude d'indices, intégrés en une unité conceptuelle. Nous avons choisi de comprendre comment deux indices cruciaux, *les visages et les voix*, sont représentés dans le cortex visuel associatif – le cortex inférotemporal – et dans l'hippocampe – zone mnésique mais aussi zone associative par excellence du fait de sa localisation anatomique centrale au sein du lobe temporal concentrant les afférences de nombreuses aires sensorielles. Chez l'humain, l'hippocampe serait impliqué dans la reconnaissance individuelle qu'elle soit initiée par la vue d'un visage ou par l'entente d'une voix personnellement connue, comme mis en évidence par IRMf (Joassin et al., 2011; Leveroni et al., 2000; Haist et al., 2001; Paller et al., 2003; Bernard et al., 2004; Rotshtein et al., 2005; Denkova et al., 2006; Ishai and Yago, 2006; Nakamura et al., 2001; von Kriegstein and Giraud, 2004). Chez le singe l'hippocampe présente également une augmentation d'activité en IRMf à la vue d'un visage ou lorsque le singe entend une vocalise (Gil-da-Costa et al., 2004; Ku et al., 2011). Bien que des cellules répondant aux visages aient été mises en évidence depuis plusieurs années dans le cortex inférotemporal, la présence de cellules homologues dans l'hippocampe n'a pas été recherchée chez le singe. Par ailleurs, peu d'études ont tenté de comprendre dans quelle mesure ces neurones contribuent à la représentation d'individus connus et si ils peuvent être modulés par les voix – indices qui correspondraient à des visages sonores des individus (Belin, 2004). Dans l'étude décrite dans le **chapitre B (§ Expériences et Résultats)**, nous avons donc enregistré l'activité des neurones hippocampiques et l'activité de neurones de TE pendant que le singe regardait des visages d'individus connus ou inconnus (humains ou singes) et entendait des voix connues ou inconnues (humaines ou de singes). Ces activités étaient comparées à celles pendant lesquelles le singe regardait ou entendait des sons d'objets connus ou inconnus. Nous nous sommes demandé si des neurones distinguaient les stimuli sociaux des stimuli non-sociaux et les stimuli connus des stimuli inconnus. L'ensemble de cette étude permet d'apporter des éléments de réponse sur la façon dont les représentations d'individus sont sous tendues par les aires cérébrales de façon redondante et complémentaire.



Question 3 : Peut-on lier la nature des représentations hippocampiques à la nature de leur substrat cellulaire ?

Des parallèles sont souvent établis entre les activités neuronales observées chez différentes espèces. Cependant la *nature* des différents types cellulaires présents dans l'hippocampe du singe a été peu étudiée par rapport à celles des cellules hippocampiques des rongeurs et de l'humain. Dans l'étude décrite dans le **chapitre C (§ Expériences et Résultats)**, en analysant les propriétés électrophysiologiques des neurones hippocampiques, nous avons voulu caractériser et distinguer les interneurons des cellules pyramidales. Nous avons mené cette caractérisation dans les différents champs hippocampiques afin de les comparer entre eux. Les différentes populations neuronales mises à jour ont ensuite été comparées à celles observées chez le rat et chez l'humain. Par ailleurs, une question importante est la compréhension d'une part, des contributions relatives des différents types cellulaires à la représentation des stimuli et d'autre part, des interactions entre ces différents types cellulaires. Par exemple, lorsque la sélectivité des cellules est étudiée, il est intéressant de comprendre comment chaque sélectivité peut être façonnée par les inputs de la cellule et par ses interactions avec les cellules voisines. Pour mieux comprendre le rôle des types cellulaires sous-tendant les réponses aux stimuli, nous avons donc examiné l'activité de chaque neurone en réponse aux stimuli visuels et acoustiques en fonction de ses propriétés électrophysiologiques. Ce dernier volet de la thèse permet de descendre d'un niveau supplémentaire dans l'échelle d'analyse des représentations faciales et vocales. En effet après les avoir étudié au niveau comportemental, puis au niveau de leur représentation dans deux aires cérébrales, nous nous intéressons à leur représentation au niveau encore plus fin de la nature des différents types cellulaires formants des microcircuits fonctionnels.



Expériences et résultats

A. Reconnaissance individuelle cross-modale par les macaques Rhésus

Résumé : Reconnaissance de l'identité faciale et vocale par le singe, prémisses du concept de Personne chez le primate non-humain.

La reconnaissance d'un individu particulier se produit lorsque nous réactivons les liens entre les entrées sensorielles actuelles et la représentation précédemment formée de cette personne. Cette reconnaissance peut être réalisée en identifiant, séparément ou simultanément, des éléments distincts tels le visage, la silhouette, ou la voix comme appartenant à une seule personne. Chez les humains, ces différents attributs sont intégrés en une représentation conceptuelle complexe de l'identité individuelle. Dans cette étude, nous avons testé si les macaques rhésus (*Macaca mulatta*) possèdent également une représentation cognitive de l'identité en évaluant s'ils présentent une reconnaissance individuelle cross-modale. En outre, nous avons évalué à la fois la reconnaissance individuelle des congénères familiers et celle d'humains personnellement connus des singes testés. Dans un paradigme mesurant les temps de regard librement alloués, nous avons constaté que, pour les deux espèces, les singes appariaient spontanément les visages des individus connus à leur voix. Ce résultat démontre que les macaques rhésus possèdent une représentation cross-modale cognitive des individus qui s'étend des congénères aux humains, révélant le potentiel adaptatif de la reconnaissance de l'identité pour les individus socio-écologiquement pertinents.

Cette étude a été publiée sous le titre :

Sliwa J, Duhamel J-R, Pascalis O, Wirth S (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. **Proceedings of the National Academy of Sciences of the USA** 108: 1735-1740

Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans

Julia Sliwa^a, Jean-René Duhamel^a, Olivier Pascalis^b, and Sylvia Wirth^{a,1}

^aCentre de Neurosciences Cognitive, Centre National de la Recherche Scientifique, Université Lyon I, 69675 Bron, France; and ^bLaboratoire de Psychologie et Neurocognition, Centre National de la Recherche Scientifique, Université Pierre Mendès France, 38040 Grenoble, France

Edited by Nikos K. Logothetis, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, and approved December 15, 2010 (received for review June 15, 2010)

Recognition of a particular individual occurs when we reactivate links between current perceptual inputs and the previously formed representation of that person. This recognition can be achieved by identifying, separately or simultaneously, distinct elements such as the face, silhouette, or voice as belonging to one individual. In humans, those different cues are linked into one complex conceptual representation of individual identity. Here we tested whether rhesus macaques (*Macaca mulatta*) also have a cognitive representation of identity by evaluating whether they exhibit cross-modal individual recognition. Further, we assessed individual recognition of familiar conspecifics and familiar humans. In a free preferential looking time paradigm, we found that, for both species, monkeys spontaneously matched the faces of known individuals to their voices. This finding demonstrates that rhesus macaques possess a cross-modal cognitive representation of individuals that extends from conspecifics to humans, revealing the adaptive potential of identity recognition for individuals of socioecological relevance.

cross-species | vocal communication | nonhuman primates | picture recognition

In humans, both faces and voices convey information about identity, providing some of the many cues we use to recognize individuals we know (1). The multifaceted nature of identity code suggests that a complex cognitive representation binds semantic information with information of different sensory modalities. In rhesus monkeys, however, it has typically been assessed only via single-modality information. For instance, rhesus monkeys can discriminate between calls of two conspecifics in a playback experiment using a spontaneous habituation–discrimination paradigm (2). They are also able to discriminate faces of two conspecifics in a match-to-sample task (3) and monkey faces or human faces in visual paired comparison tasks (4–6). These observations demonstrate that monkeys can *discriminate* idiosyncratic characteristics (“individual A is different from B”) for their own or other species. However, they do not provide evidence of *individual recognition* (“this is individual A, this is B”). In comparison with discrimination, individual recognition requires an additional associative level that allows retrieval of information belonging to a specific individual. In rhesus monkeys, coarse recognition processes such as of their own species, kin, gender, reproductive status, or hierarchy are well documented (2, 7–12). However, these rudimentary recognition abilities fail to account for some sophisticated behaviors that are observed in rhesus macaques’ societies. In particular, rhesus macaques live in large groups and maintain elaborate social relations involving, e.g., nonkin alliances during aggressive interactions, fight interference, reciprocal support, friendly grooming, and reconciliation (13–15). Such an organization would benefit well from the finest-grain individual recognition based on a cognitive multimodal representation of identity.

Individual recognition can be revealed in animals by demonstrating the existence of a cognitive representation of individuals supported by the integration of cues from multiple sensory modalities. Only a few studies have applied this approach to investigate individual recognition in species other than humans and

chimpanzees (16, 17) and none of them focused on rhesus monkeys. Moreover, these studies explored individual recognition of peers (18, 19), but did not examine whether this capacity extended to other species. Given the complexity of individual recognition, the question of whether it can be applied to both one’s own and another species remains open. When individual recognition was tested across species, e.g., dogs recognizing humans (20) or squirrel monkeys recognizing humans (21), the studies remained inconclusive regarding individual voice identification and failed to demonstrate face–voice association for more than one highly familiar human individual. Thus, the species-specific nature of individual recognition is still unclear. Yet, the cross-species issue is of special interest because it provides information about the properties of individual recognition by testing its adaptability. For instance, in human infants, discrimination of faces is known to specialize during the first year of life for conspecifics or even more drastically for individuals of their own socio-ethnic group (22–25). However, this faculty can be maintained with longer exposure to the other species or groups (e.g., ref. 26). Similarly in rhesus monkeys, performance at processing faces is poorer for other species than for their own (4, 6), but this asymmetry can be reversed in favor of humans if infant monkeys first experience human faces (27). Thus, it is of great interest to evaluate whether the multimodal process of individual recognition also generalizes to highly relevant information acquired throughout life, such as information about human individuals for laboratory monkeys.

In the present study, we aimed to evaluate both the individual recognition of conspecifics and the individual recognition of humans by adult rhesus monkeys. Our subjects had daily exposure to both rhesus monkey and human individuals from infancy and were familiarized with both the humans and other rhesus monkeys serving as stimuli in our experiment via recent real life daily exposure (housing “roommates,” caregivers, and researchers). Individual recognition was then investigated through a cross-modal approach. Our experiment assessed whether rhesus macaques would spontaneously match two attributes of familiar individuals, i.e., match a voice to the photograph of its corresponding face. In this design, voice–face matching is supported by memories of previous interactions. This process entails two components. First, can monkeys match a voice with an appropriate pictorial content? And concomitantly, is the picture of a face sufficient to elicit recall of the appropriate individual? These observations will substantially contribute to our knowledge of rhesus macaques’ aptitude for multimodal representations of others. Only affiliation calls such as coos and grunts were used as representing monkey-sound stimuli because their formant structure is a reliable marker

Author contributions: J.S., J.-R.D., O.P., and S.W. designed research; J.S. and S.W. performed research; J.S. analyzed data; and J.S., J.-R.D., O.P., and S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: swirth@isc.cnrs.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008169108/-DCSupplemental.

of individual identity (28). Human speech was used as human-sound stimuli (Fig. 1A). Still color photographs of known individuals (Fig. 1A) were used. We did not use movies, because indexical matching on the basis of the redundancy between vocal-tract resonance and its physical shape is possible when face stimuli are presented for real (physical presence) or on a forward-played movie, whereas it is impossible when presented as a still image or on a reversed-played movie (29, 30).

Six subjects were tested in a free scanning task on the basis of a preferential looking paradigm. After hearing a voice sample while fixating on a central spot of light, the monkey freely explored a pair of face pictures, only one of which matched the voice (Fig. 1B). Eye movements were restrained to a virtual window around the pair of images represented by the black area in Fig. 1B and the subject was rewarded for maintaining its gaze within this explorative window during the stimulus presentation. Importantly, the animal was not reinforced to gaze at one picture over the other. We then compared the preferential-looking time. We predicted that, if monkeys have a cross-modal representation of known individuals, they would preferentially attend to the picture matching the voice.

Results

Do Monkeys Spontaneously Look Longer at a Face When It Matches a Voice Previously Heard? To assess whether animals linked the heard voice sample to one of the face photographs, we examined looking times for all face pictures depending on their congruence with the preceding voice (Fig. 2A). This congruence accounts for the intrinsic variability in looking times across faces that may bias the time allocated to each of the two images. Looking behavior can be influenced by factors such as dominance or sexual coloration of the pictured individuals (9, 12). We confirmed that our pictures yielded different looking times with a one-way analysis of variance (ANOVA) with face identity as a factor (Table S1 and Fig. 3). By comparing looking time for a face preceded by its matching voice to that for the same face when it was preceded by a different voice, we captured the effect of voice–face concordance on the time allocated to each image. Results for each face and each pair were concatenated and the group difference was then compared with zero (absence of voice–face concordance effect). This average difference was significantly greater than zero [$t(103) = 4.48, P = 9 \times 10^{-6}$], revealing that animals look significantly longer at a face when it is preceded by its voice than when it is preceded by a different voice (Fig. 2B). This finding was replicated when we tested whether the proportion of time spent looking at the matching face, relative to the other face in the pair, was greater than expected by chance, despite differences in looking times for individual pictures (SI Materials and Methods, Text S1). Furthermore, we ensured that this effect was not driven by gender matching rather than identity matching (SI Materials and Methods, Text S2).

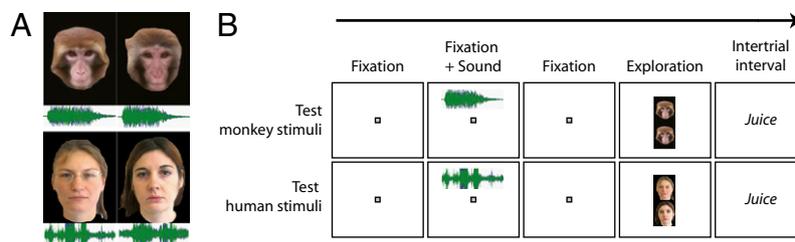


Fig. 1. Stimuli and experimental paradigm. (A) Examples of test stimuli for monkey (Upper) and human (Lower). Green diagrams represent spectrograms of “coo” vocalization and of voice samples of the individuals. (B) Test trials: The voice of a known individual is followed by the presentation of two pictures of known faces, one of which matches the preceding voice. The animals fixate for the 2 s during which the voice sample is played and explore freely during the picture presentation (1.5 s). Animals were rewarded for maintaining gaze within the boundary of the virtual window represented by the black area, regardless of their exploration pattern.

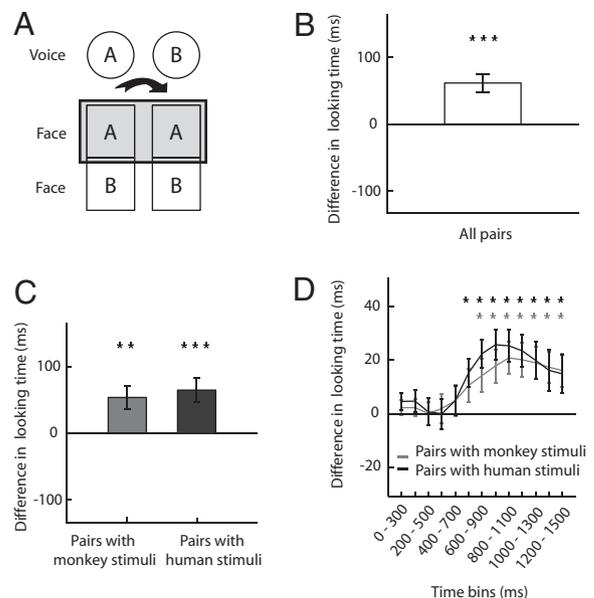


Fig. 2. Monkeys match the identity of the voice with the identity of the face. (A) For each pair AB, the box represents the relevant comparison: difference in looking time for face A when preceded by the congruent voice A compared with when preceded by the noncongruent voice B. (B) Mean looking time difference between the congruent condition and the noncongruent condition for all of the pairs ($n = 104$ pairs). The abscissa indicates chance expectation. Errors bars represent SEM ($*P < 0.05, **P < 0.01, ***P < 0.001$). (C) The same as B but separately analyzed for stimulus pairs with monkey ($n = 32$ pairs) and human ($n = 72$ pairs). (D) The same as C but analyzed across time with a sliding window of 300 ms moved by 100-ms steps for stimulus pairs with monkey ($n = 32$ pairs for each time point) and human ($n = 72$ pairs for each time point) ($*P < 0.05$; see Table S2 for exact P values).

Do Monkeys Exhibit a Cross-Modal Recognition of Both Conspecifics and Humans? The looking time for a given face when it was preceded by its matching voice, compared with when it was not, was significantly different for both monkey stimuli trials [$t_M(31) = 3.18, P = 0.0017$] and human stimuli trials [$t_H(71) = 3.52, P = 0.0004$] (Fig. 3C).

Change Over Time. To evaluate the point in time relative to the picture onset, at which animals started to look significantly more at the face matching the voice, we analyzed looking times obtained in sliding 300-ms windows moving in 100-ms steps. A significant effect was observed only from 800 ms after picture onset until the end of the visual presentation for trials with human stimuli and 900 ms for trials with monkey stimuli (Table S2 and Fig. 3D). Surprisingly, the voice–face concordance appears as a late effect.

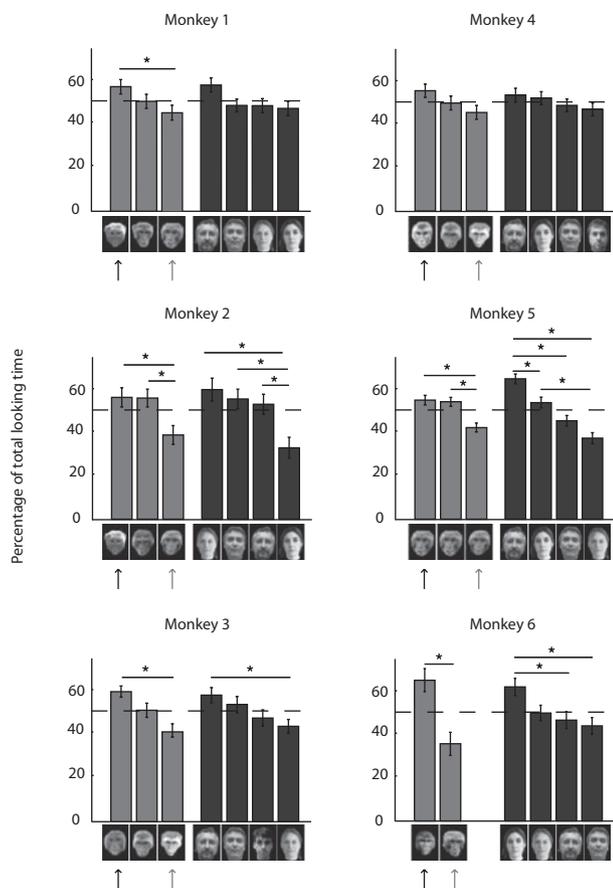


Fig. 3. Preferential looking time across individual pictures for each tested subject. Mean percentage of looking time is shown for each face picture within a pair. The dashed line indicates chance expectation. Errors bars represent SEM. Intragroup comparisons were calculated using a multiple-comparison test with a Bonferroni adjustment (*, 95% confidence intervals). Solid arrows indicate pictures of the closest monkey to each tested subject. Shaded arrows indicate pictures of the most agonistic monkey to each tested subject.

Robustness of the Results Across Individuals. To assess individual differences, the data were analyzed independently for the six tested subjects. Five of six monkeys showed significantly greater looking time for faces when preceded by matching voice relative to when they were not (Table S3 and Fig. S14). Thus, the group effects described in the global analysis were not driven by isolated individuals. An ANOVA on looking times using the identity of the subject as a random-effect variable confirmed the above results, with a significant voice–face concordance factor [$F = 15.69$, $df = 1$, $P = 0.001$], enabling us to extrapolate the matching behavior to a population larger than the six subjects. We also tested whether cross-modal recognition of humans was relevant at the individual level. When considering separately trials with human stimuli and trials with monkey stimuli, three of six monkeys showed significantly greater looking time for faces matching the voice for monkey stimuli and five of six monkeys showed significantly greater looking time for faces matching the voice for human stimuli (Table S3 and Fig. S1B). This result indicates a heterogeneous behavior among individuals. Further, we tested whether this heterogeneity reflects common variations observed in a population. We also conducted two ANOVAs on looking times using the identity of the subject as a random effect but conducted separately for monkey stimuli trials [$F_M = 5.64$, $df = 1$, $P = 0.018$] and human stimuli trials [$F_H = 9.68$, $df = 1$, $P = 0.002$]. This result

confirmed that the group effects were not driven by individuals' effects for trials with either monkey or human stimuli.

Examining Individual Face Photograph Preferences. Once the nature of the face–voice concordance factor was examined, we analyzed face identity, the other factor that ruled preferential looking times. Mean viewing times on face pictures were analyzed independently of voice–face concordance by summing viewing times for each face over trials preceded by the face's own voice and over trials preceded by another voice. As shown by the one-way ANOVA on the looking times performed for each tested subject, there was a strong effect of face identity factor as five of six monkeys presented significant face preferences (Table S1 and Fig. 3). To further evaluate what could account for differences in viewing times for individual faces, face pictures were arranged in decreasing order of looking time, separately for monkey and human stimuli (Fig. 3). Then, for each subject, a multiple comparison of the looking times on each face was performed, separately for trials with human and monkey stimuli (Table S1 and Fig. 3). For five of six subjects, the monkey face yielding the longest looking time was that of the monkey they were paired with or that of their neighbor, when not paired. The monkey face least looked at was that of the individual with whom they maintained the most agonistic relation, as assessed through daily observation of interactions mainly comprising threats. For three monkeys, it was the dominant monkey of the colony room and for two monkeys, it was an aggressive female. Some differences also appeared for human individual faces. It is more difficult to draw a conclusion about human individuals that would parallel the one with monkey individuals because of the difference in the nature of the interactions. Nonetheless, it appeared that the person most looked at was the researcher/experimenter in charge of the animal.

Discussion

Recently, a growing interest in determining the neural bases of social cognition has emerged (e.g., refs. 31 and 32). This new field often relies on the rhesus monkey as a model because of the close homology of its brain with the human brain. However, drawing parallels between human and rhesus monkey social cognition requires the development of behavioral approaches and testing procedures adapted to each species (33). In this study, we contributed to the goal of characterizing how rhesus monkeys process socially relevant stimuli in an experimental setting by demonstrating four key points. First, we show the existence of individual recognition in rhesus monkeys comprising at least two elements of identity (vocal and visual). Second, individual recognition extends adaptably from conspecifics to personally known humans. Third, rhesus macaques exhibit preferential bias toward pictures of certain individuals over others, likely reflecting their social interaction history. Fourth, our data indicate that rhesus macaques perceive pictures as being, if not equivalent, at least connected or related to real life stimuli.

Evidence for the Existence of Individual Cross-Modal Recognition.

Overall, rhesus monkeys spontaneously looked longer at a face that matched the voice previously heard. This spontaneous cross-modal identity matching suggests that rhesus monkeys possess a cognitive representation of individuals, allowing them to link two perceptual cues. The cross-modal nature of the task is crucial. In non-cross-modal tasks, matching can be done at the perceptual level: for example, matching different views of the same face by means of common features perceived through mental rotation (34) or recognizing the different vocalizations of an individual through the presence of unchanged physical features such as formants. On the contrary, in our task, monkeys had to rely on information in memory to match one stimulus, the voice, with the other, the face. Finally, the static nature of the stimuli also affirms this conclusion. As opposed to video displays, in which faces and voices share a common dynamic that enables matching the vocal

sample to the vocal-tract movements, in a static display, their combination cannot be performed at a perceptual level but rather at a conceptual one. This result indicates that monkeys have learned through daily experience, the face–voice identity associations for familiar individuals. The relative complexity of this process explains why individual recognition has long been thought to be a uniquely human capacity, which has been reported to develop ~4 mo of age (35). However, it has recently been demonstrated in a few other species (16–19) and might be more widespread in animals (36, 37). Evidence for individual recognition by rhesus macaques raises the issue of the adaptive value of this very precise type of social recognition during interactions with conspecifics. In rhesus macaques, individual recognition might be required for some particularly sophisticated behaviors (13–15); moreover, such recognition could support third-party knowledge in those contexts wherein awareness of the peculiar relationships of others is useful in maintaining amicable relationships with a minimum of hostility and stress in complex groups (38).

Plasticity and Adaptive Potential of Individual Recognition. In the present experiment, cross-modal individual recognition extends to human stimuli, indicating that this process is not species specific but also applies to socially relevant individuals from other species who become familiar in an animal's life. It is of interest to consider the role of perceptual narrowing in the capacities observed here. Human and macaque infants possess a broad perceptual tuning at birth that later narrows for socio-ecologically relevant signals (25). Studies with monkeys have reported both narrowing for face perception (27, 5, 7) and broad face–voice dynamics perception in infants (39). Amodal attributes shared by faces and voices (tempo, collocation, or synchrony) precisely enhance association of modality-specific features (characterizing age, gender, and identity) (35, 40). In this context, then, the presence of cross-species recognition in our study could be explained by such a broad amodal face–voice perception being present at birth and persisting because of the incentive value of human individuals. This hypothesis is consistent with recent evidence in infant Japanese macaques showing the influence of exposure to conspecifics and humans on forming multisensory associations (41, 42). Whereas infant Japanese macaques raised in large groups were able to match a monkey voice to a monkey face but not a human voice to a human face, young animals with daily exposure to humans were able to match voice and face in both conditions. The latter experiments tackled coarse multisensory representation of “species,” and we show here that more fine-grained multisensory representation of “identity” may also be influenced by experience with humans.

Difference in Individual Recognition for Humans and Monkeys. Surprisingly, some monkeys were equally good or better at matching human face–voice identity relative to conspecifics face–voice identity. However, we did not make direct comparisons between trials with monkey stimuli and those with human stimuli, because of the inherent inequality between human and monkey audio sets. Human speech contains more information about speaker identity than monkey coo vocalizations, making human speakers easier to recognize. Whereas monkey “coos” and human speech both contain uniform cues for individual discrimination, only human speech contains dynamic cues changing over the duration of an utterance that represent additional robust information for speaker identity (43). As a result, similarities in performance across stimulus sets could be driven by differences in the complexity of the two stimulus sets, rather than by the incentive value of one of these two groups of familiar individuals.

What Do Monkeys Perceive in the Pictures of Familiar Individuals?

Previous experiments in capuchin monkeys were conducted with pictures of familiar individuals in social discrimination tasks. They revealed that 2D images of familiar conspecifics are interpreted by

capuchin monkeys as representing reality (44) and therefore could be well suited to investigations of individual recognition. In studies with rhesus monkeys, pictures of familiar individuals were also used in social discrimination tasks (e.g., refs. 12 and 45). However, it was unclear whether rhesus macaques relied on pictorial features, such as prominent facial muscles that represent secondary sexual characteristics in dominant males, or on the social knowledge underpinned by them (i.e., on previously acquired knowledge of the hierarchy between the tested monkey and the one in the picture). Indeed, we know little about how rhesus monkeys perceive photographs, beyond the facts that they perceive features and colors of photographs as humans do (46), that they are highly motivated to attend to visual stimuli involving conspecifics (47), and that they find social stimuli rewarding (47). Pictures are both concrete objects with features and colors and a representation of something other than themselves, and it is unclear whether monkeys perceive a connection between the reality and the content of the picture (48, 49). Here we presented faces as cropped photographs that included head, neck, and hairs, on a black background. Compared with real faces, the stimuli were reduced along several physical dimensions, such as size, stereoscopy, and motion parallax cues. Despite the reduced informational content, monkeys extracted relevant visual identity cues, first, to match them with auditory identity recording and second, to drive preferential looking time. Thus, pictures provided at least a sufficient cue for the recall of the individual they represent, indicating that rhesus monkeys establish a connection between the reality and the content of the pictures and use it to orient their behavior. This result implies that face photographs of familiar individuals are relevant and adapted stimuli for use in behavioral or neurophysiological laboratory tests.

About Preferences of Individual Face Photographs Over Others.

Rhesus monkeys displayed significant differences in their looking times across individual face pictures, favoring, among monkeys, the individual with whom they maintained the closest relation (their pair or neighboring mate) or, among humans, their main caregiver. This result suggests that experience and interactions shaped subjects' preferences. This pattern of preference contrasts with findings showing that rhesus monkeys accept lower rewards as a payoff for looking at pictures of dominant monkeys (12). However, the results of the latter study also suggest that the initial orientation toward a face and total duration of the viewing time might reflect two separate processes: Monkeys might choose to look initially at a dominant peer but spend more total time looking at other monkeys. According to the authors, orientation enables one to gain visual information that might be necessary for future social behavior, whereas total viewing duration represents a commitment into a specific relationship with the individual in the picture. This interpretation is consistent with our data, as subjects spent more time looking at individuals with whom they felt safest and avoided eye contact with monkeys with whom they usually had agonistic interactions. This behavior may also be driven by the fact that we used only affiliation calls and not threat calls, which might have led to a longer viewing time toward the agonistic monkeys.

From Voice–Face Identity Association Toward the Concept of a “Person.”

In humans, a complex cognitive representation binds semantic information with information of different sensory modalities. This information includes at least name, face, voice, silhouette, odor, gait, and feel of the surface of the skin (50–52). The concept of a person also implies knowledge of interaction history with a given individual and expectations about its future behavior (53). We demonstrate that rhesus macaques possess at least a multifaceted internal model of other individuals that includes face–voice association and face–photograph association for each individual. Our experiment revealed other (but not all) aspects of the concept of a person, by exhibiting strong preferences for

certain individuals over others. These data offer unique insight into the complexity of conceptual thinking about other individuals by rhesus monkeys. A simple twofold face–voice association is much less complex than the concept of a person as a whole; yet it is constitutive of this concept. This association might represent a basic building block of cognition shared across a wide range of species and may be an evolutionary precursor of the concept of a person found in humans. As such, insight into this skill is gained from investigation in a bottom–up perspective of animal and human cognition (33).

Conclusion. The present findings advance our understanding of face–voice integration in rhesus monkeys. Prior studies showed that rhesus monkeys can *i*) perform lip–speech association by matching face dynamics to corresponding vocalization, *ii*) associate the number of voices heard to the number of faces seen, and *iii*) link vocal-tract resonance to corresponding physical shape (54–56). We now demonstrate the existence of a cross-modal voice–face identity association that encompasses both familiar peers and familiar humans. Moreover, our results suggest the existence of a cognitive representation of identity based on memories of previous interactions. Consequently, behavioral studies such as this set the stage for further neurophysiological investigations of social cognition.

Materials and Methods

Subjects. Six rhesus monkeys (*Macaca mulatta*, three females, 6–11 y old) were used. Animals were socially housed in rooms of four individuals after they arrived between 2–5 y ago at the Centre de Neurosciences Cognitives. They were born and bred during 2–3 y in small rhesus macaque groups with daily exposure to both conspecifics and humans. All experimental procedures were in accordance with the local authorities (Direction Départementale des Services Vétérinaires, Lyon, France) and the European Community standards for the care and use of laboratory animals [European Community Council Directive (1986), Ministère de l'Agriculture et de la Forêt, Commission Nationale de l'Expérimentation Animale].

Stimuli. Visual and auditory test stimuli were color photographs and audio recordings of individuals familiar to the subjects (Fig. 1A). Familiar simian individuals were the three rhesus macaques housed in the same room as the subjects for 2–5 y before the test. The individuals were adult 6- to 16-y-old monkeys. Familiar human individuals were the two caregivers and two experimenters working with the animals on a daily basis during the same 2- to 5-y period. Each stimulus set was specific to each animal to ensure a high familiarity with individuals presented in stimuli. Front view photographs were cropped to include only the face, head, and neck and then adjusted to a common size and presented on a black background (GNU Image Manipulation Program; www.gimp.org). Humans were presented without mask and goggles as the colony rooms are equipped with windows through which animals see the experimenter and staff without their protections. There was one photograph per familiar individual, leading to four human and three monkey photographs per animal tested. Photographs shown in Figs. 1 and 3 are provided for illustrative purposes but do not correspond to the actual stimuli, which included for all animal photographs the head implants (head posts and recording chambers). Photographs used otherwise resembled the illustrative ones in all respects (color, size, and gaze direction). Auditory stimuli consisted of audio recordings of 2-s duration, each containing either a sample of coo vocalization (894 ± 300 ms) or a human speech extract (877 ± 380 ms). Human speech extracts consisted of small sentences or words in French such as “bonjour tout le monde” (“hello everybody”) or “voilà” (“here it is”). Six audio samples from each familiar individual were presented, leading to 24 human voice stimuli and 18 vocalizations per animal tested. The mean sound pressure level for the duration of each audio sample was calculated. Then the 42 audio samples for each subject were normalized for this mean acoustic intensity (MATLAB; mathWorks). Visual and auditory training stimuli were fractals (Fractal Explorer; fractals.da.ru) and synthetic abstract audio samples.

Preferential Looking Paradigm. Before testing, animals first learned to complete exploration trials with abstract audio and visual stimuli (*SI Materials and Methods, Text S3* and *Fig. S2*). The goal of the training task was to maintain the animal's gaze centered in the middle of the screen during the audio playback and allow it to freely explore two images presented together on the screen

after the audio playback, within the boundary of a virtual window comprising both stimuli. Once animals could complete 150 trials, training stimuli were replaced by test stimuli representing familiar individuals (Fig. 1A and B). Size, positions, and timings of stimuli presentation remained unchanged between training and test task. Juice reward was maintained in the intertrial period to ensure the monkeys were motivated to complete trials. Two types of test trials were randomly interleaved: trials with stimuli representing familiar rhesus macaques and trials with stimuli representing familiar humans. In each trial a voice or vocalization playback was followed by the presentation of two known faces, only one of which matched the voice of the playback. The subject could freely explore these face photographs during 1.5 s as long as its gaze was maintained within the boundaries of a virtual window around the pair of photographs corresponding to the black area in Fig. 1B. The design was counterbalanced such that each face appeared equally often in the upper and lower part of the screen, and each face was equally associated with a voice–vocalization record. We limited testing to three sessions of 75 completed trials with monkey stimuli and 75 completed trials with human stimuli per subject on successive days to preserve spontaneous behavior and prevent any stereotyped behavior from occurring during sessions. Importantly at no point in the task did we encourage the animal to associate a face photograph with a voice. On the contrary, we just observed their spontaneous behavior in looking at their choice of face photograph. Animals were rewarded equally regardless of whether they looked longer at the matching face, at the nonmatching face, or at both faces equally. Further, animals were rewarded even if they chose to look at no photograph by keeping their gaze in the middle of the screen.

Procedure. For training and testing, a subject was head restrained and placed in front of, and at eye level to, an LCD screen (Dell Computers) situated at 56 cm. The subject's gaze position was monitored with an infrared eye tracker (ISCAN) with a 250-Hz sampling rate. Eye movement samples were recorded and stored using REX Software (57), which also served for real-time control of stimulus presentation, eye position monitoring, and reward delivery. The calibration procedure is detailed in *SI Materials and Methods, Text S4*.

Data Analysis. Data were analyzed with custom-written scripts and the Statistics Toolbox in Matlab (MathWorks). Looking at the upper face in the pair was defined as when the eyes' coordinates were located inside the upper part of the virtual exploration window ($12^\circ \times 12^\circ$) that contained the cropped face photograph plus $\approx 4^\circ$ of black surround. The looking time for a particular face was the amount of time the gaze was present inside of this window and was automatically computed with the Matlab routines. When the gaze was not on the upper face, it was either on the lower face or in a $12^\circ \times 2^\circ$ central black area between the two faces that was not considered for the analyses. If the gaze left the exploration window, the trial was ended prematurely and was not analyzed.

To analyze whether a particular face photograph affected the results, a one-way ANOVA was performed on the looking time on each face for each subject separately. No overall ANOVA could be used because each stimulus set was specific to each animal to ensure a high familiarity with individuals presented in stimuli. To analyze the modulation of face preferences by the voice identity, for each face of each pair, the mean μ_i and variance σ_i^2 of $D_i = (\text{looking time spent on face, when matching the voice}) - (\text{looking time spent on face, when not matching the voice})$ were calculated. Before this analysis, we tested that the variances of both groups were similar with a Bartlett's test for equal variances. A t test was performed on the concatenation of all means μ_i to assess whether the group mean was different from zero. Further, a t test was also performed separately for trials with monkey stimuli and human stimuli. The change over time of the modulation of face preferences by the identity of the voice was analyzed with a sliding window of width 0.3 s moving every 0.1 s. A t test on the D_i calculated at each time step was performed every 0.1 s.

To analyze individual performances, the modulation of face preferences by the voice identity was analyzed for each subject as described previously, by calculating for each face of each pair the mean μ_i and variance σ_i^2 of D_i . The group mean and variance (μ_G, σ_G^2) of all of the pairs were inferred with an expectation-maximization (EM) algorithm that enables us to find their maximum-likelihood (ML) estimates via an iterative method alternating expectation (E) and maximization (M) steps. In the E-step, μ_i and σ_i^2 were estimated by computing their ML estimates knowing the group mean and variance:

$$\mu_i = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_i^2} \hat{\mu}_i + \frac{\sigma_i^2}{\sigma_G^2 + \sigma_i^2} \mu_G, \sigma_i^2 = \frac{\sigma_G^2 \sigma_i^2}{\sigma_G^2 + \sigma_i^2}.$$

In the M-step, the group mean and variance were reevaluated, maximizing their likelihood knowing the face-pair means and variances. This maximization reduces to the following equation:

$$\mu_G = \frac{1}{n} \sum_i \mu_i; \sigma_G^2 = \frac{1}{n} \sum_i [\sigma_i^2 + (\mu_G - \mu_i)^2].$$

The new μ_G and σ_G^2 parameters are then used in the next E-step. This process quickly converges toward the ML values of μ_G and σ_G^2 (58). This estimate weights the contribution of face-pair results. The EM method was used to take into account the behavioral variability of the subjects for each face pair without having to classify and remove any outlier data. A *t* test was then performed to compare the group (μ_G, σ_G^2) difference with a null difference.

Finally, a three-way ANOVA was performed on the percentage of looking time with the following factors: voice–face concordance, face preference (preferred, neutral, or less preferred as assessed by a paired *t* test on each pair of faces), and identity of the tested subject as a random effect. To analyze preferences among individuals' faces, a multiple comparison with a Bonferroni adjustment of Student's *t* critical values was performed on the

looking time for each face for each subject, for trials with human and monkey stimuli, respectively.

ACKNOWLEDGMENTS. We thank animal care staff and experimenters from the Centre de Neuroscience Cognitive who were photographed and recorded; J.-L. Charieau and F. Héran for animal care; P. Baraduc and P. Vindras for useful discussions on statistical analyses; V. Chambon for helpful discussion on the data, and L. Parsons for help with the English editing of the manuscript. Research was supported by a Marie Curie reintegration grant and a salary grant from the Fondation pour la Recherche Médicale (to S.W.), a PhD grant from Centre National de la Recherche Scientifique and the Direction Départementale de l'Armenet under the directorship of P. Balzagette and the Fondation pour le Recherche Médicale (to J.S.), and National Institutes of Health Grant R01 HD046526 (to O.P.).

- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11:535–543.
- Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free ranging rhesus monkeys. *Anim Behav* 51:1007–1015.
- Parr LA, Winslow JT, Hopkins WD, de Waal FBM (2000) Recognizing facial cues: Individual discrimination by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta). *J Comp Psychol* 114:47–60.
- Gothard KM, Erickson CA, Amaral DG (2004) How do rhesus monkeys (Macaca mulatta) scan faces in a visual paired comparison task? *Anim Cogn* 7:25–36.
- Gothard KM, Brooks KN, Peterson MA (2009) Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn* 12:155–167.
- Dahl CD, Logothetis NK, Hoffman KL (2007) Individuation and holistic processing of faces in rhesus monkeys. *Proc Biol Sci* 274:2069–2076.
- Pascalis O, Bachevalier J (1998) Face recognition in primates: A cross-species study. *Behav Process* 43:87–96.
- Waitt C, et al. (2003) Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proc Biol Sci* 270(Suppl 2):S144–S146.
- Waitt C, Gerald MS, Little AC, Krauselburd E (2006) Selective attention toward female secondary sexual color in male rhesus macaques. *Am J Primatol* 68:738–744.
- De Waal FBM, Luttrell LM (1985) The formal hierarchy of rhesus macaques: An investigation of the bared-teeth display. *Am J Primatol* 9:73–85.
- Gouzoules H, Gouzoules S, Tomaszycki M (1998) Agonistic screams and the classification of dominance relationships: Are monkeys fuzzy logicians? *Anim Behav* 55:51–60.
- Deaner RO, Khera AM, Platt ML (2005) Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Curr Biol* 15:543–548.
- Kaplan JR (1978) Fight interference and altruism in rhesus monkeys. *Am J Phys Anthropol* 49:241–250.
- De Waal FBM, Yoshihara D (1983) Reconciliation and redirected affection in rhesus monkeys. *Behaviour* 85:224–241.
- De Waal FBM, Luttrell LM (1988) Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethol Sociobiol* 9: 101–118.
- Bauer HR, Philip M (1983) Facial and vocal individual recognition in the common chimpanzee. *Psychol Rec* 33:161–170.
- Kojima S, Izumi A, Ceugniet M (2003) Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates* 44:225–230.
- Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proc Natl Acad Sci USA* 106:947–951.
- Bovet D, Deputte BL (2009) Matching vocalizations to faces of familiar conspecifics in grey-cheeked mangabeys (*Lophocebus albigena*). *Folia Primatol (Basel)* 80:220–232.
- Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. *Anim Cogn* 10:17–21.
- Adachi I, Fujita K (2007) Cross-modal representation of human caretakers in squirrel monkeys. *Behav Process* 74:27–32.
- Pascalis O, de Haan M, Nelson CA (2002) Is face processing species-specific during the first year of life? *Science* 296:1321–1323.
- Kelly DJ, et al. (2009) Development of the other-race effect during infancy: Evidence toward universality? *J Exp Child Psychol* 104:105–114.
- Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci USA* 103:6771–6774.
- Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cogn Sci* 13:470–478.
- Pascalis O, et al. (2005) Plasticity of face processing in infancy. *Proc Natl Acad Sci USA* 102:5297–5300.
- Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad Sci USA* 105:394–398.
- Rendall D, Owren MJ, Rodman PS (1998) The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *J Acoust Soc Am* 103: 602–614.
- Lachs L, Pisoni DB (2004) Crossmodal source identification in speech perception. *Ecol Psychol* 16:159–187.
- Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003) "Putting the face to the voice": Matching identity across modality. *Curr Biol* 13:1709–1714.
- Ghazanfar AA, Santos LR (2004) Primate brains in the wild: The sensory bases for social interactions. *Nat Rev Neurosci* 5:603–616.
- Irki A, Sakura O (2008) The neuroscience of primate intellectual evolution: Natural selection and passive and intentional niche construction. *Philos Trans R Soc Lond B Biol Sci* 363:2229–2241.
- de Waal FB, Ferrari PF (2010) Towards a bottom-up perspective on animal and human cognition. *Trends Cogn Sci* 14:201–207.
- Wang G, Obama S, Yamashita W, Sugihara T, Tanaka K (2005) Prior experience of rotation is not required for recognizing objects seen from different angles. *Nat Neurosci* 8:1768–1775.
- Bahrick LE, Hernandez-Reif M, Flom R (2005) The development of infant learning about specific face-voice relations. *Dev Psychol* 41:541–552.
- Johnston RE, Bullock TA (2001) Individual recognition by use of odors in golden hamsters: The nature of individual representations. *Anim Behav* 61:545–557.
- Seyfarth RM, Cheney DL (2009) Seeing who we hear and hearing who we see. *Proc Natl Acad Sci USA* 106:669–670.
- Silk JB (2007) Social components of fitness in primate groups. *Science* 317:1347–1351.
- Zangenehpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ (2009) Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE* 4:e4302.
- Bahrick LE, Lickliter R (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 36:190–201.
- Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates* 47:350–354.
- Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2009) Plasticity of ability to form cross-modal representations in infant Japanese macaques. *Dev Sci* 12: 446–452.
- Fitch WT (2000) The evolution of speech: A comparative review. *Trends Cogn Sci* 4: 258–267.
- Pokorny JJ, de Waal FBM (2009) Monkeys recognize the faces of group mates in photographs. *Proc Natl Acad Sci USA* 106:21539–21543.
- Shepherd SV, Deaner RO, Platt ML (2006) Social status gates social attention in monkeys. *Curr Biol* 16:R119–R120.
- Waitt C, Buchanan-Smith HM (2006) Perceptual considerations in the use of colored photographic and video stimuli to study nonhuman primate behavior. *Am J Primatol* 68:1054–1067.
- Anderson JR (1998) Social stimuli and social rewards in primate learning and cognition. *Behav Process* 42:159–175.
- Bovet D, Vauclair J (2000) Picture recognition in animals and humans. *Behav Brain Res* 109:143–165.
- Fagot J, Thompson RKR, Parron C (2010) How to read a picture: Lessons from nonhuman primates. *Proc Natl Acad Sci USA* 107:519–520.
- Cutting JE, Kozlowski LT (1977) Recognizing friends by their walk: Gait perception without familiarity cues. *Bull Psychon Soc* 9:353–356.
- Porter RH, Balogh RD, Cernoch JM, Franchi C (1986) Recognition of kin through characteristic body odors. *Chem Senses* 11:389–395.
- Kaitz M (1992) Recognition of familiar individuals by touch. *Physiol Behav* 52: 565–567.
- Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 8:1611–1618.
- Ghazanfar AA, Logothetis NK (2003) Neuroperception: Facial expressions linked to monkey calls. *Nature* 423:937–938.
- Ghazanfar AA, et al. (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17:425–430.
- Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15:1034–1038.
- Hays AV, Richmond BJ, Optican LM (1982) A UNIX-based multiple-process system for real-time data acquisition and control. *WESCON Conference Proceedings* (National Eye Institute, Bethesda), Vol 2, pp 1–10.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data with the EM algorithm. *J R Stat Soc Ser A* 39:1–38.

Supporting Information

Sliwa et al. 10.1073/pnas.1008169108

SI Materials and Methods

Text S1: Comparison of the Time Spent Looking at the Face Matching the Voice Relative to the Other Face. Data were analyzed with MATLAB custom-written scripts and the Statistics Toolbox (MathWorks). Here, we analyzed the data without taking into account preferences that monkeys displayed for individual faces. Therefore, we assessed whether the proportion of time spent on matching face was above chance level (50%). To this end, we performed a *t* test on data for all trials for all monkeys pooled together. To compare the number of saccades and duration of the longest fixation on the matching picture compared with the nonmatching picture, a paired-sign test was performed. Saccades were defined as increases in gaze velocity $>30^\circ/\text{s}$ with minimum amplitude of 1° . Fixations were defined as the gaze position in the interval between two saccades.

Monkeys looked at the matching face significantly longer compared with the other face (match, 658 ms, 51.8%; nonmatch, 612 ms, 48.2%). Although this difference is relatively small, it differed significantly from chance [one-sample *t* test, $t(1,726) = 2.31$, $P = 0.02$]. Moreover, the number of saccades and the duration of the longest fixation were higher on the matching face than on the nonmatching face (respectively $P = 0.014$ and $P = 0.045$, sign tests).

Text S2: Is the Matching Effect Driven by Gender Characteristics or Identity Characteristics? The limited amount of individuals we could use as stimuli in our experiment forced us to present male and female monkeys and humans. Here we tested whether the perceived face–voice matching effect was driven by gender matching or by identity matching. To do so, we separated trials comparing same gender faces from trials comparing different gender faces. Then, as in the main analysis, we tested the modulation of face preferences by the identity of the voice, for each face of each pair.

This analysis revealed that monkeys spend more time looking at a face when preceded by its voice in same gender trials [$t(49) = 3.54$, $P = 0.0009$] as well as in different gender trials [$t(53) = 2.94$, $P = 0.005$]. When analyzing separately trials with monkey stimuli from trials with human stimuli, it appears that results are similar to the joint results: in same gender monkey [$t(21) = 2.14$, $P = 0.04$] and human [$t(27) = 2.96$, $P = 0.006$] trials and in different gender monkey [$t(9) = 2.49$, $P = 0.03$] and human [$t(43) = 2.23$, $P = 0.031$] trials.

Text S3: Training Paradigm. Animals first learned to complete exploration trials with abstract audio and visual stimuli. Visual training stimuli were presented in pairs subtending $10^\circ \times 24^\circ$. The virtual exploration window around the pair was $12^\circ \times 26^\circ$ of visual angle. It was composed of the two pictures (in upper and lower parts) surrounded by $\approx 4^\circ$ black surround and a central $12^\circ \times 2^\circ$ black area between the two pictures considered as belonging neither to one nor to the other. Auditory training and test stimuli were presented at an intensity of 55–65 dB (A-weighted) sound pressure level (SPL) at the subject's ear, measured with a Brüel and Kjær 2239A Integrating Sound Level Meter (www.bksv.com) from two speakers located 56 cm in front of the subject and symmetrically 45 cm apart.

The goal of the training task was to maintain the animal's gaze on the middle of the screen during the audio playback and allow it to freely explore two images presented together on the screen after the audio playback. The subject was to start by directing its gaze to a small white square presented on the center of the screen. Then it was required to fixate ($\pm 1^\circ$ – 4°) while a 2-s audio sample was played. After a delay of 0.2 s, two pictures were presented vertically and equally spaced from the center (Fig. S2). The subject could freely explore these pictures during 1.5 s, as long as its gaze was maintained within the boundaries of a virtual window around the pair of pictures corresponding to the black area in Fig. S2. In a pilot study, we tested conditions in which we did not reward keeping gaze inside of the virtual window around the pair of pictures. In this case, the monkey completed fewer trials than we needed to perform statistical tests regarding the richness of our stimuli sets. Thus, reward was given to ensure monkeys' attention to the computer monitor where visual stimuli were displayed and not to provide reinforcement for a particular exploration pattern. This training paradigm was pursued until the monkeys could readily complete 150 trials.

Text S4: Eye Tracking Calibration Procedure. Each session started by a calibration of signal offset and gain using an analog amplifier with modulating gain and offset connected in series to the ISCAN and REX Software. To do so, the monkey was rewarded for fixating on a small spot ($\pm 1^\circ$ – 4°) appearing at the center of the screen (offset calibration) and then equally distributed 15° in four directions: left (abscissa gain), right, up (ordinate gain), and down. Eye movements were not filtered for the analyses.

Table S1. Face preferences

Monkey	ANOVA on preferences for some individual faces over others	Significant confidence intervals
M1	$F_1 = 2.36$, $df = 6$, $P = 0.029$	$CI_{0.95}(\mu_{1-3}) = [0.65; 23.77]$
M2	$F_2 = 4.65$, $df = 6$, $P = 0.0001$	$CI_{0.95}(\mu_{1-3}) = [2.62; 32.75]$ $CI_{0.95}(\mu_{2-3}) = [2.63; 32.03]$ $CI_{0.95}(\mu_{4-7}) = [8.22; 46.47]$ $CI_{0.95}(\mu_{5-7}) = [4.63; 41.21]$ $CI_{0.95}(\mu_{6-7}) = [2; 38.95]$
M3	$F_3 = 4.3$, $df = 6$, $P = 0.0003$	$CI_{0.95}(\mu_{1-3}) = [7.38; 28.6]$ $CI_{0.95}(\mu_{4-7}) = [0.98; 28.27]$
M4	$F_4 = 1.34$, $df = 6$, $P = 0.24$	None
M5	$F_5 = 15.65$, $df = 6$, $P = 0$	$CI_{0.95}(\mu_{1-3}) = [5.57; 20.15]$ $CI_{0.95}(\mu_{2-3}) = [4.82; 19.37]$ $CI_{0.95}(\mu_{4-5}) = [18.55; 37.17]$ $CI_{0.95}(\mu_{4-6}) = [10.37; 28.96]$ $CI_{0.95}(\mu_{4-7}) = [7.49; 25.99]$ $CI_{0.95}(\mu_{5-7}) = [1.91; 20.33]$
M6	$F_6 = 5.34$, $df = 5$, $P = 9.10^{-5}$	$CI_{0.95}(\mu_{1-2}) = [14.53; 45.36]$ $CI_{0.95}(\mu_{4-6}) = [3.53; 33.25]$ $CI_{0.95}(\mu_{4-7}) = [0.51; 30.83]$

Table S2. Change over time

Time course bins	Overall	Monkey stimuli	Human stimuli
0–300	$t(103) = 1.62$, $P = 0.054$	$t(31) = 0.8$, $P = 0.21$	$t(71) = 1.42$, $P = 0.08$
100–400	$t(103) = 1.33$, $P = 0.094$	$t(31) = 0.7$, $P = 0.24$	$t(71) = 1.15$, $P = 0.13$
200–500	$t(103) = 0.13$, $P = 0.45$	$t(31) = 0.003$, $P = 0.50$	$t(71) = 0.13$, $P = 0.46$
300–600	$t(103) = 0.13$, $P = 0.45$	$t(31) = 0.3$, $P = 0.36$	$t(71) = -0.01$, $P = 0.5$
400–700	$t(103) = 1.16$, $P = 0.12$	$t(31) = 0.8$, $P = 0.21$	$t(71) = 0.89$, $P = 0.19$
500–800	$t(103) = 3.3$, $P = 0.0005$	$t(31) = 1.6$, $P = 0.053$	$t(71) = 2.91$, $P = 0.002$
600–900	$t(103) = 4.77$, $P = 3 \times 10^{-6}$	$t(31) = 2.2$, $P = 0.016$	$t(71) = 4.21$, $P = 4 \times 10^{-5}$
700–1,000	$t(103) = 5.46$, $P = 2 \times 10^{-7}$	$t(31) = 2.8$, $P = 0.004$	$t(71) = 4.67$, $P = 6 \times 10^{-6}$
800–1,100	$t(103) = 5.29$, $P = 3 \times 10^{-7}$	$t(31) = 3.3$, $P = 0.0009$	$t(71) = 4.25$, $P = 3 \times 10^{-5}$
900–1,200	$t(103) = 4.78$, $P = 3 \times 10^{-6}$	$t(31) = 3.0$, $P = 0.002$	$t(71) = 3.81$, $P = 0.0001$
1,000–1,300	$t(103) = 3.82$, $P = 0.0001$	$t(31) = 3.1$, $P = 0.002$	$t(71) = 2.87$, $P = 0.003$
1,100–1,400	$t(103) = 3$, $P = 0.002$	$t(31) = 2.9$, $P = 0.003$	$t(71) = 2.14$, $P = 0.018$
1,200–1,500	$t(103) = 2.85$, $P = 0.003$	$t(31) = 2.6$, $P = 0.006$	$t(71) = 2.05$, $P = 0.022$

Table S3. Individual specificities

Voice–face match	M1	M2	M3	M4	M5	M6
Overall	$t_1(17) = 16.9$, $P = 2 \times 10^{-12}$	$t_2(17) = 4.29$, $P = 0.0002$	$t_3(17) = 1.59$, $P = 0.065$	$t_4(17) = 4.78$, $P = 8 \times 10^{-5}$	$t_5(17) = 2.91$, $P = 0.005$	$t_6(13) = 7.63$, $P = 2 \times 10^{-6}$
Monkey stimuli	$t_1(5) = 1.33$, $P = 0.12$	$t_2(5) = 1.07$, $P = 0.17$	$t_3(5) = 0.032$, $P = 0.48$	$t_4(5) = 2.09$, $P = 0.045$	$t_5(5) = 6.49$, $P = 0.0006$	$t_6(2) = 7.21$, $P = 0.043$
Human stimuli	$t_1(11) = 19.42$, $P = 4 \times 10^{-10}$	$t_2(11) = 4.36$, $P = 0.0006$	$t_3(11) = 2.06$, $P = 0.031$	$t_4(11) = 4.32$, $P = 0.0006$	$t_5(11) = 1.05$, $P = 0.16$	$t_6(11) = 5.51$, $P = 9 \times 10^{-5}$

A.3. Analyses supplémentaires non publiées

Additional issues raised by referees.

1. “The analysis of difference scores for each specific pair is a good approach to control for possible differences in absolute looking times. Unfortunately, however, it does not control for possible regression to the mean effects that might have contributed to the significant effect of the test face factor in the initial ANOVA. That is, those faces that elicit relatively little attention, might elicit a lot more attention when followed by their matching vocalization whereas those faces that elicit a lot of attention prior to the vocalization may not elicit much more attention because of a ceiling effect. Is there any evidence of regression to the mean in the data set?”

We addressed the issue of regression to the mean, by performing additional statistical analyzes. We separated the data into two parts: a first one that comprises pairs of faces in which the amount of attention for the two pictures was significantly different (irrespective of the voice preceding the pair) and a second one that comprises pairs of faces in which the amount of attention for the two pictures was not significantly different (irrespective of the voice preceding the pair). This has been done on the basis of paired student test, performed on each pair of face pictures, as previously explained in the manuscript. Then for each group we conducted the main analysis described previously in the manuscript. It appears that monkeys look significantly longer at a face when it is preceded by its voice than when it is not, for pairs of faces eliciting a different amount of attention [$t(51)=3,12$, $p=0,003$] as well as for faces eliciting a similar amount of attention [$t(51)=2,43$, $p=0,02$]. There was no significant difference between these two groups [$t(102)=0,11$, $p=0,91$]. This reveals that we do not observe regression to the mean in the data.

To corroborate these results we tested directly the interaction between voice-face concordance effect and face preference effect with an analysis of variance as suggested by reviewer 2 in the question below. As previously explained in the manuscript we performed an ANOVA with three factors 1) specific animal tested (as a random factor), 2) attention elicited by the specific face in a pair (preferred/not-preferred/neutral on the basis of paired t-test comparisons of each pair of faces) and 3) voice-face concordance, which revealed the importance of Face preference and Voice-face concordance to explain monkeys viewing behavior. We now added second-order interactions in this model, which revealed that the two main factors, face preferences [$F=220.5$, $df=2$, $p=3.10^{-10}$] and voice-face concordance [$F=14.2$, $df=1$, $p=0.005$] remained significant, whereas the animal-tested factor, and the second-order interactions (and especially Concordance x Preference) were not. This contributes to conclude that we do not observe regression to the mean.

2. “Why wasn’t test-animal included as a factor in the initial overall ANOVA that included test face as a factor? Given that the initial one-way ANOVA indicated that the different faces did, indeed, elicit different amounts of attention, it would be interesting to determine whether this effect was modulated by the specific animal tested. Moreover, perhaps rather than

conducting separate analyses with the various factors, why don't the authors first conduct an overall analysis that also includes species (monkey faces, human faces) as another factor to determine whether there are any second-order or third-order interactions? These could then be followed up with simple effects analyses where appropriate. Indeed, the fact that the individual analyses showed that there was a difference in matching of monkey and human faces across different animals suggests that an initial overall ANOVA with all the factors included (test-face, test-animal, and test-face-species) would be the most appropriate way to begin the analyses.”

To ensure a high familiarity with individuals presented in stimuli, stimulus sets were specific for each animal. Therefore we did not use a crossed design that would have enabled to perform an overall ANOVA with test-face, test-animal, and test-face-species as factors, as suggested in this question. However, as presented in the manuscript, we performed an overall ANOVA with three factors 1) specific animal tested (as a random factor), 2) face preference, 3) voice-face concordance. Which revealed that face preference and voice-face concordance significantly explain monkeys behavior, whereas individuals do not. Faces were classified as preferred/not-preferred/neutral on the basis of paired t-test comparisons of each pair of faces presented together.

As suggested by reviewer 2 we now added a fourth factor (Species) to the model as well as second-order interactions. The result remains similar for factors Voice-face concordance [$F=17.321$, $df=1$, $p=0.009$] and Face preference [$F=280.7$, $df=2$, $p=4.10-13$]. Species was not a significant factor [$F=0.83$, $df=1$, $p=0.38$], neither were the second order interactions.

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Résumé

Les visages sont des stimuli essentiels aux interactions sociales complexes des primates. En enregistrant simultanément l'activité neuronale dans l'hippocampe et le cortex inférotemporal chez le singe, nous avons mis en évidence l'existence de cellules pour lesquelles les stimuli optimaux sont des visages. Ces cellules étaient situées non seulement dans le cortex inférotemporal comme précédemment décrit par plusieurs études (pour revue : Tsao and Livingstone, 2008), mais aussi dans l'hippocampe. La sélectivité de ces cellules se révèle relativement plus faible dans l'hippocampe que dans TE, toutefois, des sélectivités pour des sous-dimensions des visages sont représentées dans les deux aires. Premièrement, une grande proportion de cellules des deux aires présente un taux de décharge modulé par l'espèce des visages (homme ou singe). Deuxièmement, TE présente une plus grande proportion de cellules sensibles au point de vue des visages que l'hippocampe. Troisièmement, l'étude des latences d'activité neuronale montre que les réponses aux images d'individus personnellement connus sont plus rapides que celles observées pour des images familières d'individus inconnus. Enfin, dans les deux structures, les réponses aux stimuli acoustiques sont beaucoup plus faibles que celles aux stimuli visuels. En particulier, nous n'avons observé aucune correspondance entre l'identité faciale et vocale des stimuli dans l'une ou l'autre des structures. En résumé, ces résultats suggèrent que les codes pour la représentation des individus sont exprimés dans TE et dans l'hippocampe, de façon redondante et complémentaire. Cette représentation multiple pourrait contribuer à la reconnaissance des individus.

Cette étude est en cours de préparation pour une publication sous le titre :
Sliwa J, Duhamel J-R, Planté A, Wirth S. Comparing face-cells properties in the monkey hippocampus and TE.

B.1. Article n°2: Comparing face-cells properties in the monkey hippocampus and TE

Julia Sliwa, Jean-René Duhamel, Aurélie Planté, Sylvia Wirth

Centre de Neurosciences Cognitives, CNRS-Université de Lyon, UMR5229, 69675 Bron, France

Key-words: Faces, voices, medial temporal lobe, inferotemporal cortex, primates, single-unit, neuronal latencies, *Macaca mulatta*

Pages : 17

Figures : 12

Table : 1

Comparing face-cells properties in the monkey hippocampus and TE

Abstract

Faces are crucial stimuli for the sophisticated social interactions exhibited by primates. By simultaneously recording neuronal activity from the hippocampus and the inferotemporal cortex, we observed the existence of face preferring neurons not only in the inferotemporal cortex as previously described by several studies (for a review see: Tsao and Livingstone, 2008), but also in the hippocampus. Tuning of these cells was relatively broader in the hippocampus than in TE, however in both structures fine tuning toward facial sub-dimensions was also observed. First a large amount of these cells modulated their firing rate according to the species nature of the faces, in both structures. Second, more cells in TE were sensitive to the point-of-view of the faces than in the hippocampus. Third, a faster response was observed for personally known individuals than to unknown individuals in the hippocampus and to a lesser degree in TE, showing that personally relevant knowledge about other individuals speeds facial processing. Finally in both structures responses to acoustic stimuli were much weaker than those to visual stimuli. In particular no correspondence between facial and vocal information was observed in either structure. In summary these results suggest that codes for representation of individuals are replicated in TE and the hippocampus in a redundant as well as complementary manner. This multifaceted representation might well contribute to the recognition of individuals.

Key-words

Faces, voices, medial temporal lobe, inferotemporal cortex, *Macaca mulatta*, primates, single-unit, neuronal latencies

Introduction

Faces are processed by a dedicated and distributed neural system constituted of several areas preferentially activated by faces over other objects in both monkey (Tsao et al., 2003; Pinsk et al., 2005; Tsao et al., 2008; Ku et al., 2011) and human brains (Kanwisher et al., 1997; McCarthy et al., 1997; Haxby et al., 2000; Ishai et al., 2005). Faces are complex stimuli carrying essential social information about an individual's identity, gender, age, dominance status, species, kinship, familiarity, sexual state, attractiveness, health state, etc. and also information related to communication through language, emotions or biological motion including gazing (Ghazanfar and Santos, 2004; Leopold and Rhodes, 2010). Thus the wealth of information processing that faces require might differentially implicate distinct face-responsive areas. Specificities of several of the face-responding areas have been investigated through neuronal recordings either directly guided by fMRI activations (Tsao et al., 2006; Freiwald et al., 2009; Freiwald and Tsao, 2010; Bell et al., 2011) or based on earlier electrophysiological mapping carried in monkeys (Sugase et al., 1999; Gothard et al., 2007; Eifuku et al., 2010; Eifuku et al., 2011). They revealed for instance that facial representations are encoded into a face-map in a norm-based manner (Leopold et al., 2006; Freiwald et al., 2009) and become increasingly view-invariant along the temporal lobe's anteroposterior axis (Perrett et al., 1985; De Souza et al., 2005; Freiwald and Tsao, 2010) and dorsoventral axis (Hasselmo et al., 1989; Eifuku et al., 2004).

At the furthest end of the temporal visual stream, the hippocampus is one of the areas presenting enhanced activity to faces in functional imaging studies (Ishai et al., 2005; Ku et al., 2011). Interestingly voices also elicit an enhanced activation of the hippocampus when perceived separately (Gil-da-Costa et al., 2004) or together with faces (Holdstock et al., 2010; Joassin et al., 2011; Love et al., 2011). Since voices represent the principal auditory cues indicating social information about an individual and provide communication signals, they mirror facial information (Belin et al., 2004; Belin, 2006). And indeed monkeys perceived them as associated to faces in many ways (Ghazanfar and Logothetis, 2003; Jordan et al., 2005; Adachi et al., 2006; Ghazanfar et al., 2007; Adachi and Hampton, 2011; Sliwa et al., 2011).

Therefore we were interested in understanding how faces and voices are represented in the macaque hippocampus at the neuronal level. First, we wondered whether the existence of face-cells can be demonstrated. And if so we aimed at analyzing them in two directions: Is their tuning broad or does it represent finer dimension? Are they activated by voices also? To investigate these issues, we simultaneously recorded neuronal activity from the hippocampus and the inferotemporal cortex in order to compare both areas' contribution to face-selectivity, face characterization and face-voice association, while probed by a large dataset in which faces were varying along five dimensions: identity, species (human/monkey), gender (female/male), point-of-view (-30°/30°/front) and familiarity (personally known/unknown). This set was mirrored by an acoustic set presenting voices with the same properties excepting the point-of-view dimension.

Materials and Methods

Subjects. Two adult rhesus monkeys (*Macaca mulatta*, 8.5 kg, monkey Y and 13 kg, monkey O) participated in this study. They were socially housed in rooms of three or four individuals since they arrived at the Centre de Neurosciences Cognitives. All experimental procedures were in accordance with the regulations of local authorities (Direction Départementale des Services Vétérinaires, Lyon, France) and the European Community standards for the care and use of laboratory animals (European Community Council Directive, 1986, Ministère de l'Agriculture et de la Forêt, Commission Nationale de l'expérimentation animale).

Surgical preparation. Animals were prepared for chronic recording of single-neuron activity in the right hippocampus and right TE (**Fig. 1A**). Anesthesia was induced with Zoletil 20 (6 mg/kg) and maintained under isoflurane (2.5%) during positioning of cilux head-restraint post and recording chamber (Crist Instruments, Damascus, MD). Animals were given atropine (0.25mg/kg) to prevent excessive salivation. Analgesia was provided by a presurgical buprenorphine injection (0.2mg/kg). The position of the recording chamber for each animal was calculated using stereotaxic coordinates derived from each animal presurgical anatomical magnetic resonance images (MRI). Postsurgical MR images were used to finely monitor recording locations within the hippocampus and TE during each experiment (**Fig. 1A**).

Stimuli. Visual and auditory stimuli were color photographs and audio-recordings of individuals and objects familiar to the subjects (including subject itself) or not personally familiar to the subjects but which pictures have been presented during 1 month prior to the experiment on a daily basis to the monkeys. Monkeys were also familiarized with a set of fractal images (Fractal Explorer, fractals.da.ru) and synthetic abstract audio-samples. Thus this second group of stimuli was not personally familiar but was visually familiar to the monkeys. To simplify interpretation of

group assignment further in the manuscript, we will refer to these stimuli as “unknown”, in contrast to the stimuli picturing personally familiar stimuli which will be referred to as “known”. Familiar simian individuals were the two or three adult rhesus macaques housed in the same room as the subjects for 2 to 3 years prior to test. Familiar human individuals were one caregiver and two experimenters working with the animals on a daily basis during the same 2 to 3 year period. Familiar objects were objects present in the colony rooms (pole, primate chair and hosepipe). Each stimulus set was specific to each animal to ensure a high familiarity with individuals presented in stimuli. Front view photographs were cropped to only include the face, head and neck, and then adjusted to a common size and presented on a black background (GNU Image Manipulation Program, www.gimp.org). Humans were presented without mask and goggles as the colony rooms were equipped with windows through which animals could see the experimenter and staff without their protections. There were three pictures per individual (front, three-quarter-view to the left, three-quarter-view to the right) and one picture per objects and fractals, leading to 18 human, 18/21 monkey, 6 object and 3 fractal pictures per animal tested. Pictures shown in **Fig. 1** and **Fig. 5-12** are provided for illustrative purposes but do not correspond to the actual stimuli, which included for all animal pictures the head implants (head posts and recording chambers). Pictures used otherwise resemble the illustrative ones in all respects (color, size, gaze direction). Auditory stimuli consisted of audio-recordings of 2s duration, each containing either a sample of ‘coo’ vocalization ($894 \text{ ms} \pm 300$) or a human speech extract ($877 \text{ ms} \pm 380$). Human speech extracts consist of small sentences or words in French such as “Bonjour tout le monde”/“Hello everybody”, or “Voilà”/“Here it is”. Three audio-samples from each individual and one per object and synthetic abstract audio-samples were presented; leading to 18 human voice stimuli, 18 monkey vocalizations, 6 object and 3 synthetic abstract audio-samples per animal tested. The mean sound pressure level for the duration of each audio-sample was calculated. Then the 42 audio-samples for each subject were normalized for this mean acoustic intensity (MATLAB, MathWorks, Natick, MA).

Behavioral procedure. Subject was head restrained and placed in front of, and at eye level to, a LCD screen (Dell Computers) situated at 56 cm. The subject’s gaze position was monitored with an infrared eye-tracker (ISCAN) with a 250 Hz sampling rate. Behavioral paradigms, visual displays, eye position monitoring and reward delivery were under the control of a computer running a real-time data acquisition system (REX) (Hays et al., 1982). Visual stimuli were presented at $10^\circ \times 13^\circ$ on the center of the screen (**Fig. 1B**). The virtual exploration window, comprised of the picture surrounded by approximately 2° black surround, measured $12^\circ \times 15^\circ$ of visual angle. Auditory stimuli were presented at an intensity of 50-65 dB (A-weighted) sound pressure level (SPL) at the subject’s ear as measured with a Brüel and Kjær 2239A Integrating Sound Level Meter (www.bksv.com) from two speakers located at 56 cm in front of the subject and symmetrically 45 cm apart. The subject was to start by directing its gaze to a small white square presented on the center of the screen. Then it was required to fixate ($\pm 1^\circ$ - 4°) during 0.5 s. Two types of test trials were randomly interleaved: trials with visual stimuli and trials with auditory stimuli. In visual trials, after the 0.5 s fixation, a picture was presented at the center of the screen. The subject could freely explore the picture during 2 s, so long as its gaze was maintained within the boundaries of a virtual window around the picture corresponding to the black area in figure 1B. In audio trials, after the 0.5 s fixation, the subject was required to continue fixating ($\pm 1^\circ$ - 4°) during 2 s while an audio-sample was played. During inter-trial periods, juice reward was given to ensure monkeys’ attention to the computer monitor where visual

stimuli were displayed and their motivation to complete trials and not to provide reinforcement for a particular exploration pattern.

Electrophysiological Recordings. Single-neuron activity was recorded extracellularly with tungsten quartz insulated tetrodes (1.5-2 M Ω ; Thomas Recording, Giessen, Germany) or tungsten single microelectrodes (1–2 M Ω ; Frederick Haer Company, Bowdoinham, ME) which were lowered into the hippocampus with a NAN microdrive (Plexon Inc., Dallas, TX). The microelectrodes were inserted through 24-gauge stainless steel guide tubes set inside a delrin grid (Crist Instruments, Damascus, MD) adapted to the recording chamber. The electrode signal was preamplified using a Plexon system (Plexon Inc., Dallas, TX) and digitized using RecorderEV3 software (Mediane System, Le Pecq, France). Continuous signal was sampled at 20 kHz and stored with the RecorderEV3 software along with behavioral and stimulation data for off-line analysis. Single-units were sorted offline using the Offline Sorter software (Plexon Inc., Dallas, TX). We used a semi-manual sorting method such that three selected parameters (principal components, waveform patterns, and amplitude of the peak in each electrode of the tetrode) allowed us to separate units from background activity and yield well-isolated clusters. Because with tetrodes the signal is recorded simultaneously by four adjacent microwires, the location of the neurons can be triangulated, and neurons presenting similar amplitude on one wire can be distinguished easily from one another by comparing their recorded amplitude on the three others wires. Thus the use of tetrodes enable to rely on several additional dimensions to correctly separate the different units from each other as compared to the available ones when using conventional single electrodes and provide a superior quality of single-unit isolation (Gray et al., 1995).

Cells selectivity analysis. Data were analyzed with custom written scripts and the Statistics Toolbox in Matlab (MathWorks, Natick, MA). Neurons with very low firing rates (<0.4Hz, silent cells, (Thompson and Best, 1989) were not included in the following analysis due to their lack of reliability when tested with the following methods. The peak of the stimulus activity was defined separately for each stimulus modality, as the extremum of their trial-averaged normalized responses. Then stimulus activity was defined as the response amplitude of the firing rate in a 500ms window centered on the time of the peak stimulus activity, separately for auditory and visual modalities. This method was used to take into account the use of dynamic acoustic stimuli, which can lead to various time course responses in different neurons. A response was considered significant if it was larger than the mean plus 4 standard deviations of the baseline activity and significantly different from baseline activity ($p < 0.05$, t-test).

General trend in activity of the responsive units' population was first assessed by using a principal component analysis with observations being the number of stimuli from the set and variables being each responsive unit's z-score for a given stimulus. Z-scores for each stimulus for each neuron were calculated as the mean of the response divided by its standard deviation. Stimuli z-scores were projected on the first two principal components and then clustered using a Gaussian mixture model, which is based on an expectation-maximization (EM) algorithm. Neurons were first classified according to their responses to the two general dimensions along which the stimuli were varying: Individual-Object and Visual-Auditory. Neurons were first defined as face-, object-, voice- or sound-responsive if their activity to at least one stimulus of these categories was significant. A neuron could belong from zero to four of these categories. Neurons were then defined as face-, object-, voice- or sound-selective if additionally their activity was not significant for stimuli of the other category in the same modality. The selectivity

of their response was assessed by quantifying the number of stimuli eliciting an activity greater than the half-maximum activity of that cell (Perrodin et al., 2011) and also by computing a sparseness index S (Rolls and Tovee, 1995) as follows:

$$S = \frac{(\sum_{i=1}^N r_i / N)^2}{\sum_{i=1}^N r_i^2 / N}$$

with r_i being the response of the neuron to the i^{th} stimulus, and N being the total number of stimuli.

Further only selective and responsive neurons presenting an enhanced firing rate for their preferred stimuli, and not a suppressed/inhibited firing rate, were included in the analyses.

Stimuli varied along several sub-dimensions such as: Identity, Monkey-Human, Familiar-Unknown, Male-Female, and Face-view (0°, 30°, -30°). To take these into account, first, a generalized linear model (GLM) with four factors (Familiar-Unknown, Monkey-Human, Male-Female, and Face-view (0°, 30°, -30°)) was performed on visual subset of stimuli for each face-responsive neuron. Similarly, a generalized linear model (GLM) with three factors (Familiar-Unknown, Monkey-Human, and Male-Female) was performed on acoustic subset of stimuli for each voice-responsive neuron. Second, a generalized linear model (GLM) with one factor (Identity) was performed separately on visual or acoustic subsets of stimuli for each face-responsive or voice-responsive neuron respectively. Visual invariance, i.e. a preferred firing to the three pictures of a given person, was quantified with a receiver operating characteristics (ROC) analysis. The hit rate corresponded to the median number of spikes for each of the three pictures of an individual, and the false positive rates to the median number of spikes elicited by each of the other visual stimuli of the set. Confidence intervals were estimated using a bootstrap of the stimuli responses.

Comparing cells' selectivity and latency across brain regions. The number of cells responding for the different dimensions and sub-dimensions was compared across both brain regions with χ^2 -tests. To calculate response latencies, spike trains were smoothed by convolution with a Gaussian kernel ($\sigma=10\text{ms}$) to obtain the spike density function (SDF). The baseline activity was defined as the average activity during 500ms preceding stimulus onset during fixation. The latency of the response was defined as the time point when the SDF between 60 and 1000 ms first exceeded a level +2 standard deviations from the baseline activity, for the optimal stimulus. Then latencies of activation were compared between the two regions with Kolmogorov-Smirnoff tests and analyses of variance.

Results

Cells responsiveness for modalities. The activity of 343 neurons was recorded in two monkeys in the hippocampus (188 cells: 99 in monkey O and 89 in monkey Y; **Fig 1A**) and in area TE (155 cells: 82 in monkey O and 73 in monkey Y; **Fig 1A**), while monkeys were passively perceiving color photographs or audio-recordings of individuals and objects familiar to them (including themselves) or not personally familiar (**Fig. 1B**). Naturalistic stimuli elicited robust responses throughout the sampled regions, with 74 (39%) hippocampal neurons selective to at least one stimulus in either modality ($p<0.05$, median FR>+4SD of the baseline and with a firing rate superior to 0.4Hz) and 59 (38%) inferotemporal neurons (**Fig. 1C**). Neurons were classified as

visual, auditory, and audio-visual (**Fig. 1C, 1D**). In the hippocampus, 32 cells (17%) were visual, 30 (16%) were audio-visual and 11 (6%) were auditory. In TE, 37 cells (24%) were visual, 22 (14%) were audio-visual, and 1 (1%) was auditory. At the level of the population, visual and auditory stimuli were clearly segregated by neuronal activities in both regions, as assessed by a clustering with a Gaussian mixture model of the projection of the stimulus related z-scores, defined as activities' amplitude divided by their standard deviation, on their first two principal components (**Fig. 1F**). Dynamics of neuronal activities also presented dissimilarities across modalities in both regions, with more neurons reaching their peak of activity early in the stimulus presentation period for visual stimuli while peaks of activity were widely distributed along the auditory stimuli presentation (**Fig. 1E**).

Cells responsiveness for faces and voices. The Bayes information of the Gaussian mixture clustering was minimized by a three-component model in TE revealing that, not only auditory and visual stimuli but also face and object stimuli are segregated by neuronal activities. In HPC, the best clustering was obtained along the dual visuo-auditory dimension. However, one can note that points corresponding to objects (crosses) were not included in the visual cluster suggesting that visual stimuli were not coded in a homogeneous manner and may be parted further in faces and objects (**Fig. 1F**). On the contrary, voice and sound stimuli clustered together in either region, indicating a major role of both regions in segregation between visual stimuli rather than auditory stimuli. Further we defined neurons as face-, or object-responsive if their activity to at least one stimulus of these categories was significant. Even though at population level auditory segregation was not a main effect, we quantified the number of voice- and sound-responsive neurons to draw a parallel with visual activities. Thus a neuron could belong from zero to four of these categories. In the hippocampus, 59 (31%) and 40 (21%) neurons were respectively face- and voice-responsive (**Fig 2A, B**). In TE, 56 (36%) and 22 (14%) were respectively face- and voice-responsive (**Fig 2A, B**). Alongside 42 (22%) and 25 (13%) neurons were respectively object- and sound-responsive in the hippocampus while 47 (30%) and 12 (8%) were respectively object- and sound-responsive in TE. None of these proportions differ between the hippocampus and TE.

Cells selectivity for faces and voices. Face only selectivity was defined with a stricter criterion: face-selective cells did not respond to objects and vice versa, alongside voice-selective cells did not respond to sounds and vice versa. Hence, 20 (11%) and 17 (9%) neurons were respectively face- and voice-selective in the hippocampus (**Fig 2A**). Conversely 3 (2%) and 2 (1%) neurons were object- and sound-selective respectively in the hippocampus. In TE, 10 (6%) and 10 (6%) were respectively face- and voice-selective and 1 (0.6%) and 0 were respectively object and sound-selective. Thereby in both regions, face- and voice-responding cells that were not face- and voice-selective also responded to other stimuli from the same or from the other modality. In the following sections, only selective and responsive neurons presenting an enhanced firing rate for their preferred stimuli, and not a suppressed/inhibited firing rate, were analyzed.

Properties of face-responsive neurons. To further investigate the selectivity of face-responsive cells, we used a generalized linear model analysis (GLM) with four social factors characterizing face stimuli in our experimental set: 1. Monkey-Human, 2. Known-Unknown, 3. Male-Female, 4. Face-view (0°, 30°, -30°). We found that, 16 (34%) of the 47 hippocampal neurons and 26 (60%) of the 48 inferotemporal neurons responded in a differential fashion to at least one of these four factors (**Table 1**). Similarly for face-selective cells, 4 (31%) out of 13 hippocampal neurons and 5

(63%) out of 8 inferotemporal neurons responded in a differential fashion to at least one of these four factors (**Table 1**). In both structures, the Monkey-Human factor led to the largest proportion of cells for which activity was significantly explained by the GLM. The proportion of responses to the four factors as a whole differed significantly between HPC and TE (responsive cells: $p=10^{-5}$, selective cells: $p=4.10^{-9}$, χ^2 test). This effect was mainly driven by a higher proportion of responses to Face-view factor in TE than in HPC ($p=0.012$, χ^2 test in face-responsive cells).

Properties of voice-responsive neurons. To further investigate the selectivity of voice-responsive cells, we used a generalized linear model analysis (GLM) with three social factors characterizing voice stimuli in our experimental set: 1. Monkey-Human, 2. Known-Unknown, 3. Male-Female. We found that only 3 (9%) of the 32 hippocampal neurons and 2 (10%) of the 20 inferotemporal neurons responded in a differential fashion to at least one of these three factors (**Table 1**). These proportions were significantly smaller than the proportions observed for face-responsive neurons (HPC: $p=0.018$, TE: $p=8.10^{-8}$, χ^2 -test). Similarly for voice-selective cells, only 2 (14%) out of 14 hippocampal neurons and none of the 10 inferotemporal neurons responded in a differential fashion to at least one of these four factors (**Table 1**). These proportions were significantly smaller than the proportions observed for face-selective neurons in TE ($p=7.10^{-12}$, χ^2 -test) but not HPC ($p=0.31$, χ^2 -test).

Sparseness of neurons' activity. We analyzed whether face- and voice-selective and responsive cells in both regions were active for most stimuli or for few stimuli by computing a sparseness index, which ranges from 0 (highly-selective or sparse coding) to 1 (non-selective or dense coding). For both types of responsive cells, the mean sparseness index was high in the hippocampus (0.85, 0.76) and in TE (0.75, 0.83) indicating that cells were responding to several visual or acoustic stimuli and not just to isolated stimuli (**Fig. 2C-E**). Similarly for both types of selective cells, the mean sparseness index was high in the hippocampus (0.83, 0.83) and in TE (0.67, 0.85) (**Fig. 2C-E**).

The selectivity of face and voice-responsive cells in the hippocampus and TE was also quantified and compared across regions and modalities by quantifying the number of stimuli eliciting an activity greater than the half-maximum activity of that cell. Face-responsive cells responded to an average of 55% of the visual stimuli in the hippocampus, i.e. 25/45 of the stimuli elicited responses greater than the half-maximum response (**Fig. 2D**). In TE face-responsive cells responded to an average of 36% visual stimuli, i.e. 14/45 of the visual stimuli elicited responses greater than the half-maximum response, which is less than in the hippocampus ($p=10^{-4}$, Wilcoxon test). The same was observed for face-selective cells in both regions (TE: 28%, HPC: 56%, $p=0.014$, Wilcoxon test) (**Fig. 2D**). On the contrary both regions presented similar low selective voice-responsive (Average % of acoustic stimuli eliciting a response in TE: 53%, in HPC: 57%) and voice-selective cells (TE: 53%, HPC: 56%) (**Fig. 2E**). In TE, the average number of visual stimuli eliciting a response was lower than the average number of acoustic stimuli eliciting a response when comparing responsive cells ($p=0.024$, Wilcoxon test) or selective cells ($p=0.036$, Wilcoxon test).

Invariant coding of individuals. A receiver operating characteristics (ROC) analysis was performed, to test invariance of facial identity representation throughout point-of-views for each of the twelve individuals. For each neuron, we calculated the best area under the curve (AUC) of the twelve AUC. Invariance of neuronal activities didn't differ significantly between

brain regions ($AUC_{best,HPC}=0.85$, $AUC_{best,TE}=0.87$, $p>0.05$, two-sided Wilcoxon rank sum test). Each neuron's best AUC was then plotted against each neuron's selectivity (**Fig. 3 left**) or sparseness (**Fig. 3 right**). This resulted in two maps where invariance and selectivity/sparseness of neuronal populations from both brain areas were superposed. Exemplar neurons with different selectivity and invariance rate from the hippocampus are presented in **Figures 5 to 8**. And for comparison, exemplar neurons with different selectivity and invariance rate from TE are presented in **Figures 9 to 12**.

Latencies of neuronal responses to different stimuli categories. A 3-way ANOVA with factors Areas/Species/Familiarity was performed on face-responsive neurons from both brain regions. It showed a main effect of the Familiarity factor ($F=10.32$, $df=1$, $P=0.0016$, **Fig. 4**). There was no main effect of Area ($F=1.73$, $df=1$, $P=0.1896$) and Species factors ($F=0.55$, $df=1$, $P=0.46$) and no significant interactions ($P>0.05$). A 2-way ANOVA with factors Species/Familiarity was also performed on face-responsive neurons separately in each brain regions. It showed again a main effect of the Familiarity factor for face-responsive cells from HPC, but not for face-responsive cells from TE. Moreover a 2-way ANOVA with factor Views/Areas was separately performed and yielded no significant effect or interactions ($P>0.05$).

Discussion

This study shows the existence of face preferring neurons not only in the inferotemporal cortex as previously described by several studies (Tsao and Livingstone, 2008), but also in the hippocampus. The comparison of the cells in the two structures revealed both similarities and differences. Tuning of these cells was relatively broader in the hippocampus than in TE, however in both structures fine tuning toward facial sub-dimensions was also observed. The largest proportion of these cells was found to modulate their firing rate according to the species nature of the faces – monkey or human – in both structures. Second, more cells in TE were sensitive to the point-of-view of the faces than in the hippocampus. Third, a faster response was observed for pictures of personally known individuals than to pictures of unknown individuals (yet, the pictures were familiar) in the hippocampus and to a lesser degree in TE, showing that personal knowledge about other individuals speeds facial processing. Finally in both structures, responses to acoustic stimuli were much weaker than those to visual stimuli. In particular, we found no evidence for cross modal binding between stimuli as no correspondence between facial and vocal information was observed in either structure. In summary these results suggest that these structures could play complementary roles in recognition of individuals.

Response to faces in the hippocampus. First, we report the existence of face-cells in the hippocampus. It shows that hippocampal cells in monkeys code for social visual stimuli and categorize them accordingly, even when monkeys are not actively asked to distinguish these stimuli from others. This conclusion is also reinforced by the existence of non-face preferring cells that are silent to faces but responsive to objects. These results highlight that hippocampal neurons can perform a categorization of a cognitive space constituted of non-spatial visual social stimuli. It can be wondered whether the reported face-selectivity is protocol specific or whether it could last in time. A “remapping” study using similar and different stimuli would possibly provide useful insight into this issue (Colgin et al., 2008). On the contrary our stimulus-set is similar to a set used recently in monkey fMRI (Ku et al., 2011), which enables us to discuss our results in light of their findings. The authors report an enhanced fMRI activity for face stimuli

compared to object stimuli in an area within the monkey anterior hippocampus, among other areas. Therefore, because our neurophysiologic recordings also targeted anterior hippocampus, our recorded face-cells might enclose cells that specifically contributed to the fMRI enhanced activity observed by Ku et al. (2011) Indeed, face-cells displayed topographically clustered organization at the small scale (Article 3, companion article), which might well lead to extended patches of same stimulus preference at the scale of millimeters (Dahl et al., 2009), and further possibly be translated into an enhanced fMRI activity. We also recorded face-cells in the infero-temporal cortex. Existence of neurons presenting preferred responses to faces is long-established in the infero-temporal lobe (Tsao and Livingstone, 2008). However their simultaneous recordings, with the same stimulus set in the same monkeys with the same material and in the exact same protocol conditions, enables us to more directly compare their properties. A first general observation that can be drawn is that general tuning of hippocampal face-cells was broader in the hippocampus than in TE. Moreover at the population level, responses to faces clustered more strikingly relative to objects in TE than in the hippocampus. These observations were well expected since the inferotemporal cortex is considered as specifically dedicated to specific higher visual processing, which is not the case of the hippocampus known to be a more associative area. We consider that the evidence for face-cells in the hippocampus, despite the strong assumption on brain areas specificities in information processing, indicates that face processing might be so important in primates that it might also have colonized at least part of hippocampal processing resources.

Point of view. Another difference we observed between facial processing in the hippocampus and in TE concerns their neurons' processing of faces' point-of-view. Indeed we found that more, but not all, neurons from TE responded differentially to facial point-of view than in the hippocampus. It has been previously described in TE that facial representations become increasingly view-invariant along the temporal lobe's anteroposterior axis (Perrett et al., 1985; De Souza et al., 2005; Freiwald and Tsao, 2010) and dorsoventral axis (Hasselmo et al., 1989; Eifuku et al., 2004). In our study face-cells have been recorded in the superior part of TE, i.e. in the interior inferior bank of STS sulcus, mainly between 10-11mm on the antero-posterior axis (AP, relative to the interaural line, **Fig. 2B**) in one monkey and 6-10mm AP in the other. Given the heterogeneity between monkey brains and in particular between face-patches locations (Tsao et al., 2008), we can only state that our recordings have been performed in or in between locations of MF and AF face-patches. In these regions face-cells have been shown to be sensitive to facial views, contrary to face-cells from the anterior AM patch, which are view-invariant. Our results are therefore in accordance with previous findings (Freiwald and Tsao, 2010). Interestingly hippocampal face-neurons are mainly lacking this property, indicating either that hippocampus receives its facial inputs from a region representing faces in a view-invariant manner, i.e. more anterior, or that the hippocampal neurons actively associate different views together. We also analyzed latencies of neuronal responses to the different views. However they didn't differ between face-views as could have been suggested from a previous study carried in STS (Eifuku et al., 2004).

Effect of memory on the speed of stimulus processing. Latencies of neuronal responses differed with faces' familiarity. Known human and monkey faces elicited shorter latencies of responses than unknown human and monkey faces. This effect was still observed when testing hippocampal neurons alone. It was no longer observed when testing inferotemporal neurons alone; however the interaction between familiarity factor and brain regions was not significant.

Thus it did not enable us to conclude on a difference between the two regions (Nieuwenhuis et al., 2011). Nevertheless, it supposes that the found global effect was mainly driven by the difference in latency occurring in the hippocampus. All things being equal, this effect is interesting since it denotes that personally familiar faces elicit something more than visually familiar pictures of unknown individuals, and that this supplemental component speeds their processing. This component is not related to a difference in repeated experience with the pictorial material, since both pictures of personally known and unknown individuals have been seen by monkeys hundreds of times prior to the recording sessions. On the contrary, this component might encompass some of the following elements that define personal knowledge in humans: emotional responses, memory of associated distinctive features from other modalities, knowledge of prior real-world interaction history, theory of mind, factual knowledge about social status, etc (Haxby et al., 2000; Gobbini and Haxby, 2007). The speeded processing for personally familiar individuals by the hippocampus might arise from a fast processing path following subcortical structures such as the amygdala. Indeed the amygdala, along with other structures, have been shown to be specifically activated by faces of personally familiar individuals in fMRI (Shah et al., 2001; Sugiura et al., 2001; Rotshtein et al., 2005; Denkova et al., 2006; Gobbini and Haxby, 2007). On the contrary, we did not observe that more hippocampal neurons were activated by personally familiar faces than unfamiliar faces, as suggested by electrophysiological studies carried in humans (Viskontas et al., 2009), neither that processing of individuals' identity in TE in a view-invariant manner depended on the familiarity with the presented faces (Eifuku et al., 2011).

Coding for species. Another factor that was significantly modulating facial representations in both regions was that of species. A large amount of neurons in both regions were distinguishing between monkey and human faces, either preferring one type of faces or another. Such a distinction has also been observed in the monkey amygdala (Gothard et al., 2007), and monkey temporal cortex (Sigala et al., 2011). On the contrary we did not observe differences in response latencies between monkey and human faces, which is consistent with previous findings in the temporal cortex (Kiani et al., 2005).

Facial identity. In both structures, we observed neurons signaling individuals' identity through a view-invariant coding. However, most neurons responded to more than one individual. This result appears therefore in contradiction with recent findings from human hippocampal neurons (Quiroga et al., 2005; Waydo et al., 2006; Quiroga et al., 2008) suggesting that representation of individuals relies on ultra-sparseness through the selectivity of a few cells. This finding may convey the idea that neighboring cells are silent which allows this ultra-sparse signal to be read out. However, our recordings, together with previous human studies (Fried et al., 1997; Kreiman et al., 2000; Fried et al., 2002), show that neighboring cells are not silent and also show selective properties. In fact our results show that a much larger proportion of cells than 10% (40%!) do respond to stimuli and exhibit a medium range of selectivity (0.5 or 0.6). Those cells also include pyramidal cells (companion article, Article 3). Thus it raises the two following issues: how can we account for the fact that some cells respond ultra-sparsely, but concurrently to cells that respond to stimuli in a less selective way? Can the brain tune to only relevant cells, or is it using whole information including that of less selective cells, which code other aspect of the stimuli than its mere identity?

Weak responses to voices in the anterior hippocampus. We investigated whether the principal auditory cues indicating person identity, which are voices, are processed similarly to faces by hippocampal neurons. This approach was supported by our previous behavioral finding that rhesus monkeys, similarly to humans, link vocal and facial identity cues into one cognitive representation (Sliwa et al., 2011), by fMRI findings showing an integration of identity throughout modalities by hippocampal structure (Joassin et al., 2011; Love et al., 2011) and finally by findings of enhanced PET activities for voices in the monkey hippocampus (Gil-da-Costa et al., 2004). However, fewer cells were responding to acoustic stimuli than to visual stimuli and we only observed weak hippocampal neuronal responses to voices, which were moreover not mirroring visual selectivity of the cells, as observed by the small amount of responses to the GLM analysis on auditory stimuli. Neuroanatomical studies carried in monkeys present that whereas higher visual areas are densely and directly connected to the hippocampus; higher auditory areas are more sparsely and only indirectly connected to the hippocampus through polymodal areas (Mohedano-Moriano et al., 2007; Mohedano-Moriano et al., 2008; Munoz-Lopez et al., 2010). Thus the hippocampus might represent the furthest end of the temporal visual stream but not of the temporal auditory stream, which might explain its weaker response to voices and auditory stimuli in general.

Weak responses to voices in TE. We postulated that voices might be represented in the temporal cortex, based on imagery studies carried in monkeys and humans showing activation of inferotemporal cortical visual areas by voices of unknown monkeys (Gil-da-Costa et al., 2004) or of known humans respectively (von Kriegstein et al., 2005; von Kriegstein and Giraud, 2006; von Kriegstein et al., 2006; von Kriegstein et al., 2008; Blank et al., 2011). However voices and auditory stimuli in general elicited weak responses in TE, even weaker than in the hippocampus. This could not be explained by the fact that monkeys wouldn't perceive the association between heard voices and seen faces since our and others previous studies have shown that monkey spontaneously perform cross-modal individual recognition of both known monkeys and known humans (Adachi and Fujita, 2007; Adachi and Hampton, 2011; Sliwa et al., 2011). Since BOLD signal might be only weakly correlated to single unit activity (Sirotin and Das, 2009), an analysis of local field potential signals might reveal a population coding of voices in those regions (Logothetis et al., 2001; Viswanathan and Freeman, 2007; Logothetis, 2008).

What is the functional meaning of the hippocampal activity for faces? In summary we present that facial information, which follows a path along the temporal lobe, eventually reaches the hippocampus. Once it reached the hippocampus, the information is globally less selective, less view-invariant, but has kept important information about species. The hippocampus colors this facial information with a signal about personal familiarity that might arise from another structure, perhaps a subcortical one such as the amygdala which is anatomically close and also represents facial information (Gothard et al., 2007). Future studies might investigate the functional role of the hippocampal activity specific to faces. Is it mainly related to the encoding of upcoming information related to the viewed face? Or does it play a role in the recall of other information related to the pictured individual? Subsequent studies using active tasks and inactivation techniques might contribute to answer these issues.

Figures legends

Figure 1. A. Location of the recording sites and physiological properties of hippocampal cells. Top: Three dimensional representation of the monkey brain highlighting the hippocampi (in red) and the boundaries of the recording chamber (vertical lines). Bottom: selected MRI sections along the anteroposterior axis showing the boundaries of the recording chamber and the track of the electrode (middle section). **B.** Visual and acoustic stimuli presentation task. Monkeys initiated trials by fixating a central fixation spot for 500ms. Following this, a sound was played on acoustic trials while the monkey kept fixating the spot. On visual trials, a picture was displayed which the animal could explore within its boundaries. **C.** Proportion of responsive cells in the HPC (top) and TE (bottom). **D.** Single cells examples in the HPC (top row) and in TE (bottom) of visually selective cells (left panel), audiovisual cells (middle panels) and auditory cells (right). **E.** Distribution of peak latencies in HPC (left) and TE (right) for visual (red) and acoustic (blue) stimuli. **F.** Projection and clustering of population responses to stimuli based on each neurons' z-scores for each event in HPC (left) and TE (right).

Figure 2. A. Proportions of responsive cells in HPC and TE. **B.** Functional maps of the recording sites in TE and the HPC in both monkeys. Medio-lateral (ML), dorso-ventral (DV) and anteroposterior (AP) coordinates are counted in mm from interaural line. Plain circles indicate sites where neurons with visual responses were recorded. Empty circles indicate sites where neurons with auditory responses were recorded. Circles colors follow the key presented in A. **C.** Normalized responses of each face-responsive cell from HPC (left) and TE (right) to each of the face stimuli to the maximum and minimum firing rate of each cell. Cells are sorted according to their mean normalized firing rate (y-axis) and responses to stimuli are sorted within each line, according to their normalized firing rate (x-axis). **D.** Left panels: Proportions of cells relative to the percentage of faces eliciting responses larger than half maximum in HPC and TE for the selective cells (top) and the responsive cells (bottom). Vertical lines represent the averages. Right panels: Average of the Sparseness index for selective cells (top) and responsive cells (bottom). Errors bars represent SEM (*P < 0.05). **E.** Same as in B, but for voices.

Figure 3. Invariance vs. selectivity for the population of cells from HPC (blue) and TE (red). Left histogram depicts distribution of best AUC for each face-responsive cell. Bottom middle histogram depicts distribution of selectivity calculated as the percentage of stimuli eliciting a response superior to half the maximum of the cell's response (same as in **Fig. 2.D**). Bottom right histogram depicts distribution of sparseness index as defined in the Method section. Invariance is growing when moving up in the scatter plot and selectivity is growing when moving left.

Figure 4. A. Single cell example illustrating the visual latency (red line) calculated to four classes of stimuli: known humans, known monkeys, unknown humans and unknown monkeys. Latency was calculated as the first data point for which the cell's activity exceeded the mean baseline activity plus 2SDV (green line) for 30 consecutive milliseconds. **B.** Latencies for classes of stimuli (human/monkey and known/unknown) in HPC and in TE. The boxplot borders represent the 25th and 75th percentile and the red line indicates the median.

Figures 5. Exemplar face-responding neuron from the hippocampus, showing high invariance and high selectivity. Beneath each picture, is represented a raster of the neuron's activity during baseline (from -1 to 0 s, time calculated from stimulus onset) and presentation (from 0 to 2s) of this picture. Each line represents a single trial and each tick represents an action potential. The blue ticks represent the stimulus onset at 0 ms and the green ticks represent stimulus offset à 2

s. Beneath each raster, is represented a peristimulus histogram (PSTH) summing neuron's activity among trials. The scale is represented in Hz (i.e. in number of spikes per seconds).Neuron's activities for acoustic stimuli are not represented.

Figures 6. Example of a face-responding neuron from HPC, showing high invariance and medium selectivity, same conventions as in Figure 5.

Figures 7. Example of a face-responding neuron from HPC, showing high invariance and low selectivity, same conventions as in Figure 5.

Figures 8. Example of a face-responding neuron from HPC, showing low invariance and high selectivity, same conventions as in Figure 5.

Figures 9. Example of a face-responding neuron from TE, showing high invariance and high selectivity, same conventions as in Figure 5.

Figures 10. Example of a face-responding neuron from TE, showing high invariance and medium selectivity, same conventions as in Figure 5.

Figures 11. Example of a face-responding neuron from TE, showing high invariance and low selectivity, same conventions as in Figure 5.

Figures 12. Example of a face-responding neuron from TE, showing low invariance and high selectivity, same conventions as in Figure 5.

Table 1

	% of face-responsive cells		% of face-selective cells		% of voice-responsive cells		% of voice-selective cells	
	<i>HPC</i>	<i>TE</i>	<i>HPC</i>	<i>TE</i>	<i>HPC</i>	<i>TE</i>	<i>HPC</i>	<i>TE</i>
<i>Known-Unknown</i>	6.39	18.75	7.69	25	3.12	0	0	0
<i>Monkey-Human</i>	21.28	39.58	15.38	37.5	6.25	5	14.29	0
<i>Female-Male</i>	8.51	16.67	0	12.5	0	5	0	0
<i>View (0°,30°,-30°)</i>	6.38	24.99	7.69	25				
<i>None</i>	65.96	39.58	69.23	37.5	90.62	90	85.71	100

References

- Adachi I, Fujita K (2007) Cross-modal representation of human caretakers in squirrel monkeys. *Behav Processes* 74:27-32.
- Adachi I, Hampton RR (2011) Rhesus monkeys see who they hear: spontaneous cross-modal memory for familiar conspecifics. *PLoS ONE* 6:e23345.

- Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates* 47:350-354.
- Belin P (2006) Voice processing in human and non-human primates. *Philos Trans R Soc Lond B Biol Sci* 361:2091-2107.
- Belin P, Fecteau S, Bedard C (2004) Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 8:129-135.
- Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RBH, Ungerleider LG (2011) Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *The Journal of Neuroscience* 31:12229-12240.
- Blank H, Anwander A, von Kriegstein K (2011) Direct structural connections between voice- and face-recognition areas. *The Journal of Neuroscience* 31:12906-12915.
- Colgin LL, Moser EI, Moser MB (2008) Understanding memory through hippocampal remapping. *Trends Neurosci* 31:469-477.
- Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924-11932.
- De Souza WC, Eifuku S, Tamura R, Nishijo H, Ono T (2005) Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *J Neurophysiol* 94:1252-1266.
- Denkova E, Botzung A, Manning L (2006) Neural correlates of remembering/knowing famous people: An event-related fMRI study. *Neuropsychologia* 44:2783-2791.
- Eifuku S, De Souza WC, Tamura R, Nishijo H, Ono T (2004) Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J Neurophysiol* 91:358-371.
- Eifuku S, Nakata R, Sugimori M, Ono T, Tamura R (2010) Neural correlates of associative face memory in the anterior inferior temporal cortex of monkeys. *The Journal of Neuroscience* 30:15085-15096.
- Eifuku S, De Souza WC, Nakata R, Ono T, Tamura R (2011) Neural representations of personally familiar and unfamiliar faces in the anterior inferior temporal cortex of monkeys. *PLoS One* 6:e18913.
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845-851.
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187-1196.
- Fried I, MacDonald KA, Wilson CL (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18:753-765.
- Fried I, Cameron KA, Yashar S, Fong R, Morrow JW (2002) Inhibitory and excitatory responses of single neurons in the human medial temporal lobe during recognition of faces and objects. *Cereb Cortex* 12:575-584.
- Ghazanfar AA, Logothetis NK (2003) Neuroperception: facial expressions linked to monkey calls. *Nature* 423:937-938.
- Ghazanfar AA, Santos LR (2004) Primate brains in the wild: the sensory bases for social interactions. *Nat Rev Neurosci* 5:603-616.
- Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17:425-430.
- Gil-da-Costa R, Braun A, Lopes M, Hauser MD, Carson RE, Herscovitch P, Martin A (2004) Toward an evolutionary perspective on conceptual representation: species-specific calls activate

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

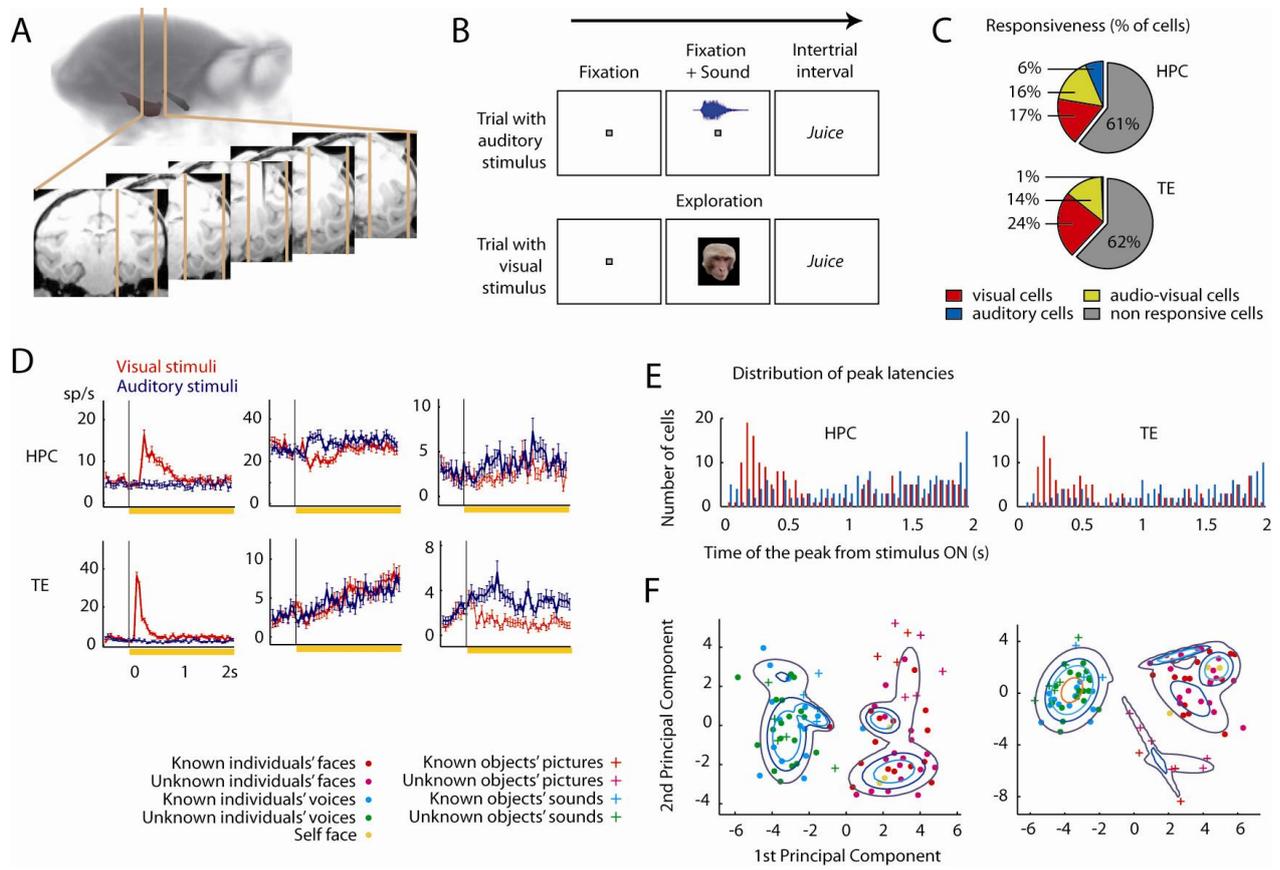
- visual and affective processing systems in the macaque. *Proc Natl Acad Sci U S A* 101:17516-17521.
- Gobbini MI, Haxby JV (2007) Neural systems for recognition of familiar faces. *Neuropsychologia* 45:32-41.
- Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG (2007) Neural responses to facial expression and face identity in the monkey amygdala. *Journal of Neurophysiology* 97:1671-1683.
- Gray CM, Maldonado PE, Wilson M, McNaughton B (1995) Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *Journal of Neuroscience Methods* 63:43-54.
- Hasselmo ME, Rolls ET, Baylis GC (1989) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res* 32:203-218.
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223-233.
- Holdstock JS, Crane J, Bachorowski JA, Milner B (2010) Equivalent activation of the hippocampus by face-face and face-laugh paired associate learning and recognition. *Neuropsychologia* 48:3757-3771.
- Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network. *Brain Res Bull* 67:87-93.
- Joassin F, Pesenti M, Maurage P, Verreclt E, Bruyer R, Campanella S (2011) Cross-modal interactions between human faces and voices involved in person recognition. *Cortex* 47:367-376.
- Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15:1034-1038.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302-4311.
- Kiani R, Esteky H, Tanaka K (2005) Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol* 94:1587-1596.
- Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci* 3:946-953.
- Ku S-P, Tolia S, Logothetis Nikos K, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70:352-362.
- Leopold DA, Rhodes G (2010) A comparative view of face perception. *J Comp Psychol* 124:233-251.
- Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572-575.
- Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* 453:869-878.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150-157.
- Love SA, Pollick FE, Latinus M (2011) Cerebral correlates and statistical criteria of cross-modal face and voice integration. *Seeing and Perceiving* 24:351-367.
- McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience* 9:605-610.
- Mohedano-Moriano A, Martinez-Marcos A, Pro-Sistiaga P, Blaizot X, Arroyo-Jimenez MM, Marcos P, Artacho-Pérula E, Insausti R (2008) Convergence of unimodal and polymodal

- sensory input to the entorhinal cortex in the fascicularis monkey. *Neuroscience* 151:255-271.
- Mohedano-Moriano A, Pro-Sistiaga P, Arroyo-Jimenez MM, Artacho-Pérula E, Insausti AM, Marcos P, Cebada-Sánchez S, Martínez-Ruiz J, Muñoz M, Blaizot X, Martínez-Marcos A, Amaral DG, Insausti R (2007) Topographical and laminar distribution of cortical input to the monkey entorhinal cortex. *Journal of Anatomy* 211:250-260.
- Munoz-Lopez M, MohedanoMoriano A, Insausti R (2010) Anatomical pathways for auditory memory in primates. *Frontiers in Neuroanatomy* 4.
- Nieuwenhuis S, Forstmann BU, Wagenmakers E-J (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci* 14:1105-1107.
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985) Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc Lond B Biol Sci* 223:293-317.
- Perrodin C, Kayser C, Logothetis Nikos K, Petkov Christopher I (2011) Voice cells in the primate temporal lobe. *Current Biology* 21:1408-1415.
- Pinsk MA, DeSimone K, Moore T, Gross CG, Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America* 102:6996-7001.
- Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87-91.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.
- Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73:713-726.
- Rotshtein P, Henson RNA, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* 8:107-113.
- Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001) The neural correlates of person familiarity. *Brain* 124:804-815.
- Sigala R, Logothetis NK, Rainer G (2011) Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. *J Neurophysiol* 105:2740-2752.
- Sirotin YB, Das A (2009) Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* 457:475-479.
- Sliwa J, Duhamel JR, Pascalis O, Wirth S (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc Natl Acad Sci U S A* 108:1735-1740.
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869-873.
- Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, Ito K, Fukuda H (2001) Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *NeuroImage* 13:877-890.
- Thompson LT, Best PJ (1989) Place cells and silent cells in the hippocampus of freely-behaving rats. *The Journal of Neuroscience* 9:2382-2390.
- Tsao DY, Livingstone MS (2008) Mechanisms of face perception. *Annual Review of Neuroscience* 31:411-437.
- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences* 105:19514-19519.

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670-674.
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989-995.
- Viskontas IV, Quiroga RQ, Fried I (2009) Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proc Natl Acad Sci U S A* 106:21329-21334.
- Viswanathan A, Freeman RD (2007) Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nat Neurosci* 10:1308-1312.
- von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. *PLoS Biol* 4:e326.
- von Kriegstein K, Kleinschmidt A, Giraud AL (2006) Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cereb Cortex* 16:1314-1322.
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL (2005) Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci* 17:367-376.
- von Kriegstein K, Dogan O, Gruter M, Giraud AL, Kell CA, Gruter T, Kleinschmidt A, Kiebel SJ (2008) Simulation of talking faces in the human brain improves auditory speech recognition. *Proc Natl Acad Sci U S A* 105:6747-6752.
- Waydo S, Kraskov A, Quiroga R, Fried I, Koch C (2006) Sparse representation in the human medial temporal lobe. *J Neurosci* 26:10232-10234.

Figure1 :



B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure2 :

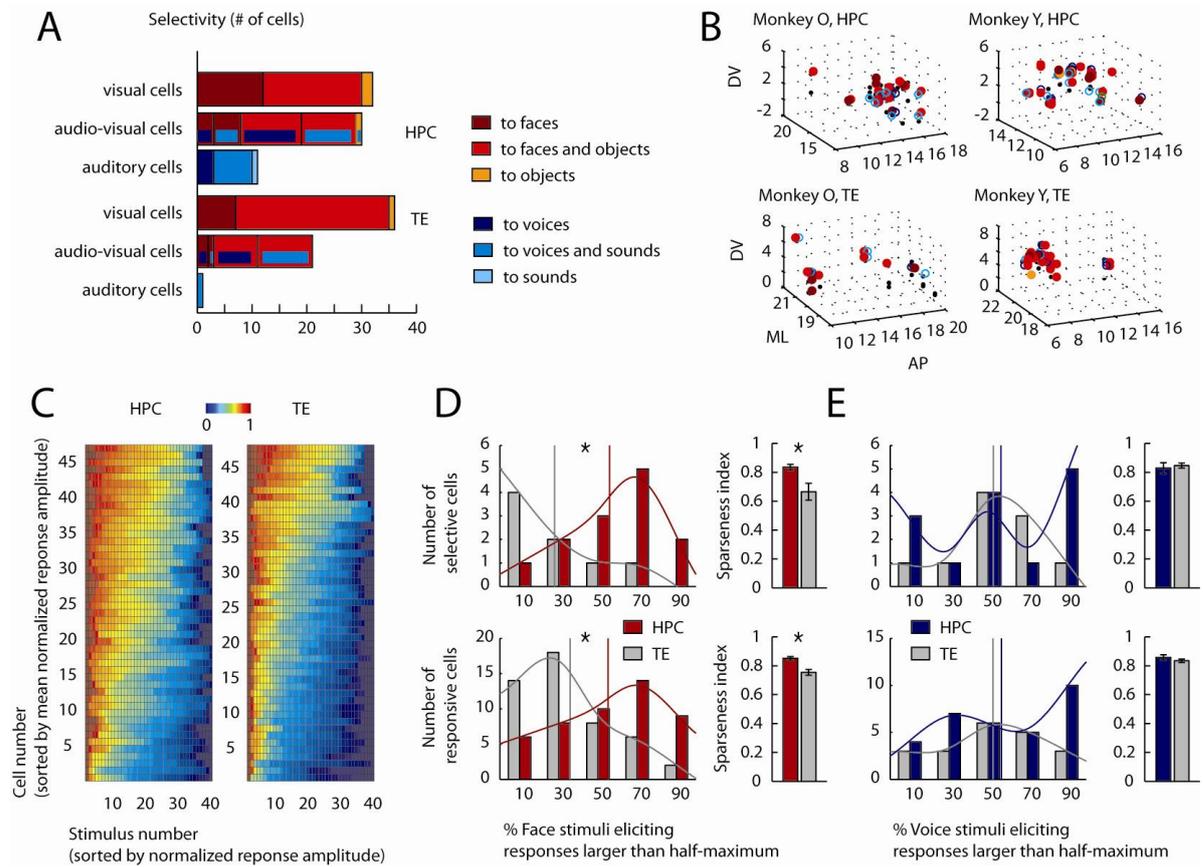
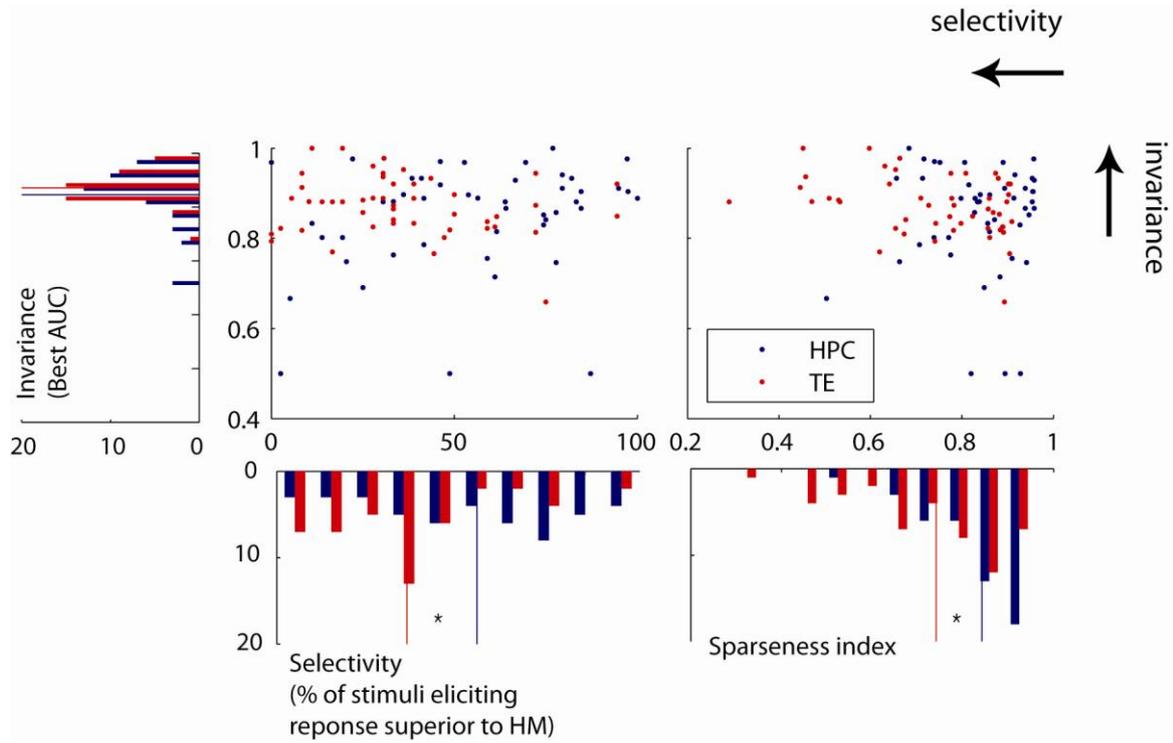


Figure 3



B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure 4

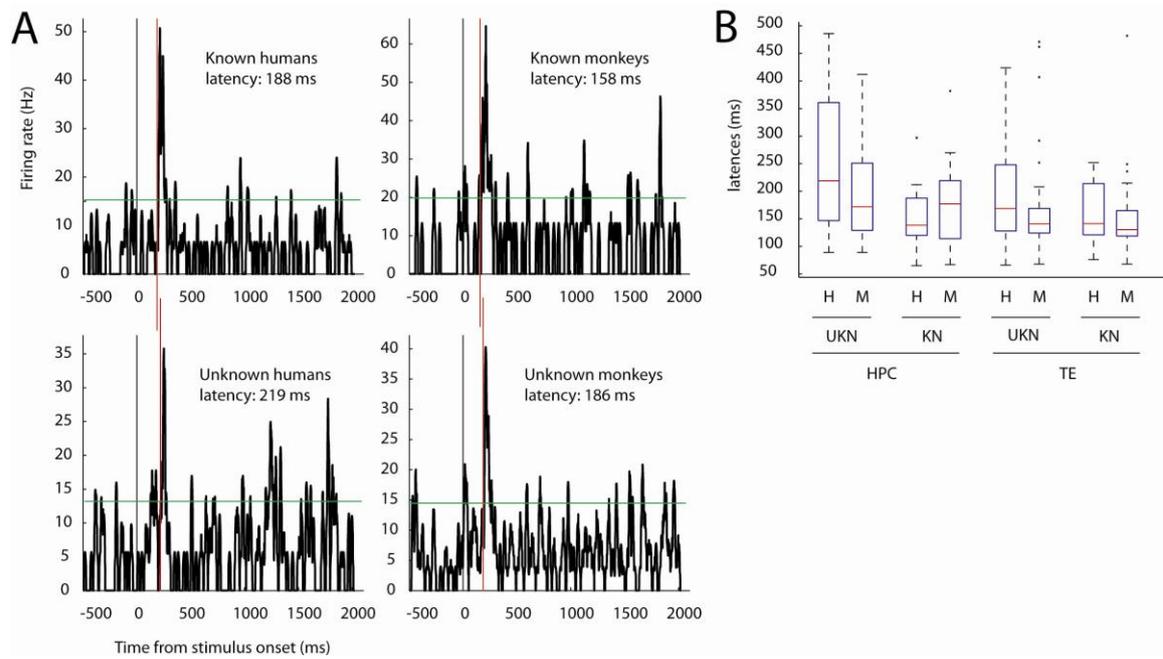
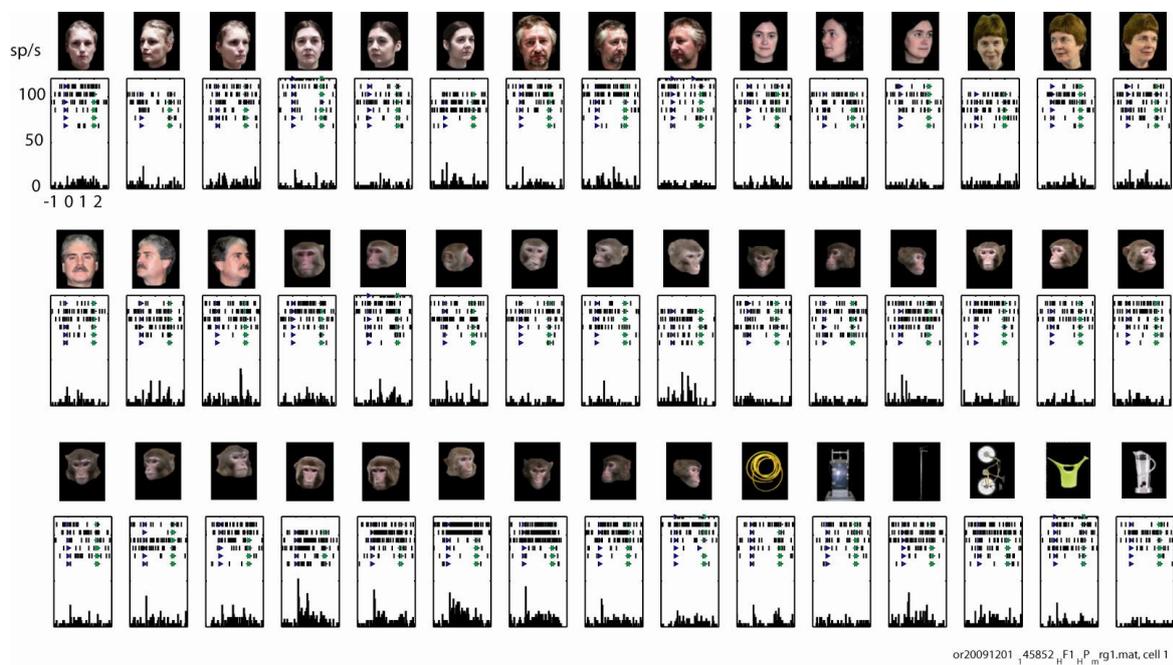


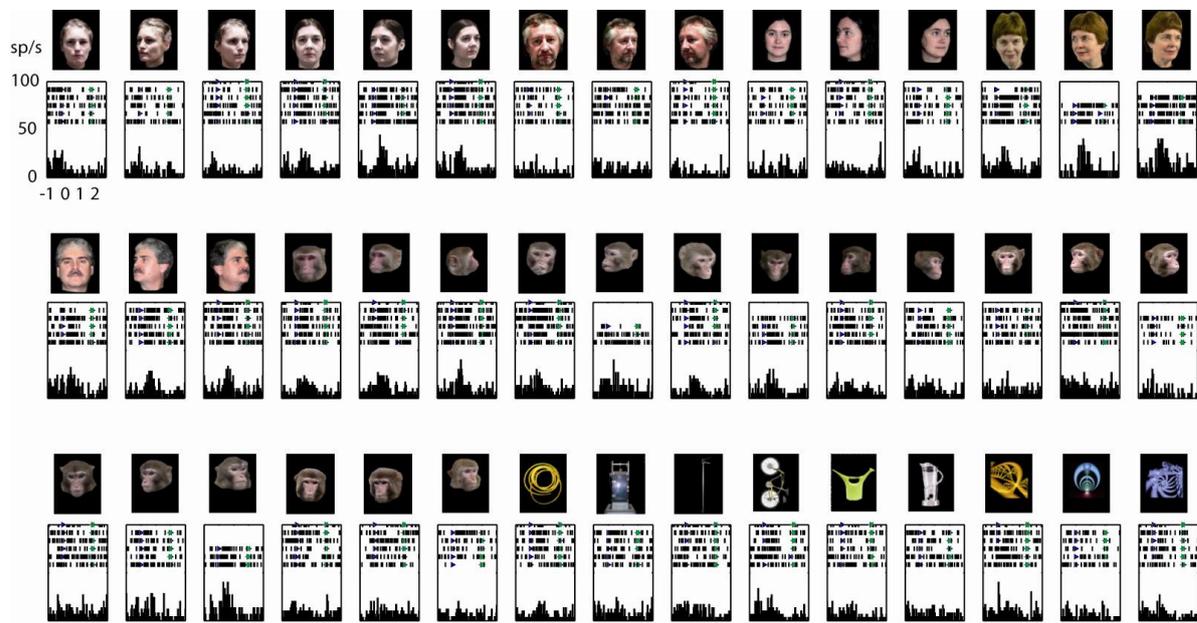
Figure 5



or20091201_45852_F1_P_rg1.mat, cell 1

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure6



yo20090130_F_H_P_m_rg.mat, cell 1

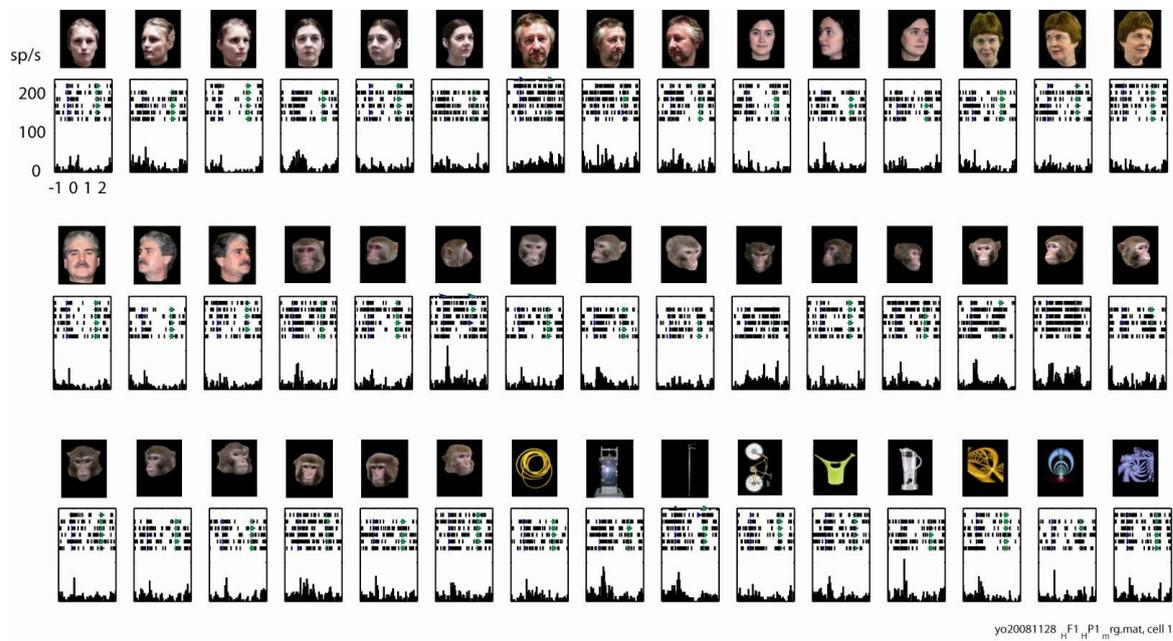
Figure 7



or20100901_60405_F1_P_rg.mat, cell 1

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure 8



yo20081128_h1h1m_rg.mat, cell 1

Figure 9



yo20100407 60138 F21E units.mat, cell 4

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure 10

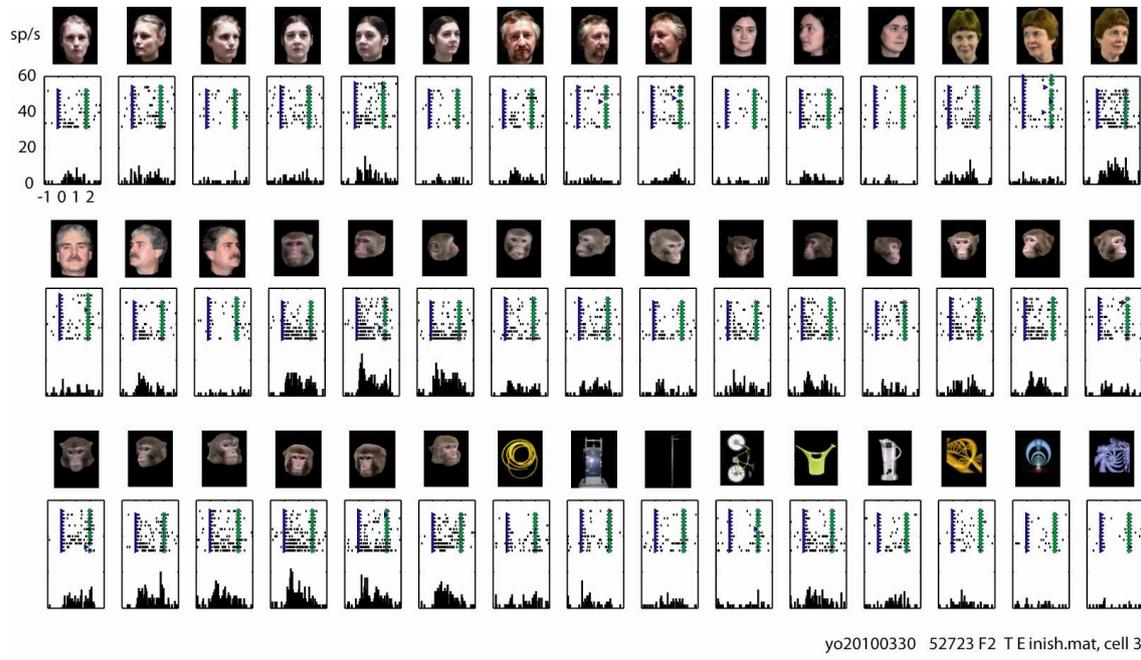
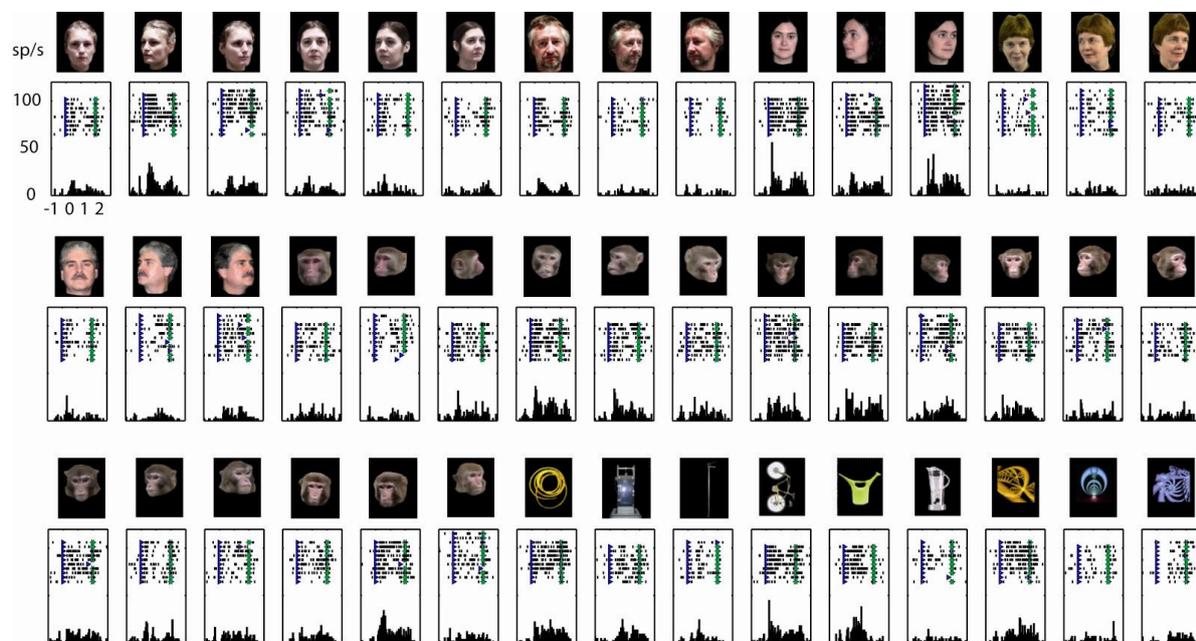


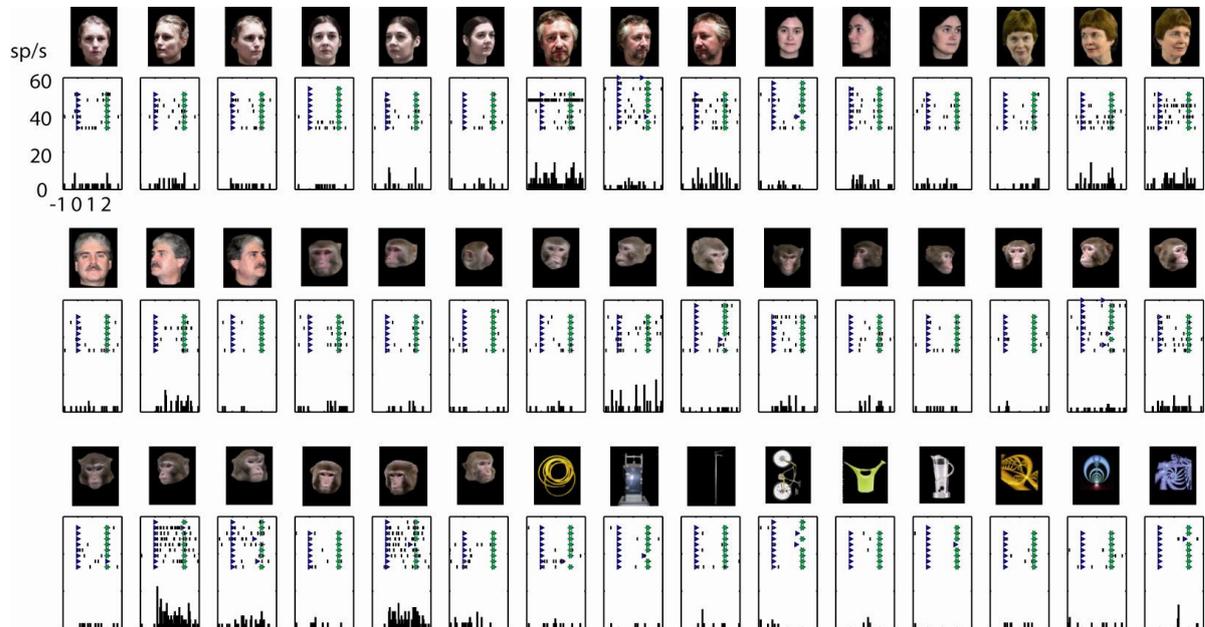
Figure 11



yo20100323 54219 1 F2Hs ET rgjs.mat, cell 3

B. Réponses neuronales aux visages : comparaison de l'hippocampe et de l'aire temporale chez le singe

Figure 12



C. Dissection des microcircuits dans l'hippocampe du singe: lien entre les réponses des cellules et leur signature électrophysiologique *in vivo*

Résumé

L'hippocampe est le locus d'interrelations denses entre les cellules pyramidales et une population importante et hétérogène d'interneurons. Des études *in vitro* et *in vivo*, conduites presque exclusivement chez les rongeurs, ont montré que la nature des neurones peut être identifiée grâce aux poids inhibiteur ou excitateur que ces neurones exhibent dans leurs interactions couplées avec d'autres cellules, ainsi que grâce à d'autres caractéristiques électrophysiologiques. Ces distinctions ont permis de construire des modèles réalistes des circuits locaux mis en jeu lors du traitement de l'information, dans lesquels le rôle régulateur des neurones GABAergiques sur l'activité des cellules pyramidales est pris en compte. Afin de fournir un cadre anatomique commun à la comparaison des fonctions hippocampiques à travers les espèces, nous avons mené une caractérisation des différents types cellulaires et de leur connectivité en réseau dans l'hippocampe du singe (*Macaca mulatta*). Nous avons montré que les cellules pyramidales peuvent, tout comme chez le rat, être distinguées des interneurons par leurs caractéristiques électrophysiologiques et par la nature de leur interactions couplées excitatrices. Nous avons ensuite relié la sélectivité des cellules, exprimées lors d'une tâche de présentation de stimuli visuels et acoustiques, à leur nature. Nos résultats montrent que les interneurons répondent à plus de stimuli mais présentent une plus faible sélectivité. À l'inverse, les cellules pyramidales répondent à peu de stimuli et présentent une sélectivité plus forte. Ces résultats suggèrent qu'une plus faible proportion de cellules pyramidales participe au codage de l'information, mais que le contenu informatif résultant est plus contrasté. Ces résultats sont semblables à ceux obtenus chez le rat et plus récemment chez l'Homme, plaidant en faveur de la conservation de la nature de l'hippocampe entre les espèces, non seulement au niveau morphologique mais aussi au niveau des dynamiques cellulaires. Enfin, nos résultats montrent que l'orientation des réponses aux stimuli visuels ou auditifs est préservée au sein des voxels de pénétrations, suggérant que des cellules voisines partagent une sélectivité sensorielle commune.

Dans l'ensemble, les résultats apportent des informations importantes sur l'organisation spatiale et fonctionnelle des microcircuits de l'hippocampe.

Cette étude est en cours de préparation pour une publication sous le titre :

Sliwa J, Duhamel J-R, Planté A, Wirth S. Dissection of micro-circuitry in monkey hippocampus: Linking cells' response properties to their electrophysiological features *in vivo*.

C. Dissection des microcircuits dans l'hippocampe du singe: lien entre les réponses des cellules et leur signature électrophysiologique *in vivo*

C.1. Article n°3 : Dissection of microcircuitry in monkey hippocampus: Linking cells' response properties to their electrophysiological features *in vivo*

Julia Sliwa, Jean-René Duhamel, Aurélie Planté, Sylvia Wirth

Centre de Neurosciences Cognitives, CNRS-Université de Lyon, UMR5229, 69675 Bron, France

Key-words: Interneurons, pyramidal cells, extracellular recordings, CA3, CA1, network interactions

Pages : 13

Figures : 4

Dissection of microcircuitry in monkey hippocampus: Linking cells' response properties to their electrophysiological features *in vivo*

Abstract

The hippocampus is marked by a strong interplay between principal cells and a large heterogeneous population of interneurons. *In vitro* and *in vivo* studies, conducted almost exclusively in rodents, provided a distinction between these categories of cells based on differential excitatory or inhibitory weights in pair-interactions and other physiological signatures. These distinctions further allowed building realistic models of the local circuitries involved during information processing whereby the regulatory role of GABAergic neurons on principal cells can be taken into account. In order to provide a common anatomical framework to discuss hippocampal functions across species, we conducted a characterization of the different types of neurons and their network connectivity in the monkey (*Macaca mulatta*) hippocampus. We show that putative pyramidal cells, as in the rat, can be distinguished *in vivo* from putative interneurons in the monkey hippocampus based on their electrophysiological properties and the nature of their paired interactions. We then linked the cells' classes to cells' functional response properties, exhibited in a task during which monkeys attended visual and acoustic stimuli. Interneurons were found globally more responsive to stimuli compared to principal cells, albeit their response was less selective. On the contrary, pyramidal cells responded to fewer stimuli and were found more selective. The results suggest that fewer pyramidal cells contribute to information coding, but their information content is more strongly contrasted. This study yields similar results to the ones obtained in rats and more recently in humans, pleading for a conserved nature of the hippocampus across species not only at the morphological level but also at the level of its cellular dynamics. Our results also show that response orientation to visual or auditory stimuli is preserved within voxels of penetrations, suggesting that neighboring cells share common sensory sensitivity. Taken together, these results provide valuable information on both the spatial and functional organization of microcircuits in the primate hippocampus.

Key words

Interneurons, pyramidal cells, extracellular recordings, CA3, CA1, network interactions

Introduction

The hippocampus evolved toward a structure supporting spatial information and episodic memory in humans and other mammals (Bird and Burgess, 2008; Clayton and Russell, 2009). In order to compare the functional properties of hippocampal neurons across species, from rodents to humans via monkeys, it is important to rely on a common framework concerning their anatomical underpinnings. In rats a significant work has been accomplished to characterize neuronal activities in the hippocampus from their electrophysiological features *in vitro* (Csicsvari et al., 1998) and *in vivo* (Csicsvari et al., 1999; Henze et al., 2000; Frank et al., 2001; Klausberger et al., 2003; Klausberger and Somogyi, 2008). These studies demonstrated that

electrophysiological properties of a neuron and of its network interactions enable to derive the neuron's nature (pyramidal cell or interneuron), its spiking property (regular, bursting or in-between firing) and its activity (inhibitory or excitatory neuron). In humans a similar classification has been conducted *in vivo*, based on hippocampal recordings obtained from patients with intractable epilepsy implanted for clinical reasons (Viskontas et al., 2007; Ison et al., 2011). However such a characterization has not been carried out for hippocampal cells from the monkey, even though monkeys are an interesting species for investigating hippocampal functions, as exemplified by numerous studies using diverse cognitive tasks (Rolls et al., 1989; Tamura et al., 1992b; Tamura et al., 1992a; Ono et al., 1993; O'Mara et al., 1994; Eifuku et al., 1995; Nakamura and Kubota, 1996; Nishijo et al., 1997; Xiang and Brown, 1998; Hori et al., 2003; Wirth et al., 2003; Hampson et al., 2004; Yanike et al., 2004; Rolls et al., 2005; Wirth et al., 2009; Jutras and Buffalo, 2010; Naya and Suzuki, 2011). Monkey hippocampus benefits from more numerous afferents from secondary visual cortical areas than the rodent hippocampus, which in counterpart receives more abundant olfactory projections (Insausti et al., 1997; Insausti and Amaral, 2008; Mohedano-Moriano et al., 2008). Therefore, monkeys represent an appealing link regarding comparative studies on the role of hippocampus in spatial and episodic memory, because they rely as much on spatial as on visual information in their natural habitat and yet lack some of the complexity of human episodic memorization (Hampton and Schwartz, 2004; Hampton et al., 2005). We propose to provide a depiction of macaques' hippocampal neurons *in vivo*, based on their electrophysiological features, and to discuss the similarities and dissimilarities with the two other species.

We recorded neurons from different hippocampal subfields in two adult awake monkeys (*Macaca mulatta*, 8.5 kg, monkey Y and 13 kg, monkey O) exploring visual stimuli or hearing auditory stimuli, and used methods developed for extracellular feature analyses to describe their nature and firing properties. We present that each neuron can be characterized by: the hippocampal subfield it lies in, its putative pyramidal or interneuron nature and its regular or bursty spiking. Moreover in a companion article (Article 2) we demonstrated the existence of selective hippocampal cells for faces, objects and finer dimensions among faces (species and familiarity). Thus, we also investigate links between cells' selectivity and their physiological properties, in order to understand how microcircuits cooperate to represent complex information. To continue this exploration, we study spatial organization of the different types of responses in two hippocampal subfields (CA1, CA3) and at the smaller scale of recording sites (Dahl et al., 2009). Thereby such a process enriches existing knowledge on different types of cells' properties across species, for instance by enabling comparison of response properties of selective place cells in rodents to selective visual cells in primates.

Materials and Methods

Subjects. Two adult rhesus monkeys (*Macaca mulatta*, 8.5 kg, monkey Y and 13 kg, monkey O) participated in this study. They were socially housed in rooms of three or four individuals since they arrived at the Centre de Neurosciences Cognitives. All experimental procedures are in accordance with the local authorities (Direction Départementale des Services Vétérinaires, Lyon, France) and the European Community standards for the care and use of laboratory animals (European Community Council Directive, 1986, Ministère de l'Agriculture et de la Forêt, Commission Nationale de l'expérimentation animale).

Surgical preparation. Stimuli. Behavioral procedure. The procedures used have been described elsewhere (Article 2).

Electrophysiological Recordings. Single-neuron activity was recorded extracellularly with tungsten quartz insulated tetrodes (1.5-2 M Ω ; Thomas Recording, Giessen, Germany) or tungsten single microelectrodes (1–2 M Ω ; Frederick Haer Company, Bowdoinham, ME) which were lowered into the hippocampus with a NAN microdrive (Plexon Inc., Dallas, TX). The microelectrodes were inserted through 24-gauge stainless steel guide tubes set inside a delrin grid (Crist Instruments, Damascus, MD) adapted to the recording chamber. The electrode signal was preamplified using a Plexon system (Plexon Inc., Dallas, TX) and digitized using RecorderEV3 software (Mediane System, Le Pecq, France). Continuous signal was sampled at 20 kHz and stored with the RecorderEV3 software along with behavioral and stimulation data for off-line analysis. Single-units were sorted offline using the Offline Sorter software (Plexon Inc., Dallas, TX). We used a semi-manual sorting method such that three selected parameters (principal components, waveform patterns, and amplitude of the peak in each electrode of the tetrode) allowed us to separate units from background activity and yield well-isolated clusters. Because with tetrodes the signal is recorded simultaneously by four adjacent microwires, the location of the neurons can be triangulated, and neurons presenting similar amplitude on one wire can be distinguished easily from one another by comparing their recorded amplitude on the three others wires (**Fig. 1 B-C**). If for example, only channel 4 had been recorded (as could be the case with a single electrode), cells 1 and 2 would have been nearly impossible to distinguish (**Fig 1 C**). Thus the use of tetrodes enable to rely on several additional dimensions to correctly separate the different units from each other as compared to the available ones when using conventional single electrodes and provide a superior quality of single-unit isolation (Gray et al., 1995; Ecker et al., 2010). This led us to sort and further consider the activity of 261 hippocampal neurons recorded in the two behaving monkeys (141 in monkey O and 120 in monkey Y). Each neuron was characterized by: the hippocampal subfield it lied in, its putative pyramidal or interneuron nature, and finally by its bursting or regular spiking activity.

Cells location. Postsurgical MR images were used to finely monitor recording locations within the hippocampus during each experiment (**Fig. 1A**), and were compared to two monkey brain atlases (Paxinos et al., 2000; Saleem and Logothetis, 2006) to identify the hippocampal subfield (CA1, CA3, dentate gyrus and subiculum) recorded from. Extracellular activity was recorded at each location and single-units were separated off-line. For cells that have been recorded simultaneously with others on a single tetrode, we counted the number of cells presenting same modality responses to investigate potential spatial organization of modality preferences. Same analysis was conducted on stimulus selectivity.

Cells characterization. Classifications of neurons' nature and spiking activity were achieved by analyzing their firing rate and their waveform pattern, with custom written scripts and the Statistics Toolbox in Matlab (MathWorks, Natick, MA). Cells were separated according to their nature into putative pyramidal cells and interneurons in two steps. First cells belonging to pairs of neurons recorded together were analyzed to unravel their network interaction. Information derived from cross-correlograms enabled to identify excitatory cells as being putative pyramidal neurons and inhibitory cells as being putative interneurons. Second these identified neurons were analyzed and mapped together with the rest of the neurons on two different planes according to their waveform pattern and firing rate. The unidentified neurons were then

classified into putative pyramidal cells and putative interneurons using their feature and firing rate resemblance to the identified cells.

Network interactions were analyzed using 1ms bin-width cross-correlograms of 461 pairs of neurons, each pair being recorded by the same tetrode (**Fig. 1D-F** black histograms, (Bryant et al., 1973; Constantinidis et al., 2001; Barthó et al., 2004; Tamura et al., 2004). Cross-correlograms were normalized to the number of action potentials of each neuron; they thus reflect the discharge probability at each time point of the first cell given the firing of the second cell of the pair. Asymmetrical significant peaks appearing within -5 to -1ms or 1 to 5 ms were interpreted as putative excitatory monosynaptic connections (**Fig. 1 D-E**). Conversely asymmetrical significant troughs appearing within -5 to -2ms or 2 to 5 ms were interpreted as putative monosynaptic inhibitory connections (**Fig. 1 E-F**). The first ± 1 ms bins were not considered in the trough calculation because of the presence of an artificial trough at 0 ms bin in the cross-correlograms due to indivisibility of neurons with superimposed action potentials. A peak or trough was considered significant when at least one of the bins exceeded respectively 99.9th or 0.1th percentile of the control number of spikes (calculated between -2000 ms and -50 ms and 50 ms and 2000 ms) (Barthó et al., 2004). Symmetrical troughs or peaks, which arise from a modulation/drive of both cells by a third cell, were not considered in the analysis because they do not provide information about neurons' own activity, but on their reaction following a third cell activity. Finally neurons exciting another neuron were categorized as putative pyramidal cells whereas neurons inhibiting another neuron were categorized as putative interneurons.

Further, these identified neurons and the others neurons were analyzed according to their firing rate and waveform pattern. From these two independent parameters, two pieces of information can be derived. First trough-to-peak and half-width durations of the each cell's waveform pattern were calculated (**Fig. 1 H**, the highest amplitude waveform from all the four wires recordings was considered for the pattern analysis). Each cell was represented in a two-dimensional space with trough-to-peak and half-width duration as axes (**Fig. 1 H**, Barthó et al., 2004). Second each cell was represented in a two-dimensional space with trough-to-peak and average firing rate as axes (**Fig. 1 H**, Wilson et al., 1994; Frank et al., 2001). On these maps, the nature of the identified neurons was superimposed. Based on their respective distribution, a cutoff was estimated in order to part the populations. Finally, the nature of the rest of the neurons not characterized by the cross-correlogram analysis, were assigned to one of the populations on the basis of their electrophysiological (trough-to peak and average firing rate) similarity to the identified neurons.

Cells firing pattern. Finally cells' firing properties were categorized into bursting, regular spiking or unclassified spiking, considering ISI histograms and autocorrelograms (**Fig. 1 D-F**, red and blue histograms). First pyramidal and interneuron cells were classified as bursting if the maximum peak of the autocorrelogram between 3 and 6 ms was superior to 50% of the maximum bin value of the first 50 ms (Barthó et al., 2004). Then when the autocorrelogram mean, defined as its expected value, was superior to 12 ms, interneurons were classified as regular spiking (Csicsvari et al., 1998; Csicsvari et al., 1999). Following, when the mode of the ISI histogram, assessed by calculating the maximum of its gamma distribution fit, was superior to 35 ms, pyramidal cells were classified as regular spiking. Other pyramidal cells were labeled "unclassified".

Functional properties of characterized cells. Finally for all the cells being assigned either a putative pyramidal nature or a putative interneuron nature, we counted the number of cells

presenting same modality responses to investigate potential organization of modality preferences according to cell type. Same analysis was conducted on stimulus type. Following selectivity of both types of cells was assessed using a sparseness statistic S (Vinje and Gallant, 2000) adapted from a sparseness index SI (Rolls and Tovee, 1995), defined as follows:

$$S = \frac{1-SI}{1-1/N}, \text{ where } SI = \frac{(\sum_{i=1}^N r_i / N)^2}{\sum_{i=1}^N r_i^2 / N},$$

with r_i being the response of the neuron to the i^{th} stimulus, and N being the total number of stimuli. Using this index, a cell responding to all stimuli, thus showing a weak selectivity would have an index close to zero, whereas a selective cell responding to few stimuli, would have an index closer to one.

Results

Characterizing interneurons and pyramidal cells. Cells identified as excitatory or inhibitory through the cross-correlogram clustered into two groups (**Fig. 1H**). Interestingly cells having an inhibitory activity were also characterized by waveform of shorter trough-to peak and half-amplitude duration and higher firing rate. On the contrary cells having an excitatory activity were characterized by waveform of longer trough-to peak and half-amplitude duration and smaller firing rate. This classification allowed us to part the population into two groups (red and blue crosses, **Fig. 1H**). Following this process, we examined neurons that could not be identified as inhibitory or excitatory based on their cross-correlograms either because they were not recorded in a pair, or because the cross-correlogram could not provide evidence of a monosynaptic connection with another neuron. Each of these neurons was then assigned into one of the two groups based on its waveform pattern and firing rate characteristics (Wiebe and Staubli, 1999, 2001). Two populations of cells were therefore separated: putative pyramidal cells and putative interneurons. Firing rates (pyramidal cells: 1.57 ± 2.40 Hz, interneurons: 11.43 ± 20.62 Hz, $P=0.0039$, t-test) and trough-to-peak times (pyramidal cells: 0.73 ± 0.20 ms, interneurons: 0.31 ± 0.09 ms, $P<0.0001$, t-test) were significantly different between the putative interneurons and pyramidal cells. Interneurons represented 9.7% of the total number of recorded cells.

Characterizing cells firing pattern. Pyramidal cells presented a bursting activity in 29% of the cases. They presented a regular spiking activity in 16% of the cases. The rest was considered unclassified. The number of cells according to their nature and type of spiking activity was compared across the different anatomical subfields. Interneurons represented 9.7% of the total number of recorded cells (8.4% of CA1, 10.4% of CA3, **Fig. 1G**). Pyramidal cells presented a bursting activity in 34% and 27% of the cases in CA1 and CA3 respectively (**Fig. 1G**). They presented a regular spiking activity in 17% and 16% of the cases in CA1 and CA3 respectively (**Fig. 1G**). In the two subfields analyzed, CA1 and CA3, the percentage of cells of each category was similar ($p=0.65$, χ^2 test).

Responsiveness of interneurons and pyramidal cells. Proportionally more interneurons (77%) were responding to either one or both modalities than pyramidal cells (32%) (**Fig. 2A**, left panel). Proportion of cells responding to the different modalities also varied within these responding cells: if visual neurons represented respectively 45% and 41% of the responding interneurons

and pyramidal cells, auditory cells represented only 10% of the responding interneurons while they represented 41% of the responding pyramidal cells, and on the contrary audio-visual cells represented 45% of the responding interneurons while they represented only 18% of the pyramidal cells. These proportions were significantly different ($p=0.018$, χ^2 test). On the contrary proportions of responses to faces, faces and objects, objects, voices, voices and sounds, and sounds were not significantly different across cell types (visual categories: $p=0.20$, auditory categories: $p=0.64$, χ^2 test, **Fig. 2A**, middle and right panels).

Selectivity of interneurons and pyramidal cells. Pyramidal cells were more selective than interneurons as assessed by computing sparseness statistic ($S_{\text{pyr}}=0.31 \pm 0.21$, $S_{\text{int}}=0.15 \pm 0.11$, $p=9.10 \cdot 10^{-5}$, Wilcoxon test, **Fig. 3A**). They also responded to less stimuli than interneurons ($S_{\text{pyr}}=1.29 \pm 2.27$, $S_{\text{int}}=3.19 \pm 2.65$, $p=0.0017$, t-test). Among pyramidal cells, regular spiking cells ($S=0.20 \pm 0.09$) were less selective than bursting ($S=0.32 \pm 0.18$), and unclassified ones ($S=0.34 \pm 0.24$) as assessed by an analysis of variance on sparseness statistics with type of firing rate as a factor ($F=4.43$, $df=2$, $P=0.014$, **Fig. 3A**).

Functional connectivity and response correlation. Stimulus preference between presumed interneurons and their target neurons could be examined to determine whether target neurons exhibit an inverse preference to the one exhibited by the interneuron. However, this will not necessarily be the case because the contribution from an input from a single neuron to the generation of spikes in a postsynaptic neuron is generally weak, at least in the neocortex (Constantinidis et al., 2001; Constantinidis and Goldman-Rakic, 2002; Tamura et al., 2004). Here we present an exemplar pair of neurons with inhibitory linkage exhibiting overall stimulus preferences that inversely correlated with each other (**Fig. 4**). Correlation between stimulus preferences should be more systematically investigated by quantification through Pearson's correlation coefficients.

Subfield organization of modality and categories' preferences. Similarly to the results obtained for the entire hippocampus, when considering only cells from CA1 or CA3, proportionally more interneurons (100% and 72% respectively) were responding to either one or both modalities than pyramidal cells (42% and 30% respectively) (**Fig. 2B**). Proportion of cells responding to the different modalities also varied within these responding cells in CA3 ($p=0.006$, χ^2 test), while proportions of responses to faces, faces and objects, objects, voices, voices and sounds, and sounds were not significantly different across cell types in both sub-regions ($p>0.05$ and $p>0.05$, χ^2 tests, **Fig. 2B**). In CA3, pyramidal cells were more selective than interneurons as assessed by computing sparseness statistic ($S_{\text{pyr}}=0.35 \pm 0.22$, $S_{\text{int}}=0.18 \pm 0.13$, $p=0.004$, Wilcoxon test, **Fig. 3B**). In CA1 they were not statistically different, mainly due to the fact that not enough interneurons were recorded in CA1 to perform an appropriate comparison. The different types of pyramidal cells did not differ in selectivity when comparing them within sub-regions (CA3: $p=0.06$, CA1: $p=0.51$, ANOVA). Moreover we compared selectivity of pyramidal cell in CA1 to those in CA3. Pyramidal cells in CA3 presented a higher sparseness statistics ($S=0.35$) than cells in CA1 ($S=0.27$), which could indicate that they encode stimuli in a more selective way than pyramidal cells in CA1. However this difference was above the level of significance ($p=0.06$, t-test).

Spatial organization of modality preferences at the scale of penetration. To determine whether the modality preference of individual neurons is spatially organized, we compared preferences of single units recorded along the same tetrode/electrode penetration. Individual penetrations

were spaced (750 μ m) on a recording grid, and often several units were recorded with the same tetrode. By counting modality preferring co-occurring cells in one recording, we found that cells with opposite modality responses, i.e. cells with solely visual response with cells with solely auditory response, were rarely co-localized (1 of 48 penetrations, **Fig. 2C**). This shows that neurons preferring different modalities occur spatially separated. Most penetrations yielded: 1) only units preferring same modality, 2) modality preferring and bimodal units together, 3) only bimodal units (**Fig. 2C**), showing that the modality preference in the hippocampus is spatially organized, with unimodal and bimodal neurons co-occurring along the same penetrations but neurons preferring distinct modalities being spatially separated. Results were similar in both CA1 and CA3 (**Fig. 2C**).

Spatial organization of categories' preferences at the scale of penetration. Similarly to the previous analysis, we aimed at determining whether face-cells co-occur in individual penetrations or whether they are colocalized with neurons presenting other selectivity. Again, by counting selectivity preferring co-occurring cells in one recording, we found that cells with opposite stimulus responses, i.e. cells selective for faces only and cells selective for objects only, were never co-localized (0 of 41 penetrations, **Fig. 2C**). Most penetrations yielded: 1) only units preferring faces, 2) face preferring and face-objects responding units together, 3) only face-objects units (**Fig. 2C**). These distributions show that stimulus preference in the hippocampus is locally preserved, with facial and facial-objects neurons co-occurring along the same penetrations but neurons preferring distinct stimuli being spatially separated. Results were similar in both CA1 and CA3 (**Fig. 2C**).

Discussion

We conducted a characterization of the different subtypes of neurons and their network connectivity in the monkey hippocampus. We also investigated links between cells' selectivity and their physiological properties, in order to understand how microcircuits cooperate to represent complex information in two hippocampal subfields (CA1, CA3) and at the smaller scale of recording sites.

First, using complementary methods, we showed that putative pyramidal cells can be distinguished from putative interneurons in the monkey hippocampus based on their electrophysiological properties. Moreover two types of electrical firing in pyramidal cells can be observed: regular or in bursts. Similarly to studies conducted in rats, the majority of inhibitory neurons, i.e. putative interneurons, also presented narrower waveforms and a fast-spiking, while excitatory neurons, i.e. putative pyramidal cells, presented larger waveforms and a lower spiking rate. However it should be noted that a smaller fraction of inhibitory neurons had larger waveforms and might reflect the specific characteristic of Ivy cells interneurons with large waveforms and slow firing rates (Fuentealba et al., 2008). In that respect it should be highlighted that only the cross-correlogram analysis enabled to unravel the inhibitory nature of these cells otherwise resembling pyramidal-cells. This remark might also be applied more generally, because in other areas of the monkey brain, existence of sub-populations of large pyramidal cells have been reported resembling putative interneurons in their fast spiking rate and narrow waveforms (Vigneswaran et al., 2011). Taken together, these observations show that the concurrent use of complementary measures allows better disambiguating of cell types when assessed through their electrophysiological properties. This study yields similar results to the

one obtained in rats and humans (Csicsvari et al., 1999; Frank et al., 2001; Viskontas et al., 2007; Ison et al., 2011), pleading for a conserved nature of the hippocampus across species not only at the morphological level but also at the level of its cellular dynamics. Such a classification study had not been performed before, yet it is of compelling interest in the hippocampus partly because it has been proposed that the different types of neurons play distinct roles in information coding: pyramidal cells being devoted to spatial information coding and interneurons to non-spatial information (Christian and Deadwyler, 1986; O'Keefe, 1999). In rats it has been demonstrated that both types of cells participate in representing as well spatial as non-spatial information (Wiebe and Staubli, 1999; Wood et al., 1999; Wiebe and Staubli, 2001). In other species however this information is missing, which precludes conclusion on this issue central to hippocampus functioning. Assessing cells properties provides a common framework for comparing studies carried out in various animal models; thereby the present results could be used as a reference tool for future studies using recordings of monkey hippocampal neurons.

Second, we used the classification generated to dissect the functional properties of hippocampal micro-circuits in monkeys, drawing on our finding of these cells selectivity to visual and auditory stimuli such as faces and objects (Article 2, companion article). We first observed that the incidence of responding cells was higher within interneurons than pyramidal cells. Second that, in both types of cells a similar amount of visual responsive neurons was observed while there were opposite amounts of auditory and audio-visual responsive neurons. Third, selectivity to stimuli was higher in bursting and non bursting pyramidal cells relative to interneurons. Finally the functional connectivity could be assessed in exemplar neurons, and a thorough analysis of the selectivity correlation between pairs of connected cells should be performed.

The fact that selectivity is lower and responsiveness higher for interneurons than for pyramidal cells, suggest that pyramidal cells contribute to stimulus representation in a sparser way at the level of the population but with a higher informational content. Within this context, interneurons might play an important role in sharpening pyramidal cells selectivity through direct inhibitions and this would be reflected on their higher responsiveness. They could contribute to the formation of response properties of target neurons either in a stimulus-specific manner (Tamura et al., 2004) or by controlling the activity levels of the neuronal networks. A lower selectivity of the interneurons compared to pyramidal cells has also been observed in human (Ison et al., 2011) and rat hippocampus (Frank et al., 2001; Marshall et al., 2002). In human hippocampus, the authors also reported the existence of ultra-selective cells in the hippocampus, which do not constitute the main part of the pyramidal cells we report in our study (Quiroga et al., 2005; Gelbard-Sagiv et al., 2008; Quiroga et al., 2009). Because we performed recordings in an unsupervised manner we couldn't expect to observe such highly selective pyramidal cells, which are very sparsely distributed in the human hippocampus (Waydo et al., 2006; Quiroga et al., 2008). On the other hand, the population of pyramidal cells we report here might be more representative of the majority of pyramidal neurons present in the monkey hippocampus, which can have homologues in the human brain as reported in earlier studies performed in human implanted epileptic patients in an unsupervised manner (Fried et al., 1997; Kreiman et al., 2000; Fried et al., 2002).

Third, we observed that modality and stimulus preferences are similar to a great extent between adjacent hippocampal neurons, co-occurring within recordings sites. Even though the hippocampus is not organized in a columnar manner similarly to the temporal cortex, precisely because it is constituted of a unique layer of pyramidal cells, there might exist a clustered

organization within it. This topographical organization at the small scale might well lead to extended patches of same modality or same stimulus preference at the scale of millimeters (Dahl et al., 2009), which could possibly be translated into an enhanced activity for face-stimuli in fMRI as observed in an area within the monkey hippocampus (Ku et al., 2011). Finally, we wondered whether selectivity and responsiveness would differ among hippocampal subfields. In CA1 and CA3, selectivity and responsiveness of interneurons and pyramidal cells were similar to the one reported for the whole hippocampal structure. Pyramidal cells from CA3 only slightly overtook interneurons in their selectivity, but in a non-significant manner, not allowing us to conclude on a distinction between pattern-completion and pattern-separation type of activities between CA1 and CA3 similar to the one observed in rats and mice (Guzowski et al., 2004).

In summary the present study enriches existing knowledge on different types of cells' properties across species, and thus enables comparison of response properties of selective place cells in rodents to selective visual cells in primates. Because we provide a common framework for comparing studies carried out in various animal models; the present results could be used as a reference tool for future studies using recordings of monkey hippocampal neurons.

Figures legends

Figure 1. Location of the recording sites and physiological properties of hippocampal cells. **A.** Top: Three dimensional representation of the monkey brain highlighting the hippocampus (in red) and the boundaries of the recording chamber (vertical lines). Bottom: selected MRI sections along the antero-posterior axis showing the boundaries of the recording chamber and the track of the electrode (middle section). **B.** Projections of the peak amplitude of 4 units recorded from a tetrode. Each subplot represents the projection of one electrode over another. Each cluster represents a unit. **C.** Waveforms corresponding to the clusters presented in B. Each row represents one unit while each column corresponds to one electrode of the tetrode. **D.** Excitatory connections between pyramidal cells (left) and interneurons (right). Each row represents a different pair. From left to right: autocorrelogram of the excitatory cell, cross correlogram between the two cells and autocorrelogram of the inhibitory cell. Superposed on the crosscorrelograms are the wave forms of the cells. **E.** reciprocal connection between an interneuron (left) and a pyramidal cell (right). Same convention as in D. **F.** Inhibitory connection from an interneuron on a pyramidal cell. Same convention as in D. **G.** Proportion of cell types in the CA1 and C3 fields of the hippocampus. **H.** Classification of cells into pyramidal (red) or interneurons (blue). Left: separation (dotted line) based on trough to peak duration and half-amplitude duration. Crosses indicate cells for which an excitatory (red) or inhibitory connection could be assessed by the crosscorrelogram. Right: trough to peak duration against firing rate. Otherwise, same as in left.

Figure2. Responsiveness of the cells by type, region and penetration. **A.** Response properties by cell types. Left: proportion of cells responding to visual and/or auditory stimuli in interneurons and pyramidal cells. Middle: distribution of visual stimuli preference in responding interneurons and pyramidal cells. Right: distribution of auditory stimuli preference in responding interneurons and pyramidal cells. **B.** Responsiveness by cell type ad by CA1 and CA3 subregions of the hippocampus. Left, middle and right, same conventions as in A. **C.** Distribution of modality responses (left) and stimuli preferences (right) by penetrations in CA1 and CA3 fields of the hippocampus.

Figure 3. Selectivity by cell types and subregions of the hippocampus. **A.** Left: Distribution of the sparseness statistics for pyramidal cells and interneurons. Middle: average of the sparseness statistics for the interneurons and pyramidal cells. Right. Average sparseness statistic for bursting, regular spiking and unclassified cells. Errors bars represent SEM (*P < 0.05, **P < 0.01). **B.** Top row: Distribution of the sparseness statistics for pyramidal cells and interneurons in CA1 (top row) and CA3 (bottom row). Same conventions as in A.

Figures 4. Example of the activities of a pair of connected neurons (neuron A inhibiting neuron B). Beneath each picture, is represented a raster of the neuron's activity during baseline (from -1 to 0 s, time calculated from stimulus onset) and presentation (from 0 to 2s) of this picture. Each line represents a single trial and each tick represents an action potential. The blue ticks represent the stimulus onset at 0 ms and the green ticks represent stimulus offset à 2 s. Beneath each raster, is represented a peristimulus histogram (PSTH) summing neuron's activity among trials. The scale is represented in Hz (i.e. in number of spikes per seconds). Neurons' activities for acoustic stimuli are not represented. In red are presented inverse activities for several stimuli. When neuron A lowers its inhibition, it enables neuron B to fire. Neuron B is also modulating its activity in response to other inputs, since it presents an enhanced firing to several stimuli for which neuron A does not inhibit its activity.

References

- Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology* 92:600-608.
- Bird CM, Burgess N (2008) The hippocampus and memory: insights from spatial processing. *Nat Rev Neurosci* 9:182-194.
- Bryant HL, Marcos AR, Segundo JP (1973) Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs. *Journal of Neurophysiology* 36:205-225.
- Christian EP, Deadwyler SA (1986) Behavioral functions and hippocampal cell types: evidence for two nonoverlapping populations in the rat. *J Neurophysiol* 55:331-348.
- Clayton NS, Russell J (2009) Looking for episodic memory in animals and young children: prospects for a new minimalism. *Neuropsychologia* 47:2330-2340.
- Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of Neurophysiology* 88:3487-3497.
- Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *The Journal of Neuroscience* 21:3646-3655.
- Csicsvari J, Hirase H, Czurko A, Buzsáki G (1998) Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* 21:179-189.
- Csicsvari J, Hirase H, Czurko A, Mamiya A, Buzsáki G (1999) Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat. *J Neurosci* 19:274-287.
- Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924-11932.

C. Dissection des microcircuits dans l'hippocampe du singe: lien entre les réponses des cellules et leur signature électrophysiologique *in vivo*

- Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327:584-587.
- Eifuku S, Nishijo H, Kita T, Ono T (1995) Neuronal activity in the primate hippocampal formation during a conditional association task based on the subject's location. *Journal of Neuroscience* 15:4952-4969.
- Frank LM, Brown EN, Wilson MA (2001) A comparison of the firing properties of putative excitatory and inhibitory neurons from CA1 and the entorhinal cortex. *Journal of Neurophysiology* 86:2029-2040.
- Fried I, MacDonald KA, Wilson CL (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18:753-765.
- Fried I, Cameron KA, Yashar S, Fong R, Morrow JW (2002) Inhibitory and excitatory responses of single neurons in the human medial temporal lobe during recognition of faces and objects. *Cereb Cortex* 12:575-584.
- Fuentealba P, Begum R, Capogna M, Jinno S, Marton L, Csicsvari J, Thomson A, Somogyi P, Klausberger T (2008) Ivy cells: a population of nitric-oxide-producing, slow-spiking GABAergic neurons and their involvement in hippocampal network activity. *Neuron* 57:917-929.
- Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322:96-101.
- Gray CM, Maldonado PE, Wilson M, McNaughton B (1995) Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *Journal of Neuroscience Methods* 63:43-54.
- Guzowski JF, Knierim JJ, Moser EI (2004) Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron* 44:581-584.
- Hampson RE, Pons TP, Stanford TR, Deadwyler SA (2004) Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. *Proc Natl Acad Sci U S A* 101:3184-3189.
- Hampton RR, Schwartz BL (2004) Episodic memory in nonhumans: what, and where, is when? *Curr Opin Neurobiol* 14:192-197.
- Hampton RR, Hampstead BM, Murray EA (2005) Rhesus monkeys (*Macaca mulatta*) demonstrate robust memory for what and where, but not when, in an open-field test of memory. *Learning and Motivation* 36:245-259.
- Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular features predicted by extracellular recordings in the hippocampus *in vivo*. *Journal of Neurophysiology* 84:390-400.
- Hori E, Tabuchi E, Matsumura N, Tamura R, Eifuku S, Endo S, Nishijo H, Ono T (2003) Representation of place by monkey hippocampal neurons in real and virtual translocation. *Hippocampus* 13:190-196.
- Insausti R, Amaral DG (2008) Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. *The Journal of Comparative Neurology* 509:608-641.
- Insausti R, Herrero MT, Witter MP (1997) Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus* 7:146-183.
- Ison MJ, Mormann F, Cerf M, Koch C, Fried I, Quiñero R (2011) Selectivity of pyramidal cells and interneurons in the Human Medial Temporal Lobe. *J Neurophysiol*.
- Jutras MJ, Buffalo EA (2010) Recognition memory signals in the macaque hippocampus. *Proc Natl Acad Sci U S A* 107:401-406.

- Klausberger T, Somogyi P (2008) Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* 321:53-57.
- Klausberger T, Magill PJ, Marton LF, Roberts JDB, Cobden PM, Buzsaki G, Somogyi P (2003) Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature* 421:844-848.
- Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci* 3:946-953.
- Ku S-P, Tolias Andreas S, Logothetis Nikos K, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70:352-362.
- Marshall L, Henze DA, Hirase H, Leinekugel X, Dragoi G, Buzsaki G (2002) Hippocampal pyramidal cell-interneuron spike transmission is frequency dependent and responsible for place modulation of interneuron discharge. *J Neurosci* 22:RC197.
- Mohedano-Moriano A, Martinez-Marcos A, Pro-Sistiaga P, Blaizot X, Arroyo-Jimenez MM, Marcos P, Artacho-Pérula E, Insausti R (2008) Convergence of unimodal and polymodal sensory input to the entorhinal cortex in the fascicularis monkey. *Neuroscience* 151:255-271.
- Nakamura K, Kubota K (1996) The primate temporal pole: its putative role in object recognition and memory. *Behav Brain Res* 77:53-77.
- Naya Y, Suzuki WA (2011) Integrating What and When across the primate medial temporal lobe. *Science* 333:773-776.
- Nishijo H, Ono T, Eifuku S, Tamura R (1997) The relationship between monkey hippocampus place-related neural activity and action in space. *Neurosci Lett* 226:57-60.
- O'Keefe J (1999) Do hippocampal pyramidal cells signal non-spatial as well as spatial information? *Hippocampus* 9:352-364.
- O'Mara SM, Rolls ET, Berthoz A, Kesner RP (1994) Neurons responding to whole-body motion in the primate hippocampus. *J Neurosci* 14:6511-6523.
- Ono T, Eifuku S, Nakamura K, Nishijo H (1993) Monkey hippocampal neuron responses related to spatial and non-spatial influence. *Neurosci Lett* 159:75-78.
- Paxinos G, Huang X-F, Toga AW (2000) The Rhesus monkey brain in stereotaxic coordinates. San Diego; London; Boston: Academic press.
- Quian Quiroga R, Kraskov A, Koch C, Fried I (2009) Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr Biol* 19:1308-1313.
- Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87-91.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.
- Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73:713-726.
- Rolls ET, Xiang J, Franco L (2005) Object, space, and object-space representations in the primate hippocampus. *J Neurophysiol* 94:833-844.
- Rolls ET, Miyashita Y, Cahusac PM, Kesner RP, Niki H, Feigenbaum JD, Bach L (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J Neurosci* 9:1835-1845.
- Saleem KS, Logothetis NK (2006) Atlas of the rhesus monkey brain in stereotaxic coordinates a combined MRI and histology. London [u.a.: Academic Press.

C. Dissection des microcircuits dans l'hippocampe du singe: lien entre les réponses des cellules et leur signature électrophysiologique *in vivo*

- Tamura H, Kaneko H, Kawasaki K, Fujita I (2004) Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. *Journal of Neurophysiology* 91:2782-2796.
- Tamura R, Ono T, Fukuda M, Nishijo H (1992a) Monkey hippocampal neuron responses to complex sensory stimulation during object discrimination. *Hippocampus* 2:287-306.
- Tamura R, Ono T, Fukuda M, Nakamura K (1992b) Spatial responsiveness of monkey hippocampal neurons to various visual and auditory stimuli. *Hippocampus* 2:307-322.
- Vigneswaran G, Kraskov A, Lemon RN (2011) Large identified pyramidal cells in macaque motor and premotor cortex exhibit "thin spikes": Implications for cell type classification. *The Journal of Neuroscience* 31:14235-14242.
- Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273-1276.
- Viskontas IV, Ekstrom AD, Wilson CL, Fried I (2007) Characterizing interneuron and pyramidal cells in the human medial temporal lobe *in vivo* using extracellular recordings. *Hippocampus* 17:49-57.
- Waydo S, Kraskov A, Quiñero R, Fried I, Koch C (2006) Sparse representation in the human medial temporal lobe. *J Neurosci* 26:10232-10234.
- Wiebe SP, Staubli UV (1999) Dynamic filtering of recognition memory codes in the hippocampus. *J Neurosci* 19:10562-10574.
- Wiebe SP, Staubli UV (2001) Recognition memory correlates of hippocampal theta cells. *J Neurosci* 21:3955-3967.
- Wilson FA, O'Scalaidhe SP, Goldman-Rakic PS (1994) Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proceedings of the National Academy of Sciences* 91:4009-4013.
- Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300:1578-1581.
- Wirth S, Avsar E, Chiu CC, Sharma V, Smith AC, Brown E, Suzuki WA (2009) Trial outcome and associative learning signals in the monkey hippocampus. *Neuron* 61:930-940.
- Wood ER, Dudchenko PA, Eichenbaum H (1999) The global record of memory in hippocampal neuronal activity. *Nature* 397:613-616.
- Xiang JZ, Brown MW (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* 37:657-676.
- Yanike M, Wirth S, Suzuki WA (2004) Representation of well-learned information in the monkey hippocampus. *Neuron* 42:477-487.

Figure 1

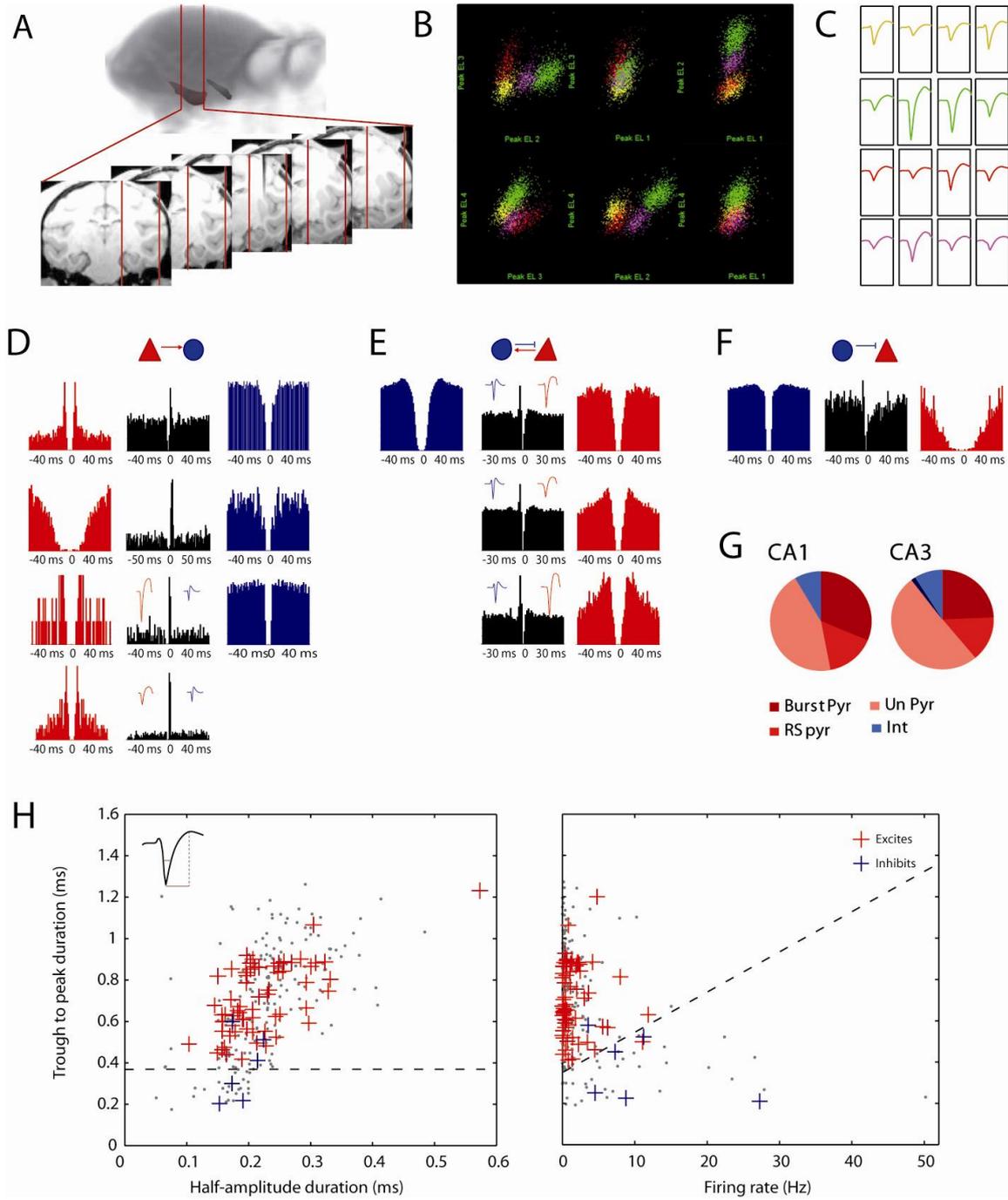


Figure 2

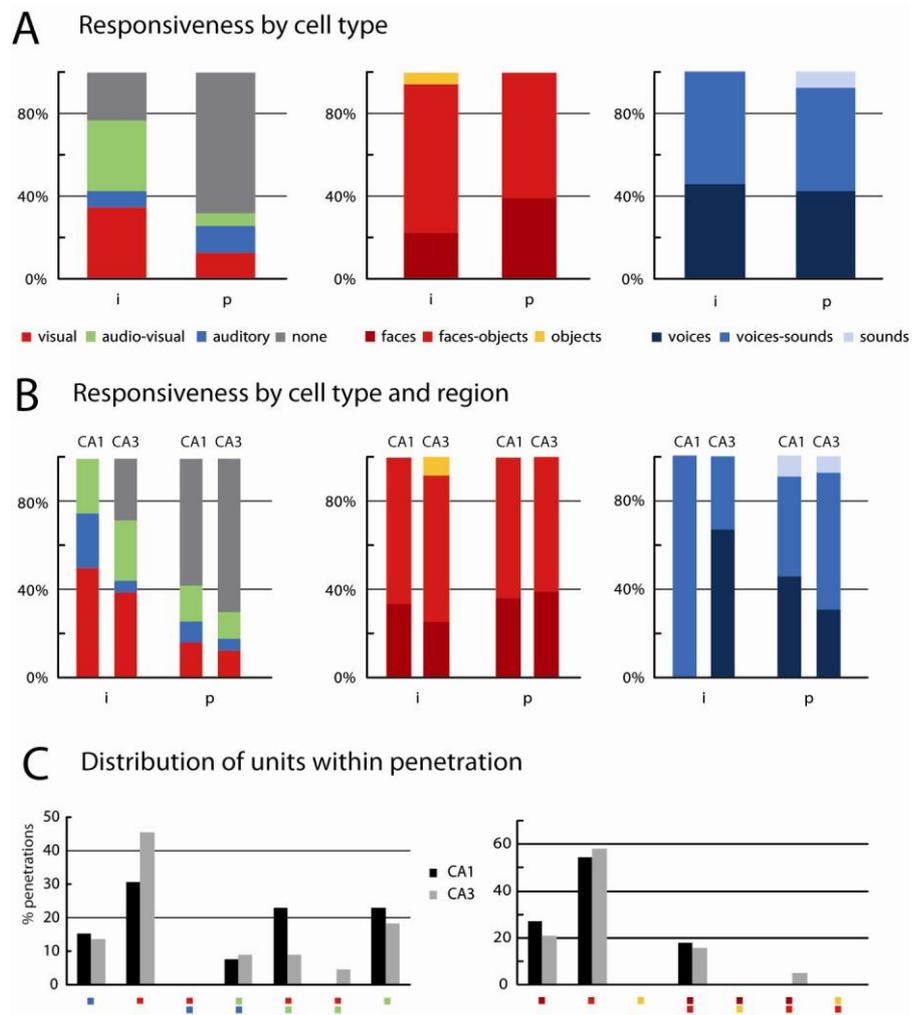


Figure 3

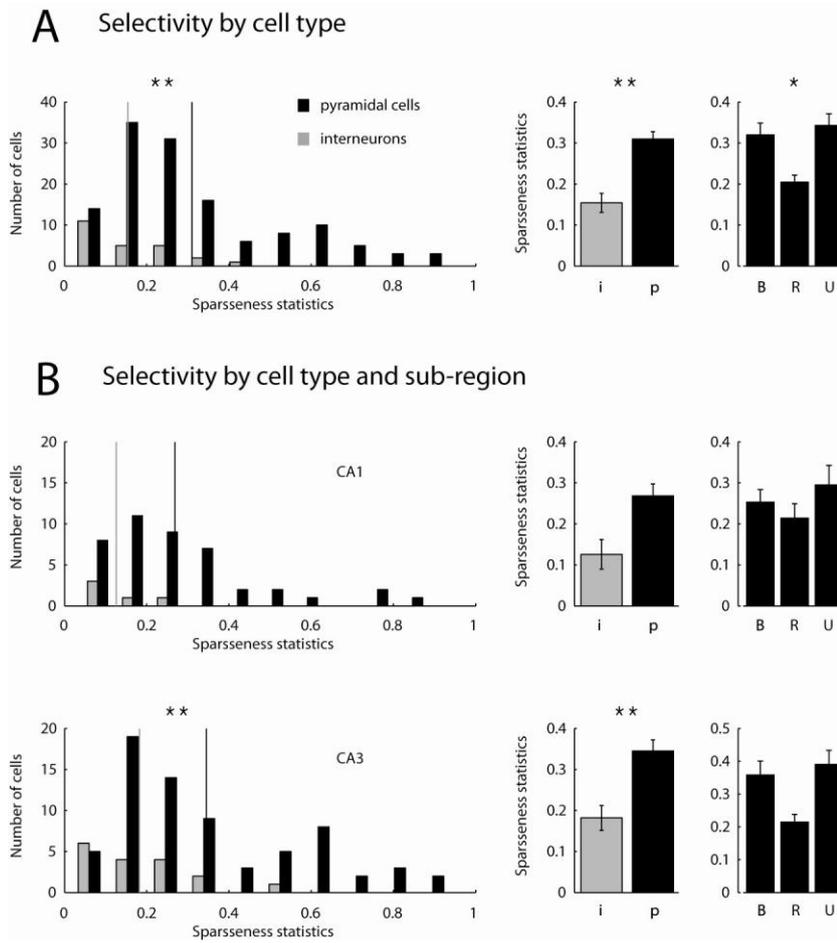


Figure 4



Discussion générale

L'Homme est un animal intensément social. Pour évoluer dans des sociétés complexes, il a développé des facultés mentales avancées pour identifier, mémoriser et utiliser les stimuli sociaux de son environnement comme les visages et les voix. La reconnaissance individuelle a été étudiée à trois niveaux complémentaires au cours de cette thèse : depuis son existence comportementale chez le macaque rhésus jusqu'à son substrat neural et neuronal. Une première étude comportementale a mis en évidence que les singes, comme les humains, possèdent la faculté de reconnaître individuellement les personnes connues, sur la base de leur voix et de leur visage qu'ils associent spontanément (**Expériences et résultats. §A**). Deuxièmement, ce projet a mis en lumière comment les stimuli sociaux – les visages – sont représentés de façon redondante et complémentaire par les neurones de différentes aires cérébrales du lobe temporal (**Expériences et résultats. §B**). Enfin, l'une de ces aires, l'hippocampe, est une structure qui a évolué à des degrés divers chez différents mammifères pour soutenir la mémoire autobiographique et les représentations spatiales. Nous avons ici caractérisé les différents types de neurones et leur micro-connectivité afin de fournir un cadre commun pour comparer les fonctions de l'hippocampe à travers les espèces et d'étudier comment ces microcircuits coopèrent pour représenter des informations complexes (**Expériences et résultats. §C**). Dans la suite nous allons décrire l'originalité, les limites et les perspectives associées à ces trois axes de recherche.

A.1. Etendue du concept de Personne chez le macaque rhésus et d'autres espèces animales

Les macaques rhésus perçoivent et discriminent les identités faciales et vocales (Rendall et al., 1996; Parr et al., 2000; Gothard et al., 2004; Hauser, 2007; Gothard et al., 2009). Nous démontrons qu'ils possèdent de plus la faculté de les associer et de retrouver l'information d'une

modalité à partir de l'autre, c'est-à-dire la faculté de reconnaître individuellement leurs congénères mais aussi des individus humains familiers (Sliwa et al., 2011). Cette mise en évidence permet 1) d'approcher l'existence d'un concept de Personne chez le singe, 2) de démontrer que les individus d'une autre espèce ayant une importance socio-écologique peuvent également bénéficier de ce type de traitement cognitif, 3) de valider l'utilisation de photographies de visages d'individus connus dans les expériences de cognition sociale chez le singe.

Ces trois résultats nouveaux ajoutent à la compréhension des relations sociales qu'entretiennent les singes rhésus entre eux. Ils démontrent que la capacité cognitive fine et complexe qu'est la reconnaissance individuelle n'est pas l'apanage de l'humain mais est partagée par un plus grand nombre de primates. Ainsi les macaques aussi pourraient faire l'usage d'un concept de Personne. Par ailleurs la flexibilité de ce trait cognitif qui peut être utilisé pour reconnaître des humains dans notre étude montre l'importance de l'exposition et de l'expérience dans son acquisition. Une originalité supplémentaire de notre étude réside dans le fait qu'elle a été menée en laboratoire avec des méthodes objectives d'analyse des mouvements oculaires assistée par ordinateur. Les résultats observés sont ainsi indépendants de la fiabilité d'observateurs humains, qui peuvent constituer une limite méthodologique, par exemple dans le cas d'éthogrammes réalisés en double aveugle par notations de vidéos. Par ailleurs ceci place notre étude en complément d'observations éthologiques réalisées en conditions naturelles, qui sont nécessaires pour comprendre le comportement des espèces, mais peuvent néanmoins se heurter à une surinterprétation anthropomorphique. Enfin notons que les résultats que nous avons obtenus ont depuis été confirmés également en laboratoire par un autre groupe de recherche utilisant une méthode originale de perturbation cross-modale de choix d'appariement actif entre une vidéo et une image d'individu (**Figure 37**, Adachi and Hampton, 2011).

Notre étude et celle d'ADACHI et HAMPTON (2011) démontrent donc que les macaques rhésus sont capables d'associer les identités faciales et vocales au niveau individuel. Elles permettent d'approcher l'existence d'un concept de Personne chez le singe qui serait constitué au moins de ces deux éléments d'identité. Chez l'Homme ce concept s'étoffe de divers autres éléments ayant trait à l'apparence physique de l'individu (odeur, touché, forme visuelle du corps statique ou en mouvement) et aux connaissances épisodiques et sémantiques qui lui sont associées (nom, informations biographiques, personnalité, nature des interactions passés, **Introduction. §A.1.**). Bien que n'ayant pas étudié ces divers éléments de façon systématique, nous avons néanmoins observé que la nature des interactions passées est également associée à l'identité des individus par les singes, comme en témoigne l'orientation préférentielle de leur attention visuelle vers les photos des individus avec lesquels ils entretiennent des relations privilégiées quotidiennes.

Une première perspective complémentaire à cette étude serait donc d'étudier l'ampleur du concept d'individu chez le singe. Cette question pourrait tout d'abord porter sur l'étendue et la plasticité des représentations multisensorielles en testant de façon cross-modale l'utilisation par les singes des diverses autres modalités par lesquelles les individus peuvent être reconnus (odeur, toucher, corps statique et dynamique, visage dynamique). Des études menées chez d'autres

espèces pourraient constituer des fondements pour ces investigations. Par exemple les chevaux (*Equus caballus*) sont capables d'associer l'identité visuelle du corps de leur congénère en mouvement à son hennissement, montrant que la reconnaissance individuelle est réalisée chez cette espèce à partir d'autres attributs physiques visuels différents du visage statique (Proops et al., 2009). Le hamster doré (*Mesocricetus auratus*) intégrerait plusieurs odeurs individuelles différentes de ses congénères (oreilles, sécrétions vaginales ou glandes du ventre, mais non l'urine) en un concept d'individu (Johnston, 1993; Johnston and Peng, 2008). Bien qu'on ne puisse pas exclure que ces différentes odeurs partagent au moins quelques composés chimiques communs, les auteurs expliquent que ceux-ci seraient trop faiblement concentrés pour permettre aux hamsters de réaliser une association perceptive. Les hamsters utiliseraient donc la mémoire d'une association antérieure créée entre ces deux odeurs, c'est-à-dire feraient preuve de reconnaissance individuelle. Enfin, bien que plus controversé (Barton, 2006), les dauphins (*Tursiops truncatus*, Janik et al., 2006) et les perroquets (*Forpus conspicillatus*, Wanker et al., 2005) utiliseraient des signatures acoustiques individuelles pour se présenter auprès des autres individus ou pour appeler leurs congénères. Ces « noms » auditifs seraient indépendants des caractéristiques acoustiques de leur voix et seraient associés par leurs congénères à leurs attributs physiques visuels individuels.



Figure 37: Tâche active de choix d'appariement (gauche) et résultats de la perturbation cross-modale des deux singes testés (droite). Les deux singes ont appris à indiquer en touchant un écran tactile quelle photo correspond à l'individu présenté précédemment en vidéo. Dans une seconde étape, une vocalise est jouée en playback entre la présentation vidéo et la présentation des photos. L'analyse des essais incorrects d'appariement dévoile que les singes M1 et M2 touchent plus souvent l'image correspondant à l'identité de la vocalise jouée. D'après Adachi and Hampton, 2011.

D'autre part, la nature des connaissances des singes rhésus quant à leurs interactions passées avec leurs congénères reste à étudier. Par exemple nous pourrions comprendre si des éléments passés permettent aux singes de prévoir le comportement futur des individus et d'interagir avec eux adéquatement. Une étude éthologique réalisée sur des babouins (*Papio cynocephalus ursinus*)

apporte un premier élément de réponse à cette question (Cheney and Seyfarth, 1997). Après un combat entre deux individus, les chercheurs ont joué en playback des vocalises de réconciliation mimant la voix soit de l'adversaire passé ayant gagné le combat, soit d'un individu contrôlé n'ayant pas participé au combat. Les vocalises de l'adversaire passé avait pour effet l'approche de la victime vers l'ancien opposant et des interactions amicales entre eux, non effectuées lorsque la voix contrôlée était jouée. Cette étude met en évidence que le babouin victime avait créé une association cognitive entre la voix de son opposant et la mémoire de l'interaction passée, et qu'il utilise cette information pour guider son comportement présent de réconciliation.

Enfin dans notre étude nous étions limités par le nombre d'individus connus personnellement des singes ayant participé à l'étude, cependant en conditions naturelles les groupes de macaques peuvent atteindre jusqu'à deux cent individus (**Introduction. §B.1.**), la connaissance de l'ampleur cette fois-ci du nombre d'individu pouvant être reconnus individuellement par les macaques est également une question importante qui mériterait d'être étudiée.

A.2. Représentation des individus dans le lobe temporal

Les neurones de plusieurs régions du cortex temporal représentent sélectivement ou préférentiellement les visages par rapport à d'autres objets visuels (pour revue : Tsao and Livingstone, 2008). Nous démontrons qu'une population de neurones hippocampiques présente également cette particularité. Ainsi les informations relatives aux visages sont représentées par les neurones hippocampiques ainsi que par les neurones inférotemporaux de façon redondante et complémentaire démontrant que ces deux structures participent aux représentations des visages. Les individus sont représentés de façon plus invariante et catégorielle par les neurones de l'hippocampe que par les neurones inférotemporaux, mais non pour autant multimodale. Au contraire les visages sont représentés de façon plus perceptive, c'est-à-dire en tenant compte du point de vue par les neurones du cortex inférotemporal.

Ces résultats montrent que les cellules hippocampiques chez le singe codent les stimuli sociaux et les catégorisent sans que le singe n'exécute une tâche de catégorisation active de stimuli. Ces nouveaux résultats révèlent le substrat neuronal probable sous-tendant l'activité IRMf sélective aux visages enregistrée dans l'hippocampe chez le singe (Ku et al., 2011). Les patrons de sélectivité observés, suggèrent que les neurones hippocampiques catégorisent l'espace cognitif créé par l'ensemble des stimuli présentés dans cette expérience à laquelle se superpose des informations relatives aux catégories auxquelles appartiennent ces stimuli.

Pour tester plus en avant l'hypothèse de « carte cognitive » chez le singe avec ce type de stimuli, il serait intéressant comme perspective à notre étude de mener une étude de « remapping » (Colgin et al., 2008). Une façon simple de le faire serait de présenter dans un premier temps un premier set de stimuli pendant l'enregistrement et en deuxième partie de l'expérience de présenter un second set de stimuli au singe. On pourrait alors observer si les cellules hippocampiques changent de sélectivité. Par ailleurs il serait alors intéressant d'observer

si les cellules gardent une certaine spécificité ou peuvent en changer entièrement. Par exemple on pourrait faire l'hypothèse que les neurones sélectifs aux visages lorsque confrontés au premier set de stimuli pourraient dans le deuxième set garder une spécificité aux visages mais répondre à une autre identité. L'hypothèse opposée pourrait être que ces neurones sélectifs aux visages dans le premier set deviendraient sélectifs à des stimuli complètement différents, par exemple des lieux de l'espace ou à des stimuli auditifs.

Le fait que nous n'observions qu'un faible nombre de neurones présentant une activité aux stimuli auditifs dans l'hippocampe, ayant de plus des activités auditives globalement peu soutenues, peu sélectives et non corrélées aux activations visuelles, remet cependant en cause profondément l'existence de neurones multisensoriels visio-auditifs dans la région d'enregistrement et plaide pour une spécialisation sensorielle à dominante visuelle de cette région. Il serait alors peu probable dans l'expérience de « remapping » proposée ci-dessus d'observer des changements drastiques de spécificité des neurones. Ceci amène à s'interroger sur les spécificités sensorielles au sein de l'hippocampe et à considérer leur distribution le long de l'axe rostro-caudal.

A.2.1. Dissociation antéropostérieure dans l'hippocampe et représentation des voix

La faiblesse de l'activité des neurones hippocampiques aux stimuli auditifs dans notre étude, amène à nous poser la question de l'existence d'une spécificité neuroanatomique et fonctionnelle ségréguée le long de l'axe rostro-caudal de l'hippocampe. En effet, les afférences et efférences hippocampiques directes montrent une organisation topographique le long de l'axe rostro-caudal. Les régions les plus postérieures chez le singe et l'Homme (homologues des régions dorsales chez le rat) se projettent sur le cortex cingulaire postérieur alors que les parties les plus antérieures présentent de plus denses projections vers le cortex préfrontal médian, orbitofrontal et sur l'amygdale. Les gradients d'entrées et sorties suggèrent que l'hippocampe est organisé fonctionnellement selon son axe rostro-caudal. Cette interprétation a été confirmée par deux analyses de profils d'expression génétique dénombrant pas moins de neuf régions distinctes organisées le long de l'axe rostro-caudal de l'hippocampe chez la souris (Thompson et al., 2008; Dong et al., 2009). La différenciation anatomique fait écho à des différenciations fonctionnelles observées chez le rat (Jung et al., 1994; Moser and Moser, 1998; Kjelstrup et al., 2008; Bast et al., 2009), le singe (Colombo et al., 1998) et l'humain (Small et al., 2001; Peters et al., 2007a; Peters et al., 2007b). En se fondant sur ces différentes observations, un éventail d'études a établi que la partie postérieure de l'hippocampe serait probablement plutôt impliquée dans les processus mnésiques et cognitifs alors que la partie antérieure serait plus impliquée dans des mécanismes émotionnels, de stress, d'intégration sensori-motrice ou dirigés vers un but (Bannerman et al., 2004; Fanselow and Dong, 2010; Poppenk and Moscovitch, 2011; Small et al., 2011).

Cette hétérogénéité fonctionnelle serait-elle doublée d'une dissociation des processus sensoriels le long de l'axe rostro-caudal ?

Les études neuroanatomiques menées chez les singes rhésus (*Macaca mulatta*) et cynomolgus (*Macaca fascicularis*) montrent que les projections des cortex unisensoriels visuel et auditif vers le complexe hippocampique (hippocampe, cortex entorhinal et cortex périrhinal et parahippocampique) diffèrent effectivement. Cependant cette distinction n'est pas principalement topographique.

En effet, les cortex unisensoriels associatifs à la fois visuels (TE) et auditifs (STG) font principalement des projections indirectes vers la région caudale de l'hippocampe (**Figure 38**) (Mohedano-Moriano et al., 2007; Mohedano-Moriano et al., 2008). Les aires les plus antérieures des cortex associatifs auditifs (STG), i.e. le pôle temporal et les aires Ts1-3 qui nous intéressent plus particulièrement car elles participent à la représentation aux niveaux neural et neuronal des identités vocales (Petkov et al., 2008; Perrodin et al., 2011), feraient des projections indirectes vers les parties médiane et caudale de l'hippocampe et dans une moindre mesure vers la partie rostrale (Kosel et al., 1982; Amaral et al., 1983; Insausti et al., 1987; Lavenex et al., 2004) cités par (Munoz-Lopez et al., 2010).



Figure 38: Diagramme des projections corticales des aires associatives unimodales visuelles (rouge) et auditives (vert pâle à gauche), du pôle temporal (vert pomme à droite) et des cortex polymodaux (orange) vers les sous-régions entorhinales. La largeur des flèches indique la densité des projections. Dans chaque cas les terminaisons hippocampiques présumées sont indiquées sur une vue schématique de l'hippocampe présentant ses différentes portions (caudale, rostrale et intermédiaire). D'après Mohedano-Moriano et al., 2007 et Mohedano-Moriano et al., 2008.

Cependant, contrairement aux connections réciproques fortes et directes qu'entretient la partie *caudale* du complexe hippocampique avec les aires visuelles inférotemporales, les connexions réciproques avec les aires auditives secondaires sont considérablement plus modestes. De même les projections auditives vers la partie *rostrale* du complexe hippocampique

sont bien moins denses que les entrées polymodales, telles celles de TPO (Insausti and Amaral, 2008; Mohedano-Moriano et al., 2008). Ainsi, alors que la « voie ventrale » visuelle décrite par MISHKIN et UNGERLEIDER (1982) est directement reliée à l'hippocampe, tel n'est pas le cas du domaine auditif. En effet la voie d'entrée des projections auditives vers l'hippocampe est largement indirecte et nécessite au moins deux relais synaptiques dans des régions polymodales du sillon temporal supérieur : TAA et TPO (Munoz-Lopez et al., 2010). Dans ces régions les informations auditives s'intègrent à des informations d'autres modalités sensorielles et rejoignent seulement alors le système hippocampique.

D'un point de vue fonctionnel, la perception et le rappel de visages ou de voix appris en isolation impliqueraient préférentiellement l'hippocampe postérieur chez l'Homme (Small et al., 2001; Peters et al., 2007a; Peters et al., 2007b). De même chez le singe, une étude par TEP indique une activation de l'hippocampe postérieur élicitée par les vocalises des congénères (Gilda-Costa et al., 2004), cependant celle menée en IRMf sur la perception des visages pointait un rôle de l'hippocampe antérieur (Ku et al., 2011). Lorsque les visages sont appris en association avec des noms, l'activation hippocampique se déplacerait alors vers une région intermédiaire (Small et al., 2001) ou serait située dans la partie antérieure (Sperling et al., 2003; Chua et al., 2007). On remarque ici que le caractère multisensoriel de l'hippocampe a la plupart du temps été étudié en utilisant des associations visage-nom (Paller et al., 2000; Paller et al., 2003; Sperling et al., 2003; Zeineh et al., 2003; Kirwan and Stark, 2004; Chua et al., 2007; Tsukiura and Cabeza, 2008) et rarement des associations voix-visage (Shah et al., 2001; Holdstock et al., 2010; Joassin et al., 2011) qui pourraient ne pas être équivalentes, et que la région hippocampique activée selon l'axe rostro-caudal est rarement identifiée comme telle et varie amplement entre les études. Ainsi les études électrophysiologiques chez l'humain présentent l'activité de neurones invariants à l'identité des visages et des noms mais ne testent pas l'invariance à l'identité vocale (Quiñones Quiroga et al., 2009). Il est envisageable que ces cellules puissent ne pas s'activer pour l'identité vocale car elles sont situées de façon très antérieure sur l'axe rostro-caudal (Fried et al., 1997; Kreiman et al., 2000; Fried et al., 2002), comme dans notre étude. En effet dans l'étude de JOASSIN et coll. (2011), la région hippocampique présentant un signal d'intégration superadditif des signaux vocaux et faciaux était située de façon intermédiaire voire postérieure sur l'axe rostro-caudal.

En conclusion, ces remarques adjointes à nos résultats, appellent à la réalisation d'une méta-analyse des différentes études menées en IRMf avec comme objectif de caractériser les activations élicitées par les voix et les visages appris seuls ou en association le long de l'axe rostro-caudal. Par ailleurs une perspective intéressante à notre étude serait d'effectuer des enregistrements systématiques le long de l'axe rostro-caudal de l'hippocampe afin de traquer les spécificités sensorielles des neurones tout du long et plus précisément de comprendre l'évolution de l'activité neuronale en réponse à la perception de voix et de visages. Une autre façon d'aborder ce problème serait d'utiliser l'IRMf chez le singe pour observer des gradients d'activité aux visages et aux voix et de réaliser des enregistrements électrophysiologiques dans les régions activées uniquement par l'un des deux stimuli ainsi que dans les régions activées par les deux stimuli séparément ou présentés conjointement.

A.2.2. Faiblesse de la représentation des voix dans le cortex inférotemporal

Enfin notons que si nous postulions que les voix pourraient être représentées dans l'hippocampe car cette structure est multimodale, nous postulions également qu'une représentation des voix pourrait être observée dans le cortex inférotemporal par les neurones répondants aux visages par effet feedback, comme ceci a été postulé par les études menées chez l'Homme en IRMf (von Kriegstein et al., 2005; von Kriegstein and Giraud, 2006; von Kriegstein et al., 2006a; von Kriegstein et al., 2008; Blank et al., 2011) et chez le singe par PET-scan (Gil-da-Costa et al., 2004). En effet ces études mettent en évidence chez le singe, que la perception de vocalises de congénères active en TEP la région visuelle associative du cortex inférotemporal TE. De façon similaire chez l'Homme, l'activation en IRMf de l'aire FFA par la perception de voix connues est observée lorsque celles-ci ont été préalablement apprises avec la vue d'un visage. Or cette caractéristique concerne la moitié des stimuli auditifs présentés dans notre étude, ce qui participe d'ailleurs à son originalité. En effet nous avons fait le choix d'utiliser un set de stimuli peu exploités dans les études réalisées jusqu'ici en électrophysiologie chez le macaque, constitué de représentations visuelles ou auditives d'individus et d'objets rencontrés par les singes quotidiennement dans leur environnement de vie et comparé à des stimuli similaires mais non personnellement connus auxquels les singes ont été familiarisés. Notre première étude ainsi que celle d'ADACHI et coll. (2011) ont au préalable démontré que les singes avaient appris la correspondance entre les stimuli vocaux et leur visage naturellement par leur interactions quotidiennes (Adachi and Hampton, 2011; Sliwa et al., 2011). Pour autant nous n'observons pas de codage des voix qui soit similaire au codage des visages par les neurones inférotemporaux enregistrés. Les activités spécifiques aux visages que nous observons sont caractérisées par une augmentation brusque et importante du taux de décharge par rapport au taux de décharge de base de la cellule, alors que les activités aux stimuli auditifs présentent rarement ce type de modulation et jamais de façon symétrique aux visages et invariante à l'identité de l'individu. Cette apparente divergence appelle des analyses supplémentaires de nos signaux d'enregistrement qu'il serait intéressant de réaliser. En effet les signaux BOLD enregistrés par IRMf ne sont que faiblement corrélés aux activités des neurones unitaires (Sirotnin and Das, 2009 ; mais voir Mukamel et al., 2005). Au contraire une plus forte corrélation a été mise en évidence entre les signaux BOLD et les activités populationnelles appelées « potentiels de champs locaux » (LFP) ou encore les activités de plusieurs neurones concaténées, appelées MUA (multiple units activity) (Logothetis et al., 2001; Viswanathan and Freeman, 2007; Logothetis, 2008). Ainsi il serait intéressant d'étudier si les voix connues ne seraient pas distinctes des voix inconnues au niveau de leur représentation par ces autres signaux électrophysiologiques. Les faibles modulations d'activité des neurones unitaires que nous observons pourraient en effet être constitutives ou refléter la présence d'une activité populationnelle sous-jacente qu'il serait intéressant d'étudier à l'aide des LFP. Ces analyses seront menées prochainement.

A.3. Continuités et discontinuités évolutives dans le fonctionnement de l'hippocampe

Chez le rat, les différents types cellulaires peuvent être distingués à partir de leurs caractéristiques électrophysiologiques. Dans la dernière étude présentée ici, nous avons utilisés ces techniques pour identifier les neurones hippocampiques du macaque. Nous montrons que les analyses des interactions entre paires de cellules permettent d'inférer la nature du type cellulaire. Par extension, nous assignons l'ensemble des cellules enregistrées à l'une des deux classes de cellules en se référant cette fois-ci à la forme des potentiels d'action. Une quantification de la taille des deux groupes de neurones suggère que la probabilité d'enregistrer des interneurons par rapport à la probabilité d'enregistrer des cellules pyramidales est stable entre les espèces testées. En effet dans les autres études réalisées chez le rat et l'humain ainsi que dans la notre menée chez le singe, cette chance avoisine les 10%.

Une étude mettant en évidence une telle caractérisation n'avait pas été entreprise précédemment dans l'hippocampe du singe. La connaissance de la nature des différents neurones permet d'étudier comment les microcircuits coopèrent pour représenter des informations complexes.

Cette étude fournit un cadre commun pour comparer les fonctions de l'hippocampe à travers les espèces. En effet, chez le rat ce sont principalement les activités spatiales qui ont été étudiées, alors que chez l'Homme et le singe, les tâches cognitives ont plutôt tenté de mimer des tâches mnésiques comme en témoigne l'invention et l'utilisation du DNMTS par exemple. Notre étude permet donc d'établir un langage commun pour comprendre quels types cellulaires seraient spécifiques ou non de telle ou telle fonction. Cependant, il est important de noter, comme indiqué précédemment, que chez le rat ce sont les régions postérieures, i.e. dorsales, qui sont principalement étudiées, alors que ce sont les régions antérieures, i.e. ventrales, qui sont principalement étudiées chez le macaque car plus larges et plus accessibles (Tamura et al., 1992; Eifuku et al., 1995; Xiang and Brown, 1998; Wirth et al., 2003; Yanike et al., 2004; Wirth et al., 2009; Naya and Suzuki, 2011). Peu d'études se sont aventurées plus postérieurement chez le singe (Rolls et al., 1989; O'Mara et al., 1994; Colombo et al., 1998). De même, chez l'humain, ce sont souvent les régions antérieures qui sont implantées pour des raisons cliniques car elles sont souvent proches de foyers épileptiques intractables. Comme décrit précédemment, les régions antérieures et postérieures pourraient jouer des rôles largement différents et il serait important de prendre également cet aspect en compte pour comparer les fonctions hippocampiques entre les espèces.

Nos résultats montrent que les cellules sélectives aux visages observées dans la seconde étude sont principalement des cellules pyramidales. Chez le rat, les cellules pyramidales montrent une forte sélectivité aux lieux dans l'hippocampe dorsal, mais également aux odeurs ou aux associations (Wiebe and Staubli, 1999). Cependant, les interneurons, même s'ils déchargent plus fortement, peuvent également coder de façon sélective ces trois types de stimuli (Wiebe and

Staubli, 2001). De façon similaire, chez l'humain les cellules pyramidales présentent une forte sélectivité pour les stimuli (Ison et al., 2011). Comparativement, la sélectivité des interneurons est cependant plus faible (Ison et al., 2011). Ainsi, la caractérisation cellulaire menée dans notre troisième étude, de même que le fait que nous ayons enregistré l'activité neuronale hippocampique à des endroits homologues de ceux étudiés chez l'humain, nous permettent de comparer directement nos résultats avec les études humaines.

A.3.1. Nature de la sélectivité des neurones hippocampiques aux visages chez le singe et l'Homme

Une première question que l'on peut se poser concerne la sélectivité des neurones hippocampiques enregistrés dans notre seconde étude. Pourquoi dans notre étude les neurones sélectifs aux visages tendent à être sélectifs pour plusieurs individus de façon invariante alors que chez l'Homme ceux-ci seraient sélectifs pour des individus uniques ? Pour mener à bien cette comparaison faisons un détour pour saisir l'ensemble des études menées chez l'Homme. Chez l'Homme, deux salves d'études ont eu lieu. La première a bénéficié d'un enregistrement non supervisé des neurones et a utilisé des sets fixes de stimuli établis avant l'expérience, identiques pour tous les sujets (Fried et al., 1997; Kreiman et al., 2000; Fried et al., 2002). Cette méthode est similaire à celle que nous avons utilisée à deux différences près. Premièrement, afin de maintenir l'attention des sujets pendant qu'ils réalisaient la tâche, les expérimentateurs leur demandaient d'indiquer activement si l'image représentait un visage ou un objet (Kreiman et al., 2000) ou si le stimulus avait déjà été présenté (Fried et al., 1997; Fried et al., 2002) au contraire les singes dans notre étude devaient uniquement regarder ou écouter passivement les stimuli. Deuxièmement, notre set de stimuli a été présenté à l'identique un plus grand nombre de fois aux deux singes qu'il ne l'a été aux humains, ce qui a pu créer une habituation plus importante aux différents stimuli. Ces premières études chez l'Homme ont mis en évidence l'existence de neurones spécifiques à certaines catégories d'objets ou de visages. Dans ce contexte, les auteurs ont observé incidemment un petit nombre de neurones très sélectifs et invariants à l'identité visuelle. Une deuxième salve d'études s'est alors spécifiquement attelée à trouver ces cellules très sélectives dans l'hippocampe et à les étudier (Quiroga et al., 2005; Mormann et al., 2008; Quiroga et al., 2008b; Quiroga et al., 2009; Viskontas et al., 2009; Ison et al., 2011). Pour cela, les auteurs ont modifié leur paradigme pour chercher et caractériser systématiquement ces neurones en utilisant deux sets de stimuli consécutifs. Dans un premier set, un grand nombre d'images variées d'individus et de lieux était présenté. Les données d'enregistrements étaient alors rapidement analysées pour déterminer les stimuli préférés des neurones, dont l'activité était enregistrée en continu. Un deuxième set était ensuite présenté au sujet comportant un grand nombre d'images représentant des variations du stimulus préféré ainsi que d'autres images uniques représentant d'autres individus ou lieux. Les sujets devaient indiquer si le stimulus présenté représentait un visage ou non. Dans ces conditions, les neurones présentaient une

grande invariance allant jusqu'à représenter non seulement les différentes photos d'un même individu mais également son nom écrit ou prononcé.

Ainsi, une première observation que l'on peut tirer de l'ensemble de ces études humaines est que les neurones hippocampiques humains ne sont pas tous extrêmement sélectifs. Des neurones extrêmement sélectifs peuvent être observés chez l'humain mais ils ne représentent pas l'ensemble de la population neuronale. Ainsi, nos résultats se positionnent entre les résultats observés dans la première salve d'études réalisées chez les humains et la seconde. Bien entendu il existe des différences entre nos approches comme précisé ci-dessus, cependant nos résultats se rejoignent également en plusieurs points. Plus précisément, notre étude se rapproche de la seconde salve d'études humaine car le set de stimuli utilisé présente également des individus connus personnellement des singes. On aurait donc pu s'attendre à observer des neurones plus sélectifs même sans les avoir recherchés systématiquement. Plusieurs hypothèses explicatives peuvent être avancées pour comprendre cette divergence : 1. Il s'agit d'une habituation aux stimuli. 2. Il s'agit d'une différence de protocole. 3. Il s'agit d'une différence entre les espèces.

La question de l'habituation a été étudiée dans deux des études menées chez l'Homme. Dans une première étude dédiée au sujet, les auteurs ont mis en évidence que le taux de décharge pour les stimuli préférés s'amenuisait au fil des répétitions de ces stimuli (Pedreira et al., 2010). L'habituation pourrait donc rendre compte d'une baisse de sélectivité générale au cours de la session ou à travers les sessions mais ne répond cependant pas à la question de sélectivités multiples. Des sélectivités multiples développées au cours de la présentation ont cependant été incidemment observée dans la première étude de la seconde salve, même si elles n'ont pas fait l'objet d'une exploration systématique (Quiroga et al., 2005). En effet, il est rapporté dans cette étude que des neurones ayant une préférence pour un individu, pouvait acquérir au cours de la session une préférence pour un second stimulus. Les auteurs ont interprété cette observation au vu des activités associatives que peuvent présenter les neurones hippocampiques (Wirth et al., 2003). Les neurones auraient alors associé des stimuli n'ayant pas de rapport au vu de l'expérience ultérieure des sujets humains avec ces stimuli mais ayant acquis un lien du fait d'avoir été présenté ensemble au sein d'un set commun pendant une expérience spécifique de tests cognitifs réalisés par l'équipe de chercheurs. Une telle interprétation pourrait expliquer les sélectivités multiples observées dans notre expérience. Du fait de la longue exposition des singes testés aux stimuli, les singes pourraient avoir associé les stimuli uns avec les autres dans leur esprit, ce qui serait reflété par l'activité de leurs neurones hippocampiques. En effet, ces sélectivités multiples représenteraient à la fois une réalité antérieure à la session – ici le fait que trois vues d'un même visage représentent le même individu – et à la fois une réalité liée à la session de test – ici le fait que les stimuli soient toujours présentés ensembles dans un laps de temps réduit correspondant à la session d'enregistrement.

La question de la différence de protocole pourrait également jouer un rôle. En effet, lors des sessions avec des sujets humains, un seul individu était présenté sous différents angles et à l'aide de différentes photos, alors que les autres individus présentés ne l'étaient qu'à travers des

exemplaires de photos uniques. Cette multiplication de stimuli pour un individu pourrait renforcer l'attention vers les stimuli représentant une personne dans différentes situations. Ces stimuli pourraient alors bénéficier d'une sélectivité particulière, qui s'expliquerait par une différence attentionnelle dont un deuxième individu ne pourrait bénéficier car il n'aurait pas été présenté à travers de multiples points de vue.

Enfin, il était demandé aux humains de catégoriser les stimuli perçus en stimuli humains ou non humains. Cette catégorisation pourrait s'être inconsciemment doublée d'une identification sémantique avec accès au nom de ces différents individus. Les capacités de caractérisation propre à l'hippocampe pourraient alors avoir été incidemment spécifiquement renforcées par cette tâche. Dans notre étude, une telle tâche n'était pas requise des singes testés, la perception des stimuli étant entièrement passive.

Toutefois, nos résultats mettent en évidence qu'un code pour les stimuli sociaux est présent depuis le lobe temporal jusqu'à l'hippocampe. La présence ubiquitaire de ce code plaide pour une représentation en réseau dans laquelle l'individu est représenté grâce à l'ensemble des informations et non pas par une petite portion des neurones. Enfin, nos résultats suggèrent aussi, que les correspondances entre identité faciale et vocale, ne se font pas au sein des aires visuelles unisensorielles. Des études visant à préciser cet aspect seront à envisager.

A.4. Conclusion et retombées potentielles dans le domaine de la santé

Le traitement des visages et des voix est perturbé dans un éventail de troubles psychiatriques et neurologiques, comme la prosopagnosie, l'autisme ou la schizophrénie. Dans ce projet de doctorat, nous avons testé si les primates peuvent constituer un modèle valable pour l'étude de la perception des individus. Nous avons mis en évidence que comme les humains, les singes rhésus reconnaissent individuellement leurs congénères familiers. Lorsqu'ils ont été élevés par des humains, cette faculté s'étend à la reconnaissance individuelle des individus humains. Cela démontre d'une part la validité des singes rhésus en tant que modèle pour l'étude des voix et des visages aussi bien au niveau perceptif que conceptuel et d'autre part que le cerveau est capable de s'adapter de façon flexible pour reconnaître les individus d'autres espèces lorsque ceux-ci ont une importance socio-écologique. Par la suite nous avons proposé une approche unique des corrélats neuraux de représentations faciales en utilisant ce modèle animal. Ce projet a mis en lumière la manière dont les connaissances sociales sont représentées de façon de plus en plus abstraite par les neurones de différentes aires cérébrales du lobe temporal. Enfin, l'une de ces aires, l'hippocampe, est une structure qui a évolué à des degrés divers chez différents mammifères pour soutenir la mémoire autobiographique et les représentations spatiales. Nous avons ici caractérisé les différents types de neurones et leur micro-connectivité afin de fournir un cadre commun pour comparer les fonctions de l'hippocampe à travers les espèces et d'étudier comment ces microcircuits coopèrent pour représenter des informations complexes. Ainsi, ce projet sur les

primates jette les bases pour développer de meilleurs modèles des déficiences sociales humaines. Il démontre que la dissection de compétences cognitives complexes, comme la reconnaissance individuelle, peut être réalisée chez cette espèce avec l'avantage essentiel de bénéficier de plusieurs niveaux d'investigation complémentaires.

Annexes

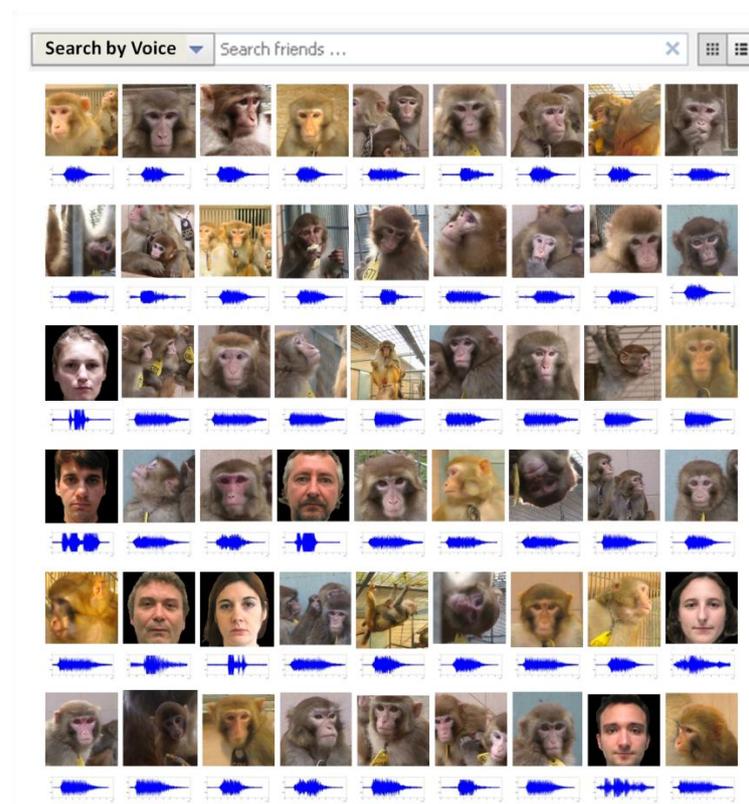


Figure 39: Réseau social imaginaire d'un singe rhésus (*Macaca mulatta*). Les singes rhésus, comme les humains, reconnaissent de façon individuelle leurs congénères. Pour chaque individu, ils associent leur voix (oscillogramme bleu) à leur visage. Lorsqu'ils sont élevés en présence d'humains, leur aptitude à la reconnaissance individuelle s'étend aux humains. La reconnaissance fine est donc une faculté partagée par un éventail de primates et pourrait servir de fondement au maintien d'un réseau social sophistiqué. Photographies réalisées à la Station de Primatologie du CNRS, Rousset sur Arc, France.

Annexe A. Paradigmes des « visages composites » et du « renversement des visages »

Deux paradigmes expérimentaux ont historiquement mis en évidence la spécificité du traitement perceptif des visages par rapport aux objets: le paradigme des *visages composites* et celui du *renversement des visages*.

Dans le cadre du paradigme des visages composites, la partie haute d'un visage connu est alignée sur la partie basse d'un second visage connu mais différent. Il devient alors extrêmement difficile d'identifier les individus auxquels appartiennent ces deux parties de visages. Par contre lorsque la partie haute est légèrement décalée de la partie basse, alors l'identification des deux portions de visage est plus simple (**Figure 40**). Cet effet, dénommé « effet du visage composite », montre que l'alignement des deux parties de visages fait émerger une nouvelle configuration qui perturbe la reconnaissance des différents détails du visage composite (Young et al., 1987). Cette expérience met en évidence que le traitement d'un visage est *holistique* : les détails sont traités en interdépendance à l'intérieur du cadre du visage. Ils ne sont pas perçus, analysés ou mémorisés de façon indépendante mais sont tributaires de la configuration du visage auquel ils appartiennent.



Figure 40 : Paradigme des visages composites. Lorsqu'ils sont alignés les visages de Bill Clinton et Jesse Jackson sont difficilement identifiables séparément (gauche). Illusion de Thatcher (gauche): Lorsque les visages sont présentés à l'envers l'inversion des détails n'est pas perceptible. Elle sauté aux yeux dans un visage à l'endroit. D'après Thompson, 1980 et Tsao and Livingstone, 2008.

De façon similaire le renversement des visages altère le traitement des informations configurales et componentielles. Si des visages et des objets sont présentés deux fois à des sujets témoins, lors de la seconde présentation les visages sont mieux reconnus que les objets. Ceci reproduit la rapidité de traitement des informations contenues dans un visage, mentionnée précédemment (§ A.3). Par contre si la seconde fois les stimuli sont présentés à l'envers, les visages sont moins bien reconnus que les objets également présentés à l'envers (Yin, 1969). Ainsi il semblerait que lorsque les visages sont présentés à l'envers, ils redeviennent des stimuli complexes et très homogènes dont la perception est difficile. Cette expérience démontre que le traitement rapide et fiable de l'identité d'un visage est réalisé uniquement si celui-ci présente une *configuration de premier ordre* spécifique d'un visage de base (représentée par les deux yeux en haut et la bouche en bas, les traits étant plus foncés que le contour).

Par la suite et pour aller plus loin, des études ont manipulé soit les distances entre les traits soit le détail des traits sur des visages présentés à l'envers ou à l'endroit. Les deux opérations de manipulation sont difficilement perceptibles sur un visage à l'envers alors qu'elles sautent aux yeux sur un visage à l'endroit (**Figure 40**) (Thompson, 1980; Kemp et al., 1990; Goffaux and Rossion, 2007). Ceci montre d'une part que, comme dans le paradigme des « visages composites », les composants du visage sont traités en interdépendance à l'intérieur du cadre du visage. D'autre part ceci montre que, non seulement les composants du visage, mais également leur configuration (également appelée configuration de second ordre) sont traités en interdépendance dans le cadre du visage (configuration de premier ordre). Ainsi le traitement rapide et holistique des informations contenues dans un visage est spécifique à la configuration de premier ordre du visage à l'endroit.

Annexe B. Une propension innée à détecter les visages ?

La vision du nouveau-né est extrêmement limitée car ses yeux ne peuvent pas encore accommoder et que seules les fréquences spatiales basses et les contrastes élevés sont traités par son système visuel en développement. Ainsi la perception en général, et celle des visages en particulier, est nettement plus pauvre chez le nouveau-né que chez l'adulte (**Figure 41**).

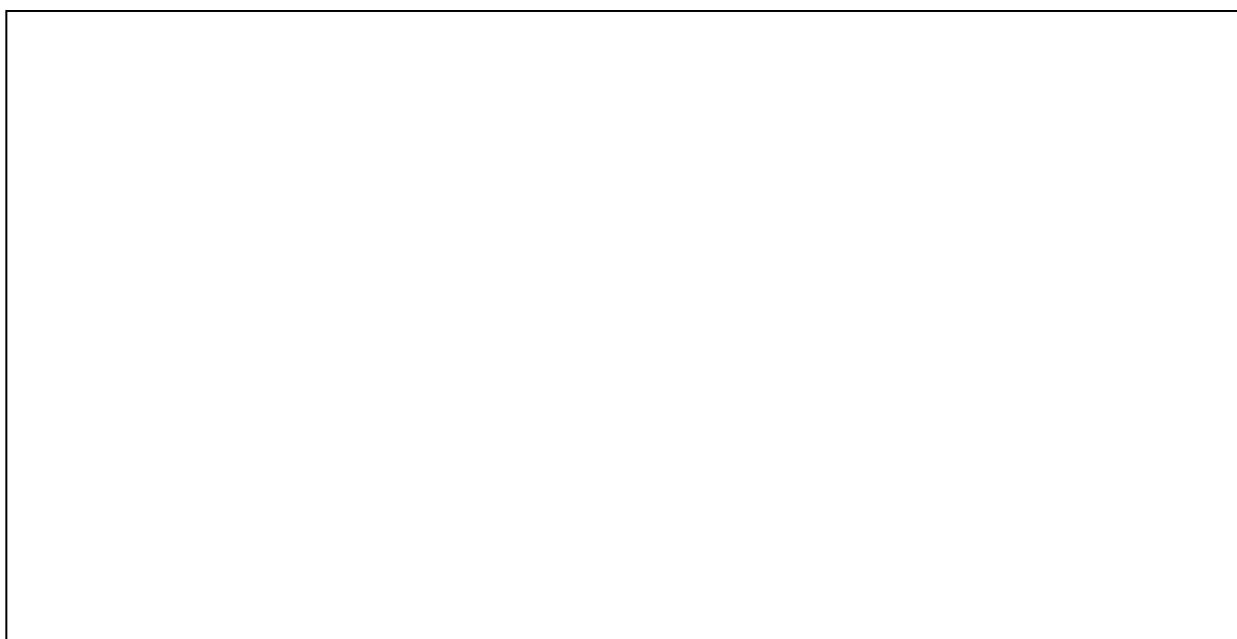


Figure 41 : Visages tels qu'ils seraient perçus par un nouveau-né à une distance de 2m (gauche) : visage neutre, orientant le regard vers la gauche et les yeux fermés. La partie haute du visage au niveau des yeux est saillante par rapport au contour, dans les trois conditions. Préférences visuelles des nouveau-nés, testées par la durée de regard (droite haut) : 1>2, 3>4, 5>6, 7>8, 9>10 et (droite bas) : 1>2, 3>4, 5>6. D'après Viola Macchi et al., 2004; Johnson, 2005 et De Schonen, 2009.

La *configuration de premier ordre* spécifique d'un visage de face est suffisante pour attirer préférentiellement l'attention du nouveau-né. Elle présente les deux yeux en haut et la bouche en bas d'un ovale, les traits étant plus foncés que le contour et ayant ou n'ayant pas de détails apparents. Cependant ce pattern n'est pas nécessaire ; des propriétés plus simples attirent également le regard du nouveau-né, comme l'horizontalité des contrastes plutôt que leur verticalité et une asymétrie de répartition des contrastes favorisant le haut de la figure plutôt que le bas, avec toujours toutefois des détails plus sombres que leur contour (**Figure 41**) (Viola Macchi et al., 2004; Farroni et al., 2005). Or la structure des visages de face à l'endroit présente

précisément ces deux propriétés d'horizontalité et d'asymétrie des contrastes, avec notamment des yeux très saillants (Haith et al., 1977; Farroni et al., 2002; Farroni et al., 2006; Gava et al., 2008). Au contraire les visages à l'envers ou encore désaxés (comme dans les visages composites décalés) qui n'ont pas ces caractéristiques, n'attirent pas l'attention des nouveau-nés.

Ainsi la configuration de premier ordre utilisée à l'âge adulte pour analyser finement l'identité, est celle qui dès la naissance attire l'attention des enfants vers les visages. Au cours du développement, le traitement des différentes informations contenues dans un visage (identité, émotions, etc) se construit donc sur la base de ce pattern très simple. Ainsi après l'avoir renversé, diverses études ont déformé le visage de diverses façons, soit en préservant le pattern de base (par exemple par floutage ou allongement), soit en le perturbant (négatif, inversion) (Harmon and Julesz, 1973; Sinha, 2002; Gilad et al., 2009). Elles ont montré que les visages sont traités de façon holistique tant que le pattern de base est préservé (**Figure 42**). Cependant dès que le pattern de base est perturbé le processus de traitement holistique n'est plus réalisé, indiquant le lien nécessaire entre détection et analyse des visages. Dans ces cas l'analyse devient plus difficile et suivrait l'analyse dédiée aux autres stimuli visuels, comme les objets.



Figure 42 : Propriétés de la perception et de la détection des visages. Contrairement aux systèmes automatisés de reconnaissance des visages, les observateurs humains peuvent supporter des dégradations par floutage (haut gauche), par compression (haut milieu) mais non qui perturbent le pattern de base (haut droite). La propension naturelle à détecter des visages donne lieu à des fausses détections qui peuvent prendre un caractère mystique (bas). D'après Sinha, 2002.

Notons par ailleurs que la propension naturelle et innée à détecter ce pattern spécifique mais néanmoins extrêmement simple pourrait expliquer la tendance encore présente chez l'adulte à percevoir des visages même là où ils sont absents, par exemple dans les formes créées par les nuages (**Figure 42**). L'hypothèse qu'il y aurait certainement un fort avantage adaptatif à détecter des visages trop souvent plutôt que pas assez, peut être émise sur la base du coût que représenterait une fausse détection par rapport à la non-détection d'un individu (Sinha, 2002).

Annexe C. Différents modèles de représentation neurale du concept

Plusieurs possibilités ont été proposées pour expliquer comment les populations neuronales distantes codent une représentation conceptuelle unique.

A.1.1. L'*unité gnostique* comme représentation du concept, dans le cadre d'un codage hiérarchique

La première proposition postule un modèle hiérarchique ascendant (*feedforward* ou *bottom-up*). A chaque niveau les neurones réalisent un traitement plus complexe qu'au niveau précédent. L'extrême pointe de la hiérarchie serait occupée par des *unités gnostiques*, c'est-à-dire des neurones à l'activité spécifique et invariante à tous les attributs d'un concept donné et qui ne seraient pas actifs pour d'autres concepts (Konorski, 1967).



Figure 43: Diagramme du modèle de traitement visuel (gauche) et auditif (droite) sériel ascendant dans une architecture neuronale hiérarchiquement organisée. Le contenu des représentations porté par les neurones individuels croît en complexité lorsque l'on progresse dans la hiérarchie. Les champs récepteurs s'élargissent, les représentations combinent les attributs codés par les neurones sous-jacents, l'invariance augmente. D'après Rousselet et al., 2004; Rolls and Stringer, 2006 et Rauschecker and Scott, 2009.

Des modèles hiérarchiques ont été proposés pour l'analyse des visages (Wallis and Rolls, 1997; Riesenhuber et al., 2004) et des objets (Riesenhuber and Poggio, ; Felleman and Van Essen, 1991; VanRullen, 2003; Rousselet et al., 2004; Rolls and Stringer, 2006). Ils comprennent une succession de couches cellulaires aux champs récepteurs de taille croissante, mimant les caractéristiques des cellules de V1 décrites par HUBEL et WIESEL (Hubel and Wiesel, 1963a, b). Dans ces modèles, les afférences des cellules simples vers des cellules plus complexes aboutissent

à une augmentation de la complexité des attributs traités par chaque cellule puis à une invariance à la position dans le champ visuel, à la taille sur la rétine, à la rotation et à un certain degré de changement, assurant la constance du visage ou de l'objet dans le temps et l'espace (**Figure 43 gauche**).

Plusieurs tentatives ont été émises pour adapter ce schéma à la perception auditive (**Figure 43 droite**, Husain et al., 2004; Rauschecker and Scott, 2009). La difficulté à démontrer un codage hiérarchique, surtout dans les premières aires corticales où l'analyse est précoce, provient de la difficulté à définir un objet sonore et ses contours par rapport à l'environnement auditif (Näätänen et al., 2001; Griffiths and Warren, 2004). Cependant des neurones simples du cortex auditif primaire décomposant les scènes auditives en divers composants acoustiques ont été observés (deCharms et al., 1998). De plus ils se projettent sur des neurones plus complexes (Rauschecker et al., 1995; Tian and Rauschecker, 2004). Enfin des aires sélectives à l'identité des objets sonores ont été observées dans le lobe temporal (Zatorre et al., 2004) et frontal (Adams and Janata, 2002; Binder et al., 2004; Russ et al., 2007) montrant une analyse croissante de la complexité le long de la voie ventrale.

Le long du lobe temporal, les représentations des visages (Freiwald and Tsao, 2010) et des voix (Kikuchi et al., 2010) deviennent progressivement invariantes. Ces mesures électrophysiologiques plaident pour un type de codage hiérarchique ascendant. Dans cette perspective la complexité du traitement neuronal dépasserait le cadre unisensoriel et serait finalement atteinte par les unités gnostiques, aussi surnommées *neurones grand-mères*. Dans le cas d'un individu ce neurone représenterait son visage, sa voix, son nom, son odeur, sa personnalité, etc. Il stockerait donc l'ensemble du concept et chaque concept serait stocké par ce type de neurone extrêmement invariant et abstrait. Par leur description ces neurones ressemblent à première vue à ceux observés par QUIROGA et coll. (2005, 2009, **Figure 21**) mais nous verrons par la suite qu'ils ne peuvent pas être qualifiés de « neurones grand-mères » (Quiroga et al., 2005; Quiroga et al., 2009). Une objection majeure peut-être apportée à l'existence d'unités gnostiques situées à l'extrême pointe d'une hiérarchie ascendante. En effet ce modèle est limité par *l'explosion combinatoire* qu'il générerait, c'est-à-dire à l'impossibilité de représenter l'ensemble des concepts connus par le cerveau (Eichenbaum, 1993; von der Malsburg, 1999), par ailleurs il limite la possibilité de regrouper les différents concepts en catégories ou encore de les comparer entre eux.

A.1.2. L'assemblée neuronale comme représentation du concept, dans les cadres de codage connexionniste ou temporel par synchronisation

La seconde proposition postule que chaque attribut serait codé par des neurones distincts et que l'ensemble du concept serait représenté par une *assemblée neuronale*, ou *cognit*, c'est-à-dire un groupe de neurones lointains actifs simultanément (Eichenbaum, 1993; Fuster, 2006).

Le modèle de l'assemblée neuronale se base sur les principes énoncés par HEBB (1949), selon lesquels les groupes de neurones interconnectés et qui déchargent simultanément à la perception

d'un stimulus, deviennent plus fortement associés qu'avant (Hebb, 1949). Les phénomènes de *potentialisation à long terme*, renforçant les synapses entre deux neurones, s'observent effectivement dans l'ensemble du cortex et dans l'hippocampe pendant l'apprentissage. Ainsi, un seul attribut du stimulus devrait pouvoir à lui seul éliciter l'activation de l'ensemble du réseau et par là permettre le rappel du concept complet. De plus, chaque neurone peut faire partie de plusieurs assemblées neuronales. Ainsi, le nombre d'assemblées neuronales potentielles équivaut à l'ensemble des combinaisons possibles entre tous les neurones ; il dépasse donc largement le nombre d'unités gnostiques potentielles.

Dans ce modèle, les neurones d'une même assemblée sont interconnectés anatomiquement. La *synchronisation temporelle* des activités neuronales a été proposée comme alternative à la connectivité anatomique (von der Malsburg, 1999; Varela et al., 2001). Dans cette perspective, pour pouvoir être actifs simultanément, les neurones d'une assemblée se calent sur l'activité oscillatoire des potentiels de champs du cerveau, qui peut être synchrone (Gray et al., 1989; Gray and Singer, 1989). Ainsi, la synchronisation des neurones représentant les différents attributs de la personne donnerait lieu à une image unique du concept dans son ensemble (Hoffman and McNaughton, 2002; Sehatpour et al., 2008; Canolty et al., 2010).

Le *liage* par synchronie caractérise la façon dont sont codées les représentations par des groupes de neurones distants, cependant lors de la reconnaissance d'un individu tous les attributs de l'individu ne sont pas présents en même temps et sa reconnaissance dépend de la complétion d'un pattern incomplet avec les connaissances antérieures. La réactivation de ces connaissances dépend donc de la réactivation de l'assemblée de neurones constituant ce concept.

Annexe D. Cartes cognitives et modèles top-down de reconnaissance

Différentes manières, non mutuellement exclusives, peuvent être envisagées pour réactiver l'ensemble de l'assemblée neuronale. D'une part, la perception de l'un des attributs peut permettre à elle seule de réactiver le reste de l'assemblée neuronale. D'autre part, la réactivation globale peut-être réalisée par un *concentrateur* (ou *hub*) supplémentaire du réseau (**Figure 44 droite**). Ainsi, la perception de l'attribut activerait le hub, qui à son tour permettra la réactivation de l'ensemble du réseau. La seconde théorie repose sur l'existence de traitements descendants (*feedback* ou *top-down*), mis en évidence par les études de connectivités et des connexions anatomiques (par exemple : Moeller et al., 2008; Nestor et al., 2011).

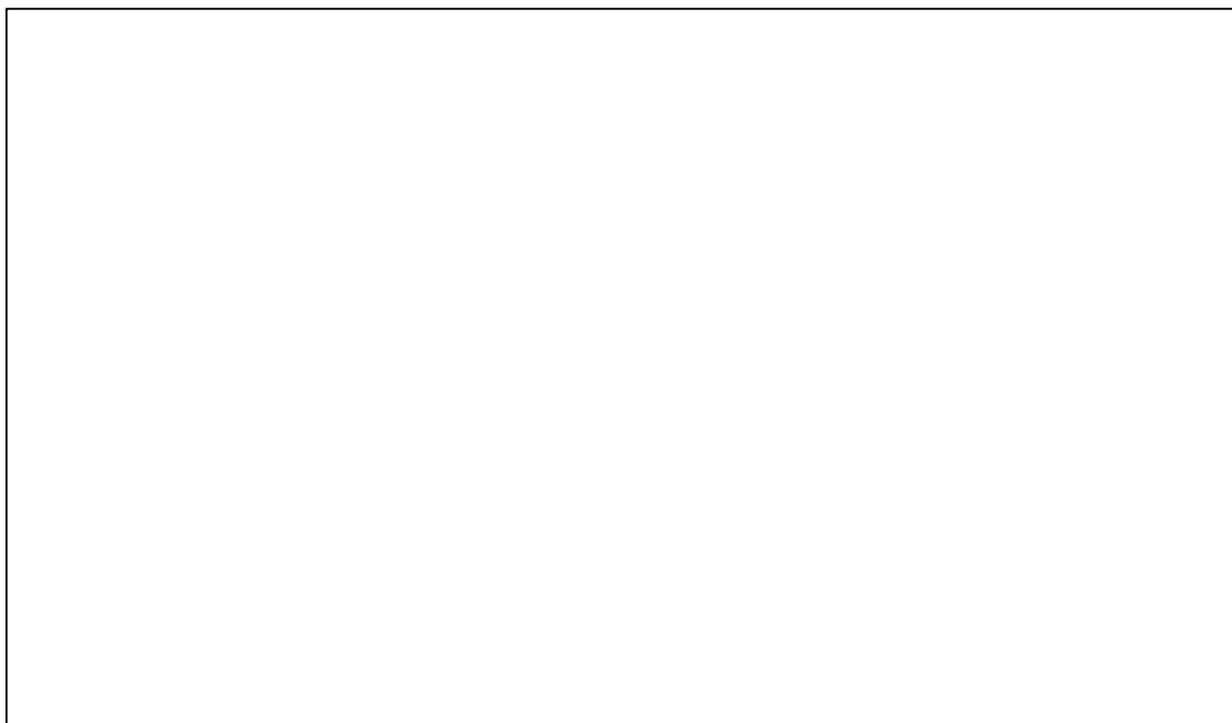


Figure 44 : Schématisation de la théorie de hiérarchie inverse pour l'analyse visuelle (gauche haut) et auditive (gauche bas). Dans ces modèles le traitement antérograde (feedforward) est essentiellement implicite et les informations de « bas-niveau » ne deviennent accessibles que tardivement à l'observateur grâce aux connexions rétrogrades (feedback). Schématisation de la réactivation d'une assemblée neuronale sans (droite haut) et avec (droite bas) l'aide d'un hub, ici le cortex temporal antérieur. D'après Hochstein and Ahissar, 2002; Patterson et al., 2007; Shamma, 2008 et Ahissar et al., 2009.

Dans les voies sensorielles, une théorie dite de *hiérarchie inverse* a été proposée supposant une complémentarité entre les voies ascendantes et descendantes (Hochstein and Ahissar, 2002; Bar

et al., 2006; Rogers and Patterson, 2007; Shamma, 2008; Ahissar et al., 2009). Les voies ascendantes réaliseraient un filtrage adaptatif des signaux d'entrée ; les voies descendantes porteraient des signaux prédictifs devant être comparés avec l'entrée (**Figure 44 gauche**). Cette théorie propose que de nombreux aspects de la cognition et du comportement ne soient pas dirigés par les stimuli de façon réflexe, mais soient des anticipations issues d'expériences antérieures. Cette théorie se limite cependant aux processus unisensoriels.

D'autres modèles top-down postulent l'existence d'une communication entre divers systèmes sensoriels. Dans ce cas, les influences descendantes proviennent de régions cérébrales se comportant comme des *concentrateurs*. Par exemple, dans le cas du concept d'individu, les neurones du concentrateur constitueraient des cartes cognitives pour chaque individu par convergence d'information lors de l'encodage. L'activation de ces cartes modulerait ensuite les aires sous-jacentes, où les différents attributs de la personne (voix, visage, etc.) seraient représentés lors du rappel (**Figure 44 droite**). Ces signaux top-down lieraient et réactiveraient les contenus de bas niveau. DAMASIO a proposé que le cortex antérieur chez l'humain joue ce rôle pour la réactivation de concepts (Damasio, 1989 ; Patterson et al., 2007; Meyer and Damasio, 2009). Il a appelé ce processus : *convergence-divergence*. De façon similaire, l'hippocampe encode et rappelle les souvenirs épisodiques (Battaglia et al., 2011). D'ailleurs le terme de carte cognitive a d'abord été utilisé pour décrire son fonctionnement. Enfin les régions préfrontales et pariétales, représentant les buts de l'action, modulent les circuits sensorimoteurs (Desimone and Duncan, 1995 ; Miller and Cohen, 2001; Buschman and Miller, 2007).

Au niveau neuronal, les régions de haut-niveau pourraient influencer les systèmes sensoriels de plus bas niveau en modulant le taux de décharges des neurones (Desimone and Duncan, 1995 ; Miller and Cohen, 2001) ou leur codage temporel (Fries et al., 2001; Bichot et al., 2005). Par exemple les synchronies oscillatoires locales pourraient refléter l'influence des signaux top-down naissant des régions associatives de haut niveau (Damasio, 1989).

Annexe E. Bases neuronales des associations voix-visage amodales

A.1.3. Zones d'intégration pour les associations amodales

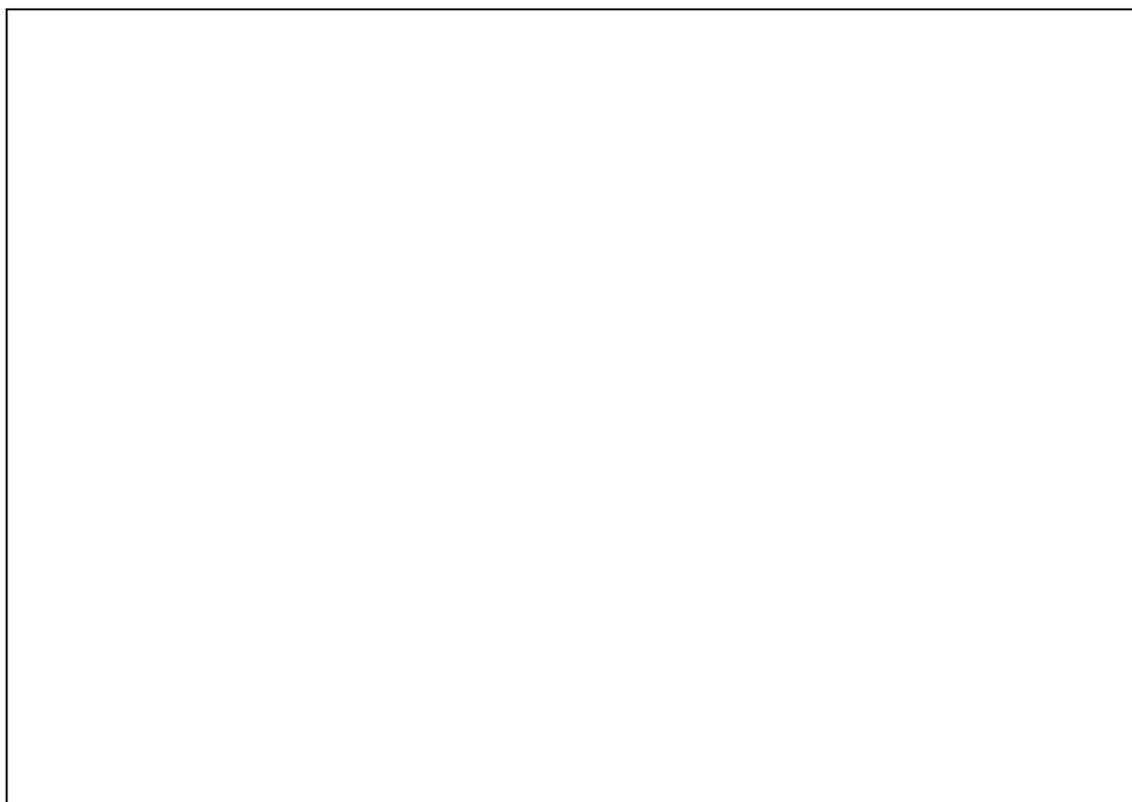


Figure 45 : Différentes associations entre les voix et les visages et leurs bases cérébrales, schématisées à partir du modèle de Bruce et Young augmenté. D'après Campanella and Belin, 2007.

Plusieurs régions du cerveau sont considérées comme multisensorielles, sur la base de leurs connexions avec plusieurs aires unimodales (Falchier et al., 2002), de l'activité de leurs neurones pour les stimuli multisensoriels (Duhamel et al., 1998; Avillac et al., 2005) ou des déficits causés par leur lésion. Au niveau du cortex, ces aires incluent principalement : le sillon temporal supérieur, le sillon intrapariétal et des régions du cortex frontal et préfrontal (Ghazanfar and Schroeder, 2006). A ces aires peuvent s'ajouter plusieurs zones sous-corticales, notamment l'amygdale (Calvert, 2001). Le sillon temporal supérieur semble particulièrement intéressant pour l'intégration dynamique des voix et des visages pour la communication. En effet cette zone est activée par les visages dynamiques (§ C.2.2.) (Ghazanfar et al., 2010; Pitcher et al., 2011) ou ayant

un mouvement implicite (Perrett et al., 1982; Perrett et al., 1984, 1985; Hasselmo et al., 1989; Perrett et al., 1991; Hietanen et al., 1992; Perrett et al., 1992; Hoffman and Haxby, 2000).

Ainsi plusieurs études ont mis en évidence l'implication du sillon temporal supérieur dans l'intégration audiovisuelle du langage (**Figure 45 et 46**, Calvert, 2001; Amedi et al., 2005). Par exemple CALVERT et coll. (2000) ont spécifiquement recherché les aires répondant à trois critères : 1) qu'elles soient actives par chacune des modalités isolément, 2) qu'elles présentent une interaction positive lors d'une présentation congruente audiovisuelle [AV>A+V] et 3) une interaction négative lors d'une présentation audiovisuelle incongruente [AV<A+V] (Calvert et al., 2000). Ces critères sont analogues à ceux utilisés en électrophysiologie pour dévoiler l'intégration à l'échelle du neurone, à l'exception du dernier. Cette étude mit en évidence que seul le STS postérieur gauche répond à ces trois critères. Chez le singe, plusieurs études ont mis en évidence que cette région est constituée de neurones de plusieurs types présentant des réponses visuelles, auditives, multimodales super-additives et multimodales sub-additives pour des voix et visage congruents (Chandrasekaran and Ghazanfar, 2009; Dahl et al., 2009a, 2010a). Ces neurones seraient organisés en clusters sensoriels entremêlés (Beauchamp et al., 2004).

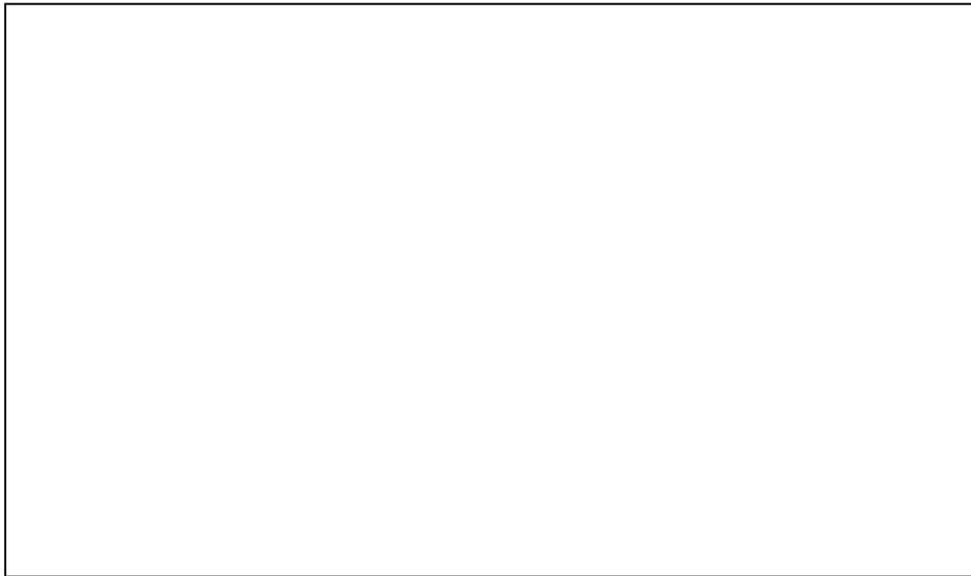


Figure 46 : Zones d'intégrations de la parole audio-visuelle mises en évidence par plusieurs études. Les aires principalement impliquées sont STS, STG et MTG. Les étoiles représentent également les aires impliquées dans l'intégration des lettres et des sons. D'après Amedi et al., 2005.

L'amygdale semble également intéressante car elle participe à la reconnaissance des émotions faciales chez l'humain (Breiter et al., 1996; Morris et al., 1996; Phillips et al., 1997; Ishai et al., 2004) et le singe (Kuraoka and Nakamura, 2006; Gothard et al., 2007) et s'active pour les vocalises émotionnelles non-linguistiques (Sander and Scheich, 2001; Fecteau et al., 2007). En effet, il a été mis en évidence qu'elle présente une activation renforcée lorsque l'association voix-visage est congruente pour une émotion de peur (Dolan et al., 2001). Cette intégration a également été mise en évidence au niveau neuronal (**Figure 47 haut**, Kuraoka and Nakamura,

2007). Notons que la congruence pour d'autres émotions, comme la joie et la colère, entraîne une activation super-additive dans d'autres régions, comme le gyrus temporal moyen, l'hippocampe ou le gyrus frontal inférieur (Johnstone et al., 2006; Campanella and Belin, 2007; Park et al., 2010).

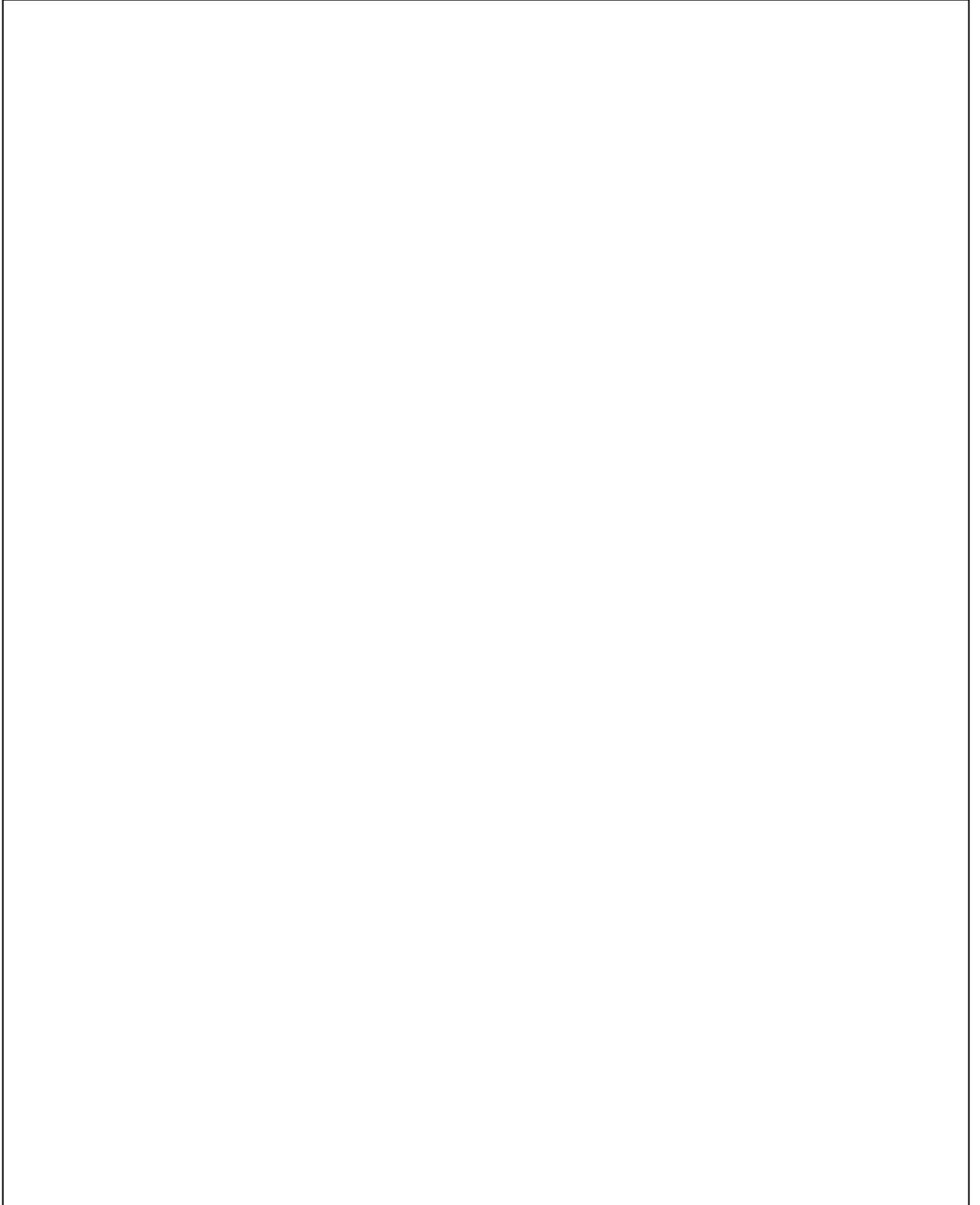


Figure 47 : Exemples d'intégrations multisensorielles Voix+Visage au niveau neuronal dans l'amygdale (haut), le cortex préfrontal ventro-médian (milieu) et le middle latéral belt (ML) (bas). Au niveau de l'amygdale l'intégration concerne les cris de peur. Au niveau du cortex préfrontal deux exemples d'intégration sont présentés : par augmentation et par suppression d'activité. Au niveau du lateral belt, l'intégration Voix+Visage est comparée à Voix+Disk. Une intégration par augmentation et par suppression d'activité sont présentées. D'après Kuraoka and Nakamura, 2006; Sugihara et al., 2006 et Ghazanfar et al., 2008.

Enfin, des neurones intégrant les informations faciales et vocales ont également été mis en évidence dans une autre région multisensorielle du cortex chez le singe, le cortex préfrontal ventrolatéral (**Figure 47 milieu**, Sugihara et al., 2006). Ceci indique que d'autres régions que le STS pourraient être engagées dans l'intégration des associations amodales voix-visages.

En conclusion, de multiples zones polymodales intègrent les voix et les visages dynamiques. Ces différentes aires sont impliquées dans des tâches cognitives différentes : la perception de la peur, de la joie ou d'autres émotions et enfin de la parole. Modulent-elles en retour les aires unisensorielles impliquées dans la perception de ces stimuli à travers des canaux sensoriels spécialisés ?

A.1.4. Activité des zones unimodales en réponse à des stimulations cross-modales ou multimodales

Très précocement dans la hiérarchie du réseau cortical dédié au traitement des informations auditives, une modulation des activités aux sons par la vision est observée (Ghazanfar, 2009; Kayser et al., 2009). Par exemple, le cortex auditif est activé ou modulé par des visages dynamiques. Ainsi CALVERT et coll. (2000) ont mis en évidence que la lecture sur les lèvres, même sans voix audible, active le cortex auditif (Calvert et al., 1997; Calvert and Campbell, 2003). Au niveau neuronal également, des réponses visuelles à des visages et des objets présentés sans sons ont été enregistrées dans le cortex auditif (Hoffman et al., 2007a).

De plus, lorsque le locuteur est vu en plus d'être entendu, les activations dans les cortex unisensoriels visuels et auditifs sont augmentées par rapport à leur activation pour une modalité seule (Calvert et al., 1999). Ceci est en accord avec l'effet comportemental connu sous le nom de *Cocktail party*, qui montre que l'adjonction de la vue à l'ouïe augmente la perception de la voix d'une intensité équivalente à 10-20 dB (Ross et al., 2007). Au niveau neuronal, les enregistrements de potentiels de champs locaux (LFP) dans le noyau et le belt du cortex auditif ont révélé que l'activité pour les voix est modulée par la présence d'un visage dynamique (Ghazanfar et al., 2005; Ghazanfar et al., 2008). Par ailleurs, la spécificité de l'intégration voix-visage a été testée en remplaçant l'image de visage par une image de rond ayant la circonférence de l'ouverture de la bouche du singe et bougeant de façon dynamique et congruente avec le son. Les réponses intégratrices au niveau du latéral belt et du cortex primaire étaient alors moins importantes que celles observées en présentant la voix associée à un vrai visage dynamique

(**Figure 47 bas**, Ghazanfar et al., 2005; Ghazanfar et al., 2008). Ceci démontre la spécificité de l'intégration voix-visage par rapport à d'autres intégrations audiovisuelles.

Il serait intéressant de savoir si, dans les cortex visuels primaires, une telle modulation de l'activité neuronale par des stimulations provenant d'autres modalités est également observée. En effet, ceci prouverait ou réfuterait le strict parallélisme entre les traitements visuels et auditifs.

Les activités intégratrices observées dans le cortex auditif sont-elles des rétroprojections de l'activité intégratrice de STS ? La cohérence entre les activités du cortex auditif et celles de l'aire multisensorielle STS a été étudiée chez le singe. L'interaction entre le STS et le cortex auditif s'est révélée augmentée dans la bande de fréquence gamma, pendant la présentation simultanée de voix et de visages, par rapport à une présentation de l'un des stimuli seulement (Ghazanfar et al., 2008). Ainsi, les activités visuelles et intégratrices du cortex auditifs pourraient provenir de feedbacks de régions multisensorielles, telles STS. Cependant une autre possibilité, non exclusive et en faveur de laquelle de plus en plus d'indices expérimentaux s'accumulent, serait que les cortex unisensoriels soient connectés directement entre eux, de façon plus intense que supposé précédemment (Cappe and Barone, 2005; Ghazanfar and Schroeder, 2006; Smiley and Falchier, 2009; Falchier et al., 2010). Les connexions réciproques précoces pourraient avoir comme corrélat fonctionnel la synchronisation des cortex unimodaux pour l'analyse des objets multisensoriels, i.e. la plupart des objets de notre environnement (Ghazanfar and Chandrasekaran, 2007; Lakatos et al., 2007).

Bibliographie

- Adachi I, Hampton RR (2011) Rhesus monkeys see who they hear: spontaneous cross-modal memory for familiar conspecifics. *PLoS ONE* 6:e23345.
- Adachi I, Chou DP, Hampton RR (2009a) Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Current Biology* 19:1270-1273.
- Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates* 47:350-354.
- Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2009b) Plasticity of ability to form cross-modal representations in infant Japanese macaques. *Dev Sci* 12:446-452.
- Adams RB, Janata P (2002) A comparison of neural circuits underlying auditory and visual object categorization. *NeuroImage* 16:361-377.
- Adolphs R (2003) Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci* 4:165-178.
- Adolphs R (2009) The social brain: neural basis of social knowledge. *Annu Rev Psychol* 60:693-716.
- Afraz S-R, Cavanagh P (2008) Retinotopy of the face aftereffect. *Vision Research* 48:42-54.
- Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC (2001) Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32:537-551.
- Ahissar M, Nahum M, Nelken I, Hochstein S (2009) Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:285-299.
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences* 98:12301-12306.
- Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences* 4:267-278.
- Amaral DG, Insausti R, Cowan WM (1983) Evidence for a direct projection from the superior temporal gyrus to the entorhinal cortex in the monkey. *Brain Research* 275:263-277.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559-571.
- Anderson JR (1998) Social stimuli and social rewards in primate learning and cognition. *Behavioural Processes* 42:159-175.
- Andics A, McQueen JM, Petersson KM, Gál V, Rudas G, Vidnyánszky Z (2010) Neural mechanisms for voice recognition. *NeuroImage* 52:1528-1540.
- Andrews TJ, Ewbank MP (2004) Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage* 23:905-913.
- Avillac M, Deneve S, Olivier E, Pouget A, Duhamel JR (2005) Reference frames for representing visual and tactile locations in parietal cortex. *Nat Neurosci* 8:941-949.
- Bachorowski JA, Owren MJ (1999) Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J Acoust Soc Am* 106:1054-1063.
- Bahrnick LE, Lickliter R (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 36:190-201.
- Bahrnick LE, Lickliter R (2002) Intersensory redundancy guides early perceptual and cognitive development. *Adv Child Dev Behav* 30:153-187.
- Bahrnick LE, Netto D, Hernandez-Reif M (1998) Intermodal perception of adult and child faces and voices by infants. *Child Dev* 69:1263-1275.
- Bahrnick LE, Hernandez-Reif M, Flom R (2005) The development of infant learning about specific face-voice relations. *Dev Psychol* 41:541-552.

- Bannerman DM, Rawlins JNP, McHugh SB, Deacon RMJ, Yee BK, Bast T, Zhang WN, Pothuizen HHJ, Feldon J (2004) Regional dissociations within the hippocampus - memory and anxiety. *Neuroscience & Biobehavioral Reviews* 28:273-283.
- Bar M, Neta M, Linz H (2006) Very first impressions. *Emotion* 6:269-278.
- Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liegeois-Chauvel C (2008) Spatio temporal dynamics of face recognition. *Cereb Cortex* 18:997-1009.
- Bargh JA, Chen M, Burrows L (1996) Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71:230-244.
- Barrett L, Henzi P, Rendall D (2007) Social brains, simple minds: does social complexity really require cognitive complexity? *Philosophical Transactions of the Royal Society B: Biological Sciences* 362:561-575.
- Bartels A, Zeki S (2000) The neural basis of romantic love. *Neuroreport* 11:3829-3834.
- Bartels A, Zeki S (2004) The neural correlates of maternal and romantic love. *Neuroimage* 21:1155-1166.
- Barton JJS, Press DZ, Keenan JP, O'Connell M (2002) Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology* 58:71-78.
- Barton RA (2006) Animal communication: do dolphins have names? *Curr Biol* 16:R598-599.
- Basile BM, Hampton RR (2011) Monkeys recall and reproduce simple shapes from memory. *Curr Biol* 21:774-778.
- Bast T, Wilson IA, Witter MP, Morris RGM (2009) From rapid place learning to behavioral performance: a key role for the intermediate hippocampus. *PLoS Biol* 7:e1000089.
- Battaglia FP, Benchenane K, Sirota A, Pennartz CMA, Wiener SI (2011) The hippocampus: hub of brain network communication for memory. *Trends in Cognitive Sciences* 15:310-318.
- Bauer HR, Philip MM (1983) Facial and vocal individual recognition in the common chimpanzee. *Psychological Record* 33:161-170.
- Bauer RM (1984) Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia* 22:457-469.
- Baumann O, Belin P (2010) Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research* 74:110-120.
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190-1192.
- Bekoff M, Allen C (1997) Cognitive ethology: slayers, skeptics, and proponents. In: *Anthropomorphism, Anecdotes, and Animals: The Emperor's New Clothes?* (Mitchell RW, Thompson N, Miles L, eds), pp 313-334. New York: State University of New York Press.
- Belin P (2006) Voice processing in human and non-human primates. *Philos Trans R Soc Lond B Biol Sci* 361:2091-2107.
- Belin P, Zatorre RJ (2000) 'What', 'where' and 'how' in auditory cortex. *Nat Neurosci* 3:965-966.
- Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14:2105-2109.
- Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res* 13:17-26.
- Belin P, Fecteau S, Bedard C (2004) Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 8:129-135.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309-312.
- Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RBH, Ungerleider LG (2009) Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *Journal of Neurophysiology* 101:688-700.
- Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RBH, Ungerleider LG (2011) Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *The Journal of Neuroscience* 31:12229-12240.
- Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, Battaglia FP, Wiener SI (2010) Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. *Neuron* 66:921-936.

- Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161-1165.
- 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.
- Bernard FA, Bullmore ET, Graham KS, Thompson SA, Hodges JR, Fletcher PC (2004) The hippocampal region is involved in successful recognition of both remote and recent famous faces. *NeuroImage* 22:1704-1714.
- Best CT, McRoberts GW, LaFleur R, Silver-Isenstadt J (1995) Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development* 18:339-350.
- Bichot NP, Rossi AF, Desimone R (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529-534.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295-301.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 10:512-528.
- Bird CM, Burgess N (2008) The hippocampus and memory: insights from spatial processing. *Nat Rev Neurosci* 9:182-194.
- Blank H, Anwender A, von Kriegstein K (2011) Direct structural connections between voice- and face-recognition areas. *The Journal of Neuroscience* 31:12906-12915.
- Bonte M, Valente G, Formisano E (2009) Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *The Journal of Neuroscience* 29:1699-1706.
- Bovet D, Vauclair J (2001) Judgment of conceptual identity in monkeys. *Psychon Bull Rev* 8:470-475.
- Braeutigam S, Bailey AJ, Swithenby SJ (2001) Task-dependent early latency (30-60 ms) visual processing of human faces and other objects. *Neuroreport* 12:1531-1536.
- Brannon EM, Terrace HS (1998) Ordering of the numerosities 1 to 9 by monkeys. *Science* 282:746-749.
- Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Strauss MM, Hyman SE, Rosen BR (1996) Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17:875-887.
- Brown MW, Aggleton JP (2001) Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neurosci* 2:51-61.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369-384.
- Bruce V, Young A (1986) Understanding face recognition. *Br J Psychol* 77 (Pt 3):305-327.
- Bruckert L, LiÅ©nard J-S, Lacroix A, Kreutzer M, Leboucher Gr (2006) Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences* 273:83-89.
- Bruyer R, Laterre C, Seron X, Feyereisen P, Strypstein E, Pierrard E, Rectem D (1983) A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain Cogn* 2:257-284.
- Bukach CM, Bub DN, Gauthier I, Tarr MJ (2006) Perceptual expertise effects are not all or none: spatially limited perceptual expertise for faces in a case of prosopagnosia. *Journal of Cognitive Neuroscience* 18:48-63.
- Bukach CM, Le Grand R, Kaiser MD, Bub DN, Tanaka JW (2008) Preservation of mouth region processing in two cases of prosopagnosia. *J Neuropsychol* 2:227-244.
- Burton AM, Bruce V, Johnston RA (1990) Understanding face recognition with an interactive activation model. *Br J Psychol* 81 (Pt 3):361-380.
- Burton AM, Young AW, Bruce V, Johnston RA, Ellis AW (1991) Understanding covert recognition. *Cognition* 39:129-166.
- Burwell RD, Witter MP, Amaral DG (1995) Perirhinal and postrhinal cortices of the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus* 5:390-408.
- Buschman TJ, Miller EK (2007) Top-Down versus Bottom-Up control of attention in the prefrontal and posterior parietal cortices. *Science* 315:1860-1862.

- Byrne RW (1997) Machiavellian Intelligence. *Evolutionary Anthropology*:172-180.
- Calder AJ, Young AW (2005) Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci* 6:641-651.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110-1123.
- Calvert GA, Campbell R (2003) Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience* 15:57-70.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology* 10:649-657.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619-2623.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593-596.
- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11:535-543.
- Canolty RT, Ganguly K, Kennerley SW, Cadieu CF, Koepsell K, Wallis JD, Carmena JM (2010) Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies. *Proceedings of the National Academy of Sciences* 107:17356-17361.
- Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience* 22:2886-2902.
- Carney, R, Temple, C M (1993) Prosopagnosia? a possible category-specific anomia for faces. Basingstoke, ROYAUME-UNI: Taylor and Francis.
- Cézilly F (2004) Cognition et écologie comportementale. In: L'éthologie cognitive (Vauclair J, Kreutzer M, eds), pp 25-47: Ophrys.
- Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *J Neurophysiol* 101:773-788.
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436.
- Cheney D, Seyfarth R, Smuts B (1986) Social relationships and social cognition in nonhuman primates. *Science* 234:1361-1366.
- Cheney DL, Seyfarth RM (1997) Reconciliatory grunts by dominant female baboons influence victims' behaviour. *Anim Behav* 54:409-418.
- Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Naatanen R (1998) Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1:351-353.
- Christian EP, Deadwyler SA (1986) Behavioral functions and hippocampal cell types: evidence for two nonoverlapping populations in the rat. *J Neurophysiol* 55:331-348.
- Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA (2007) Evidence for a specific role of the anterior hippocampal region in successful associative encoding. *Hippocampus* 17:1071-1080.
- Cipolotti L, Shallice T, Chan D, Fox N, Scahill R, Harrison G, Stevens J, Rudge P (2001) Long-term retrograde amnesia... the crucial role of the hippocampus. *Neuropsychologia* 39:151-172.
- Clayton NS, Dickinson A (1998) Episodic-like memory during cache recovery by scrub jays. *Nature* 395:272-274.
- Clayton NS, Russell J (2009) Looking for episodic memory in animals and young children: prospects for a new minimalism. *Neuropsychologia* 47:2330-2340.
- Cohen L, Dehaene S (2004) Specialization within the ventral stream: the case for the visual word form area. *NeuroImage* 22:466-476.
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff M-A, Michel F (2000) The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients *Brain* 123:291-307.
- Cohen YE, Hauser MD, Russ BE (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biology Letters* 2:261-265.
- Cohen YE, Theunissen F, Russ BE, Gill P (2007) Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol* 97:1470-1484.

- Cohen YE, Russ BE, Gifford GW, Kiringoda R, MacLean KA (2004) Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. *The Journal of Neuroscience* 24:11307-11316.
- Colgin LL, Moser EI, Moser MB (2008) Understanding memory through hippocampal remapping. *Trends Neurosci* 31:469-477.
- Collins SA, Missing C (2003) Vocal and visual attractiveness are related in women. *Animal Behaviour* 65:997-1004.
- Colombo M, Fernandez T, Nakamura K, Gross CG (1998) Functional differentiation along the anterior-posterior axis of the hippocampus in monkeys. *Journal of Neurophysiology* 80:1002-1005.
- Connor CE (2010) A new viewpoint on faces. *Science* 330:764-765.
- Corkin S (2002) What's new with the amnesic patient H.M.? *Nat Rev Neurosci* 3:153-160.
- Corkin S, Amaral DG, Gonzalez RG, Johnson KA, Hyman BT (1997) H. M.'s medial temporal lobe lesion: findings from magnetic resonance imaging. *J Neurosci* 17:3964-3979.
- Cowey A, Weiskrantz L (1975) Demonstration of cross-modal matching in rhesus monkeys, *Macaca mulatta*. *Neuropsychologia* 13:117-120.
- Cutting JE, Kozlowski LT (1977) Recognizing friends by their walk: gait perception without familiarity cues. *Bulletin of the psychonomic society* 9:353-356.
- Dahl CD, Logothetis NK, Hoffman KL (2007) Individuation and holistic processing of faces in rhesus monkeys. *Proceedings of the Royal Society B: Biological Sciences* 274:2069-2076.
- Dahl CD, Logothetis NK, Kayser C (2009a) Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924-11932.
- Dahl CD, Logothetis NK, Kayser C (2010a) Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Front Integr Neurosci* 4:10.
- Dahl CD, Wallraven C, Bühlhoff HH, Logothetis NK (2009b) Humans and macaques employ similar face-processing strategies. *Current Biology* 19:509-513.
- Dahl CD, Logothetis NK, Bühlhoff HH, Wallraven C (2010b) The Thatcher illusion in humans and monkeys. *Proc Biol Sci* 277:2973-2981.
- Damasio AR (1989) Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33:25-62.
- Damasio AR, Tranel D, Damasio H (1990) Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience* 13:89-109.
- Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A (2004) Neural systems behind word and concept retrieval. *Cognition* 92:179-229.
- Damjanovic L, Hanley JR (2007) Recalling episodic and semantic information about famous faces and voices. *Mem Cognit* 35:1205-1210.
- Daselaar SM, Fleck MS, Cabeza R (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology* 96:1902-1911.
- Dasser V (1987) Slides of group members as representations of the real animals (*Macaca fascicularis*). *Ethology* 76:65-73.
- De Gelder B, Rouw R (2000) Structural encoding precludes recognition of face parts in prosopagnosia. *Cognitive Neuropsychology* 17:89-102.
- de Gelder B, Frissen I, Barton J, Hadjikhani N (2003) A modulatory role for facial expressions in prosopagnosia. *Proceedings of the National Academy of Sciences* 100:13105-13110.
- De Schonen S (2009) Percevoir un visage dans la petite enfance. *Evolution psychiatrique* 74:27-54.
- De Souza WC, Eifuku S, Tamura R, Nishijo H, Ono T (2005) Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *J Neurophysiol* 94:1252-1266.
- De Waal FBM, Yoshihara D (1983) Reconciliation and redirected affection in rhesus monkeys. *Behaviour* 85:224-241.
- de Waal FBM, Luttrell LM (1988) Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology* 9:101-118.
- de Waal FBM, Ferrari PF (2010) Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences* 14:201-207.
- Deacon T (1989) The neural circuitry underlying primate calls and human language. *Human Evolution* 4:367-401.

- Deaner RO, Khera AV, Platt ML (2005) Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Current Biology* 15:543-548.
- DeCasper AJ, Fifer WP (1980) Of human bonding: newborns prefer their mothers' voices. *Science* 208:1174-1176.
- DeCasper AJ, Prescott PA (1984) Human newborns' perception of male voices: preference, discrimination, and reinforcing value. *Dev Psychobiol* 17:481-491.
- deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439-1444.
- Dehaene S, Pegado F, Braga LW, Ventura P, Filho GN, Jobert A, Dehaene-Lambertz G, Kolinsky Rg, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. *Science* 330:1359-1364.
- Delorme A, Richard G, Fabre-Thorpe M (2000) Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res* 40:2187-2200.
- DéMonet J-F, Chollet F, Ramsay S, Cardebat D, Nespoulous J-L, Wise R, Rascol A, Frackowiak R (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753-1768.
- Denkova E, Botzung A, Manning L (2006) Neural correlates of remembering/knowing famous people: An event-related fMRI study. *Neuropsychologia* 44:2783-2791.
- Dépy D, Fagot J, Vauclair J (1999) Processing of above/below categorical spatial relations by baboons (*Papio papio*). *Behavioural Processes* 48:1-9.
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18:193-222.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051-2062.
- Distler C, Boussaoud D, Desimone R, Ungerleider LG (1993) Cortical connections of inferior temporal area TEO in macaque monkeys. *J Comp Neurol* 334:125-150.
- Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences* 98:10006-10010.
- Dong H-W, Swanson LW, Chen L, Fanselow MS, Toga AW (2009) Genomic-anatomic evidence for distinct functional domains in hippocampal field CA1. *Proceedings of the National Academy of Sciences* 106:11794-11799.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470-2473.
- Dufour V, Pascalis O, Petit O (2006) Face processing limitation to own species in primates: a comparative study in brown capuchins, Tonkean macaques and humans. *Behav Processes* 73:107-113.
- Duhamel JR, Colby CL, Goldberg ME (1998) Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J Neurophysiol* 79:126-136.
- Duzel E, Penny WD, Burgess N (2010) Brain oscillations and memory. *Curr Opin Neurobiol* 20:143-149.
- Ecklund-Flores L, Turkewitz G (1996) Asymmetric headturning to speech and nonspeech in human newborns. *Dev Psychobiol* 29:205-217.
- Eger E, Schyns PG, Kleinschmidt A (2004) Scale invariant adaptation in fusiform face-responsive regions. *NeuroImage* 22:232-242.
- Eichenbaum H (1993) Thinking about brain cell assemblies. *Science* 261:993-994.
- Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* 30:123-152.
- Eichenbaum H, Kuperstein M, Fagan A, Nagode J (1987) Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *J Neurosci* 7:716-732.
- Eifuku S, Nishijo H, Kita T, Ono T (1995) Neuronal activity in the primate hippocampal formation during a conditional association task based on the subject's location. *Journal of Neuroscience* 15:4952-4969.
- Eifuku S, De Souza WC, Tamura R, Nishijo H, Ono T (2004) Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J Neurophysiol* 91:358-371.

- Eifuku S, Nakata R, Sugimori M, Ono T, Tamura R (2010) Neural correlates of associative face memory in the anterior inferior temporal cortex of monkeys. *The Journal of Neuroscience* 30:15085-15096.
- Eifuku S, De Souza WC, Nakata R, Ono T, Tamura R (2011) Neural representations of personally familiar and unfamiliar faces in the anterior inferior temporal cortex of monkeys. *PLoS One* 6:e18913.
- Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. *Nature* 425:184-188.
- Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA (2000) Remembering episodes: a selective role for the hippocampus during retrieval. *Nat Neurosci* 3:1149-1152.
- Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453:1102-1106.
- Elliott RC (1977) Cross-modal recognition in three primates. *Neuropsychologia* 15:183-186.
- Ellis AW, Young AW, Critchley EMR (1989) Loss of memory for people following temporal lobe damage. *Brain* 112:1469-1483.
- Ellis HD, Young AW (1990) Accounting for delusional misidentifications. *Br J Psychiatry* 157:239-248.
- Ellis HD, Lewis MB (2001) Capgras delusion: a window on face recognition. *Trends Cogn Sci* 5:149-156.
- Ellis HD, Jones DM, Mosdell N (1997a) Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol* 88 (Pt 1):143-156.
- Ellis HD, Young AW, Quayle AH, De Pauw KW (1997b) Reduced autonomic responses to faces in Capgras delusion. *Proc Biol Sci* 264:1085-1092.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598-601.
- Etcoff NL (1984) Selective attention to facial identity and facial emotion. *Neuropsychologia* 22:281-295.
- Ettlinger G, Blakemore CB (1967) Cross-modal matching in the monkey. *Neuropsychologia* 5:147-154.
- Evans JJ, Hegggs AJ, Antoun N, Hodges JR (1995) Progressive prosopagnosia associated with selective right temporal lobe atrophy. A new syndrome? *Brain* 118 (Pt 1):1-13.
- Fabre-Thorpe M (2003) Visual categorization: accessing abstraction in non-human primates. *Philos Trans R Soc Lond B Biol Sci* 358:1215-1223.
- Fabre-Thorpe M, Richard G, Thorpe SJ (1998) Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9:303-308.
- Fagot J, Wasserman E, Young M (2004) Catégorisation d'objets visuels et concepts relationnels chez l'animal. In: L'éthologie cognitive (Vauclair J, Kreutzer M, eds), pp 117-136: Ophrys.
- Fairhall SL, Ishai A (2007) Effective connectivity within the distributed cortical network for face perception. *Cereb Cortex* 17:2400-2406.
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749-5759.
- Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF (2010) Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb Cortex* 20:1529-1538.
- Fanselow MS, Dong H-W (2010) Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* 65:7-19.
- Fant G (1960) Acoustic theory of speech production. La Haye, Netherlands: Mouton & Co.
- Farroni T, Menon E, Johnson MH (2006) Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology* 95:298-308.
- Farroni T, Csibra G, Simion F, Johnson MH (2002) Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences* 99:9602-9605.
- Farroni T, Johnson MH, Menon E, Zulian L, Faraguna D, Csibra G (2005) Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America* 102:17245-17250.
- Fecteau S, Armony JL, Joanette Y, Belin P (2004) Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage* 23:840-848.
- Fecteau S, Armony JL, Joanette Y, Belin P (2005) Sensitivity to voice in human prefrontal cortex. *Journal of Neurophysiology* 94:2251-2254.

- Fecteau S, Belin P, Joanette Y, Armony JL (2007) Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage* 36:480-487.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1-47.
- Fischer J (2004) Emergence of individual recognition in young macaques. *Animal Behaviour* 67:655-661.
- Fitch WT (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J Acoust Soc Am* 102:1213-1222.
- Fitch WT (2000) The evolution of speech: a comparative review. *Trends Cogn Sci* 4:258-267.
- Fitch WT, Fritz JB (2006) Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J Acoust Soc Am* 120:2132-2141.
- Fize D, Cauchoix M, Fabre-Thorpe M (2011) Humans and monkeys share visual representations. *Proc Natl Acad Sci U S A* 108:7635-7640.
- Flom R, Bahrick LE (2010) The effects of intersensory redundancy on attention and memory: infants' long-term memory for orientation in audiovisual events. *Dev Psychol* 46:428-436.
- Foldiak P, Xiao D, Keysers C, Edwards R, Perrett DI (2004) Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog Brain Res* 144:107-116.
- Fooden J (1982) Ecogeographic segregation of macaque species. *Primates* 23:574-579.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322:970-973.
- Formisano E, Kim D-S, Di Salle F, van de Moortele P-F, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859-869.
- Frankland PW, Bontempi B (2005) The organization of recent and remote memories. *Nat Rev Neurosci* 6:119-130.
- Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845-851.
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187-1196.
- Fried I, MacDonald KA, Wilson CL (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18:753-765.
- Fried I, Cameron KA, Yashar S, Fong R, Morrow JW (2002) Inhibitory and excitatory responses of single neurons in the human medial temporal lobe during recognition of faces and objects. *Cereb Cortex* 12:575-584.
- Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560-1563.
- Frith CD, Frith U (1999) Interacting minds--A biological basis. *Science* 286:1692-1695.
- Fuster JM (2006) The cognit: A network model of cortical representation. *International Journal of Psychophysiology* 60:125-132.
- 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.
- Galaburda AM, Pandya DN (1983) The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey. *J Comp Neurol* 221:169-184.
- Gallagher HL, Frith CD (2003) Functional imaging of "theory of mind". *Trends in Cognitive Sciences* 7:77-83.
- Gauthier I, Logothetis NK (2000) Is face recognition not so unique after all? *Cogn Neuropsychol* 17:125-142.
- Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000a) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3:191-197.
- Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci* 2:568-573.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000b) The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience* 12:495-504.
- Gava L, Valenza E, Turati C, Schonen Sd (2008) Effect of partial occlusion on newborns' face preference and recognition. *Developmental Science* 11:563-574.

- Gavrilets S, Duenez-Guzman EA, Vose MD (2008) Dynamics of alliance formation and the egalitarian revolution. *PLoS ONE* 3:e3293.
- Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322:96-101.
- Gerald MS, Waite C, Little AC (2009) Pregnancy coloration in macaques may act as a warning signal to reduce antagonism by conspecifics. *Behavioural Processes* 80:7-11.
- Ghazanfar AA (2008) Language evolution: neural differences that make a difference. *Nat Neurosci* 11:382-384.
- Ghazanfar AA (2009) The multisensory roles for auditory cortex in primate vocal communication. *Hear Res* 258:113-120.
- Ghazanfar AA, Hauser MD (1999) The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends in Cognitive Sciences* 3:377-384.
- Ghazanfar AA, Logothetis NK (2003) Neuroperception: facial expressions linked to monkey calls. *Nature* 423:937-938.
- Ghazanfar AA, Santos LR (2004) Primate brains in the wild: the sensory bases for social interactions. *Nat Rev Neurosci* 5:603-616.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278-285.
- Ghazanfar AA, Chandrasekaran CF (2007) Paving the way forward: integrating the senses through phase-resetting of cortical oscillations. *Neuron* 53:162-164.
- Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28:4457-4469.
- Ghazanfar AA, Chandrasekaran C, Morrill RJ (2010) Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. *Eur J Neurosci* 31:1807-1817.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004-5012.
- Ghazanfar AA, Tureson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17:425-430.
- Ghio A, Pinto S (2007) Résonance sonore et cavités supralaryngées. In: Les Dysarthries (Auzou P, Rolland-Monnoury V, Pinto S, Özsancak C, eds), pp 101-110. France: Solal.
- Gifford GW, MacLean KA, Hauser MD, Cohen YE (2005) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *Journal of Cognitive Neuroscience* 17:1471-1482.
- Gil-da-Costa R, Martin A, Lopes MA, Munoz M, Fritz JB, Braun AR (2006) Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nat Neurosci* 9:1064-1070.
- Gil-da-Costa R, Braun A, Lopes M, Hauser MD, Carson RE, Herscovitch P, Martin A (2004) Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc Natl Acad Sci U S A* 101:17516-17521.
- Gilad S, Meng M, Sinha P (2009) Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences* 106:5353-5358.
- Girardeau G, Benchenane K, Wiener SI, Buzsaki G, Zugaro MB (2009) Selective suppression of hippocampal ripples impairs spatial memory. *Nat Neurosci* 12:1222-1223.
- Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84:1588-1598.
- Gobbini MI, Haxby JV (2006) Neural response to the visual familiarity of faces. *Brain Res Bull* 71:76-82.
- Gobbini MI, Haxby JV (2007) Neural systems for recognition of familiar faces. *Neuropsychologia* 45:32-41.
- Goffaux V, Rossion B (2007) Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *J Exp Psychol Hum Percept Perform* 33:995-1002.
- González J (2004) Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics* 32:277-287.

- Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends in Neurosciences* 15:20-25.
- Goren CC, Sarty M, Wu PY (1975) Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56:544-549.
- Gorno-Tempini ML, Price CJ (2001) Identification of famous faces and buildings. *Brain* 124:2087-2097.
- Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, Frackowiak RS, Tempini ML (1998) The neural systems sustaining face and proper-name processing. *Brain* 121:2103-2118.
- Gothard KM, Erickson CA, Amaral DG (2004) How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Animal Cognition* 7:25-36.
- Gothard KM, Brooks KN, Peterson MA (2009) Multiple perceptual strategies used by macaque monkeys for face recognition. *Animal Cognition* 12:155-167.
- Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG (2007) Neural responses to facial expression and face identity in the monkey amygdala. *Journal of Neurophysiology* 97:1671-1683.
- Gouzoules S, Gouzoules H, Marler P (1984) Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour* 32:182-193.
- Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc Natl Acad Sci U S A* 86:1698-1702.
- Gray CM, Konig P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334-337.
- Greenwald AG, Banaji MR (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review* 102:4-27.
- Griffin DR (1978) Prospects for a cognitive ethology. *Behavioral and Brain Sciences* 1:527-538.
- Griffiths TD, Warren JD (2004) What is an auditory object? *Nat Rev Neurosci* 5:887-892.
- Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci* 7:555-562.
- Grill-Spector K, Sayres R, Ress D (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat Neurosci* 9:1177-1185.
- Gross CG, Bender DB, Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166:1303-1306.
- Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* 35:96-111.
- Grossmann T, Oberecker R, Koch SP, Friederici AD (2010) The developmental origins of voice processing in the human brain. *Neuron* 65:852-858.
- Gruppuso V, Lindsay DS, Masson ME (2007) I'd know that face anywhere! *Psychon Bull Rev* 14:1085-1089.
- Guderian S, Düzel E (2005) Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans. *Hippocampus* 15:901-912.
- Guo K, Robertson RG, Mahmoodi S, Tadmor Y, Young MP (2003) How do monkeys view faces? - a study of eye movements. *Experimental Brain Research* 150:363-374.
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197-222.
- Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RBH (2008) Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. *Proceedings of the National Academy of Sciences* 105:5591-5596.
- Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD (2010) Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia* 48:1104-1114.
- Haist F, Gore JB, Mao H (2001) Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nat Neurosci* 4:1139-1145.
- Haith MM, Bergman T, Moore MJ (1977) Eye contact and face scanning in early infancy. *Science* 198:853-855.
- Hampson RE, Pons TP, Stanford TR, Deadwyler SA (2004) Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. *Proc Natl Acad Sci U S A* 101:3184-3189.

- Hampton RR, Schwartz BL (2004) Episodic memory in nonhumans: what, and where, is when? *Curr Opin Neurobiol* 14:192-197.
- Hampton RR, Hampstead BM, Murray EA (2004) Selective hippocampal damage in rhesus monkeys impairs spatial memory in an open-field test. *Hippocampus* 14:808-818.
- Hampton RR, Hampstead BM, Murray EA (2005) Rhesus monkeys (*Macaca mulatta*) demonstrate robust memory for what and where, but not when, in an open-field test of memory. *Learning and Motivation* 36:245-259.
- Hampton RR, Healy SD, Shettleworth SJ, Kamil AC (2002) 'Neuroecologists' are not made of straw. *Trends in Cognitive Sciences* 6:6-7.
- Hanley JR, Turner JM (2000) Why are familiar-only experiences more frequent for voices than for faces? *Q J Exp Psychol A* 53:1105-1116.
- Hanley JR, Young AW, Pearson NA (1989) Defective recognition of familiar people. *Cognitive Neuropsychology* 6:179-210.
- Hanley JR, Pearson NA, Young AW (1990) Impaired memory for new visual forms. *Brain* 113 (Pt 4):1131-1148.
- Hanley JR, Smith ST, Hadfield J (1998) I recognise you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology Section A* 51:179-195.
- Harmon LD, Julesz B (1973) Masking in visual recognition: effects of two-dimensional filtered noise. *Science* 180:1194-1197.
- Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. *Trends Cogn Sci* 11:299-306.
- Hasselmo ME, Rolls ET, Baylis GC (1989) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res* 32:203-218.
- Hauser MD (1998) Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Animal Behaviour* 55:1647-1658.
- Hauser MD (2007) When males call, females listen: sex differences in responsiveness to rhesus monkey, *Macaca mulatta*, copulation calls. *Animal Behaviour* 73:1059-1065.
- Hauser MD, Marler P (1993) Food-associated calls in rhesus macaques (*Macaca mulatta*): II. Costs and benefits of call production and suppression. *Behavioral Ecology* 4:206-212.
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223-233.
- Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* 14:6336-6353.
- Haxby JV, Ungerleider LG, Horwitz B, Maisog JM, Rapoport SI, Grady CL (1996) Face encoding and recognition in the human brain. *Proceedings of the National Academy of Sciences* 93:922-927.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425-2430.
- Haxby JV, Grady CL, Horwitz B, Ungerleider LG, Mishkin M, Carson RE, Herscovitch P, Schapiro MB, Rapoport SI (1991) Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences* 88:1621-1625.
- Hayasaka K, Fujii K, Horai S (1996) Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Mol Biol Evol* 13:1044-1053.
- Hebb DO (1949) The organization of behavior : a neuropsychological theory / D.O. Hebb. New York :: Wiley.
- Heffner RS (2004) Primate hearing from a mammalian perspective. *Anat Rec A Discov Mol Cell Evol Biol* 281:1111-1122.
- Hendelman W (2000) Atlas of functional neuroanatomy: CRC Press.
- Heywood CA, Cowey A, Rolls ET (1992) The role of the "face-cell" area in the discrimination and recognition of faces by monkeys [and Discussion]. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 335:31-38.

- Hietanen JK, Perrett DI, Oram MW, Benson PJ, Dittrich WH (1992) The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp Brain Res* 89:157-171.
- Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36:791-804.
- Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3:80-84.
- Hoffman KL, McNaughton BL (2002) Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297:2070-2073.
- Hoffman KL, Ghazanfar AA, Gauthier I, Logothetis NK (2007a) Category-specific responses to faces and objects in primate auditory cortex. *Frontiers in Systems Neuroscience* 2.
- Hoffman KL, Gothard KM, Schmid Michael C, Logothetis NK (2007b) Facial-expression and gaze-selective responses in the monkey amygdala. *Current Biology* 17:766-772.
- Holdstock JS, Crane J, Bachorowski JA, Milner B (2010) Equivalent activation of the hippocampus by face-face and face-laugh paired associate learning and recognition. *Neuropsychologia* 48:3757-3771.
- Hubel DH, Wiesel TN (1963a) Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J Neurophysiol* 26:994-1002.
- Hubel DH, Wiesel TN (1963b) Shape and arrangement of columns in cat's striate cortex. *J Physiol* 165:559-568.
- Husain FT, Tagamets MA, Fromm SJ, Braun AR, Horwitz B (2004) Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *NeuroImage* 21:1701-1720.
- Ida Gobbin M, Leibenluft E, Santiago N, Haxby JV (2004) Social and emotional attachment in the neural representation of faces. *NeuroImage* 22:1628-1635.
- Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K (1997) Vocal identification of speaker and emotion activates different brain regions. *NeuroReport* 8:2809-2812.
- Insausti R, Muñoz M (2001) Cortical projections of the non-entorhinal hippocampal formation in the cynomolgus monkey (*Macaca fascicularis*). *European Journal of Neuroscience* 14:435-451.
- Insausti R, Amaral DG (2008) Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. *The Journal of Comparative Neurology* 509:608-641.
- Insausti R, Amaral DG, Cowan WM (1987) The entorhinal cortex of the monkey: II. Cortical afferents. *The Journal of Comparative Neurology* 264:356-395.
- Ishai A (2007) Sex, beauty and the orbitofrontal cortex. *Int J Psychophysiol* 63:181-185.
- Ishai A, Yago E (2006) Recognition memory of newly learned faces. *Brain Res Bull* 71:167-173.
- Ishai A, Ungerleider LG, Haxby JV (2000) Distributed neural systems for the generation of visual images. *Neuron* 28:979-990.
- Ishai A, Haxby JV, Ungerleider LG (2002) Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage* 17:1729-1741.
- Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network. *Brain Res Bull* 67:87-93.
- Ishai A, Pessoa L, Bickle PC, Ungerleider LG (2004) Repetition suppression of faces is modulated by emotion. *Proc Natl Acad Sci U S A* 101:9827-9832.
- Ison MJ, Mormann F, Cerf M, Koch C, Fried I, Quiñero Quiroga R (2011) Selectivity of pyramidal cells and interneurons in the Human Medial Temporal Lobe. *J Neurophysiol*.
- Ives DT, Smith DR, Patterson RD (2005) Discrimination of speaker size from syllable phrases. *J Acoust Soc Am* 118:3816-3822.
- Izumi A, Kojima S (2004) Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal Cognition* 7:179-184.
- Janik VM, Sayigh LS, Wells RS (2006) Signature whistle shape conveys identity information to bottlenose dolphins. *Proc Natl Acad Sci U S A* 103:8293-8297.
- Jarvis MJ, Ettliger G (1977) Cross-modal recognition in chimpanzees and monkeys. *Neuropsychologia* 15:499-506.

- 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. *Neurosci Lett* 369:132-137.
- Joassin F, Pesenti M, Maurage P, Verreckett E, Bruyer R, Campanella S (2011) Cross-modal interactions between human faces and voices involved in person recognition. *Cortex* 47:367-376.
- Johnson MH (2005) Subcortical face processing. *Nat Rev Neurosci* 6:766-774.
- Johnson MH, Dziurawiec S, Ellis H, Morton J (1991) Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40:1-19.
- Johnston RE (1993) Memory for individual scent in hamsters (*Mesocricetus auratus*) as assessed by habituation methods. *J Comp Psychol* 107:201-207.
- Johnston RE, Peng A (2008) Memory for individuals: hamsters (*Mesocricetus auratus*) require contact to develop multicomponent representations (concepts) of others. *J Comp Psychol* 122:121-131.
- Johnstone T, van Reekum CM, Oakes TR, Davidson RJ (2006) The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience* 1:242-249.
- Jones MW, Wilson MA (2005) Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol* 3:e402.
- Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15:1034-1038.
- Judge PG, De Waal FBM (1997) Rhesus monkey behaviour under diverse population densities: coping with long-term crowding. *Animal Behaviour* 54:643-662.
- Jung MW, Wiener SI, McNaughton BL (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience* 14:7347-7356.
- Jutras MJ, Buffalo EA (2010a) Synchronous neural activity and memory formation. *Curr Opin Neurobiol* 20:150-155.
- Jutras MJ, Buffalo EA (2010b) Recognition memory signals in the macaque hippocampus. *Proc Natl Acad Sci U S A* 107:401-406.
- Jutras MJ, Fries P, Buffalo EA (2009) Gamma-band synchronization in the macaque hippocampus and memory formation. *J Neurosci* 29:12521-12531.
- Kaas JH, Hackett TA (1999) 'What' and 'where' processing in auditory cortex. *Nat Neurosci* 2:1045-1047.
- Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* 97:11793-11799.
- Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9:164-170.
- Kaitz M (1992) Recognition of familiar individuals by touch. *Physiol Behav* 52:565-567.
- Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003) "Putting the face to the voice": matching identity across modality. *Curr Biol* 13:1709-1714.
- Kanwisher N (2006) What's in a face? *Science* 311:617-618.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302-4311.
- Kaplan JR (1977) Patterns of fight interference in free-ranging rhesus monkeys. *American Journal of Physical Anthropology* 47:279-287.
- Kaplan JR (1978) Fight interference and altruism in rhesus monkeys. *American Journal of Physical Anthropology* 49:241-249.
- Kayser C, Petkov CI, Logothetis NK (2009) Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hear Res* 258:80-88.
- Kelly DJ, Quinn PC, Slater AM, Lee K, Gibson A, Smith M, Ge L, Pascalis O (2005) Three-month-olds, but not newborns, prefer own-race faces. *Dev Sci* 8:F31-36.
- Kemp R, McManus C, Pigott T (1990) Sensitivity to the displacement of facial features in negative and inverted images. *Perception* 19:531-543.
- Kikuchi Y, Horwitz B, Mishkin M (2010) Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *The Journal of Neuroscience* 30:13021-13030.
- Kim H, Adolphs R, O'Doherty JP, Shimojo S (2007) Temporal isolation of neural processes underlying face preference decisions. *Proc Natl Acad Sci U S A* 104:18253-18258.

- Kirwan CB, Stark CE (2004) Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus* 14:919-930.
- Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z (2003) Effects of experience on fetal voice recognition. *Psychol Sci* 14:220-224.
- Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser M-B (2008) Finite scale of spatial representation in the hippocampus. *Science* 321:140-143.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71:856-867.
- Kojima S, Izumi A, Ceugniet M (2003) Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates* 44:225-230.
- Konkel A, Cohen NJ (2009) Relational memory and the hippocampus: Representations and methods. *Frontiers in Neuroscience* 3.
- Konorski J (1967) Some new ideas concerning the physiological mechanisms of perception. *Acta Biol Exp (Warsz)* 27:147-161.
- Korshunov VA, Wiener SI, Korshunova TA, Berthoz A (1996) Place- and behavior-independent sensory triggered discharges in rat hippocampal CA1 complex spike cells. *Exp Brain Res* 109:169-173.
- Kosaka H, Omori M, Iidaka T, Murata T, Shimoyama T, Okada T, Sadato N, Yonekura Y, Wada Y (2003) Neural substrates participating in acquisition of facial familiarity: an fMRI study. *NeuroImage* 20:1734-1742.
- Kosel KC, Van Hoesen GW, Rosene DL (1982) Non-hippocampal cortical projections from the entorhinal cortex in the rat and rhesus monkey. *Brain Research* 244:201-213.
- Kranz F, Ishai A (2006) Face perception is modulated by sexual preference. *Current Biology* 16:63-68.
- Kraskov A, Quiroga RQ, Reddy L, Fried I, Koch C (2007) Local field potentials and spikes in the human medial temporal lobe are selective to image category. *J Cogn Neurosci* 19:479-492.
- Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci* 3:946-953.
- Kreutzer M, Vaclair J (2004) La cognition animale au carrefour de l'éthologie et de la psychologie. In: *L'éthologie cognitive* (Vaclair J, Kreutzer M, eds), pp 1-19: Ophrys.
- Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences* 104:20600-20605.
- Ku S-P, Tolias Andreas S, Logothetis Nikos K, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70:352-362.
- Kuhl PK, Tsao FM, Liu HM (2003) Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci U S A* 100:9096-9101.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606-608.
- Kuraoka K, Nakamura K (2006) Impacts of facial identity and type of emotion on responses of amygdala neurons. *Neuroreport* 17:9-12.
- Kuraoka K, Nakamura K (2007) Responses of single neurons in monkey amygdala to facial and vocal emotions. *J Neurophysiol* 97:1379-1387.
- Kusmierek P, Rauschecker JP (2009) Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102:1606-1622.
- Kuwahata H, Adachi I, Fujita K, Tomonaga M, Matsuzawa T (2004) Development of schematic face preference in macaque monkeys. *Behavioural Processes* 66:17-21.
- Lachs L, Pisoni DB (2004) Crossmodal source identification in speech perception. *Ecological Psychology* 16:159-187.
- Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.
- Lancker DRV, Kreiman J, Cummings J (1989) Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology* 11:665-674.
- Lander K, Davies R (2007) Exploring the role of characteristic motion when learning new faces. *The Quarterly Journal of Experimental Psychology* 60:519-526.

- Lander K, Hill H, Kamachi M, Vatikiotis-Bateson E (2007) It's not what you say but the way you say it: matching faces and voices. *J Exp Psychol Hum Percept Perform* 33:905-914.
- Landis T, Cummings JL, Christen L, Bogen JE, Imhof HG (1986) Are unilateral right posterior cerebral lesions sufficient to cause prosopagnosia? Clinical and radiological findings in six additional patients. *Cortex* 22:243-252.
- Latinus M, Belin P (2011) Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology* 2.
- Lavenex P, Suzuki WA, Amaral DG (2004) Perirhinal and parahippocampal cortices of the macaque monkey: Intrinsic projections and interconnections. *The Journal of Comparative Neurology* 472:371-394.
- Lehn H, Steffenach H-A, van Strien NM, Veltman DJ, Witter MP, Håberg AK (2009) A specific role of the human hippocampus in recall of temporal sequences. *The Journal of Neuroscience* 29:3475-3484.
- Leibenluft E, Gobbi MI, Harrison T, Haxby JV (2004) Mothers' neural activation in response to pictures of their children and other children. *Biol Psychiatry* 56:225-232.
- Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572-575.
- Leopold DA, O'Toole AJ, Vetter T, Blanz V (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci* 4:89-94.
- Leopold DA, Rhodes G, Müller K-M, Jeffery L (2005) The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences* 272:897-904.
- Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience* 20:878-886.
- Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci U S A* 103:6771-6774.
- Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences* 13:470-478.
- Lewkowicz DJ, Sowinski R, Place S (2008) The decline of cross-species intersensory perception in human infants: underlying mechanisms and its developmental persistence. *Brain Res* 1242:291-302.
- Livingstone M, Srihasam K, Morocz I (2010) The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Animal Cognition* 13:711-719.
- Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* 453:869-878.
- Logothetis NK, Guggenberger H, Peled S, Pauls J (1999) Functional imaging of the monkey brain. *Nat Neurosci* 2:555-562.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150-157.
- Loula F, Prasad S, Harber K, Shiffrar M (2005) Recognizing people from their movement. *J Exp Psychol Hum Percept Perform* 31:210-220.
- Lucchelli F, Spinnler H (2008) A reappraisal of person recognition and identification. *Cortex* 44:230-237.
- Lundstrom JN, Boyle JA, Zatorre RJ, Jones-Gotman M (2009) The neuronal substrates of human olfactory based kin recognition. *Hum Brain Mapp* 30:2571-2580.
- Mace MJ, Delorme A, Richard G, Fabre-Thorpe M (2010) Spotting animals in natural scenes: efficiency of humans and monkeys at very low contrasts. *Anim Cogn* 13:405-418.
- Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran J-P, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. *NeuroImage* 14:802-816.
- Maestriperi D (2007) *Macchiavellian intelligence*. Chicago; London: the University of Chicago Press.
- Maguire EA (2001) Neuroimaging studies of autobiographical event memory. *Philos Trans R Soc Lond B Biol Sci* 356:1441-1451.
- Mahajan N, Martinez MA, Gutierrez NL, Diesendruck G, Banaji MR, Santos LR (2011) The evolution of intergroup bias: perceptions and attitudes in rhesus macaques. *J Pers Soc Psychol* 100:387-405.

- Manns JR, Howard MW, Eichenbaum H (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56:530-540.
- Martin-Malivel J, Fagot J (2001) Cross-modal integration and conceptual categorization in baboons. *Behav Brain Res* 122:209-213.
- Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology* 11:194-201.
- Martinez L, Matsuzawa T (2009) Auditory-visual intermodal matching based on individual recognition in a chimpanzee (*Pan troglodytes*). *Animal Cognition* 12:71-85.
- Maviel T, Durkin TP, Menzaghi F, Bontempi B (2004) Sites of neocortical reorganization critical for remote spatial memory. *Science* 305:96-99.
- Mayes A, Montaldi D, Migo E (2007) Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences* 11:126-135.
- McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience* 9:605-610.
- McEchron MD, Disterhoft JF (1997) Sequence of single neuron changes in CA1 hippocampus of rabbits during acquisition of trace eyeblink conditioned responses. *J Neurophysiol* 78:1030-1044.
- McEchron MD, Disterhoft JF (1999) Hippocampal encoding of non-spatial trace conditioning. *Hippocampus* 9:385-396.
- McGurk H, Macdonald J (1976) Hearing lips and seeing voices. *Nature* 264:746-748.
- Mechelli A, Price CJ, Friston KJ, Ishai A (2004) Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cerebral Cortex* 14:1256-1265.
- Meyer K, Damasio A (2009) Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci* 32:376-382.
- Meyer K, Kaplan JT, Essex R, Webber C, Damasio H, Damasio A (2010) Predicting visual stimuli on the basis of activity in auditory cortices. *Nat Neurosci* 13:667-668.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24:167-202.
- Milner AD (1973) Matching within and between sense modalities in the monkey (*Macaca mulatta*). *J Comp Physiol Psychol* 83:278-284.
- Milner AD, Ettlinger G (1970) Cross-modal transfer of serial reversal learning in the monkey. *Neuropsychologia* 8:251-258.
- Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences* 6:414-417.
- Mitchell JP, Heatherton TF, Macrae CN (2002) Distinct neural systems subserving person and object knowledge. *Proc Natl Acad Sci U S A* 99:15238-15243.
- Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320:1355-1359.
- Mohedano-Moriano A, Martinez-Marcos A, Pro-Sistiaga P, Blaizot X, Arroyo-Jimenez MM, Marcos P, Artacho-Pérula E, Insausti R (2008) Convergence of unimodal and polymodal sensory input to the entorhinal cortex in the fascicularis monkey. *Neuroscience* 151:255-271.
- Mohedano-Moriano A, Pro-Sistiaga P, Arroyo-Jimenez MM, Artacho-Pérula E, Insausti AM, Marcos P, Cebada-Sánchez S, Martínez-Ruiz J, Muñoz M, Blaizot X, Martínez-Marcos A, Amaral DG, Insausti R (2007) Topographical and laminar distribution of cortical input to the monkey entorhinal cortex. *Journal of Anatomy* 211:250-260.
- Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005) The neural basis of human moral cognition. *Nat Rev Neurosci* 6:799-809.
- Mondloch CJ, Lewis TL, Budreau DR, Maurer D, Dannemiller JL, Stephens BR, Kleiner-Gathercoal KA (1999) Face perception during early infancy. *Psychological Science* 10:419-422.
- Morales JC, Melnick DJ (1998) Phylogenetic relationships of the macaques (*Cercopithecidae:Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *Journal of Human Evolution* 34:1-23.
- Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried I, Koch C (2008) Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J Neurosci* 28:8865-8872.

- Morris JS, Frith CD, Perrett DI, Rowland D, Young AW, Calder AJ, Dolan RJ (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383:812-815.
- 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* 9:555-604.
- Moscovitch M, Nadel L, Winocur G, Gilboa A, Rosenbaum RS (2006) The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr Opin Neurobiol* 16:179-190.
- Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, McAndrews MP, Levine B, Black S, Winocur G, Nadel L (2005) Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anat* 207:35-66.
- Moser M-B, Moser EI (1998) Distributed encoding and retrieval of spatial memory in the hippocampus. *The Journal of Neuroscience* 18:7535-7542.
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951-954.
- Mullennix JW, Ross A, Smith C, Kuykendall K, Conard J, Barb S (2009) Typicality effects on memory for voice: Implications for eyewitness testimony. *Applied Cognitive Psychology* 25:29-34.
- Munoz-Lopez M, MohedanoMoriano A, Insausti R (2010) Anatomical pathways for auditory memory in primates. *Frontiers in Neuroanatomy* 4.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences* 24:283-288.
- Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* 7:217-227.
- Nakamura K, Mikami A, Kubota K (1992) Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J Neurophysiol* 67:1447-1463.
- Nakamura K, Matsumoto K, Mikami A, Kubota K (1994) Visual response properties of single neurons in the temporal pole of behaving monkeys. *J Neurophysiol* 71:1206-1221.
- Nakamura K, Kawashima R, Nagumo S, Ito K, Sugiura M, Kato T, Nakamura A, Hatano K, Kubota K, Fukuda H, Kojima S (1998) Neuroanatomical correlates of the assessment of facial attractiveness. *Neuroreport* 9:753-757.
- Nakamura K, Kawashima R, Sato N, Nakamura A, Sugiura M, Kato T, Hatano K, Ito K, Fukuda H, Schormann T, Zilles K (2000) Functional delineation of the human occipito-temporal areas related to face and scene processing. *Brain* 123:1903-1912.
- 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.
- Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, Sun LD, Kato A, Carr CA, Johnston D, Wilson MA, Tonegawa S (2002) Requirement for Hippocampal CA3 NMDA Receptors in Associative Memory Recall. *Science* 297:211-218.
- Naya Y, Suzuki WA (2011) Integrating What and When across the primate medial temporal lobe. *Science* 333:773-776.
- Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences* 108:9998-10003.
- Neuner F, Schweinberger SR (2000) Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition* 44:342-366.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147-155.
- O'Kane G, Kensinger EA, Corkin S (2004) Evidence for semantic learning in profound amnesia: an investigation with patient H.M. *Hippocampus* 14:417-425.
- O'Keefe J (1976) Place units in the hippocampus of the freely moving rat. *Experimental Neurology* 51:78-109.
- O'Keefe J (1999) Do hippocampal pyramidal cells signal non-spatial as well as spatial information? *Hippocampus* 9:352-364.

- O'Mara SM, Rolls ET, Berthoz A, Kesner RP (1994) Neurons responding to whole-body motion in the primate hippocampus. *J Neurosci* 14:6511-6523.
- O'Toole AJ, Roark DA, Abdi H (2002) Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences* 6:261-266.
- O Scalaidhe SP, Wilson FAW, Goldman-Rakic PS (1997) Areal segregation of face-processing neurons in prefrontal cortex. *Science* 278:1135-1138.
- O Scalaidhe SP, Wilson FAW, Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cerebral Cortex* 9:459-475.
- Ockleford EM, Vince MA, Layton C, Reader MR (1988) Responses of neonates to parents' and others' voices. *Early Human Development* 18:27-36.
- Olsson N, Juslin P, Winman A (1998) Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied* 4:101-118.
- Olsson SB, Barnard J, Turri L (2006) Olfaction and identification of unrelated individuals: examination of the mysteries of human odor recognition. *J Chem Ecol* 32:1635-1645.
- Ono T, Nakamura K, Nishijo H, Eifuku S (1993a) Monkey hippocampal neurons related to spatial and nonspatial functions. *J Neurophysiol* 70:1516-1529.
- Ono T, Eifuku S, Nakamura K, Nishijo H (1993b) Monkey hippocampal neuron responses related to spatial and non-spatial influence. *Neurosci Lett* 159:75-78.
- Op de Beeck HP, Dicarlo JJ, Goense JB, Grill-Spector K, Papanastassiou A, Tanifuji M, Tsao DY (2008) Fine-scale spatial organization of face and object selectivity in the temporal lobe: do functional magnetic resonance imaging, optical imaging, and electrophysiology agree? *J Neurosci* 28:11796-11801.
- Orban de Xivry JJ, Ramon M, Lefevre P, Rossion B (2008) Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *J Neuropsychol* 2:245-268.
- Owren MJ, Rendall D, Bachorowski JA (2005) Nonlinguistic vocal communication. In: *Primate Psychology* (Maestriperieri D, ed), pp 359-394. Cambridge, Massachusetts: Harvard University Press.
- Paller KA, Gonsalves B, Grabowecky M, Bozic VS, Yamada S (2000) Electrophysiological correlates of recollecting faces of known and unknown individuals. *Neuroimage* 11:98-110.
- Paller KA, Ranganath C, Gonsalves B, LaBar KS, Parrish TB, Gitelman DR, Mesulam MM, Reber PJ (2003) Neural correlates of person recognition. *Learn Mem* 10:253-260.
- Park J-Y, Gu B-M, Kang D-H, Shin Y-W, Choi C-H, Lee J-M, Kwon JS (2010) Integration of cross-modal emotional information in the human brain: An fMRI study. *Cortex* 46:161-169.
- Parr LA, Maestriperieri D (2005) Nonvocal communication. In: *Primate Psychology* (Maestriperieri D, ed), pp 324-358. Cambridge, Massachusetts: Harvard University Press.
- Parr LA, Winslow JT, Hopkins WD, de Waal FB (2000) Recognizing facial cues: individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *J Comp Psychol* 114:47-60.
- Pascalis O, Bachevalier J (1998) Face recognition in primates: a cross-species study. *Behavioural Processes* 43:87-96.
- Pascalis O, de Haan M, Nelson CA (2002) Is face processing species-specific during the first year of life? *Science* 296:1321-1323.
- Pascalis O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, Nelson CA (2005) Plasticity of face processing in infancy. *Proc Natl Acad Sci U S A* 102:5297-5300.
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976-987.
- Paz R, Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2010) A neural substrate in the human hippocampus for linking successive events. *Proc Natl Acad Sci U S A* 107:6046-6051.
- Pedreira C, Mormann F, Kraskov A, Cerf M, Fried I, Koch C, Quiroga RQ (2010) Responses of human medial temporal lobe neurons are modulated by stimulus repetition. *J Neurophysiol* 103:97-107.
- Peretz I, Kolinsky Rg, Tramo M, Labrecque R, Hublet C, Demeurisse G, Belleville S (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117:1283-1301.
- Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329-342.

- Perrett DI, Hietanen JK, Oram MW, Benson PJ (1992) Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci* 335:23-30.
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1984) Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum Neurobiol* 3:197-208.
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985) Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc Lond B Biol Sci* 223:293-317.
- Perrett DI, Oram MW, Harries MH, Bevan R, Hietanen JK, Benson PJ, Thomas S (1991) Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp Brain Res* 86:159-173.
- Perrodin C, Kayser C, Logothetis Nikos K, Petkov Christopher I (2011) Voice cells in the primate temporal lobe. *Current Biology* 21:1408-1415.
- Peters J, Koch B, Schwarz M, Daum I (2007a) Domain-specific impairment of source memory following a right posterior medial temporal lobe lesion. *Hippocampus* 17:505-509.
- Peters J, Suchan B, Köster O, Daum I (2007b) Domain-specific retrieval of source information in the medial temporal lobe. *European Journal of Neuroscience* 26:1333-1343.
- Petersen SE, Fox PT, Snyder AZ, Raichle ME (1990) Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science* 249:1041-1044.
- Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4:e215.
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. *Nat Neurosci* 11:367-374.
- Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat Neurosci* 12:919-926.
- Phillips ML, Young AW, Senior C, Brammer M, Andrew C, Calder AJ, Bullmore ET, Perrett DI, Rowland D, Williams SC, Gray JA, David AS (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389:495-498.
- Pinsk MA, DeSimone K, Moore T, Gross CG, Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America* 102:6996-7001.
- Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *Journal of Neurophysiology* 101:2581-2600.
- Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage* 56:2356-2363.
- Platt ML, Ghazanfar AA (2010) Primate Neuroethology. New York: Oxford University Press.
- Pokorný JJ, de Waal FBM (2009) Monkeys recognize the faces of group mates in photographs. *Proceedings of the National Academy of Sciences* 106:21539-21543.
- Poppenk J, Moscovitch M (2011) A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. *Neuron* 72:931-937.
- Porter RH (1998) Olfaction and human kin recognition. *Genetica* 104:259-263.
- Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proceedings of the National Academy of Sciences* 106:947-951.
- Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol* 74:1192-1199.
- Quiñan Quiroga R, Kraskov A, Koch C, Fried I (2009) Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr Biol* 19:1308-1313.
- Quiroga RQ, Kreiman G, Koch C, Fried I (2008a) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87-91.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.
- Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I (2008b) Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci U S A* 105:3599-3604.
- Rämä P, Courtney SM (2005) Functional topography of working memory for face or voice identity. *NeuroImage* 24:224-234.

- Ramón y Cajal S (1899-1904) *Textura del sistema nervioso del hombre y de los vertebrados*. Madrid: Imprenta N. Moya.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91:2578-2589.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718-724.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111-114.
- Recanzone GH (2008) Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *J Neurosci* 28:13184-13193.
- Recanzone GH, Sutter ML (2008) The biological basis of audition. *Annu Rev Psychol* 59:119-142.
- Reddy L, Wilken P, Koch C (2004) Face-gender discrimination is possible in the near-absence of attention. *J Vis* 4:106-117.
- Reddy L, Reddy L, Koch C (2006) Face identification in the near-absence of focal attention. *Vision Res* 46:2336-2343.
- Remedios R, Logothetis NK, Kayser C (2009) An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *The Journal of Neuroscience* 29:1034-1045.
- Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour* 51:1007-1015.
- Rendall D, Kollias S, Ney C, Lloyd P (2005) Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *J Acoust Soc Am* 117:944-955.
- Richard A, Goldstein S, Dewar R (1989) Weed macaques: The evolutionary implications of macaque feeding ecology. *International Journal of Primatology* 10:569-594.
- Riesenhuber M, Poggio T Models of object recognition. *Nat Neurosci*.
- Riesenhuber M, Jarudi I, Gilad S, Sinha P (2004) Face processing in humans is compatible with a simple shape-based model of vision. *Proc Biol Sci* 271 Suppl 6:S448-450.
- Roberts WA, Mazmanian DS (1988) Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes* 14:247-260.
- Rockland KS, Van Hoesen GW (1999) Some temporal and parietal cortical connections converge in CA1 of the primate hippocampus. *Cereb Cortex* 9:232-237.
- Rogers TT, Patterson K (2007) Object categorization: reversals and explanations of the basic-level advantage. *J Exp Psychol Gen* 136:451-469.
- Roitblat HL, von Fersen L (1992) Comparative cognition: representations and processes in learning and memory. *Annual Review of Psychology* 43:671-710.
- Rolls E, Critchley H, Browning A, Inoue K (2006) Face-selective and auditory neurons in the primate orbitofrontal cortex. *Experimental Brain Research* 170:74-87.
- Rolls ET, Stringer SM (2006) Invariant visual object recognition: A model, with lighting invariance. *Journal of Physiology-Paris* 100:43-62.
- Rolls ET, Xiang J, Franco L (2005) Object, space, and object-space representations in the primate hippocampus. *J Neurophysiol* 94:833-844.
- Rolls ET, Miyashita Y, Cahusac PM, Kesner RP, Niki H, Feigenbaum JD, Bach L (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J Neurosci* 9:1835-1845.
- Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15-16.
- Romanski LM, Averbeck BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci* 32:315-346.
- Romanski LM, Diehl MM (2011) Neurons responsive to face-view in the primate ventrolateral prefrontal cortex. *Neuroscience* 189:223-235.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131-1136.

- Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2007) Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex* 17:1147-1153.
- Rossion B, Caldara R, Seghier M, Schuller AM, 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.
- Rotshtein P, Henson RNA, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* 8:107-113.
- Rousselet GA, Macé MJ, Fabre-Thorpe M (2004) Spatiotemporal analyses of the N170 for human faces, animal faces and objects in natural scenes. *Neuroreport* 15:2607-26011.
- Rousselet GA, Thorpe SJ, Fabre-Thorpe M (2004) How parallel is visual processing in the ventral pathway? *Trends Cogn Sci* 8:363-370.
- Russ BE, Lee YS, Cohen YE (2007) Neural and behavioral correlates of auditory categorization. *Hear Res* 229:204-212.
- Russ BE, Ackelson AL, Baker AE, Cohen YE (2008) Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol* 99:87-95.
- Rutishauser U, Schuman EM, Mamelak AN (2008) Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *Proc Natl Acad Sci U S A* 105:329-334.
- Rutishauser U, Ross IB, Mamelak AN, Schuman EM (2010) Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* 464:903-907.
- Sagiv N, Bentin S (2001) Structural encoding of human and schematic faces: holistic and part-based processes. *Journal of Cognitive Neuroscience* 13:937-951.
- Sakurai Y (1994) Involvement of auditory cortical and hippocampal neurons in auditory working memory and reference memory in the rat. *J Neurosci* 14:2606-2623.
- Sakurai Y (1996) Hippocampal and neocortical cell assemblies encode memory processes for different types of stimuli in the rat. *J Neurosci* 16:2809-2819.
- Sakurai Y (2002) Coding of auditory temporal and pitch information by hippocampal individual cells and cell assemblies in the rat. *Neuroscience* 115:1153-1163.
- Sander K, Scheich H (2001) Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research* 12:181-198.
- Saxe R, Carey S, Kanwisher N (2004) Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology* 55:87-124.
- Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci* 8:657-661.
- Schell A, Rieck K, Schell K, Hammerschmidt K, Fischer J (2011) Adult but not juvenile Barbary macaques spontaneously recognize group members from pictures. *Anim Cogn* 14:503-509.
- Schrier AM, Brady PM (1987) Categorization of natural stimuli by monkeys (*Macaca mulatta*): effects of stimulus set size and modification of exemplars. *J Exp Psychol Anim Behav Process* 13:136-143.
- Schwartz BL, Evans S (2001) Episodic memory in primates. *Am J Primatol* 55:71-85.
- Schwartz BL, Colon MR, Sanchez IC, Rodriguez IA, Evans S (2002) Single-trial learning of "what" and "who" information in a gorilla (*Gorilla gorilla gorilla*): implications for episodic memory. *Anim Cogn* 5:85-90.
- Schwartz BL, Meissner CA, Hoffman M, Evans S, Frazier LD (2004) Event memory and misinformation effects in a gorilla (*Gorilla gorilla gorilla*). *Anim Cogn* 7:93-100.
- Schwartz CE, Wright CI, Shin LM, Kagan J, Whalen PJ, McMullin KG, Rauch SL (2003) Differential amygdalar response to novel versus newly familiar neutral faces: a functional MRI probe developed for studying inhibited temperament. *Biol Psychiatry* 53:854-862.
- Schwarzlose RF, Baker CI, Kanwisher N (2005) Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience* 25:11055-11059.
- Schweinberger SR, Burton AM (2003) Covert recognition and the neural system for face processing. *Cortex* 39:9-30.
- Schweinberger SR, Herholz A, Stief V (1997a) Auditory long-term memory: Repetition priming of voice recognition. *The Quarterly Journal of Experimental Psychology A Human Experimental Psychology* 50A:498-517.

- Schweinberger SR, Herholz A, Sommer W (1997b) Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *J Speech Lang Hear Res* 40:453-463.
- Schweinberger SR, Robertson D, Kaufmann JrM (2007) Hearing facial identities. *The Quarterly Journal of Experimental Psychology* 60:1446-1456.
- Schweinberger SR, Zäske R, Walther C, Golle J, Kovács G, Wiese H (2010) Young without plastic surgery: Perceptual adaptation to the age of female and male faces. *Vision Research* 50:2570-2576.
- Schweinberger SR, Casper C, Hauthal N, Kaufmann JM, Kawahara H, Kloth N, Robertson DMC, Simpson AP, Zäske R (2008) Auditory adaptation in voice perception. *Current Biology* 18:684-688.
- Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry* 20:11-21.
- Sehatpour P, Molholm S, Schwartz TH, Mahoney JR, Mehta AD, Javitt DC, Stanton PK, Foxe JJ (2008) A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proceedings of the National Academy of Sciences* 105:4399-4404.
- Sergent J, Signoret JL (1992) Functional and anatomical decomposition of face processing: evidence from prosopagnosia and PET study of normal subjects. *Philos Trans R Soc Lond B Biol Sci* 335:55-61; discussion 61-52.
- Sergent J, Ohta S, MacDonald B (1992) Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* 115 Pt 1:15-36.
- Seyfarth RM, Cheney DL (2003a) Meaning and emotion in animal vocalizations. *Ann N Y Acad Sci* 1000:32-55.
- Seyfarth RM, Cheney DL (2003b) Signalers and receivers in animal communication. *Annu Rev Psychol* 54:145-173.
- Seyfarth RM, Cheney DL, Marler P (1980) Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210:801-803.
- Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001) The neural correlates of person familiarity. *Brain* 124:804-815.
- Shamma S (2008) On the emergence and awareness of auditory objects. *PLoS Biol* 6:e155.
- Sheffert SM, Olson E (2004) Audiovisual speech facilitates voice learning. *Percept Psychophys* 66:352-362.
- Silk JB (2007) Social components of fitness in primate groups. *Science* 317:1347-1351.
- Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004) Empathy for pain involves the affective but not sensory components of pain. *Science* 303:1157-1162.
- Sinha P (2002) Recognizing complex patterns. *Nat Neurosci*.
- Sirota A, Csicsvari J, Buhl D, Buzsaki G (2003) Communication between neocortex and hippocampus during sleep in rodents. *Proc Natl Acad Sci U S A* 100:2065-2069.
- Sirota A, Montgomery S, Fujisawa S, Isomura Y, Zugaro M, Buzsaki G (2008) Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron* 60:683-697.
- Sirotin YB, Das A (2009) Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* 457:475-479.
- Sliwa J, Duhamel JR, Pascalis O, Wirth S (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc Natl Acad Sci U S A* 108:1735-1740.
- Small SA, Schobel SA, Buxton RB, Witter MP, Barnes CA (2011) A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nat Rev Neurosci* 12:585-601.
- Small SA, Nava AS, Perera GM, DeLaPaz R, Mayeux R, Stern Y (2001) Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nat Neurosci* 4:442-449.
- Smiley JF, Falchier A (2009) Multisensory connections of monkey auditory cerebral cortex. *Hear Res* 258:37-46.
- Sperling R, Chua E, Cocchiarella A, Rand-Giovannetti E, Poldrack R, Schacter DL, Albert M (2003) Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage* 20:1400-1410.

- Squire LR, Zola-Morgan J, Clark RE (2007) Recognition memory and the medial temporal lobe: a new perspective. *Nat Rev Neurosci* 8:872-883.
- Staresina BP, Davachi L (2009) Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63:267-276.
- Stein BE, Burr D, Constantinidis C, Laurienti PJ, Alex Meredith M, Perrault TJ, Ramachandran R, Röder B, Rowland BA, Sathian K, Schroeder CE, Shams L, Stanford TR, Wallace MT, Yu L, Lewkowicz DJ (2010) Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience* 31:1713-1720.
- Steinvorth S, Wang C, Ulbert I, Schomer D, Halgren E (2010) Human entorhinal gamma and theta oscillations selective for remote autobiographical memory. *Hippocampus* 20:166-173.
- Sterck EHM, Watts DP, van Schaik CP (1997) The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology* 41:291-309.
- Stewart C-B, Disotell TR (1998) Primate evolution - in and out of Africa. *Current Biology* 8:R582-R588.
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869-873.
- Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138-11147.
- Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences* 105:394-398.
- Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, Ito K, Fukuda H (2001) Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *NeuroImage* 13:877-890.
- Suzuki WA, Amaral DG (2003) Where are the perirhinal and parahippocampal cortices? a historical overview of the nomenclature and boundaries applied to the primate medial temporal lobe. *Neuroscience* 120:893-906.
- Tamura R, Ono T, Fukuda M, Nakamura K (1992) Spatial responsiveness of monkey hippocampal neurons to various visual and auditory stimuli. *Hippocampus* 2:307-322.
- Tanaka JW (2001) The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen* 130:534-543.
- Tanaka K (2003) Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex* 13:90-99.
- Taylor MJ, Arsalidou M, Bayless SJ, Morris D, Evans JW, Barbeau EJ (2009) Neural correlates of personally familiar faces: parents, partner and own faces. *Hum Brain Mapp* 30:2008-2020.
- Thierry G, Martin CD, Downing P, Pegna AJ (2007) Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci* 10:505-511.
- Thompson CL, Pathak SD, Jeromin A, Ng LL, MacPherson CR, Mortrud MT, Cusick A, Riley ZL, Sunkin SM, Bernard A, Puchalski RB, Gage FH, Jones AR, Bajic VB, Hawrylycz MJ, Lein ES (2008) Genomic anatomy of the hippocampus. *Neuron* 60:1010-1021.
- Thompson P (1980) Margaret Thatcher: a new illusion. *Perception* 9:483-484.
- Thorpe SJ, Rolls ET, Maddison S (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res* 49:93-115.
- Thorpe SJ, Gegenfurtner KR, Fabre-Thorpe M, Bülthoff HH (2001) Detection of animals in natural images using far peripheral vision. *Eur J Neurosci* 14:869-876.
- Tian B, Rauschecker JP (2004) Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 92:2993-3013.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292:290-293.
- Tippett LJ, Miller LA, Farah MJ (2000) Prosopamnesia: a selective impairment in face learning. *Cognitive Neuropsychology* 17:241-255.
- Todorov A, Uleman JS (2002) Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. *Journal of Personality and Social Psychology* 83:1051-1065.
- Tomonaga M (1994) How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. *Primates* 35:155-165.

- Tranel D, Damasio AR (1985) Knowledge without awareness: an autonomic index of facial recognition by prosopagnosics. *Science* 228:1453-1454.
- Tranel D, Damasio AR (1988) Non-conscious face recognition in patients with face agnosia. *Behav Brain Res* 30:235-249.
- Tranel D, Damasio AR, Damasio H (1988) Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology* 38:690-696.
- Truppa V, Garofoli D, Castorina G, Piano Mortari E, Natale F, Visalberghi E (2010) Identity concept learning in matching-to-sample tasks by tufted capuchin monkeys (*Cebus apella*). *Animal Cognition* 13:835-848.
- Tsao DY, Livingstone MS (2008) Mechanisms of face perception. *Annual Review of Neuroscience* 31:411-437.
- Tsao DY, Moeller S, Freiwald WA (2008a) Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences* 105:19514-19519.
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670-674.
- Tsao DY, Schweers N, Moeller S, Freiwald WA (2008b) Patches of face-selective cortex in the macaque frontal lobe. *Nat Neurosci* 11:877-879.
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989-995.
- Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM (2007) Schemas and memory consolidation. *Science* 316:76-82.
- Tsukiura T, Cabeza R (2008) Orbitofrontal and hippocampal contributions to memory for face-name associations: the rewarding power of a smile. *Neuropsychologia* 46:2310-2319.
- Tulving E (2002) Episodic memory: from mind to brain. *Annual Review of Psychology* 53:1-25.
- Ungerleider LG (1995) Functional brain imaging studies of cortical mechanisms for memory. *Science* 270:769-775.
- Valentine T (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q J Exp Psychol A* 43:161-204.
- Van Lancker D, Kreiman J (1987) Voice discrimination and recognition are separate abilities. *Neuropsychologia* 25:829-834.
- Van Lancker DR, Canter GJ (1982) Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition* 1:185-195.
- Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH (1988) Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex* 24:195-209.
- Vanlancker DR, Kreiman J, Cummings J (1989) Voice perception deficits - Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology* 11:665-674.
- VanRullen R (2003) Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology-Paris* 97:365-377.
- Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci* 2:229-239.
- Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M (1997) Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277:376-380.
- Vatakis A, Ghazanfar AA, Spence C (2008) Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *J Vis* 8:14 11-11.
- Vauclair J, Fagot JI (1996) Categorization of alphanumeric characters by Guinea baboons: Within- and between-class stimulus discrimination. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 15:449-462.
- Viard A, Piolino P, Desgranges B, Chetelat G, Lebreton K, Landeau B, Young A, De La Sayette V, Eustache F (2007) Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: an fMRI study. *Cereb Cortex* 17:2453-2467.
- Viard A, Lebreton K, Chetelat G, Desgranges B, Landeau B, Young A, De La Sayette V, Eustache F, Piolino P (2010) Patterns of hippocampal-neocortical interactions in the retrieval of episodic autobiographical memories across the entire life-span of aged adults. *Hippocampus* 20:153-165.
- Viola Macchi C, Turati C, Simion F (2004) Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science* 15:379-383.

- Viskontas IV, Quiroga RQ, Fried I (2009) Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proc Natl Acad Sci U S A* 106:21329-21334.
- Viswanathan A, Freeman RD (2007) Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nat Neurosci* 10:1308-1312.
- Vizioli L, Rousselet GA, Caldara R (2010) Neural repetition suppression to identity is abolished by other-race faces. *Proceedings of the National Academy of Sciences USA* 107:20081-20086.
- von der Malsburg C (1999) The what and why of binding: the modeler's perspective. *Neuron* 24:95-104.
- von Kriegstein K, Giraud AL (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 22:948-955.
- von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. *PLoS Biol* 4:e326.
- von Kriegstein K, Kleinschmidt A, Giraud AL (2006a) Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cereb Cortex* 16:1314-1322.
- von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res* 17:48-55.
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL (2005) Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci* 17:367-376.
- von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD (2006b) Processing the acoustic effect of size in speech sounds. *Neuroimage* 32:368-375.
- von Kriegstein K, Smith DR, Patterson RD, Ives DT, Griffiths TD (2007) Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Curr Biol* 17:1123-1128.
- von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD (2010) How the human brain recognizes speech in the context of changing speakers. *J Neurosci* 30:629-638.
- von Kriegstein K, Dogan O, Gruter M, Giraud AL, Kell CA, Gruter T, Kleinschmidt A, Kiebel SJ (2008) Simulation of talking faces in the human brain improves auditory speech recognition. *Proc Natl Acad Sci U S A* 105:6747-6752.
- Vonk J, Macdonald SE (2002) Natural concepts in a juvenile gorilla (*Gorilla gorilla gorilla*) at three levels of abstraction. *Journal of the experimental analysis of behavior* 78:18.
- Vouloumanos A, Werker JF (2007) Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science* 10:159-164.
- Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001) Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci* 13:994-1005.
- Waite C, Buchanan-Smith HM (2006) Perceptual considerations in the use of colored photographic and video stimuli to study nonhuman primate behavior. *Am J Primatol* 68:1054-1067.
- Waite C, Gerald MS, Little AC, Kraiselburd E (2006) Selective attention toward female secondary sexual color in male rhesus macaques. *American Journal of Primatology* 68:738-744.
- Waite C, Little AC, Wolfensohn S, Honess P, Brown AP, Buchanan-Smith HM, Perrett DI (2003) Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proc Biol Sci* 270 Suppl 2:S144-146.
- Wallace P (1977) Individual discrimination of humans by odor. *Physiology & Behavior* 19:577-579.
- Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. *Progress in Neurobiology* 51:167-194.
- Wang X (2000) On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A* 97:11843-11849.
- Wang X (2007) Neural coding strategies in auditory cortex. *Hear Res* 229:81-93.
- Wang X, Kadia SC (2001) Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol* 86:2616-2620.
- Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74:2685-2706.
- Wanker R, Sugama Y, Prinage S (2005) Vocal labelling of family members in spectacled parrotlets, *Forpus conspicillatus*. *Animal Behaviour* 70:111-118.

- Warren JD, Jennings AR, Griffiths TD (2005) Analysis of the spectral envelope of sounds by the human brain. *NeuroImage* 24:1052-1057.
- Warren JD, Scott SK, Price CJ, Griffiths TD (2006) Human brain mechanisms for the early analysis of voices. *NeuroImage* 31:1389-1397.
- Wasserman E, Young M, Fagot JI (2001) Effects of number of items on the baboon's discrimination of same from different visual displays. *Animal Cognition* 4:163-170.
- Waydo S, Kraskov A, Quiñero Quiroga R, Fried I, Koch C (2006) Sparse representation in the human medial temporal lobe. *J Neurosci* 26:10232-10234.
- Webster MA, MacLin OH (1999) Figural aftereffects in the perception of faces. *Psychon Bull Rev* 6:647-653.
- Webster MA, Kaping D, Mizokami Y, Duhamel P (2004) Adaptation to natural facial categories. *Nature* 428:557-561.
- Weiskrantz L, Cowey A (1975) Cross-modal matching in the rhesus monkey using a single pair of stimuli. *Neuropsychologia* 13:257-261.
- Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49-63.
- Westhoff C, Troje NF (2007) Kinematic cues for person identification from biological motion. *Percept Psychophys* 69:241-253.
- Westmacott R, Leach L, Freedman M, Moscovitch M (2001) Different patterns of autobiographical memory loss in semantic dementia and medial temporal lobe amnesia: a challenge to consolidation theory. *Neurocase* 7:37-55.
- Westmacott R, Black SE, Freedman M, Moscovitch M (2004) The contribution of autobiographical significance to semantic memory: evidence from Alzheimer's disease, semantic dementia, and amnesia. *Neuropsychologia* 42:25-48.
- Widdig A, NÅ½rnberg P, Krawczak M, Streich WJr, Bercovitch FB (2001) Paternal relatedness and age proximity regulate social relationships among adult female rhesus macaques. *Proceedings of the National Academy of Sciences* 98:13769-13773.
- Wiebe SP, Staubli UV (1999) Dynamic filtering of recognition memory codes in the hippocampus. *J Neurosci* 19:10562-10574.
- Wiebe SP, Staubli UV (2001) Recognition memory correlates of hippocampal theta cells. *J Neurosci* 21:3955-3967.
- Wiener SI, Korshunov VA (1995) Place-independent behavioural correlates of hippocampal neurones in rats. *Neuroreport* 7:183-188.
- Wiener SI, Paul CA, Eichenbaum H (1989) Spatial and behavioral correlates of hippocampal neuronal activity. *J Neurosci* 9:2737-2763.
- Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. *Science* 308:873-876.
- Wiltgen BJ, Zhou M, Cai Y, Balaji J, Karlsson MG, Parivash SN, Li W, Silva AJ (2010) The hippocampus plays a selective role in the retrieval of detailed contextual memories. *Current Biology* 20:1336-1344.
- Winocur G, Moscovitch M, Sekeres M (2007) Memory consolidation or transformation: context manipulation and hippocampal representations of memory. *Nat Neurosci* 10:555-557.
- Winter L, Uleman JS, Cunniff C (1985) How automatic are social judgments? *J Pers Soc Psychol* 49:904-917.
- Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300:1578-1581.
- Wirth S, Avsar E, Chiu CC, Sharma V, Smith AC, Brown E, Suzuki WA (2009) Trial outcome and associative learning signals in the monkey hippocampus. *Neuron* 61:930-940.
- Wood ER, Dudchenko PA, Eichenbaum H (1999) The global record of memory in hippocampal neuronal activity. *Nature* 397:613-616.
- Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H (2000) Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* 27:623-633.
- Wright CI, Martis B, Schwartz CE, Shin LM, Fischer HH, McMullin K, Rauch SL (2003) Novelty responses and differential effects of order in the amygdala, substantia innominata, and inferior temporal cortex. *Neuroimage* 18:660-669.

- Xiang JZ, Brown MW (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* 37:657-676.
- Yamane S, Kaji S, Kawano K (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp Brain Res* 73:209-214.
- Yanike M, Wirth S, Suzuki WA (2004) Representation of well-learned information in the monkey hippocampus. *Neuron* 42:477-487.
- Yin RK (1969) Looking at upside-down faces. *Journal of Experimental Psychology* 81:141-145.
- Yonelinas AP, Kroll NEA, Quamme JR, Lazzara MM, Sauve M-J, Widaman KF, Knight RT (2002) Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat Neurosci* 5:1236-1241.
- Young AW, Hellawell D, Hay DC (1987) Configurational information in face perception. *Perception* 16:747-759.
- Young AW, Newcombe F, de Haan EH, Small M, Hay DC (1993) Face perception after brain injury. Selective impairments affecting identity and expression. *Brain* 116 (Pt 4):941-959.
- Young BJ, Fox GD, Eichenbaum H (1994) Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. *J Neurosci* 14:6553-6563.
- Young MP, Yamane S (1992) Sparse population coding of faces in the inferotemporal cortex. *Science* 256:1327-1331.
- Yovel G, Paller KA (2004) The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered. *Neuroimage* 21:789-800.
- Yukie M (2000) Connections between the medial temporal cortex and the CA1 subfield of the hippocampal formation in the japanese monkey (*Macaca fuscata*). *The Journal of Comparative Neurology* 423:282-298.
- Zaki J, Ochsner K (2009) The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences* 1167:16-30.
- Zäske R, Schweinberger SR (2011) You are only as old as you sound: auditory aftereffects in vocal age perception. *Hearing Research* In Press, Corrected Proof.
- Zäske R, Schweinberger SR, Kawahara H (2010) Voice aftereffects of adaptation to speaker identity. *Hearing Research* 268:38-45.
- Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. *The Journal of Neuroscience* 24:3637-3642.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846-849.
- Zeineh MM, Engel SA, Thompson PM, Bookheimer SY (2003) Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science* 299:577-580.
- Zhang Y-p, Shi L-m (1993) Phylogeny of rhesus monkeys (*Macaca mulatta*) as revealed by mitochondrial DNA restriction enzyme analysis. *International Journal of Primatology* 14:587-605.
- Zhong Y-M, Yukie M, Rockland KS (2005) Direct projections from CA1 to the superior temporal sulcus in the monkey, revealed by single axon analysis. *Brain Research* 1035:211-214.
- Zhong YM, Rockland KS (2004) Connections between the anterior inferotemporal cortex (area TE) and CA1 of the hippocampus in monkey. *Exp Brain Res* 155:311-319.

Travaux scientifiques

Publications à comité de lecture

Sliwa J, Duhamel J-R, Pascalis O, Wirth S (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. **Proceedings of the National Academy of Sciences of the United States of America** 108: 1735-1740

Piwkowska Z, Pospischil M, Brette R, Sliwa J, Rudolph-Lilith M, Bal T, Destexhe A (2008) Characterizing synaptic conductance fluctuations in cortical neurons and their influence on spike generation. **Journal of Neuroscience Methods** 169: 302-322.

Publications en preparation

Sliwa J, Duhamel J-R, Planté A, Wirth S. Comparing face-cells properties in the monkey hippocampus and TE.

Sliwa J, Duhamel J-R, Planté A, Wirth S. Dissection of micro-circuitry in monkey hippocampus: Linking cells' response properties to their electrophysiological features *in vivo*.

Présentations à des congrès

Sliwa J, Planté A, Duhamel J-R, Wirth S (2010) Representation of identity through voices and faces in macaque hippocampus. **Society for Neuroscience Abstracts**

Sliwa J, Duhamel J-R, Wirth S (2010) Reconnaissance cross-modale des individus : étude comportementale et bases neurophysiologiques. **Rencontres scientifiques de la DGA**, Paris

Sliwa J, Duhamel J-R, Pascalis O, Wirth S (2009) Cross-modal recognition of identity in rhesus monkeys for familiar monkeys and humans. **Society for Neuroscience Abstracts**

Sliwa J, Duhamel J-R, Ben Hamed S, Wirth S (2008) Neural correlates of memory reactivation in the monkey hippocampus. **The Plastic Brain**, Arcachon

Sliwa J, Ibos G, Duhamel J-R, Ben Hamed S (2008) Exemple de guidage par IRM anatomique des enregistrements électrophysiologiques chez le primate non-humain. **GRAMM 08**, Lyon

Sliwa J, Ben Hamed S, Duhamel J-R, Wirth S (2007) Codage de l'identité et de la familiarité visuelle dans l'hippocampe. **GDR Vision**, Lyon

Présentations à des séminaires

Champalimaud Centre for the Unknown, Lisbon, Portugal (11/2011, Z. Mainen laboratory)

Rockefeller University, NY, USA (06/2011, W. Freiwald laboratory)

Columbia University, NY, USA (06/2011, D. Salzman laboratory)

Princeton University, USA (06/2011, A. Ghazanfar laboratory)

Nathan Kline Institute, NY, USA (06/2011, C. Schroeder laboratory)

Rutgers University, NJ, USA (06/2011, G. Buzsaki laboratory)

CNRS, Centre de Neurosciences Cognitives, Lyon (01/2010)

Marine Biological Laboratories, Neuroinformatics Course, Woods Hole, MA, USA (08/2008)