



HAL
open science

**Les bases neurophysiologiques de la perception
audiovisuelle syllabique : étude simultanée en imagerie
par résonance magnétique fonctionnelle et en
électroencéphalographie (IRMf/EEG)**

Cyril Dubois

► **To cite this version:**

Cyril Dubois. Les bases neurophysiologiques de la perception audiovisuelle syllabique : étude simultanée en imagerie par résonance magnétique fonctionnelle et en électroencéphalographie (IRMf/EEG). Neurosciences [q-bio.NC]. Université de Strasbourg, 2009. Français. NNT : . tel-00462461

HAL Id: tel-00462461

<https://theses.hal.science/tel-00462461>

Submitted on 9 Mar 2010

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.



UFR des LETTRES

EA 1339 Linguistique, Langues et Parole - LILPA

Composante Parole et Cognition & Institut de Phonétique de Strasbourg (IPS)

THÈSE

présentée pour l'obtention du doctorat de
Sciences du Langage

par

Cyril Dubois

Spécialité : Phonétique Générale et Expérimentale,
Neurophonétique

**Les bases neurophysiologiques de la perception audiovisuelle syllabique :
Étude simultanée en Imagerie par Résonance Magnétique fonctionnelle
et en Électroencéphalographie (IRMf/EEG)**

Jury

Mme Marie-Noëlle Metz-Lutz

Chargée de Recherche INSERM,

Laboratoire Imagerie et Neurosciences Cognitives, CNRS, Université de Strasbourg

M. le Professeur Noël Nguyen

Laboratoire Parole et Langage & CNRS, Université de Provence

M. Jean-Luc Schwartz

Directeur de Recherche CNRS, INPG Grenoble, ICP/GIPSA-lab

M. Willy Serniclaes

Directeur de Recherche CNRS, Université Paris Descartes

M. le Professeur Rudolph Sock

IPS & EA 1339 LILPA, Composante Parole et Cognition, Université de Strasbourg

M. le Professeur Jean-Pierre Zerling

IPS & EA 1339 LILPA, Composante Parole et Cognition, Université de Strasbourg

Examinatrice

Rapporteur

Examineur

Rapporteur

Directeur de Thèse

Président

09/12/2009

UNIVERSITÉ DE STRABOURG
UFR des LETTRES
EA 1339 Linguistique, Langues et Parole - LiLPA
Composante Parole et Cognition & Institut de Phonétique de Strasbourg (IPS)

THÈSE

présentée pour l'obtention du doctorat de
Sciences du Langage
par

Cyril Dubois

Spécialité : Phonétique Générale et Expérimentale,
Neurophonétique

Les bases neurophysiologiques de la perception audiovisuelle syllabique : Étude simultanée en Imagerie par Résonance Magnétique fonctionnelle et en Électroencéphalographie (IRMf/EEG)

Jury

Mme Marie-Noëlle Metz-Lutz Chargée de Recherche INSERM, Laboratoire Imagerie et NeuroCognition, CNRS, Université de Strasbourg	Examinatrice
M. le Professeur Noël Nguyen Laboratoire Parole et Langage & CNRS, Université de Provence	Rapporteur
M. Jean-Luc Schwartz Directeur de Recherche CNRS, INPG Grenoble, ICP/GIPSA-lab	Examineur
M. Willy Serniclaes Directeur de Recherche CNRS, Université Paris Descartes	Rapporteur
M. le Professeur Rudolph Sock IPS & EA 1339 LiLPA, Composante Parole et Cognition, Université de Strasbourg	Directeur de Thèse
M. le Professeur Jean-Pierre Zerling IPS & EA 1339 LiLPA, Composante Parole et Cognition, Université de Strasbourg	Président

Remerciements

Une thèse, bien que présentée par un seul impétrant, est toujours le fruit de rencontres multiples et enrichissantes. Ce travail de recherche¹ m'a permis de grandir, tant sur le plan scientifique que sur le plan humain. Il est naturel de remercier les personnes qui ont apporté leurs contributions à sa réalisation.

Mon premier « contact » avec la phonétique s'est fait avec Monsieur le Professeur Jean-Pierre Zerling, venu présenter aux nouveaux étudiants les Sciences du Langage, et plus spécifiquement la phonétique. Ce n'est pas un hasard s'il préside le jury de cette thèse ; je lui adresse mes sincères remerciements pour son aide et ses multiples conseils au cours de toutes ces années.

Lors de mes premiers cours de phonétique et de phonologie, Monsieur le Professeur Rudolph Sock, devenu par la suite mon directeur de recherche, m'a émerveillé en me faisant découvrir la fine complexité de la parole. Je lui réitère mes remerciements pour la confiance qu'il me porte, son dévouement et ses conseils.

Je ne peux oublier feu André Bothorel, qui nous a communiqué sa passion, et une partie, hélas trop infime, de ses immenses connaissances.

François Wioland a, je pense, marqué plusieurs générations d'étudiants avec son enthousiasme et sa bonne humeur ; tout comme Gilbert Brock, toujours prompt à nous détendre avec humour, et à nous sortir de situations techniques inextricables.

Mme Péla Simon et Mme Béatrice Vaxelaire ont toujours accordé un grand intérêt à mes recherches, et je tiens à les remercier pour leur bienveillance à mon égard.

Mes études ont aussi donné lieu à des amitiés durables, Johanna et Mélanie m'ont toujours supporté et soutenu. Mélanie a, de plus, assumé la difficile tâche de relecture. Tout comme Marion, ma « collègue de bureau », qui a toujours su se rendre disponible.

Je salue l'Institut de Phonétique de Strasbourg dans son entier, Amélie, Fabrice(s), Fayssal, Véronique, Saïd, Kofi, sans oublier tout le personnel de notre UFR de tutelle, Vèrène Wodli, Claudine Jung, Chantal Weber, Maryam Hamzavi (pour ne citer qu'elles), sans lequel

¹Cette thèse a été réalisée dans le cadre d'une convention de collaboration de recherche (n° 05/901/737), passée entre la Composante Parole et Cognition de LILPA - E.A. 1339 / IPS et le Laboratoire d'Imagerie et Neurosciences Cognitives (LINC), CNRS FRE 3289, Faculté de Médecine de Strasbourg. Elle a bénéficié de financements de contrats attribués à la Composante Parole et Cognition et à l'IPS : Contrat du Ministère de la Recherche ACI TTT, Mesures et Données, 2003 – 2006 ; Contrat ANR (DOCVACIM) », 2008-2011 ; Contrat de recherche de la Maison Interuniversitaire des Sciences de l'Homme d'Alsace (MISHA) 2008 – 2012.

nous ne pourrions travailler. Je me dois aussi de rendre hommage à mes professeurs de linguistique, dont Astrid Schneider, Francine Gerhart-Krait, Claude Buridant, Thierry Revol, Martin-Dietrich Glessgen, Jean-Christophe Pellat, Catherine Schnedecker et Georges Kleiber.

J'ai une pensée pour toutes les « petites mains » de l'Université de Strasbourg, des services d'entretien à la comptabilité, en passant par l'informatique ou l'audiovisuel, *etc.* J'adresse aussi tous mes remerciements aux employés du CROUS de la cafétéria, en particulier Karine et Antoine, pour les innombrables fous rires partagés et, bien sûr, pour les cafés : sans vous, je n'aurais pas pu faire d'études !

Je tiens à remercier sincèrement tous les membres de l'Institut de Physique Biologique (IPB), en particulier Hélène Otzenberger pour son soutien indéfectible. Elle m'a initié à l'étude des potentiels évoqués et aux analyses statistiques, et s'est toujours efforcée de transmettre son savoir et son expérience en faisant preuve d'une application et d'un investissement inégalés. Daniel Gounot s'est évertué à m'enseigner les bases de l'IRM fonctionnelle ; il a aussi donné de son temps et de son énergie afin d'exploiter mes résultats expérimentaux. Une relectrice précieuse, Nathalie Heider qui, en plus de me soutenir dans mon projet, l'a rendu possible par ses talents. Corine Marrer a participé à la réalisation de l'expérience, et a toujours montré une vive curiosité pour ce travail. Christelle Nithart, ma camarade, devenue docteur n'a pas cessé de m'aider à tous les niveaux, durant les quatre années que nous avons partagées à l'IPB. Je ne dois pas oublier Mme Lamichel, sans qui, l'institut serait invivable.

Hélène Loevenbruck et Pascal Perrier font partie des rencontres décisives, je souhaite à tout étudiant de pouvoir un jour côtoyer des chercheurs comme eux.

Je me souviens aussi des rencontres faites à l'occasion de colloques, et que j'aimerais voir devenir pérennes, avec Rachid Ridouane, Cécile Fougeron, Marc Sato, Serge Pinto, Marion Dohen, Yves Laprie, Christophe Pallier.

J'adresse aussi mes remerciements à Madame Marie-Noëlle Metz-Lutz, Chargée de Recherches au CNRS, pour m'avoir initié aux neurosciences, pour avoir été un soutien sans faille, pour sa générosité, sa gentillesse et sa disponibilité. Sans elle, ce travail ne serait pas.

Je termine en remerciant Monsieur Willy Serniclaes, Directeur de Recherches au CNRS, et Monsieur le Professeur Noël Nguyen d'avoir accepté la difficile tâche de rapporteur, ainsi que Monsieur Jean-Luc Schwartz, Directeur de Recherches au CNRS, qui a bien voulu être membre du jury et examiner ce travail.

À un niveau plus intime, j'adresse une mention spéciale à mon camarade Olivier, non content de me supporter, il a participé à l'expérience, il a relu mon travail, et a toujours manifesté une vive curiosité pour celui-ci. Ce complice, des mauvaises heures comme des bonnes, m'a permis de tenir le coup et de décompresser.

Je ne dois pas oublier de remercier ma chère famille, mes parents, ma sœur, ma tante, et mes amis pour leurs encouragements, leur compréhension et leur patience.

Je ne puis finir ces remerciements qu'en évoquant ma chère et tendre. Mais pour elle, ni une ligne, ni un livre ne suffiraient à lui exprimer ma reconnaissance et lui témoigner mon amour.

Sommaire

Introduction	11
A. Partie théorique	17
A.1. De la phonétique à la neuro-imagerie : unités de bases et outils expérimentaux	19
A.2. Théories phonétiques de la perception de la parole et modèles d'accès au «lexique mental»	41
A.3. Les mécanismes cérébraux de la perception de la parole	53
A.4. Modèles cognitif et anatomo-fonctionnel de la perception de la parole	89
A.5. Conclusion et hypothèses	99
B. Partie expérimentale	109
B.1. Matériel et Méthode	111
B.2. Résultats : Données comportementales	139
B.3. Résultats : IRMf	161
B.4. Résultats : Potentiels évoqués	189
C. Discussion, Conclusion et Perspectives	205
C.1. Discussion :	207
C.2. Conclusion et Perspectives	215
Bibliographie	219
Index des auteurs	251
Index des notions	257
Index des tableaux	261
Index des figures	265
Tables des matières	269
Annexes	273

Liste des abréviations :

AB : Aire de Brodmann

AVa : Audiovisuelle animée

AVf : Audiovisuelle fixe

AMS : Aire Motrice Supplémentaire

BOLD : *Blood Oxygen Level-Dependent*

C : Consonne

CV : Consonne Voyelle

Cz : Vertex central

EEG : Électroencéphalographie

F1, F2, F3 : Premier, deuxième et troisième formants

Fz : Vertex frontal

GFI : Gyrus Frontal Inférieur

GTM : Gyrus Temporal Moyen

GTS : Gyrus Temporal Supérieur

CAP : Cortex Auditif Primaire (Gyrus de Heschl)

IRM : Imagerie par Résonance Magnétique

IRMf : Imagerie par Résonance Magnétique fonctionnelle

Lab : Paires s'opposant sur la labialité

LdA : Paires s'opposant sur les Lieux d'Articulation

Oz : Vertex occipital

PA : *Pulse Artefact* : Artefact cardiaque

PE : Potentiel évoqués

PM : *Premotor Cortex* : Cortex Prémoteur

Pz : Vertex pariétal

Spt : Région corticale située dans la scissure de Sylvius à la jonction pariéto-temporale

STI : Sillon Temporal Inférieur,

STS : Sillon Temporal Supérieur.

TEP : Tomographie par Émission de Positons

V : Voyelle

Vois : Paires s'opposant sur le Voisement

VOT : *Voice Onset Time*

Introduction

Cette thèse de phonétique générale et expérimentale, et plus précisément de neurophonétique est, par essence, interdisciplinaire tant par son sujet d'étude, les substrats neuraux de la *perception syllabique audiovisuelle*, que par ses outils, l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) couplée à l'Électroencéphalographie (EEG) et le recueil de données comportementales. À notre connaissance, il s'agit d'une première tentative de cerner la problématique linguistique et phonétique de la perception syllabique audiovisuelle à partir de ce système d'acquisition simultanée.

L'*objectif général* de cette étude est une contribution à la précision du recrutement de zones cérébrales dans les processus de perception de la parole audiovisuelle, et par là même dans la compréhension du langage articulé. La localisation des zones cérébrales impliquées dans une tâche expérimentale, de première importance pour les neurochirurgiens, est un moyen d'étudier les processus cognitifs engagés dans la perception et la compréhension de la parole. Pour le « neurocogniticien », la possibilité de recueillir des potentiels évoqués (PE), données extraites des signaux EEG, qui nous renseignent sur le déroulement temporel (le « *timing* ») de l'activité neuronale, en réponse à une stimulation, est tout aussi primordiale.

Parler est une faculté cognitive essentielle parmi toutes celles que possède l'espèce humaine ; les personnes qui souffrent de pathologies entraînant des déficits de la perception ou de la production de la parole mesurent quotidiennement l'importance de cette fonction vitale pour les interactions sociales. La facilité avec laquelle la parole est maîtrisée durant les premières années de vie est déconcertante, principalement lorsqu'elle est comparée à l'apprentissage de la lecture et de l'écriture. Au quotidien, l'exercice de la parole est plutôt aisé. Or derrière une apparente simplicité se cache évidemment une grande complexité. Celle-ci pourrait être résumée par les antonymes « variabilité » et « invariance ». Nous ignorons encore beaucoup des relations entre les signaux linguistiques perçus et le code langagier utilisé, afin de communiquer oralement.

De plus, de nombreux facteurs sont susceptibles d'influer sur la perception et la production de la parole, comme le bruit ambiant, la présence d'un auditoire, le stress ou encore la fatigue, *etc.* Ces exemples ont surtout une influence négative, mais d'autres pourront avoir un impact favorable, comme le fait de voir son interlocuteur articuler. L'idée selon laquelle la vision améliore la perception de la parole en milieu bruyant est désormais bien

acceptée par les spécialistes, mais demeure surprenante pour la plupart des locuteurs. Sumbly & Pollack (1954) avaient jadis démontré que la perception visuelle du visage du locuteur améliorait l'intelligibilité des mots en milieu bruité. Le célèbre effet « McGurk » a énormément contribué à la promotion de cette réalité, mettant en lumière le caractère irréprouvable de l'influence de la vision lors de la perception syllabique (McGurk & MacDonald, 1976). Il existe une troisième modalité opérant afin de compléter la perception de la parole : la proprioception. Il s'agit de la perception, consciente ou non, de la position relative des différentes parties du corps, les unes par rapport aux autres (Sherrington, 1906 ; Delmas, 1981). La modalité proprioceptive pourrait bien intervenir dans la relation étroite entre la production et la perception de la parole, comme semble l'indiquer, par exemple, les travaux sur les neurones miroirs (Gallese, Fadiga, Fogassi & Rizzolatti, 1996 ; Rizzolatti, Fadiga, Gallese & Fogassi, 1996).

Partant du postulat que la perception visuelle de la parole articulée exerce une influence sur la perception auditive, comment se traduit-elle au niveau du fonctionnement cérébral ? Est-ce qu'une zone corticale est spécialisée dans le traitement des indices visuels, ou alors sont-ce les zones dévolues à la perception auditive qui prennent en charge la dimension visuelle du langage articulé ? L'activité électrique du cerveau peut-elle refléter le traitement audiovisuel de la parole ?

La principale interrogation concerne le rôle des indices visuels linguistiques dans le traitement des « représentations ». Notre hypothèse liminaire et générale est que la dimension visuelle de la parole fait partie intégrante des « représentations » dites « phonologiques ». Nous aurons l'occasion d'y revenir, mais nous pouvons déjà signaler que le terme de « représentation » n'est pas sans conséquence sur la conceptualisation que nous pouvons avoir du système cognitif. D'un point de vue théorique, nous verrons que la plupart des théories phonétiques et des modèles d'accès au lexique se sont concentrés sur la modalité acoustique, traitant la dimension visuelle de la parole de façon annexe. D'un point de vue expérimental, nous avons réalisé une étude destinée à préciser l'influence des mouvements visuels linguistiquement pertinents sur le fonctionnement cérébral lié au traitement de la parole.

Ayant l'opportunité de collaborer avec l'Institut de Physique Biologique, qui dispose d'un système d'acquisition simultanée de données IRMf et EEG, nous avons pu réaliser une étude anatomo-fonctionnelle consistant à discriminer (lors de tâches de discrimination à choix forcé) des syllabes produites par une locutrice française. Notre étude met en jeu deux

modalités perceptives se différenciant par la présence ou l'absence de mouvements visuels linguistiquement pertinents. À l'aide du paradigme de la soustraction cognitive, nous souhaitons observer les zones cérébrales impliquées dans la perception visuelle de la parole. Dans la perspective d'affiner nos données, nous avons choisi de comparer des syllabes CV se différenciant en fonction d'un seul *trait distinctif*. Les tâches de discrimination présentent trois contrastes au sein des paires syllabiques : une opposition portant sur la *labialisation* vocalique (étirée vs arrondie [i y], ou selon Zerling (1992) non labialisée vs superlabialisée), et deux oppositions consonantiques portant soit sur le *voisement*, ou plus précisément sur l'un des indices de l'opposition de sonorité en français, le Délai d'Établissement du Voisement ou "*Voice Onset Time*" (VOT : sourdes vs sonores : [p b] et [t d]), soit sur les lieux d'articulation (extra vs intra-buccales : [p t] et [b d]). Nous avons retenu ces trois traits en fonction de leur apport visuel à la perception de la parole. Nous savons que l'indice de voisement est faiblement rentable sur le plan visuel, alors que les traits de labialité et de lieux d'articulation que nous avons choisis le sont davantage. Sur le plan auditif, en revanche, tous ces traits sont opératoires en français.

Parmi toutes les contraintes induites par ce type d'expérience, le scanner IRM produit un bruit important lors de son fonctionnement. Un scanner IRM est, pour simplifier, un électro-aimant qui génère deux champs magnétiques, le principal, intense et stable, et le secondaire, peu intense et bref. Le champ principal a pour but d'aimanter les noyaux de certains atomes, alors que le secondaire va faire varier la position dans l'espace (nommée « temps de relaxation ») de ces noyaux reprenant leur position initiale. C'est la commutation des gradients de l'IRM qui provoque la perturbation sonore, durant l'acquisition des images anatomo-fonctionnelles.¹ Cette variable imposée nous a incité à compléter les données comportementales acquises lors de la passation IRMf/EEG, par deux tests en dehors du scanner. Nous avons entrepris une comparaison entre un environnement exempt de toute perturbation et un environnement très bruyant, recréant le bruit des gradients de l'IRM.

Nous présentons par conséquent trois types de données (IRMf, EEG et comportementales) enregistrées au cours de trois sessions expérimentales distinctes. Les données acquises en IRMf et en EEG sont de natures différentes, outre l'aspect

¹ Un gradient est la différence de valeur physique de deux points dans un espace, pour nous : un champ magnétique. Les gradients permettent donc la différenciation de points suivant des valeurs connues dans l'espace en question. Les bobines de gradient sont des structures circulaires et parallèles, en les alimentant en courant électrique, elles vont émettre le champ magnétique secondaire, qui est perpendiculaire au champ principal.

méthodologique, l'avantage d'un recueil couplé réside dans la complémentarité des informations enregistrées.

Avant d'aborder les études sur le cerveau humain, nous revenons, au cours de notre premier chapitre, sur les notions phonétiques et sur les outils des neurosciences. Le deuxième chapitre présente les principales théories phonétiques et les modèles d'accès au lexique mental. Cette étape peut paraître, de prime abord, superflue, mais elle nous semble au contraire primordiale. La phonétique et la phonologie ont respectivement pour but l'étude des signaux linguistiques et leur fonctionnement dans le système langagier. Il existe une frontière avec le domaine de la psycholinguistique que nous nous efforcerons de dépasser, grâce à une prise en compte de divers modèles d'accès au lexique mental. Le troisième chapitre est le plus conséquent ; nous y examinons les données issues des différentes techniques qui ouvrent une fenêtre sur la neurocognition. Nous verrons que la position selon laquelle une zone corticale discrète puisse prendre en charge une fonction cognitive donnée est contestée par de nombreuses études. Les études anatomo-fonctionnelles sont nombreuses et disparates. Afin d'organiser notre réflexion, nous avons choisi de nous inscrire dans le modèle cognitif à deux voies de Hickok & Poeppel (2007), que nous abordons au cours du quatrième chapitre, principalement parce qu'il offre un cadre conceptuel relativement large. En effet, ce modèle permet la prise en compte de données pathologiques et anatomo-fonctionnelles, de certaines théories phonétiques et de modèles d'accès au lexique mental.

L'organisation de cette thèse

Le premier chapitre de notre partie théorique (A.1. De la phonétique à la neuro-imagerie : unités de bases et outils expérimentaux) est consacré aux notions éventuellement nécessaires à la bonne compréhension de notre travail de recherche, ainsi qu'aux outils et paradigmes empruntés aux neurosciences cognitives.

Le deuxième (A.2. Théories phonétiques de la perception de la parole et modèles d'accès au « lexique mental ») s'intéresse aux indices critiques pour la perception de la parole, et la manière dont ils seront utilisés pour accéder aux unités lexicales mémorisées.

Le troisième chapitre (A.3. Les mécanismes cérébraux de la perception de la parole) est constitué par notre revue de la littérature anatomo-fonctionnelle, et notre quatrième chapitre (A.4. Modèles cognitif et anatomo-fonctionnel) présente le cadre théorique que nous avons choisi comme référence.

Ces quatre premiers chapitres sont suivis d'encadrés résumant les notions les plus importantes qui y sont évoquées. Des rappels sommaires sont présentés en caractères italiques entre les sous-chapitres ; ils servent aussi de relance pour le sous-chapitre suivant.

C'est au cours du cinquième et dernier chapitre de cette partie théorique (A.5. Conclusion et hypothèses) que nous présentons une discussion, faisant office de conclusion. Cette discussion est suivie de nos hypothèses de départ.

Le premier chapitre de la partie expérimentale (B.1. Matériels et Méthodes) est consacré à la méthode que nous avons suivie et au matériel que nous avons utilisé, lors de notre phase expérimentale.

Le deuxième chapitre (B.2. Résultats Comportementaux) présente les données comportementales.

Le troisième chapitre (B.3. Résultats IRMf) expose les résultats anatomo-fonctionnels en IRMf, et notre quatrième chapitre (B.4. Résultats EEG) présente les potentiels évoqués.

La troisième et dernière partie de ce travail est constituée par la discussion générale de nos résultats. Elle est suivie de nos conclusions et de nos perspectives.

A. Partie théorique

A.1. De la phonétique à la neuro-imagerie : unités de bases et outils expérimentaux

La phonétique et la phonologie se sont employées à dégager diverses catégories, tentant ainsi de mettre au jour des dimensions stables du langage, face à une variabilité notoire présente dans les aspects physiques de la parole.

Nous avons choisi de commencer cette partie théorique en revenant sur certaines notions et certains concepts fondamentaux habituellement maniés dans le domaine des sciences de la parole avec comme objectif de résoudre cette tension entre variabilité des signaux de la parole et stabilité relative du domaine phonologique-linguistique. Nous nous excusons de ce chapitre qui pourra paraître « superflu » pour des spécialistes de la parole, mais le caractère transdisciplinaire de cette thèse nous a incité à ce détour didactique.

La nature de notre démarche expérimentale nous oblige aussi à exposer les bases des techniques utilisées dans l'étude du fonctionnement cérébral. Là encore, ces notions de base sont présentées afin d'éclairer notre propos pour les non spécialistes. Nous nous attardons en particulier sur les deux techniques utilisées lors de notre phase expérimentale.

A.1.1. Le phonème et ses traits

Le nombre de phonèmes que contient une langue est très variable, allant de moins d'une dizaine à plusieurs centaines. Le nombre de mots ou de phrases est potentiellement illimité. Un inventaire relativement restreint de phonèmes, entre trente et quarante unités, permet de créer une multitude de mots. S'il est possible de définir le morphème comme le plus petit élément porteur de sens, le phonème peut être qualifié d'élément sonore le plus bref permettant de distinguer différents mots (Calliope, 1988). Deux phonèmes s'opposent entre eux, s'il est possible d'identifier les traits pertinents qui les distinguent. Les facteurs influant sur les réalisations articulatoires et acoustiques d'un phonème sont multiples. Le sexe, l'âge, l'origine sociale, régionale sont autant d'exemples de facteurs de variabilité. Ce sont ce que (Troubetzkoy, 1939) nomme des « variantes libres ». L'interlocuteur identifie une variation notable par rapport à son idiolecte mais arrive à décoder le message linguistique, parfois au prix de quelques efforts. Il existe aussi des « variantes combinatoires » ou « allophones » qui sont les différentes réalisations d'un même phonème selon les contextes articulatoires. Les oppositions phonémiques élémentaires reposent sur le jugement perceptif des locuteurs de la

langue considérée. Par exemple, dans le cas de l'opposition entre [p t k] d'une part et [b d g] d'autre part, la présence ou l'absence de voisement (vibration des plis vocaux) est le trait pertinent qui nous permet d'opposer les deux classes. Mais un locuteur anglais et un locuteur français ne catégoriseront pas cette opposition de voisement de la même manière. De plus, toutes les langues n'utilisent pas les mêmes indices ; les tons ou les phénomènes de quantité ne sont pas présents universellement.

La notion de « faisceau de traits pertinents », présente dans de nombreux domaines scientifiques, est utilisée pour la description des phonèmes. L'examen de paires minimales permet d'assigner à chaque trait des invariants acoustiques et/ou articulatoires, ce qui ne signifie pas des valeurs constantes pour chaque phonème ou trait. Les commutations (« barre » [b a R] opposée à « gare » [g a R]) et les permutations (« rage » [R a ʒ] opposée à « jarre » [ʒ a R]) visent à dégager des paires minimales. Il s'agit d'établir quelles réalisations phonémiques permettent de créer des contrastes distinctifs. Les paires sont dites « minimales » lorsqu'un seul trait distingue deux unités lexicales (« pour » [p u R] vs. « cour » [k u R], qui se distinguent uniquement par le lieu d'articulation de l'occlusive initiale). Alors qu'une substitution d'un seul phonème ne sera pas toujours considérée comme une opposition minimale si elle fait varier plusieurs traits de façon concomitante (« sœur » [s œ R] vs. « beurre » [b œ R] qui varient sur le lieu d'articulation, sur le mode articulatoire et sur le voisement).

Les traits pertinents sont donc eux aussi des unités définies en termes d'oppositions distinctives du fait de la variabilité contextuelle. Les variations entre deux représentants de la même catégorie, par exemple les différences entre les deux [t] dans le mot « tête » [t ɛ t] ne sont généralement pas perçues, alors que l'opposition reposant sur le voisement entre [d] et [t], dans les mots « dette » et « tête », sera dans la grande majorité des cas correctement analysée. Les traits seront qualifiés de « distinctifs », « phonologiques », ou « phonétiques » selon le point de vue adopté. Ils permettent de réduire la diversité du signal acoustique à un nombre de catégories pertinentes pour la production et la perception du langage parlé. La notion de trait pertinent induit des caractéristiques communes sur les plans articulatoire, acoustique ou perceptif (Jakobson, Fant & Halle, 1952 ; Chomsky & Halle, 1968). Il n'est pas rare de faire l'expérience des variations d'un trait particulier. Par exemple, une consonne voisée peut facilement devenir non voisée selon le contexte. Dans l'exemple « robe fuchsia »

[Rɔ b f u ʃ j a], la consonne voisée [b] étant assourdie par la consonne sourde [f] sera souvent prononcée [Rɔ p f u ʃ j a].

Une distinction fondamentale est celle entre les consonnes et les voyelles. Les voyelles sont produites avec une ouverture du conduit vocal suffisamment large pour éviter la formation de bruit fricatif, tandis que les consonnes sont produites avec une fermeture complète (occlusive) ou partielle (fricative) de la cavité buccale. La théorie “*Frame then Content*” de MacNeilage & Davis (1998) pose la syllabe et le cycle ouverture - fermeture du conduit buccal dans les universaux du langage articulé. Les cycles mandibulaires influenceraient l’acquisition du système de production-perception de la parole. L’oscillation de la mâchoire inférieure serait le cadre naturel (“*a pure frame*”) sur lequel se superpose, par la suite, les mouvements antérieurs - postérieurs de la langue. Ce cycle d’ouverture(s) / fermeture(s) du conduit buccal engendre des configurations articulatoires (“*frame*”) permettant la phonation d’éléments vocaliques et consonantiques (“*content*”). Ceci met en lumière la place de la syllabe dans la parole, sujet qui aujourd’hui encore suscite de vifs débats.

Les traits et les phonèmes sont avant tout des unités distinctives. La question principale est de savoir s’ils peuvent constituer, à eux seuls, des « primitives » ; ou s’ils prennent place dans des unités plus grandes comme la syllabe.

A.1.2. La syllabe

La syllabe CV semble être présente dans toutes les langues étudiées, les autres types ne seraient que des déclinaisons de ce schéma naturel (Maddieson, 1984). Plusieurs travaux suggèrent que l’unité syllabique constitue une unité de reconnaissance de la parole (Mehler, 1981 ; Ferrand, Grainger & Segui, 1995 ; Pallier, 1994 ; Content, Kearns & Frauenfelder, 2001 ; Dumay, Frauenfelder & Content, 2002). La syllabe est souvent décrite comme l’unité intermédiaire entre le morphème et le phonème. Elle peut être constituée de deux parties : l’attaque et la rime, cette dernière se subdivisant en un noyau et une coda. De nombreuses descriptions tripartites se rencontrent, comprenant une attaque, un noyau et une coda. La figure A1 permet de visualiser ces subdivisions :

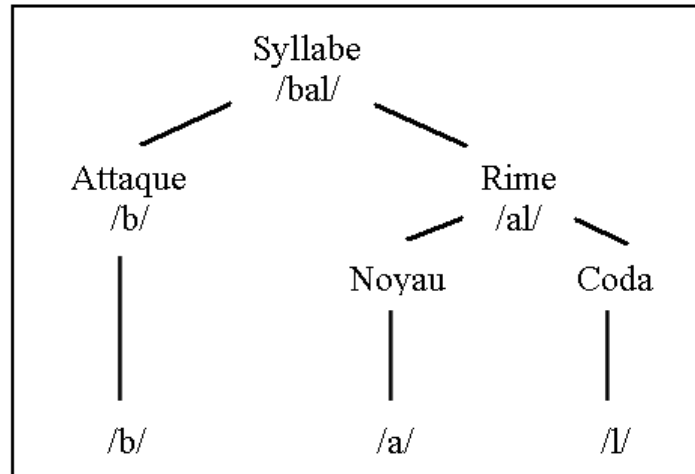


Figure A1 : Représentation schématique de la structure syllabique.

L'attaque est la partie initiale de la syllabe ; elle est composée d'une ou plusieurs consonnes. Le noyau est la partie centrale et essentielle de la syllabe ; il est le plus souvent constitué d'un élément vocoïde. La coda est la partie finale et elle peut être formée d'une ou plusieurs consonnes. L'attaque et la coda ne sont pas essentielles à la syllabe (« o » [O] dans « orange » [ORɑ̃ʒ]). En français, une syllabe peut constituer, à elle seule, un mot d'un seul morphème (« chaise » [ʃɛz]) ou encore, dans sa plus simple expression, elle peut être formée d'une seule voyelle (« à » dans la séquence « à deux pas d'ici » [adøpadisi]). Un locuteur francophone a une intuition forte de la syllabe, comme pour le mot ou la phrase. Il est relativement simple de se prononcer sur le nombre de syllabes constituant un mot et même de proposer une segmentation (le mot « caducée » [kadysɛ] est composé de trois syllabes « ca », « du » et « cée »). L'opération consistant à compter les syllabes d'un mot nous fait utiliser ce mot en mention, c'est-à-dire hors contexte. Dans l'usage, la chaîne parlée peut complexifier sensiblement le dénombrement des syllabes. Pour des mots monosyllabiques français, il existe d'importantes variations. Un mot comme « strict » [strikt] du type CCCVCC diffère fortement d'un monosyllabe (*cf. supra*) composé d'une seule voyelle. En examinant une séquence comme « une stricte attention » [ynstriktatɑ̃sjɑ̃], il est possible de constater que la consonne codaïque « t » de « strict » peut devenir l'attaque du mot suivant « attention ». Le vocabulaire diffère selon les auteurs, mais le terme de « resyllabation » est souvent considéré comme le plus approprié pour décrire ce phénomène.

En matière prosodique, c'est-à-dire au niveau suprasegmental, le français est décrit comme étant une langue à oxytonie stricte. C'est-à-dire que les phénomènes d'accentuation et d'intonation se réalisent sur la dernière syllabe de chaque groupe rythmique. Des modèles plus récents plaident en faveur de l'émergence d'un second accent fixe à l'initiale du groupe rythmique, reposant lui aussi sur la syllabe (Jun & Fougeron, 2000 ; Jun & Fougeron, 2002). Les langues à tons exploiteraient de la même manière le niveau syllabique (Martinet, 1967).

Si la définition de la syllabe demeure épineuse, son existence n'est plus guère contestée. Là encore, la principale question tient à son rôle dans les processus langagiers. Après avoir longtemps été considérée comme un phénomène essentiellement auditif, la parole est désormais acceptée comme étant multimodale. Nous allons aborder la notion de « visème » qui est la pierre de touche de notre démarche expérimentale.

A.1.3. La perception visuelle de la parole et les visèmes

L'effet découvert par McGurk & MacDonald (1976) est l'un des exemples les plus stupéfiants de l'influence d'une deuxième modalité perceptive dans la perception de la parole : la vision. McGurk & MacDonald (1976) ont utilisé des signaux auditif et visuel non congruents. Un signal acoustique de syllabes redoublées [b a b a] a été monté sur la vidéo d'un visage d'homme articulant un [g a g a]. Les sujets ont perçu [d a d a]. Lorsque la syllabe [g a] est présentée auditivement doublée avec la vidéo d'une syllabe [b a], les sujets perçoivent des percepts [b d a] ou [b g a]. Ils ont réitéré l'expérience avec les variantes sourdes des consonnes précédentes et ont trouvé des résultats concordants même si l'effet est moins prononcé en l'absence du voisement. L'effet « McGurk » se produit aussi lors de l'introduction entre les modalités visuelle et auditive d'un décalage spatial (Jun & Fougeron, 2002) ou temporel (Van Wassenhove, Grant & Poeppel, 2007). De plus, Sekiyama, Kanno, Miura & Sugita (2003) montrent que l'effet McGurk est plus important quand l'intelligibilité est dégradée (faible rapport signal sur bruit), les sujets éprouvant plus de difficultés pour percevoir la composante auditive des signaux. L'effet McGurk est un phénomène relativement robuste, il a été testé sur des mots (Dekle, Fowler & Funnel, 1992) et des phrases (McGurk, 1981), ainsi qu'en faisant varier les visages et les configurations buccales (Hietanen, Manninen, Sams & Surrakala, 2001). Rosenblum, Schmuckler & Johnson (1997) montrent que les enfants y sont sensibles, l'effet se renforçant avec le temps.

Colin & Radeau (2003) publient une revue des principales études sur l'effet McGurk et cherchent à établir quels facteurs peuvent le moduler. Ce sont principalement la saillance perceptive et le degré d'informativité qui sont susceptibles d'influer sur cet effet. Tiippana, Andersen & Sams (2004) pensent que l'attention est cruciale, la présence de distracteurs influençant la force de l'effet, alors que pour Soto-Faraco, Navarra & Alsius (2004) ce processus est incoercible.

Jusqu'alors, l'étude de Sumbly & Pollack (1954) sur l'apport des indices visuels était la plus importante découverte sur la perception audiovisuelle. Cette étude semble démontrer que la perception du visage du locuteur facilite la discrimination de la parole en milieu bruité ; le gain progressant en fonction de l'augmentation du rapport signal sur bruit. Notons que Ross, Saint-Amour, Leavitt, Javitt & Foxe (2006) n'observent pas une progression linéaire, le gain étant maximal pour un rapport signal sur bruit de -12 dB. La perception de phonèmes et de syllabes, appartenant à une même classe de visèmes, semble aussi améliorée par la présence d'indices visuels phonologiques (Schwartz, Berthommier & Savariaux, 2004), tout comme la perception des accents lexicaux (Scarborough *et al.*, 2006). Ces résultats, déjà largement débattus, sont très intéressants pour notre sujet d'étude. Ils montrent que la perception visuelle a un impact favorable, non seulement, sur l'intelligibilité, mais aussi sur la perception prélexicale. Néanmoins, l'effet McGurk nous indique que la perception des indices visuels peut aussi perturber la perception auditive.

Summerfield (1992) introduit une distinction entre les termes anglais "*lipreading*", "*audiovisual speech perception*" et "*speechreading*" qui nous semble particulièrement importante dans le cadre de ce travail. Le premier de ces termes "*lipreading*" (traduit par « lecture labiale ») recouvre la perception uniquement visuelle des mouvements articulatoires orofaciaux parmi lesquels ceux des lèvres, mais également des joues ou de la mâchoire. Le deuxième terme "*audiovisual speech perception*" (perception audiovisuelle de la parole) désigne la perception de la modalité auditive en combinaison avec la lecture labiale. Le dernier terme "*speechreading*" (qui n'a pas d'équivalent en français) décrit une perception globale de l'acte langagier comprenant d'une part la lecture labiale mais également la perception d'indices généraux tels que la position du corps, des mains et inclut éventuellement l'audition.

Les indices visuels susceptibles de contribuer à la perception de la parole ont été regroupés à l'instar des phonèmes en fonction de leur distinctivité.

Fisher (1968) introduit le terme de “*viseme*”, construit sur “*visual phoneme*”, afin de désigner la contrepartie visuelle des phonèmes. Les visèmes sont donc des unités visuelles minimales dont les traits distinctifs permettent de les discriminer les unes des autres. Généralement ce terme s’oppose à “*homophene*” (Berger, 1972) ou à « sosie labial », qui désignent des unités visuelles dont les caractéristiques communes permettent de les regrouper. Comme le rappelle Cathiard (1994) :

« [...] le visème est une unité avant tout oppositive comme le phonème ; le sosie ou homophène, une unité définie par le contenu d’une classe ».

Les distinctions les mieux perçues sont les traits directement visibles, tels la labialité ou le degré d’aperture ; alors que les articulations intra-buccales, ou encore les traits de nasalité et de sonorité sont difficilement perceptibles. Mourand-Dornier (1980) distingue quatre classes de visèmes pour des consonnes situées à l’initiale, soit :

[p b m] [f v] [ʃ ʒ] [s z t d n k g ɲ ʀ]

La reconnaissance visuelle de consonnes, placées à l’initiale et en finale, a été étudiée par Gentil (1981). Quelle que soit la position, les bilabiales [p b m], les labiodentales [f v] et les post-alvéolaires protruses [ʃ ʒ] forment trois groupes de visèmes. En position initiale, les autres consonnes forment un groupe unique ; alors qu’en position finale, il est possible de distinguer quatre groupes : [s z] [t d n] [k g ɲ] et [ʀ]. Notons que [s z] sont souvent confondus avec [t d n].

En ce qui concerne les voyelles, il est difficile de constituer des groupes de visèmes comme pour les consonnes. Heider & Heider (1940) remarquent :

“there are no two vowels which are homophenous in the same way as [p] and [m] are homophenous.”

« Il n’y a pas deux voyelles qui soient homophènes comme [p] et [m] le sont. »

Cela s’explique par le fait que les voyelles sont fortement influencées par le contexte consonantique. Néanmoins, certains résultats sont intéressants pour notre étude. Cathiard, Tiberghien & Abry (1992) obtiennent, pour un choix simple [y i], 77 et 81 % d’identification en contexte consonantique difficile, à savoir [s] et [ʃ]. Benoît, Mohamadi & Kandel (1995) ont obtenu des taux d’identification corrects de la voyelle [y] de 94 %, lorsque ce phonème n’est comparé qu’à [i] et [a], identifiées respectivement à 85 % et 77 %. Dans une étude portant sur la modalité visuelle, Lisker & Rossi (1992) ont demandé à 18 auditeurs si les voyelles proposées étaient arrondies ou non. Les voyelles du français [o] [u]

[ĩ] [ø] ont été identifiées à plus de 75 % comme arrondies, tandis que [i] [e] [ɛ] [a] [ẽ] ainsi que la semi-voyelle [w] ont été identifiées à 75 % comme non arrondies. Demeurent les voyelles [ã] [œ] [ɑ] [œ] [O], pour lesquelles les taux de reconnaissance sont inférieurs à 75 %.

Ces quelques études permettent de conclure que le nombre de visèmes pouvant être regroupés sous une même classe peuvent différer selon les conditions dans lesquelles ils sont perçus. Alors que les visèmes consonantiques regroupant les sons labiaux ou labialisés sont plus facilement discriminables, les autres sont plus aisément confondus. La labialité semble aussi être le premier facteur d'identification des visèmes vocaliques bien que celle-ci défavorise la reconnaissance des segments environnants. De nombreuses études (Cathiard, 1994 ; Vaxelaire, Sock & Hecker, 1999 ; Hecker, Vaxelaire, Sock & Cathiard, 1999 ; Roy, 2005 ; Troille & Cathiard, 2006) révèlent le lien entre la perception auditive et l'anticipation de la protrusion. Dans un mot comme « structure » ([s t R y k t y R]), le premier [y] est en quatrième position et précédé de trois consonnes. Les études cinématiques, utilisant le paradigme du dévoilement progressif, nous apprennent que la labialisation débute dès le premier phonème consonantique [s] et que les sujets exposés à la parole audiovisuelle sont capables de détecter, de façon significative, la présence ou l'absence de protrusion par anticipation.

Le visème est une unité moins spécifiée que le phonème, en ce sens que sa distinctivité est moindre. Il est envisageable que les visèmes soient traités comme des informations subsidiaires, permettant d'améliorer les processus de catégorisation.

Après ce retour sans doute accessoire sur les notions phonétiques, nous abordons maintenant les différentes sources de données exploitées dans les études neurocognitives. Nous allons nous attarder en particulier sur les techniques de l'IRMf et de l'EEG, que nous avons utilisées lors de notre expérience.

A.1.4. Les méthodes et techniques d'investigation du fonctionnement cérébral

Durant près d'un siècle et demi, les informations collectées sur l'anatomie et sur la physiologie du cerveau étaient consécutives de lésions cérébrales induisant des déficits cognitifs et / ou fonctionnels. Il s'agissait donc d'étudier des cas pathologiques, et le travail d'autopsie permettait de localiser avec plus ou moins de précision l'étendue de la ou des

lésions. L'émergence de techniques pouvant observer le fonctionnement *in vivo* du sujet sain a sensiblement fait progresser nos connaissances dans le domaine cognitif. L'imagerie cérébrale a suivi deux grandes voies : l'observation de l'activité électromagnétique et celle des flux sanguins. Entre temps, les expériences en psychologie cognitive ont utilisé des temps de réaction ou de réponse afin d'accéder au fonctionnement cérébral.

Nous exposons, dans ce qui suit, différentes techniques d'investigation dans le domaine des neurosciences cognitives.

A.1.4.1. La méthode anatomoclinique

Laennec (1781 - 1826) définit la méthode anatomoclinique comme une étude des états pathologiques reposant sur l'analyse de l'observation des symptômes ou des déficits de fonctions qui coïncident avec chaque espèce d'altération(s) d'organe(s). Morgagni (1682 - 1771), professeur de médecine à Padoue et anatomiste, peut être considéré comme le premier anatomo-pathologiste moderne. Il entreprit d'établir une relation de cause à effet entre les lésions constatées chez le cadavre et la sémiologie clinique. Un médecin français, Marc Dax (1770 - 1837), dont les travaux ont été prolongés par son fils Gustave (1815 - 1893), avait observé lors des guerres napoléoniennes, la relation existant entre la localisation des lésions corticales et les répercussions consécutives. Cependant ces constatations empiriques, aussi intéressantes soient-elles, ne signifient pas forcément qu'une aire corticale unique soit responsable d'une faculté déterminée dans une relation linéaire. En revanche, cela peut signifier que la seule lésion d'une aire corticale faisant partie d'un réseau de traitement suffit à le rendre inopérant.

A.1.4.2. La chronométrie mentale : les temps de réaction ou de réponse

La technique de chronométrie mentale vise à étudier les processus de traitement de l'information, grâce à la mesure des temps de réaction ou des temps de réponse, ainsi que les performances de réussite lors d'une tâche expérimentale. Les principales questions soulevées concernent d'une part l'existence de différents niveaux de traitements, et d'autre part le caractère discret ou continu de ces traitements. La méthode soustractive de Donders (1868) permet de réaliser des mesures indirectes des processus cognitifs et du fonctionnement du cerveau. Cette méthode consiste à proposer différentes tâches expérimentales afin de les comparer et de pouvoir évaluer la différence des temps de réaction. En proposant deux tâches

qui diffèrent en fonction du processus étudié, il est postulé que la soustraction des temps de réaction correspond à la durée du traitement. Deux problèmes principaux existent, le premier concerne le fait de pouvoir « isoler » un processus grâce à la méthode de soustraction, et le deuxième est de savoir si l'ajout d'un facteur supplémentaire ne change pas la stratégie globale de traitement.

Le principe des facteurs additifs de Steinberg (1969) stipule que si deux facteurs affectent des étapes différentes du traitement de l'information, alors leurs effets sont additifs. Si, au contraire, deux facteurs affectent au moins une étape du traitement de l'information en commun, alors une interaction de leurs effets sera attendue. Deux tâches pourront être réalisées simultanément, à la condition qu'elles reposent sur des sous-systèmes fonctionnels et / ou cérébraux distincts et indépendants. En revanche, si les deux tâches font appel au même système, alors une interférence est attendue et doit se traduire par une baisse des performances.

A.1.4.3. La méthode psychophysique : les potentiels évoqués

Le cerveau humain recèle une activité biochimique et électrique spontanée qui engendre des fluctuations de différences de potentiel entre deux points du scalp. L'enregistrement de ces fluctuations constitue l'électroencéphalogramme.

Les premières applications de l'EEG ont été médicales en permettant d'identifier des déficits perceptifs chez des patients avec une procédure rapide et sans danger. Pendant longtemps, l'analyse de l'EEG n'a concerné que les propriétés temporelles des signaux, l'analyse spatiale se réduisant à l'établissement de cartes de potentiel sur le scalp obtenues à partir d'interpolations des mesures. L'information spatiale de l'EEG n'a été utilisée pour tenter de localiser les sources primaires que très tardivement (Kavanagh, Darcey & Lehmann, 1978). En faisant la somme des tracés issus de la répétition de la même tâche, il est possible de faire apparaître des composantes spécifiques, appelées réponses ou potentiels évoqués (Dawson, 1951).

A.1.4.3.1. Courants cellulaires et potentiels de surface

Les potentiels mesurés à la surface du scalp par l'électroencéphalographie sont le reflet des courants ioniques membranaires des neurones actifs. Ces courants transmembranaires provoquent la circulation d'un potentiel d'action le long du neurone (voir figure A2). À l'échelle de la cellule, la propagation du potentiel d'action depuis le corps cellulaire le long de l'axone donne lieu à un dipôle électrique, dont les caractéristiques dépendent de la configuration de la synapse et de la cellule. Ce dipôle électrique crée autour de lui un petit potentiel électrique. Si le nombre de cellules excitées est suffisant et si elles sont orientées parallèlement (cellules pyramidales) les potentiels s'ajoutent et sont détectables sur le scalp. C'est, en fait, la sommation temporelle et spatiale des potentiels post-synaptiques des cellules pyramidales du cortex cérébral, qui donne naissance à l'EEG et aux potentiels évoqués. À l'échelle du cerveau entier, un petit volume du cerveau comportant des neurones actifs est considéré comme une région contenant une distribution complexe de dipôles mobiles. Plus on s'éloigne de ce volume, plus le potentiel enregistré ressemble à celui créé par un dipôle unique. L'activité de ce petit volume est schématisée par un dipôle électrique équivalent, considéré comme la résultante de tous les dipôles décrits à l'échelle cellulaire. Lorsque plusieurs groupes de cellules nerveuses sont activés au même moment, le potentiel recueilli sur une électrode sera la somme vectorielle des potentiels créés par chaque groupe de cellules. Les activités électriques qui contribuent aux différences de potentiels enregistrées en surface sont exclusivement les potentiels électriques liés à la transmission synaptique, actifs en groupe qu'ils soient orientés tangentiellement ou radialement par rapport à la surface du scalp. Ils ne distinguent pas les activités excitatrices et inhibitrices et intègrent les variations de potentiels, conjuguées de plusieurs sources sans les différencier.

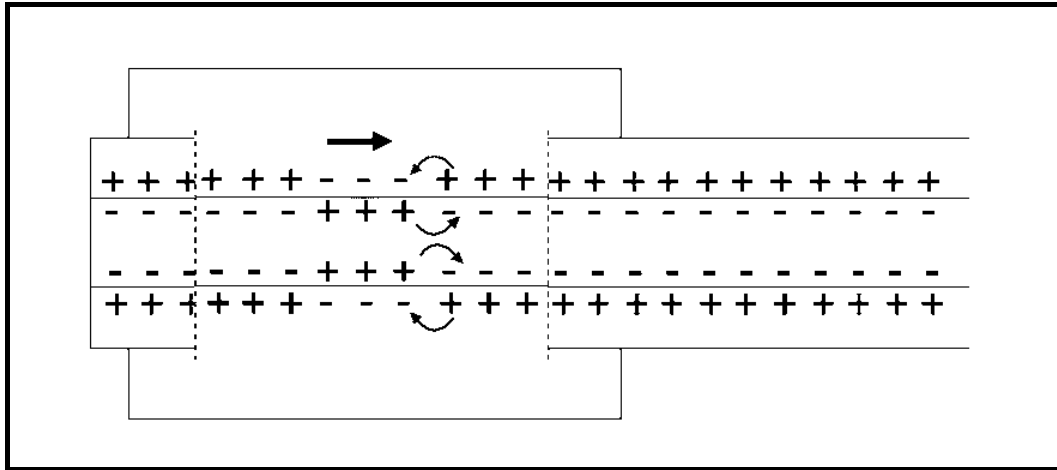


Figure A2 : Représentation schématique de la propagation d'un potentiel d'action. La flèche horizontale représente le dipôle électrique.

A.1.4.3.2. Les potentiels évoqués

Les potentiels évoqués sont des réponses des voies nerveuses et du cerveau qui se produisent soit spontanément soit à la suite d'une stimulation. Ces réponses sont spécifiques à la nature du stimulus employé pour les obtenir. Elles ont de plus la particularité d'être de très faible intensité (quelques microvolts) et sont superposées à l'activité cérébrale spontanée, beaucoup plus importante en amplitude. Elles doivent être extraites par une technique de moyennage impliquant que l'événement déclenchant le processus soit répété plusieurs fois (plusieurs dizaines voire plusieurs centaines dans l'idéal). Cela implique donc l'utilisation d'un appareillage spécialisé pour faire ressortir ces réponses consécutives (évoquées) au stimulus parmi l'activité électrique spontanée, non désirée ici. Le nombre d'électrodes utilisées varient selon les objectifs que se fixe l'étude. Quatre électrodes peuvent suffire lorsque l'intérêt n'est porté que sur l'aspect temporel des variations de l'activité électrique d'une région donnée. Pour accroître la définition spatiale des sources de courant et utiliser les possibilités offertes par la cartographie, les enregistrements se font sur 16, 32, 64, 128 voire 256 électrodes. Les électrodes sont en général placées selon une disposition normalisée appelée « système 10-20 » (voir figure A3). Ce système fait référence à l'espacement des électrodes variant de 10 ou de 20 % en fonction de la morphologie de chaque individu. Chaque électrode est identifiée par un label (*American Electroencephalographic Society*, 1991) qui indique sa position sur le scalp par une lettre (F = frontal, C = central, T = temporal, P = pariétal, O = occipital) et un chiffre pair pour les localisations droites et impair pour les

gauches. Lorsque les électrodes sont situées sur le vertex (point le plus élevé de la voûte crânienne), ce chiffre est remplacé par la lettre z (Fz, Cz, Pz, Oz).

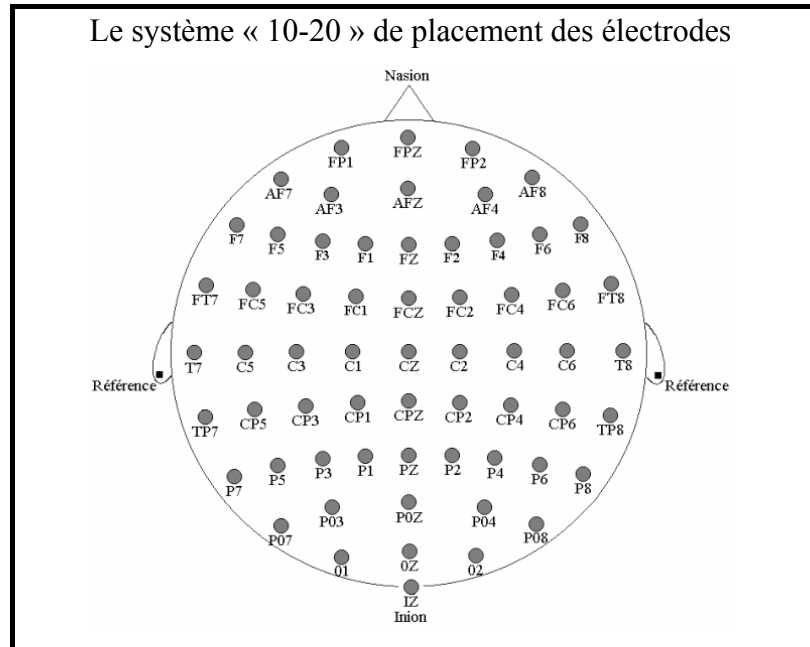


Figure A3 : Représentation schématique de l'emplacement des électrodes sur le scalp dans le système « 10-20 ».

En raison de l'amplitude très faible du potentiel lié à un événement par rapport à l'activité spontanée du cerveau, il est nécessaire d'enregistrer de nombreuses réponses évoquées par le même événement. Ainsi, des événements sont répétés à plusieurs reprises : une stimulation sonore, visuelle ou sensitive, une opération mentale ou une action, et la somme des variations d'activité électrique, synchronisées à cet événement, est effectuée. Les signaux électriques liés à l'apparition d'un événement sont extraits des signaux interférents, telles que les ondes cérébrales de l'EEG spontané, par des techniques informatiques de traitement du signal. Au cours du moyennage, les réponses évoquées restent constantes alors que les bruits diminuent avec la répétition des événements pertinents (voir figure A4). La méthode du moyennage repose sur trois présupposés. Premièrement, le signal et le bruit s'additionnent de manière linéaire pour donner l'onde enregistrée. Deuxièmement, la forme du signal est identique à chaque répétition de l'événement. Troisièmement, les ondes du bruit sont suffisamment irrégulières d'un événement à un autre de sorte qu'elles peuvent être considérées comme étant des éléments statistiquement indépendants survenant au hasard.

Le nombre de répétitions de l'événement peut varier de plusieurs dizaines à plusieurs centaines, selon que l'on s'intéresse à des réponses électriques se déroulant sur une période de temps plus ou moins longue après la présentation du stimulus.

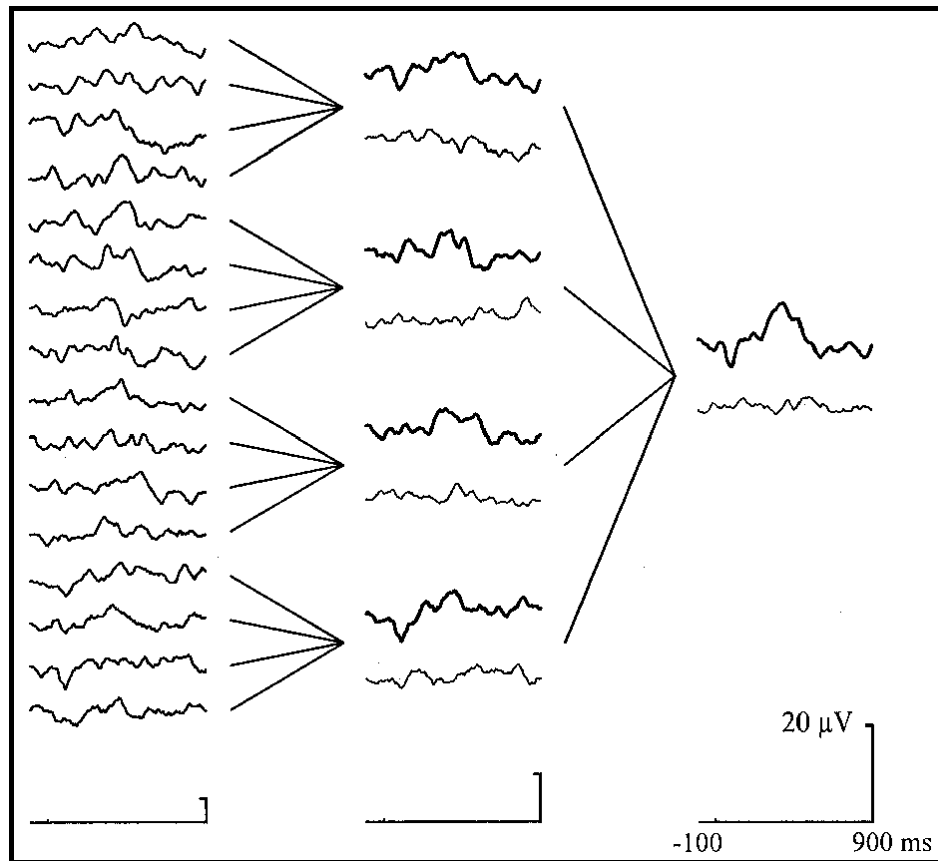


Figure A4 : Extraction des potentiels évoqués par la technique du moyennage.

Selon la nature de la stimulation employée, selon la consigne donnée au sujet et selon les opérations mentales engagées par celui-ci, les potentiels évoqués présentent une alternance d'ondes négatives et positives plus ou moins sensibles aux conditions expérimentales (voir figure A5). Les ondes caractéristiques des potentiels évoqués sont nommées selon leur polarité, une composante négative sera désignée N, une composante positive P, suivi d'un chiffre indiquant soit l'ordre d'apparition (P1, N2, P3, etc.), soit la latence d'apparition de cette composante : N100 (composante négative apparaissant environ 100 ms après le début de la stimulation) ou P200 (composante positive apparaissant environ 200 ms après le début de la stimulation). Les ondes négatives sont considérées comme marquant la prise en compte de la stimulation, alors que les ondes positives pourraient traduire la fin d'un processus (Timsit, 2001).

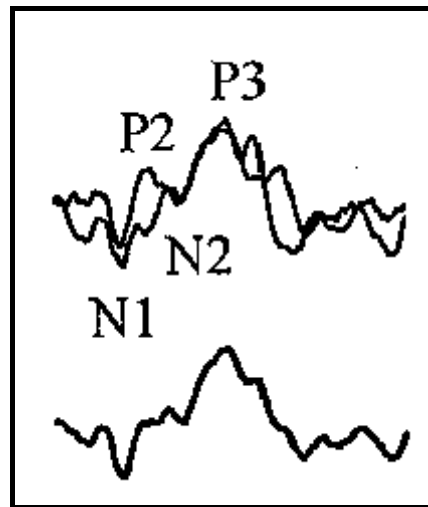


Figure A5 : Ondes négatives et positives observées sur les « grandes moyennes » issues des données EEG.

Il est possible de distinguer au sein des potentiels évoqués des composantes temporelles précoces, semi-précoces ou de latence moyenne, et des composantes tardives selon leur latence d'apparition après le début de la stimulation à laquelle elles répondent. Ces composantes ont une signification fonctionnelle différente et sont liées à la mise en jeu de sources différentes. Parmi les potentiels évoqués auditifs, des potentiels précoces apparaissent dans les 50 ms suivant le début de la stimulation. Ils sont le reflet des mécanismes perceptifs automatiques de bas niveau dépendant de l'activité physiologique dans les noyaux auditifs et leurs relais sous-corticaux. Ces potentiels sont appelés « exogènes », car leur existence, leur amplitude et leur latence sont modulées par les caractéristiques physiques du stimulus. Les potentiels évoqués exogènes reflètent l'activation d'un système sensoriel spécifique. Ils permettent d'observer l'activité neuronale engagée lors de la perception passive d'un stimulus. Ils interviennent très tôt, entre 50 ms et 150 ms, et leur amplitude est relativement faible. La plupart des potentiels évoqués auditifs précoces ou de latence moyenne ainsi que les potentiels visuels et somesthésiques sont des potentiels exogènes (Timsit, 2001). Les potentiels tardifs reflètent, quant à eux, les activités neuronales liées aux activités cognitives proprement dites. Ils sont nommés potentiels « endogènes ou cognitifs », car ils dépendent des intentions du sujet et de son activité mentale. Les potentiels cognitifs sont une réponse non obligatoire au stimulus. L'amplitude, la latence et la distribution spatiale des potentiels endogènes ou cognitifs ne dépendent pas des variations physiques des stimulations. Ils se

définissent comme dépendant des opérations mentales requises pour une tâche précise. Ils varient en fonction de phénomènes cognitifs tels que l'attention, la perception, la prise de décision, la préparation d'une réponse, le langage ou la mémoire.

A.1.4.4. L'imagerie cérébrale fonctionnelle

En 1896, Wilhelm Röntgen découvre les « rayons X » et marque ainsi la naissance de l'imagerie médicale *in vivo*. Le physicien autrichien Wolfgang Pauli, en 1924, suggère que l'intérieur du noyau atomique ait des propriétés magnétiques, causées par la rotation des charges électriques qui le composent. Un peu plus tard, en 1944, c'est au tour d'un autre physicien américain, Isidore Rabi, d'apporter une pierre à l'édifice en mesurant cette rotation ("*spin*" du noyau) et par la même de conquérir le Nobel. Le "*spin*" des nucléons détermine la propriété magnétique de tout noyau ayant un nombre impair de protons et de neutrons. En fait, le "*spin*" est le moment cinétique propre d'une particule résultant de la rotation de celle-ci sur elle-même et sur un axe. Deux ans plus tard naît la Résonance Magnétique Nucléaire (RMN ou *NMR* pour "*Nuclear Magnetic Resonance*") qui n'avait pas à la base d'application médicale précise, mais servait aux chimistes et aux géologues. Il a fallu attendre 1971, pour que le docteur Raymond Damadian démontre l'efficacité de la Résonance Magnétique Nucléaire en identifiant par ce biais des tissus cancéreux prélevés sur un de ses patients.

A.1.4.4.1. La Tomographie par Émission de Positons (TEP)

Une des techniques d'imagerie cérébrale est la Tomographie par Émission de Positons (TEP), qui utilise l'émission de positons pour tomographier, c'est-à-dire obtenir des images *in vivo* représentant la physiologie de coupes d'organes *in vivo*. La TEP mesure la physiologie cérébrale à partir d'un ensemble de rayons gamma issus de l'organe étudié. Ces signaux gamma sont produits *via* un émetteur de positons injecté au patient, dont la transition nucléaire conduit à l'émission d'une particule de même masse, mais de charge opposée à celle de l'électron : le positon. La TEP s'appuie sur l'existence d'isotopes des principaux éléments naturels constituant la matière vivante comme le Carbone 11, l'Oxygène 15, le Fluor 19 ou l'Azote 13. En effet, ces éléments radioactifs peuvent être substitués à leurs isotopes naturels dans des molécules biologiques (eau, sucres, acides aminés ou gras, *etc.*), et les rayonnements induits mesurés au décours même du métabolisme de la molécule dans l'organe étudié. Le

principal émetteur utilisé est le « Fluoro-Désoxy-Glucose », qui marque la consommation de glucose dans le cerveau.

A.1.4.4.2. L'Imagerie par Résonance Magnétique fonctionnelle

L'Imagerie par Résonance Magnétique utilise les propriétés magnétiques des protons contenus dans les tissus biologiques (en particulier l'eau et la graisse). Les protons sont soumis à un champ magnétique statique et excités par une onde de radiofréquence. Le signal détecté en IRM est celui qui est libéré lors du retour à l'équilibre (phénomène de relaxation). Il y a plusieurs phénomènes de relaxation, celui utilisé en IRMf est appelé T2* et correspond à la relaxation transversale, qui prend en compte les inhomogénéités de champ créées par les variations de la perfusion sanguine cérébrale.

L'IRM fonctionnelle est une imagerie indirecte de l'activité neuronale par le biais des variations locales de perfusion sanguine. À la fin du 19^{ème} siècle, Roy & Sherrington (1890) avaient suggéré l'existence d'une étroite relation spatiale entre l'activité neuronale et la perfusion cérébrale. L'activité neuronale entraîne une augmentation locale du métabolisme cérébral et de la consommation d'oxygène. En quelques centaines de millisecondes, ce phénomène est suivi d'une augmentation de la perfusion cérébrale locale responsable d'une élévation de l'apport en oxyhémoglobine (sang artériel) qui surpasse la consommation d'oxygène (Edelman, Wielopolski & Schmitt, 1994). En résulte une hyper-oxygénation sanguine dans et en aval du lit capillaire du territoire concerné. C'est cette hyper-oxygénation qui est détectée en IRMf.

A.1.4.4.2.1. L'effet *BOLD*

L'IRMf s'appuie sur la technique de l'IRM et mesure les fluctuations de signal RMN causés par les fluctuations de perfusion cérébrales : c'est l'effet *BOLD* ("*Blood Oxygen Level-Dependent*"; Ogawa, Lee, Kay & Tank, 1990). Cet effet est composite, traduisant simultanément l'augmentation locale de perfusion sanguine et la teneur du sang en oxyhémoglobine.

Ayant pour rôle principal de véhiculer l'oxygène dans le sang, l'hémoglobine peut se trouver sous deux formes : avec ou sans oxygène. Sans oxygène, la molécule est capable de modifier l'intensité du champ magnétique à proximité des vaisseaux, ce qui n'est pas le cas lorsqu'elle contient de l'oxygène. Lors d'activités cérébrales, les neurones des régions actives

concernées consomment de l'énergie et de l'oxygène. Le sang capillaire situé à leur abord s'appauvrit en oxygène : l'hémoglobine se libère de la molécule qu'elle transporte et devient dé-oxygénée. Lors du fonctionnement de l'IRMf, il se crée dans ces régions une petite inhomogénéité du champ magnétique : le temps de relaxation et le signal diminuent. Quelques centaines de millisecondes plus tard, le vaisseau sanguin voit son débit et sa teneur en hémoglobine oxygénée augmenter, par réaction physiologique vasodilatatrice à l'appauvrissement en oxygène, avant de revenir à la normale. En effet, il existe un couplage entre le débit sanguin cérébral et l'activité neuronale (plus l'activité neuronale augmente, plus le débit sanguin cérébral augmente) et un découplage entre le débit sanguin cérébral et la consommation d'oxygène (appelé hyperhémie fonctionnelle). Donc quand l'activité neuronale augmente, l'augmentation du débit sanguin cérébral associée est plus importante que la consommation d'oxygène : le sang est trop oxygéné, ce qui correspond à une diminution de la concentration en désoxyhémoglobine. La désoxyhémoglobine est utilisée ici comme traceur endogène car elle est paramagnétique, donc crée des inhomogénéités locales dans le champ magnétique. La diminution de sa concentration diminue les inhomogénéités, donc augmente l'intensité du signal T2*. La réponse hémodynamique, due à l'effet *BOLD*, présente un retard de 6 à 8 secondes sur le stimulus présenté et dure une vingtaine de secondes au total.

L'IRMf constitue une approche dynamique des activations cérébrales, en ce qu'elle permet l'enregistrement à haute cadence d'un grand nombre d'images couvrant le cerveau entier, en continu. Chaque élément de volume (voxel) du cerveau est donc observé avec une résolution temporelle de quelques secondes, permettant de détecter le signal *BOLD* éventuel qui s'y déroule. C'est sur ces variations du signal IRM que l'on se base pour obtenir une image tridimensionnelle du cerveau en action.

A.1.4.4.2.2. La soustraction cognitive

La méthode dite de la « soustraction cognitive » consiste à mettre en place deux conditions expérimentales en tous points identiques à l'exception du processus d'intérêt (voir figure A6). L'hypothèse sous-jacente est que le contraste des deux conditions révèle l'activité cognitive et / ou cérébrale, liée au seul processus d'intérêt. Par exemple, il est possible d'observer un groupe de sujets entendant des mots opposés à des logatomes, afin de révéler les zones cérébrales impliquées dans les traitements lexicaux. En IRMf, la soustraction du

signal cérébral des deux conditions révèle l'activité cérébrale engendrée par le processus d'intérêt.

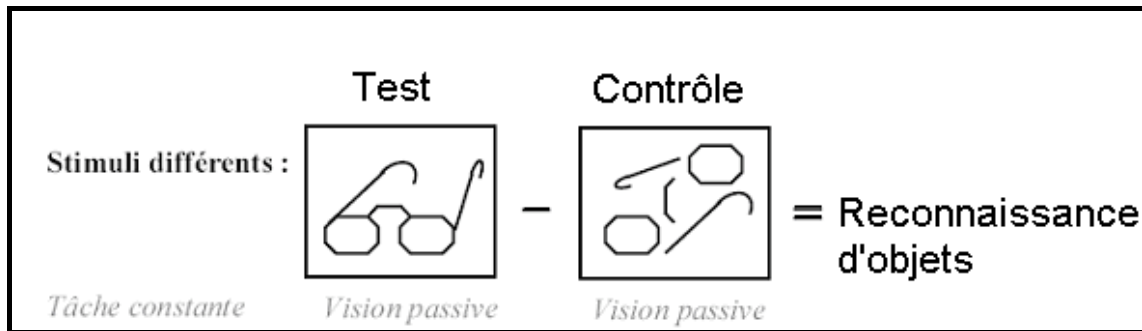


Figure A6 : Illustration de la méthode de la soustraction cognitive.

A.1.4.4.2.3. Paradigme en bloc et paradigme événementiel

Selon le type de paradigme choisi et l'enchaînement des conditions ou des tâches dans le temps, les courbes théoriques de la réponse hémodynamique et du signal *BOLD* sont établies. Elles prennent notamment en compte le décalage dans le temps entre l'activation neuronale et la réponse hémodynamique mesurée. Dans le cas de tâches répétées rapidement, il existe une sommation de leur effet qui aboutit à une réponse hémodynamique en plateau. Ce modèle est la base de l'analyse statistique, qui recherchera les pixels dont la variation de signal est liée au paradigme.

- Paradigme en bloc : les activités sont organisées en bloc de quelques dizaines de secondes qui alternent à intervalles réguliers. Au sein d'un même bloc, les réponses hémodynamiques vont se chevaucher et s'accumuler avant de former un plateau.
- Paradigme événementiel : les activités ou stimuli sont uniques ou présentés en courtes répétitions, avec un enchaînement qui peut être pseudo-aléatoire (ce qui évite le phénomène d'anticipation). Il est possible d'évaluer ainsi la réponse hémodynamique locale lors des différentes activités (voir figure A7).

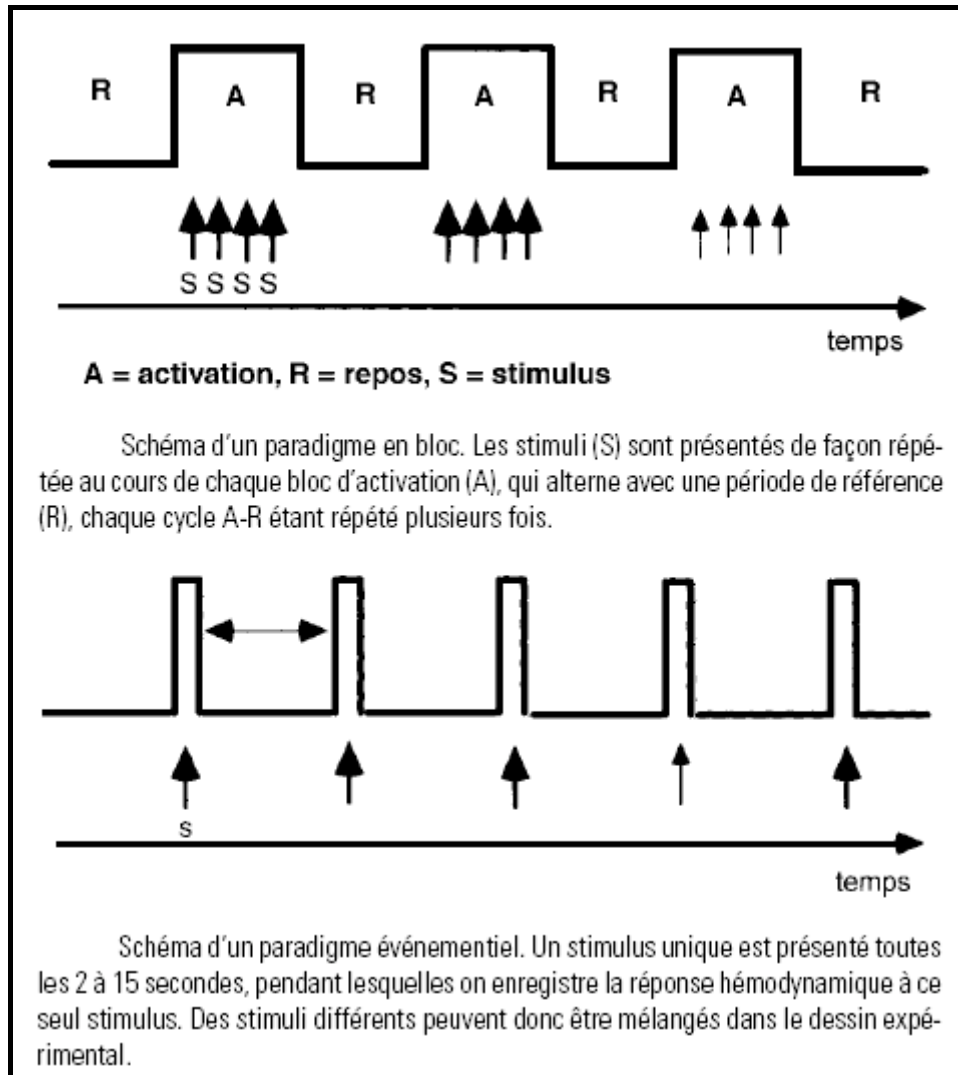


Figure A7 : Représentation schématique d'un paradigme en bloc et d'un paradigme événementiel. (Tiré de Hertz-Pannier, Lehericy, Cordoliani, Le Bihan, Marsault & Brunelle, 2000).

A.1.4.5. Le couplage de l'IRMf et de l'EEG

Les premières études couplant l'IRMf et les potentiels évoqués ont principalement été mises en œuvre afin de localiser les foyers épileptogènes (Seeck, Lazeyras, Michel, Blanke, Gericke, Ives, Delavelle, Golay, Haenggel, De Tribolet, Landis, 1998 ; Krakow, Woermann, Symms, Allen, Lemieux, Barker, Duncan, Fish, 1999). Le couplage de ces deux techniques permet aussi d'explorer le traitement de la parole chez le sujet sain. L'avantage d'un recueil simultané pour les études cognitives repose sur le fait, que les phénomènes observés sont enregistrés dans des conditions expérimentales similaires laissant penser que les processus attentionnels, sensoriels et motivationnels sont identiques (Debener, Ullsperger, Siegel,

Fiehler, Von Cramon, Engel, 2005). Le décalage lors de l'enregistrement de la réponse hémodynamique permet de corrélérer temporellement ses fluctuations avec le stimulus produisant la réponse évoquée.

Néanmoins, l'enregistrement de l'activité électrique, grâce à l'EEG, dans un champ magnétique intense, comme celui de l'IRM, est un défi tant méthodologique que technique. Les études simultanées sont encore à l'heure actuelle très rares, dans le domaine visuel (Bonmassar, Schwartz, Liu, Kwong, Dale, Belliveau, 2001) ; pour les mécanismes attentionnels (Bledowski, Prvulovic, Goebel, Zanella & Linden, 2004 ; Mulert, Jäger, Schmitt, 2004 ; Bussfeld, Pogarell, Mölle, Juckel, Hegerl, 2004). À notre connaissance, il n'y a pas d'expériences similaires à la nôtre dans le domaine de la parole. Certaines études ont évité les écueils induits par une acquisition simultanée, en réalisant des enregistrements avec les deux techniques, mais lors de sessions distinctes ; ce sont des paradigmes identiques qui sont utilisés, avec les mêmes stimuli et, bien sûr, les mêmes participants qui prennent part aux deux sessions IRMf et EEG.

Du point de vue de la phonologie, les concepts de phonèmes, de visèmes, mais aussi de traits distinctifs sont avant tout des unités de classement. La syllabe pourrait être, selon MacNeilage & Davis (1998), un cadre naturel émergeant des cycles d'ingestion, et qui pourrait fournir un support au contenu segmental.

Un seul trait distinctif peut permettre de distinguer, au niveau lexical, deux mots constitués d'une seule syllabe, (« baie » vs. « paix »). Des syllabes identiques, constituées des mêmes phonèmes et des mêmes traits, ne seront jamais tout à fait semblables. Cette tension entre différences et similarités n'est toujours pas résolue. La perception des visèmes pourrait réduire la variabilité du signal acoustique en permettant d'amorcer un nombre de catégories phonémiques plus restreint, facilitant ainsi le traitement acoustico-phonétique. Néanmoins, il n'est pas possible de mettre sur un même plan les phonèmes et les visèmes, du fait du caractère sous-spécifié de ces derniers. En effet, les catégories de visèmes présentent un nombre d'unités plus large que les catégories de phonèmes ; les visèmes sont plus proches des traits distinctifs. De plus, les indices visuels linguistiquement pertinents ne sont pas exempts de variabilité, néanmoins l'intégration de deux modalités pourrait permettre de faciliter le traitement cognitif par catégorisation.

En TEP et en IRMf, il est supposé que la réalisation d'une tâche cognitive entraîne un apport plus important de sang dans les régions corticales concernées. Pour la TEP, un inconvénient majeur est l'utilisation d'isotopes radioactifs d'une période très courte, ce qui nécessite d'œuvrer à proximité d'un cyclotron. Les deux avantages principaux par rapport aux études en IRMf viennent de la possibilité de réaliser des tâches de production de la parole et de l'absence de bruit lors de la session expérimentale. La soustraction cognitive est de première importance dans le cadre des études langagières. Il est nécessaire d'utiliser deux types de matériau expérimental suffisamment éloignés pour engendrer des activations différenciables tout en proposant un contraste qui ne soit pas sans rapport. L'application de l'EEG aux neurosciences cognitives nécessitait de faire ressortir de l'activité spontanée du cerveau celle liée à la perception d'un stimulus et au traitement d'informations requis pour accomplir une tâche. Il a fallu développer des techniques spécifiques et les progrès de l'informatique ont participé à l'essor de l'EEG. Les potentiels évoqués fournissent des informations sur le *timing* et la saillance des traitements cérébraux, mais la localisation des structures responsables demeure difficile, en particulier dans le scanner IRMf. Le couplage de l'EEG et de l'IRMf engendre des conditions particulières qui sont loin d'être sans impact sur le recueil des données expérimentales. L'enregistrement d'ondes électriques dans un appareil générant des champs magnétiques intenses est, en soit, un défi technique. Mais celui-ci s'avère des plus intéressants, parce que l'interprétation de données anatomo-fonctionnelles nécessite la prudence et requiert, dans la mesure du possible, de trouver des appuis convergents issus d'autres techniques (Wilkinson & Halligan, 2004).

A.2. Théories phonétiques de la perception de la parole et modèles d'accès au «lexique mental»

La variabilité présente à de nombreux niveaux dans les échanges langagiers pourrait être surmontée grâce à une particularité du système perceptif qui opérerait de façon catégorielle. Liberman, Harris, Hoffman & Griffith (1957) ont décrit un phénomène de ce type pour la parole qui sera plus tard baptisé « perception catégorielle » (Eimas, 1963) et qui s'oppose à la perception dite « continue ». Il a été observé, lors de tests de perception auditive portant sur des occlusives à l'initiale, que les tâches d'identification étaient moins bien réussies que les tâches de classification. Deux stimuli seront classés dans une même catégorie lors de l'identification, mais seront perçus comme distincts lors de la classification. En parole, les différences acoustiques entre les variations d'une même catégorie phonémique ne sont généralement pas perçues. *A contrario* des différences acoustiques de même amplitude sont perceptibles et pertinentes entre des sons appartenant à des catégories différentes. Ainsi, à l'intérieur d'une catégorie, la discrimination est faible, elle devient forte à la frontière phonémique (Wood, 1976). Les unités d'une même catégorie ne sont pourtant pas identiques. Il existe des représentants pertinents de la catégorie et d'autres qui le sont moins. L'« effet d'aimant perceptif » (*“magnet effect”*) pourrait tirer son origine de cette structure interne des catégories phonétiques. Un représentant de la catégorie est plus difficile à discriminer s'il est comparé au prototype de la catégorie que s'il est comparé à un exemplaire non prototypique (Kuhl, 1991). Ce qui signifie que c'est la similarité entre deux segments qui serait primordiale, la perception de frontières catégorielles pourrait être liée à la « distance » par rapport au prototype de la catégorie. La notion même de frontière catégorielle se définirait en termes de « degré de similarité » et non plus en « degré de différence ».

Les modèles d'accès au lexique mental présupposent un traitement préalable des indices linguistiquement pertinents présents dans les flux sensoriels. Le lexique mental est classiquement défini comme étant une partie de la mémoire qui contient l'information sémantique, phonologique, orthographique et syntaxique associée aux mots (Holender, 1988). Dans cette optique, selon Fayol, Gombert, Lecocq, Sprenger-Charolles & Zagar (1992), reconnaître un mot serait comme chercher dans un dictionnaire une entrée lexicale, donnant

accès aux informations sur la prononciation, la classe lexicale et le sens. Cette notion renvoie au concept des « représentations ». Mentale, phonologique, lexicale, sémantique ou orthographique, les adjectifs spécifiant la notion de représentation sont très nombreux. La reconnaissance d'un mot consisterait en un appariement entre une « représentation » construite à partir de l'information sensorielle décodée et une « représentation » stockée dans le cerveau de l'interlocuteur (Kandel & Boë, 1996). Les « représentations phonologiques » seraient des formes abstraites des traits phonétiques et non pas des « représentations acoustiques ou phonétiques » détaillées (Frost, 1998). La notion de « représentation orthographique » renvoie, quant à elle, à un encodage visuel, qui porterait sur les aspects analogiques du stimulus, le mot, et qui est nommée « forme visuelle du mot », traduction de l'anglais "*Visual Word Form*".

A.2.1. Les théories de la perception de la parole articulée

Il est possible de classer les principales théories phonétiques en fonction de l'approche qu'elles adoptent. Une subdivision est souvent effectuée en théories *motrices*, *auditives* et *interactionnelles*. Les théories auditives mettent l'accent sur la perception des invariants acoustiques de la parole sans accorder une attention particulière aux configurations articulatoires. Les théories interactionnelles cherchent à dégager un équilibre entre les versants articulatoires et auditifs. Ayant pour bases des données psycho-comportementales et neuropsychologiques, les théories motrices essayent de décrire la manière dont le signal acoustique continu est segmenté en unités discrètes renvoyant à des patrons utilisés lors de la production de la parole.

A.2.1.1. Théorie auditive : la Théorie Quantique

La Théorie Quantique de Stevens (1972) postule une organisation hiérarchique des traits segmentaux phonologiques structurant le lexique ainsi qu'une représentation directe de caractéristiques acoustiques invariantes du signal de parole. Pour Stevens, le locuteur doit connaître les patrons de coordination articulatoire afin de réaliser les exigences acoustiques et auditives de la phonologie de sa langue. Il y a donc une dimension articulatoire très présente dans cette théorie. Mais il existerait une indépendance entre la production de la parole et sa perception. La Théorie Quantique affirme que les propriétés auditives des stimuli sont les facteurs critiques de la manifestation de la perception catégorielle. Cette théorie s'apparente à

un ensemble d'explications de la perception catégorielle fondées sur des propriétés auditives innées (Stevens, 1981). Certains déplacements des articulateurs n'auront pas de conséquences acoustiques importantes pour la perception, alors que d'autres déplacements de même amplitude engendreront des différences acoustico-auditives significatives.

A.2.1.2. Les théories interactionnelles

Ces théories se fondent sur l'idée d'un couplage intrinsèque entre la perception et la production de la parole.

A.2.1.2.1. La Théorie de la Variabilité Adaptative

Lindblom (1987) plaide pour une théorie prenant en compte, à la fois, le locuteur et l'auditeur. L'idée n'est plus de rechercher des invariants. Il s'agit davantage d'identifier l'information échangée selon les conditions d'élocution. Lindblom nomme cette théorie « Théorie de la Variabilité Adaptative ». L'invariance résiderait dans la compréhension globale de l'auditeur. Le contrôle moteur opérerait en tenant compte de l'exigence de distinctivité perceptive nécessaire à l'auditeur et du principe d'économie régissant la production du locuteur. Ce concept rappelle l'opposition « principe d'économie / pression du système » présentée jadis par Martinet (1955). La variabilité des paramètres physiques de la parole respecterait les contraintes de distinctivité dans l'optique de produire un message linguistique constituant l'invariant. Le phénomène de l'hyper-articulation est le corollaire de l'exigence de distinctivité, alors que l'économie articulatoire sera rendue par un phénomène d'hypo-articulation. En milieu bruyé, le locuteur devra privilégier la distinctivité. Lindblom pense que le locuteur cherche constamment à se rapprocher d'une « diagonale idéale » entre production et perception, s'ajustant ainsi à l'environnement. La perception visuelle de la parole pourrait s'inscrire dans une même logique, le système employant une modalité supplémentaire afin d'augmenter l'intelligibilité dans un environnement défavorable.

A.2.1.2.2. La Théorie de la Perception pour le Contrôle de l'Action

La Théorie de la Perception pour le Contrôle de l'Action (Schwartz, Abry, Boë & Cathiard, 2002) vise à :

« la définition de représentations perceptives pour le contrôle de l'action et sous le contrôle de l'action. »

Le système perceptif permettrait de récupérer les « représentations phonologiques » et de contrôler l'action des articulateurs. Le *timing* des événements et l'atteinte de leurs « cibles » spectrales seraient effectués par ce système. Cette récupération n'est que partielle puisqu'il existe des gestes, engendrés par la coarticulation segmentale, qui sont inaudibles. Dans le même temps, les percepts seraient contraints par des procédures de régularisation issues de l'action du système moteur. Ces percepts ne seraient pas exclusivement de nature auditive mais mettraient en jeu une perception visuelle voire proprioceptive. Les traitements perceptifs multimodaux forgeraient les « représentations » permettant le contrôle des réalisations articulatoires.

A.2.1.2.3. La Théorie de la Viabilité en production-perception de la parole

La Théorie de la Viabilité en production-perception de la parole (Sock, 2001) remplace la notion de « représentation symbolique » par celle de « propriété émergente ». Cette théorie est issue de la Théorie de la Viabilité d'Aubin (1991) conçue dans le domaine mathématique et adaptée dans le domaine de la production et de la perception de la parole. L'information proviendrait de l'organisation sensori-motrice du système de production-perception de la parole. En effet, selon Sock, la production et la perception de la parole formeraient un système couplé de contraintes réciproques afin de mettre au jour des composantes auto-organisées régulant ledit système. D'une manière générale, nous percevons ce que nous sommes capables de produire et nous produisons ce que nous sommes capables de percevoir. C'est-à-dire que le système de production ne peut retenir des différences imperceptibles pour le niveau perceptif. Des contrastes instables ou peu rentables ont pu être éliminés lors de l'évolution du proto-langage, favorisant les traits ou phonèmes plus robustes.

A.2.1.3. Les théories motrices

Ces théories postulent les gestes articulatoires comme une base utilisée par le système perceptif pour définir des invariants.

A.2.1.3.1. La Théorie de la Perception Directe

La Théorie de la Perception Directe (Fowler, 1986, 1996) découle de la Théorie Perspective Réaliste Directe de Gibson (1979). Gibson postule qu'un système visuel et un environnement dans lequel il est possible d'évoluer sont les deux composantes essentielles du phénomène de perception. L'hypothèse de base avance l'existence d'invariances dans la topologie de la lumière ambiante spécifiant les propriétés de l'environnement et les "*affordances*" (c'est-à-dire les possibilités d'actions mises à la disposition du sujet par son environnement). Les bases de la perception sont présentes dans l'environnement, il n'est donc pas obligatoire de supposer l'existence de « représentations symboliques ». Par conséquent, Fowler (1986, 1996) a proposé que les gestes articulatoires contiennent intrinsèquement les invariants malgré la coarticulation. De plus, ce seraient les gestes qui correspondraient aux unités phonologiques en production et non leurs substrats neuraux. Les gestes phonologiques seraient les actions visibles du tractus vocal engendrant les structures analysées dans les signaux acoustiques. L'hypothèse fondamentale propose que les gestes soient des spécificateurs ("*specifiers*"), c'est-à-dire les invariants, dans le signal acoustique. L'existence de ces propriétés acoustiques invariantes rendrait possible la perception directe des propriétés phonologiques sans médiation *via* des « représentations ».

A.2.1.3.2. La Théorie Motrice de la Perception de la parole

La Théorie Motrice de la Perception de la parole (Lieberman, Cooper, Shankweiler & Studdert-Kennedy, 1967 ; Lieberman & Mattingly, 1985) avance trois points essentiels : le traitement de la parole serait spécifique, percevoir la parole serait percevoir les gestes articulatoires du conduit vocal et la perception de la parole impliquerait un accès au système moteur.

Cette théorie a pour origine une expérience sur un continuum de syllabes synthétiques /da – ga/ perçue en écoute binaurale par des auditeurs. Ceux-ci identifiaient clairement les syllabes situées aux extrémités du continuum, alors que les stimuli situés au centre de celui-ci étaient assimilés à une syllabe ambiguë entre /da/ et /ga/. Le paramètre distinctif principal entre ces deux syllabes est constitué par la trajectoire du troisième formant (F3). Les auditeurs ont été invités à participer à une deuxième phase expérimentale séparant le continuum précédent en deux parties. La première partie correspondait aux transitions de F3,

partie variable, alors que la deuxième partie, la base, correspondait à l'élément vocalique, ainsi qu'aux transitions de F1 et de F2. La perception de la transition de F3 en continuum est assimilée à un pépiement, tandis que la seconde partie est perçue comme une syllabe ambiguë. Finalement, ces deux parties ont été présentées en écoute dichotique. Les résultats sont similaires. Un pépiement était perçu dans l'oreille à laquelle était présentée la transition de F3 et une syllabe ambiguë dans l'autre oreille, où la base était envoyée. Les meilleurs représentants de la catégorie se trouvent aux extrémités du continuum. Ces résultats tendraient à prouver que les deux parties du stimulus sont préalablement intégrées avant qu'une catégorie puisse être dégagée. Un module spécialisé de perception serait à l'œuvre, afin de traiter l'information à partir des unités les plus élémentaires vers des unités plus larges (processus "*bottom-up*").

Ce module, souvent qualifié de « pré-cognitif », aurait donc la capacité d'interpréter la source acoustique en termes de gestes articulatoires sans pour autant que ces composantes soient accessibles aux modules de décodage cognitif de haut niveau.

Pour Liberman & Mattingly (1985), la coarticulation est primordiale afin de compenser l'inertie des articulateurs du conduit vocal et ainsi permet de maintenir une vitesse d'élocution stable. Le chevauchement spatial et temporel des unités rend non linéaire la correspondance entre les gestes et les segments acoustiques successifs nécessitant une capacité d'analyse « pré-cognitive ».

A.2.1.3.3. La Phonologie Articulatoire

La « Phonologie Articulatoire » a pour base la notion de « geste » prise dans une acception particulière. Pour Browman & Goldstein (1990, 2000), les gestes sont un ensemble de configurations articulatoires utilisées lors de la production, mais aussi lors de la perception de la parole. Il ne s'agit pas d'unités représentationnelles mais d'unités d'action spécifiant les articulateurs utilisés et leurs mouvements dynamiques vers une cible, la zone de constriction. La coarticulation est intégrée dans les gestes qui seront modifiés de façon à créer une synergie adaptant au mieux les différentes unités. Une notion de « rigidité » a été introduite afin de spécifier les liens entre les gestes. Les liens les plus forts unissant des gestes formeront par exemple des segments ou des attaques syllabiques. Il existe en outre une subdivision entre gestes vocaliques et consonantiques mis en place les uns par rapport aux autres. En recourant à des unités spatio-temporelles, intermédiaires entre les traits et les segments et utilisées à la

fois en production et en perception, la « Phonologie Articulatoire » tente de réduire la variabilité tant acoustique que perceptive.

L'étude, per se, des mécanismes perceptifs ne doit pas faire oublier que ceux-ci sont tournés vers un but qui est de comprendre le message délivré, mais aussi de pouvoir y répondre. De plus, la plupart des modèles d'accès au lexique mental empruntent des concepts à la phonologie et à la phonétique, afin de caractériser les unités susceptibles de constituer une entrée de ce lexique. De nombreuses études montrent l'influence des connaissances lexicales sur l'identification phonémique ou syllabique, et réciproquement (Cutler, Mehler, Norris & Segui, 1987 ; Cutler & Mehler, 1993 ; Ferrand et al., 1995), même si cette dernière demeure contestée (Pitt & McQueen, 1998).

A.2.2. Les principaux modèles d'accès au lexique mental

En général, les modèles d'accès au lexique mental conçoivent l'identification d'un mot comme un processus de discrimination entre de multiples candidats lexicaux. Nous nous inspirons ici largement de Frauenfelder qui indique, dans son chapitre consacré aux modèles de reconnaissance des mots parlés (Kolinsky, Morais & Segui, 1992), que la plupart d'entre eux envisagent un processus d'appariement entre une « représentation d'entrée » et des « représentations lexicales » conservées par le cerveau humain. Selon les modèles, les concepts de « représentation d'entrée », de « représentation lexicale » et les processus d'appariement entre les deux diffèrent. La « représentation d'entrée » est une interface entre les signaux linguistiques et la « représentation lexicale ». Il y aurait un codage des données sensorielles, afin d'amorcer l'appariement avec la « représentation lexicale ».

A.2.2.1. COHORTE

Le modèle COHORTE, dans sa version originale (Marslen-Wilson, William & Welsh, 1978), attribue un rôle prépondérant au début des mots et considère le traitement de la parole comme un processus adapté à la distribution séquentielle de l'information acoustique. Ce modèle suppose qu'à l'aide des deux ou trois premiers phonèmes d'un mot, un auditeur sélectionne une cohorte de candidats alignés avec ces sons initiaux. Un mot est reconnu lorsqu'il est le seul membre restant dans la cohorte et selon un paradigme du mode « tout ou

rien ». Le mot « éléphant » [e l e f ɑ̃] serait identifié à partir du phonème [f] qui permet d'évacuer les autres items possibles. Ce point de reconnaissance correspond au point d'unicité (PU), c'est-à-dire le moment à partir duquel un mot reste l'unique candidat à être activé. Les principales objections envers ce modèle reposent sur le fait qu'il est généralement possible de reconnaître un mot même si celui-ci subit une altération du segment initial (« Bonjour » prononcé [p ɔ̃ ʒ u R]). Les auteurs reprennent le modèle afin de résoudre ces objections (Marslen-Wilson, 1987). Ils renoncent à l'appariement « tout ou rien » lui préférant la notion de « degré d'appariement ». À l'instar du modèle « TRACE » de McClelland & Elman (1986), la sélection pourrait relever d'un processus de discrimination entre des items proches se différenciant par leur niveau d'activation dépendant du degré d'appariement avec les informations sensorielles et ses propriétés spécifiques. La « représentation langagière » du « lexique mental » n'est plus spécifiée en termes de segments phonémiques mais comme une « représentation » moins catégorielle pouvant reposer sur les « traits acoustiques » (Seidenberg & McClelland, 1989).

A.2.2.2. TRACE

Ce modèle, proposé par McClelland & Elman (1986), postule trois types d'unités organisées hiérarchiquement : les traits distinctifs, les unités phonémiques et les unités de mots. Il y a des facilitations entre deux niveaux adjacents (traits < phonèmes < mots) et des inhibitions entre deux unités de même niveau (trait - trait ; phonème - phonème). Les unités de traits stimulent les unités de phonèmes de la catégorie (par exemple la présence de voisement active tous les phonèmes voisés), puis une unité de phonèmes suffisamment stimulée active des unités de mots. Le processus d'inhibition des unités de même niveau dépend de la connexion, du niveau d'activation de l'unité inhibée, et du taux de recouvrement entre les unités concurrentes. Ce processus réduit les hypothèses à chaque niveau permettant d'éliminer les unités faiblement activées. Ce modèle propose aussi un mécanisme "*top-down*", permettant au mot d'exciter les unités des phonèmes. Tous les mots du lexique sont constamment en lice pour la reconnaissance, il n'y a pas de segmentation des flux sensoriels mais un appariement en ligne.

A.2.2.3. NAM

Le modèle NAM pour “*Neighbourhood Activation Model*” (Luce & Pisoni, 1998 ; Luce, Goldinger, Auer & Vitevitch, 2000 ; Vitevitch, Luce, Pisoni & Auer, 1999a ; Vitevitch, Luce, Pisoni & Auer, 1999b) ne tient pas compte de la directionnalité du signal de parole et aucune portion du signal acoustique n’est privilégiée. Ce modèle, conçu pour rendre compte du traitement des mots monosyllabiques, définit le voisinage lexical en référence à un traitement parallèle. Selon le modèle NAM, les voisins lexicaux correspondent à tous les mots qui peuvent être générés par addition, délétion, ou substitution d’un phonème, quelle que soit sa position. La nature des similitudes entre mots supposées pertinentes pour le traitement est intimement liée à la conceptualisation des unités de traitement des mots. Le mot étant considéré comme une séquence de phonèmes, la similarité sera estimée en référence aux phonèmes partagés entre le mot et les autres candidats lexicaux.

A.2.2.4. LAFS

Le modèle LAFS (“*Lexical Access From Spectra*”) développé par Klatt (1979) postule que l’analyse du signal acoustique permet d’accéder de manière directe aux mots qui sont considérés comme des séquences de gabarit spectraux. Définis en termes de nœuds phonémiques et de transitions entre ces nœuds, les mots ayant les mêmes phonèmes initiaux seront en concurrence, jusqu’à une ramification permettant un chemin linéaire à travers le réseau constitué. La « représentation » est ici considérée comme un réseau phonétique lexical qui substitue aux gabarits spectraux, soit les allophones appropriés, soit d’autres phonèmes, ou efface les segments superflus. Les capacités mémorielles du cerveau humain permettraient une comparaison en ligne du signal acoustique avec l’ensemble des mots mémorisés. Le système exploiterait les effets de fréquence et la répartition statistique des mots d’une langue.

A.2.2.5. SARAH

Cette approche postule que des « représentations sublexicales » de type syllabique jouent un rôle essentiel dans la reconnaissance des mots. Après l’analyse acoustico-phonétique, la syllabe intervient à un niveau de traitement intermédiaire, responsable du déclenchement du processus d’accès au lexique mental, lorsque l’accumulation

d'informations sensorielles est suffisante. La syllabe ne doit pas être considérée comme une unité d'analyse du signal mais plutôt comme une unité d'organisation des différentes sources d'information acoustique possédant, au niveau de la forme ("*Gestalt*"), des propriétés d'invariance (Dupoux, 1989).

Le modèle SARAH, pour "*Syllable Acquisition, Representation and Access Hypothesis*" de Mehler, Dupoux & Segui (1990), se distingue des autres modèles d'accès au lexique mental en cela qu'il repose sur une forte correspondance entre les processus d'accès au lexique chez l'adulte et les processus d'acquisition du langage chez l'enfant. La syllabe est considérée comme le matériau de base permettant la constitution de « représentations lexicales » chez le jeune enfant, impliquant un rôle prépondérant dans les processus d'accès au lexique au terme de la maturation. Contrairement au modèle COHORTE, le flux d'information entre le niveau de traitement acoustico-phonétique et les niveaux supérieurs n'est pas continu ; il repose sur un niveau prélexical de traitement intermédiaire qui est responsable de l'accumulation d'informations acoustiques avant leur transmission aux niveaux supérieurs. La reconnaissance des mots s'effectuerait en trois étapes : syllabique, lexicale et phonologique. Une banque « d'analyseurs syllabiques » prendrait en charge la reconnaissance du cadre syllabique. Celui-ci correspond à une réalisation élémentaire : l'unité fonctionnelle minimale impliquée dans la production de la parole. Cette unité capture l'invariance au-delà des différences inter- et intra-individuelles. Le niveau du traitement lexical débute lorsqu'une quantité critique d'information est déjà traitée : la première syllabe d'un item constituant le code d'accès. Celle-ci représente la quantité minimale d'information susceptible d'activer un grand nombre de candidats lexicaux. Enfin, les phonèmes sont dérivés des cadres syllabiques.

A.2.2.6. SHORTLIST

Ce modèle, proposé par Norris (1994), utilise le phonème comme « représentation d'entrée », ainsi que des informations sur les frontières de mots. Ces informations sont liées à la langue cible ; en anglais, la distinction entre les syllabes accentuées ("*strong*") et inaccentuées ("*weak*") influence la segmentation de la parole. Ce modèle est unidirectionnel, il n'y a pas d'interaction entre les mots et les phonèmes. Les segments les plus proches des propriétés phonologiques seront les plus activés.

Les liens entre la perception et la production de la parole sont primordiaux. La dissociation entre ces deux versants est aussi artificielle que nécessaire, sur le plan expérimental. D'un point de vue théorique, il n'est pas rare que l'un conditionne l'autre. Pour la Théorie Quantique (Stevens, 1972), le versant articulatoire n'a qu'un rôle secondaire ; alors que la Théorie Motrice de Libermann & Mattingly (1985) lui attribue une place de première importance. Les notions de « spécificateur » ou de « geste » permettent de lier la production et la perception de la parole *via* des unités intermédiaires. Les théories interactionnelles replacent l'« auditeur – locuteur » au centre des réflexions. La « diagonale idéale » de Lindblom (1987), les « cibles » de Schwartz *et al.* (2002) et les « propriétés émergentes » de Sock (2001) nous invitent à penser que le système de perception et production de la parole est adaptable et plastique. Il est possible de contrôler consciemment un ensemble de paramètres concernant le versant productif ; nos capacités d'action sur le versant perceptif, quant à elles, sont beaucoup moins nombreuses.

Les modèles d'accès au lexique mental diffèrent mais se sont tous concentrés sur la façon dont les « représentations phonologiques » entrent en contact avec les « représentations lexicales ». Tous mis à part, les modèles « COHORTE » (Marslen-Wilson *et al.*, 1978, 1987) et « LAFS » (Klatt, 1979) présupposent l'accès lexical comme une compétition entre différents candidats. Les traitements auditifs et/ou visuels sont réalisés par concordance partielle ou totale entre le stimulus et un prototype jusqu'à ce qu'un item unique demeure. Dans le cadre de la parole articulée, le processus engagé en premier devrait être celui de la segmentation des phrases en mots. La majorité des modèles suivent une analyse linguistique standard comprenant un niveau acoustique, phonétique et ou phonémique et une analyse phonologique (tous n'ont pas chacune des étapes). « TRACE » (McClelland et Elman, 1986) adopte une approche dans laquelle les niveaux d'analyse sont les traits pertinents et les phonèmes mais celui-ci ne présuppose pas de « représentations phonologiques » de la structure métrique, telles que la syllabe ou la more. « SHORTLIST » (Norris, 1994) fait intervenir ces structures suprasegmentales, mais ne précise pas un niveau explicite de traitement prélexical ; le niveau de traitement phonémique constituerait la pierre de touche de l'édifice. Dans la version la plus récente de « COHORTE » l'unité fondamentale est le trait qui accède directement aux « représentations lexicales » sans passer par un traitement phonémique ou syllabique. « SARAH » (Mehler *et al.*, 1990) défend la syllabe qui serait le principal niveau de « représentations », et la propose comme unité élémentaire déclenchant l'appariement vers les « représentations lexicales ».

Nous n'avons traité que des modèles et théories auxquels nous avons eu accès. Il existe, évidemment, de nombreux autres modèles d'accès au lexique mental, ainsi que d'autres théories de la perception de la parole.

A.3. Les mécanismes cérébraux de la perception de la parole

Malgré les progrès techniques considérables effectués depuis les travaux des pionniers du 19^{ème} siècle, l'organisation cérébrale d'une tâche cognitive donnée demeure difficile à déterminer de façon précise et reproductible. Après avoir longtemps été considérée comme l'apanage exclusif de certaines régions de l'hémisphère gauche, il semble que la parole puisse être traitée par des réseaux distribués à travers les deux hémisphères cérébraux. Les premières découvertes issues des patients aphasiques ont fortement influencé notre conception des processus cognitifs, en particulier dans le domaine de la perception / production du langage parlé, laissant penser qu'une région corticale discrète assume une fonction donnée.

Nous reprenons une perspective historique dans ce deuxième chapitre, exposant tout d'abord les avancées liminaires effectuées grâce à l'observation anatomoclinique, avant d'aborder les dernières données issues de l'imagerie fonctionnelle. Nous préciserons, autant que faire ce peut, les aires impliquées en prenant comme référence la figure 6 de l'article de Démonet, Thierry & Cardebat (2005) utilisée pour notre planche centrale. Nous avons fait ce choix afin d'éclaircir notre propos, il faut néanmoins garder à l'esprit que la classification de Brodmann n'est pas exempte de critiques. En effet, cette classification a été établie à partir de l'examen visuel de l'organisation cytoarchitectonique, laissant place à une certaine subjectivité (Amunts & Von Cramon, 2006).

A.3.1. Les apports des précurseurs

Monsieur Leborgne est aussi connu dans le monde des neurosciences que le médecin qui a observé ses troubles (Broca, 1865). L'aphasie de Broca est consécutive de lésions d'une zone corticale, située au niveau de la partie inférieure de la troisième circonvolution frontale de l'hémisphère gauche chez la plupart des individus (Aires de Brodmann 44 et 45). Après sa destruction, le sujet semble incapable de formuler oralement ses idées alors que celles-ci sont intactes dans son esprit. La compréhension serait peu affectée, mais l'expression gravement diminuée. Les aphasiques de Broca s'expriment en style télégraphique, avec une vitesse d'élocution réduite, une production articulatoire perturbée, ainsi qu'un lexique et une syntaxe appauvris. Décrite peu de temps après ces observations, l'aphasie de Wernicke (1874) est, quant à elle, engendrée par des lésions du tiers postérieur de la circonvolution temporale supérieure de l'hémisphère gauche (Aires de Brodmann 41, 42 et 22). Contrairement à

l'aphasie de Broca, cette pathologie n'entraînerait pas de déficits lexicaux ou articulatoires. Pourtant les aphasiques de Wernicke ont un débit très rapide, presque logorrhéique. Les propositions s'enchaînent sans construction logique, rendant le discours incompréhensible. L'écriture spontanée est également perturbée alors que l'écriture copiée est normale. Wernicke oppose cette aphasie « sensorielle » à celle de Broca qualifiée de « motrice ».

Wernicke postule aussi qu'il doit exister un troisième type d'aphasie provoqué par l'interruption des fibres cortico-corticales reliant les zones cérébrales attribuées à « l'émission » et à la « réception » du langage parlé. Il nomme ce troisième type : « aphasie de conduction ». Mise en évidence par la suite, l'aphasie de conduction est classiquement considérée comme un syndrome de déconnexion, impliquant des lésions du faisceau arqué ou de la partie postérieure et interne de la scissure de Sylvius. La vitesse d'élocution et la compréhension sont à peu près normales, des paraphrasies phonémiques sont observées, la répétition à haute voix et la lecture sont largement déficitaires. Anderson *et al.* (1999) avancent que les origines de cette aphasie ne sont pas liées à une rupture du faisceau arqué mais à un dysfonctionnement cortical. Damasio & Damasio (1980) constatent, quant à eux, des atteintes du gyrus supérieur temporal, du gyrus supramarginal et du faisceau arqué. Ces constatations montrent que les troubles aphasiques ne sont pas clairement circonscrits (Wilshire & McCarthy, 1996) ; déjà à la fin du 19^{ème} siècle, Lichtheim, (1885) propose un schéma illustrant les différents types d'aphasie. Sont représentés (voir figure A2) trois centres (M : centre des images motrices des mots ; A : centre des images auditives des mots ; B : “*begriffe*”, centre d'élaboration conceptuelle) et leurs connexions (aA représente la voie afférente qui transmet les informations acoustiques au centre A ; Mm est la voie efférente qui conduit les informations venant de M aux organes de la parole). L'auteur postule l'existence de sept variétés d'aphasie consécutives aux destructions sélectives des différents centres ou de leurs voies de communication (voir figure A8), dont l'aphasie de Broca (1), l'aphasie de Wernicke (2) et l'aphasie de conduction (3), et s'y ajoutent quatre troubles supplémentaires : l'aphasie transcorticale motrice (4), l'anarthrie pure (5), l'aphasie transcorticale sensorielle (6) et la surdité verbale pure (7). Nous allons décrire brièvement ces quatre troubles en adoptant un point de vue contemporain.

- L'aphasie transcorticale motrice est associée à des lésions périphériques à l'aire de Broca ou aux alentours de l'aire motrice supplémentaire. La répétition et la compréhension sont

proches de la normale mais le langage spontané est affecté, des cas de mutisme sont aussi rapportés.

- L'anarthrie pure correspond à la perte totale de l'expression orale alors que la compréhension est préservée ; ce trouble est souvent décrit comme consécutif à des lésions sises dans le sillon central, c'est-à-dire la portion recouvrant l'insula.
- L'aphasie transcorticale sensorielle résulte, le plus souvent, d'une atteinte du lobe pariétal inférieur gauche, et engendre de grandes difficultés de compréhension et des paraphrasies sémantiques, alors que la vitesse d'élocution est normale.
- La surdit  verbale pure est caract ris e par un d ficit de l'identification et de la discrimination des sons langagiers, alors que toutes les autres capacit s linguistiques sont pr serv es, ainsi que la perception des sons non langagiers.

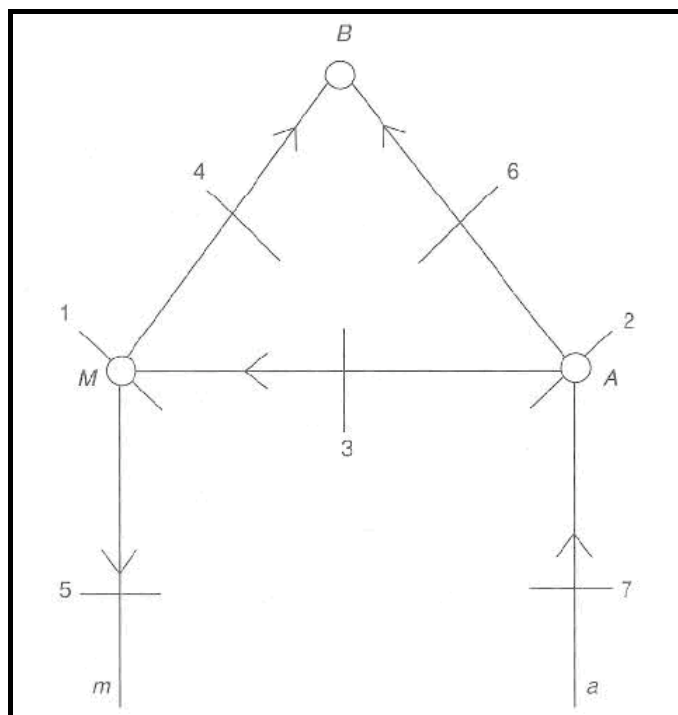


Figure A8 : Le schéma de Lichteim contenant les trois « modules » M : images motrices des mots ; A : images auditives des mots ; B : élaboration conceptuelle) et leurs connexions (aA : voie afférente ; Mm : voie efférente). Les troubles consécutifs : (1) aphasie de Broca, (2) aphasie de Wernicke (3) aphasie de conduction, (4) aphasie transcorticale motrice, (5) anarthrie pure, (6) aphasie transcorticale sensorielle, (7) la surdit  verbale pure.

A.3.1.1. L'apport de l'étude des cas pathologiques

Les premières données anatomocliniques ont permis de dégager un réseau d'aires cérébrales impliquées dans les processus du langage articulé (Price, 2000). Une partie de l'aire de Wernicke, le cortex auditif primaire (Aires de Brodmann 41 et 42), serait responsable des traitements acoustiques ; l'accès aux « images auditives des mots » serait effectué par la portion de l'aire de Wernicke en dehors du cortex auditif primaire correspondant à la partie postérieure de l'aire de Brodmann 22. Le faisceau arqué ferait le lien avec l'aire de Broca (Aires de Brodmann 44 et 45), laquelle contiendrait les « images motrices des mots » ; le cortex moteur serait chargé de la production articulatoire de la parole. Ce modèle, en partie valide, met en avant l'implication d'aires corticales, dans les processus langagiers, mais la variété des observations anatomocliniques oblige à rester prudent, à compléter ledit modèle. Ces découvertes ont pesé en faveur d'une dominance de l'hémisphère gauche dans les processus langagiers. Désormais, le concept de « latéralisation hémisphérique » est préférentiellement utilisé.

En effet, Metz-Lutz & Dahl (1984) ont observé un cas de surdit  verbale pure qui est une agnosie limit e   l'identification des symboles sonores du langage. Les l sions c r brales sont g n ralement localis es dans le gyrus temporal sup rieur dans l'h misph re gauche, ou bilat ralement. La surdit  verbale pure s'observe chez des patients dont l'examen audiom trique est quasiment normal, mais qui ne peuvent plus comprendre le langage parl , ni r p ter des mots, alors qu'ils sont encore capables de discriminer les sons de la parole des sons non langagiers. Des d ficits dans les t ches de discrimination ou d'identification de phon mes sont constat s (Saffran, Marin & Yeni-Komshian, 1976 ; Denes & Semenza, 1975), alors que les t ches d'identification des sons de l'environnement, des instruments de musique, et des cris d'animaux demeurent intactes. Cette pathologie, extr mement rare, met en exergue des r seaux dissoci s pour le traitement phonologique et le traitement s mantique (Poeppel, 2001). Hickok & Poeppel (2007) citent deux articles (Miceli, Gainotti, Caltagirone & Masullo, 1980 ; Baker, Blumstein & Goodglass, 1981), appuyant une dissociation entre le traitement syllabique et la compr hension de la parole. Selon les auteurs, la capacit    identifier ou   discriminer des syllabes est diff renciable de la capacit    comprendre des mots audibles. Il existe en effet des patients incapables de comprendre la parole, alors qu'ils r ussissent des t ches de discrimination syllabique. En outre, le d ficit d'habilit    percevoir

les sons de la parole n'a qu'un impact minime sur la compréhension auditive dans l'aphasie de Wernicke (Basso, Casati & Vignolo, 1977 ; Miceli *et al.*, 1980), et la destruction du gyrus temporal supérieur gauche n'impliquerait pas de déficits dans la compréhension auditive, mais dans la production de la parole (Damasio & Damasio, 1980). Les dommages des régions temporales supérieures, particulièrement le long du gyrus temporal moyen, ont longtemps été associés aux déficits de compréhension auditive (Damasio, 1991 ; Dronkers, Redfern & Knight, 2000 ; Bates *et al.*, 2003). Les données issues d'études de stimulation corticale directe corroborent l'implication du gyrus temporal moyen dans la compréhension auditive, mais indiquent aussi un réseau plus large pouvant impliquer la partie supérieure du lobe temporal gauche (incluant les portions antérieures) et la partie inférieure du lobe frontal (Miglioretti & Boatman, 2003). Dans un même temps, les études neuropsychologiques montrent que les dommages des aires frontales et pariétales inférieures de l'hémisphère gauche provoquent des déficits dans des tâches requérant la discrimination ou l'identification de syllabes naturelles (Blumstein, Baker & Goodglass, 1977a ; Blumstein, Cooper, Zurif & Caramazza, 1977b ; Caplan, Gow & Makris, 1995). Ces données mettent en lumière le rôle possible d'un réseau fronto-pariétal dans la perception de la parole. Notons que ces lobes sont fortement interconnectés (Wise, Boussaoud, Johnson & Caminiti, 1997).

De plus, la capacité à réaliser des tâches de discrimination ou d'identification est dissociable de la capacité à comprendre des mots présentés oralement. Jacquemot, Dupoux, Pallier & Bachoud-Lévi (2002) reportent un cas d'aphasie, suite à une ischémie du cortex temporo-pariétal de l'hémisphère gauche, induisant une incapacité à discriminer des phonèmes, alors que la discrimination lexicale est partiellement préservée. La surdit  verbale pure, les donn es issues de tests de Wada (McGlone, 1984), mais aussi de patients avec des l sions unilat rales ou ayant subi une section du corps calleux ("*split-brain*"; Zaidel, 1985), indiquent qu'il existe probablement au moins un r seau dans chaque h misph re qui peut traiter les sons langagiers suffisamment bien, afin d'acc der au lexique mental. La grande variabilit  des troubles aphasiques et des l sions c r brales qui les engendrent a contraint les observateurs   des regroupements de sympt mes en syndromes. Cette d marche classique en m decine ne doit pas laisser penser aux chercheurs que les troubles aphasiques sont cons cutifs d'atteintes pr cises entra nant syst matiquement des effets semblables. La m thode anatomoclinique a fourni les bases permettant de dissocier les r les fonctionnels de zones c r brales, n anmoins la plasticit  c r brale est telle qu'il faut accorder une attention

particulière, lors des études des patients aphasiques, au temps séparant l'apparition des symptômes liés à la lésion et la date de leur observation (Bates *et al.*, 2001). De plus, circonscrire l'étendue d'une lésion du fait de l'importante variabilité inter-individuelle de l'anatomie cérébrale n'est jamais chose aisée. Dronkers (1996) observe, à l'aide de la tomodensitométrie et de l'IRM, vingt-cinq patients souffrant d'apraxie de la parole (induisant une incapacité à produire de la parole articulée tout en conservant la perception). En le comparant à dix-neuf sujets sains, l'auteure met en lumière l'importance d'une structure profonde du cortex : l'insula de Reil (du nom du médecin qui l'a décrite). Le cortex insulaire est placé dans le sillon rolandique en-dessous de l'aire de Broca. Cette proximité anatomique pourrait expliquer, en partie, la confusion entre les rôles de Broca et ceux du cortex insulaire. L'exemple emblématique de Leborgne a été contesté en particulier par Castaigne, Lhermitte, Signoret & Abelanet (1980) cités dans l'article d'Abry, Stefanuto, Vilain & Laboissière (2002). Après examen au scanner du cerveau de Leborgne, Castaigne observe que la lésion dépassait les limites des aires 44 et 45 de Brodmann en largeur et en profondeur, affectant aussi le cortex insulaire. Dronkers (1996) pense que l'insula pourrait être spécialisée dans la planification motrice de la parole articulée.

L'aphasiologie et l'étude des cas pathologiques dans son ensemble ont permis d'ébaucher un panorama des processus cognitifs. Les techniques d'imagerie in vivo ont fourni de nouvelles données complétant les modèles disponibles, sans pour autant apporter de réponses définitives. La variabilité anatomique, mais aussi celle présente dans les tâches expérimentales, voire dans l'exploitation des données, ne facilitent pas l'analyse et la synthèse des résultats.

A.3.2. Les études en Tomographie par Émission de Positons

Parmi les premières études s'intéressant à la parole à l'aide de la Tomographie par Émission de Positons, certaines mettent en lumière un recrutement bilatéral des régions temporales supérieures. Houdé, Mazoyer & Tzourio-Mazoyer (2002) nous rappellent que Petersen, Fox, Posner, Mintun & Raichle (1988) réalisent l'une des premières expériences en TEP. L'écoute passive des stimuli auditifs engendre une augmentation des flux sanguins cérébraux régionaux dans le cortex auditif primaire (Aires de Brodmann 41 et 42), alors que la vision passive révèle que cette augmentation a lieu dans le lobe occipital. La production

orale repose sur les gyri précentraux bilatéraux, l'aire motrice supplémentaire (Aire de Brodmann 6) et l'insula antérieure gauche (située dans le sillon central). L'aire motrice supplémentaire est située dans la partie médiane de l'aire 6, et elle serait principalement impliquée dans la planification de mouvements complexes, ainsi que dans la coordination de mouvements impliquant les deux mains. Dans le domaine de la parole, cette aire corticale participerait au contrôle de la rythmicité de la parole propositionnelle et automatique, et au contrôle de la source glottique (Jonas, 1981). Petersen *et al.* (1988) n'ont pas observé d'activations correspondant aux aires de Broca et de Wernicke. Ces résultats seront contestés, peu de temps après, par l'étude de Wise *et al.* (1991) qui relève, pour des tâches similaires, des activations dans les aires de Broca et de Wernicke. Selon ces auteurs, ces différences proviennent de la condition de référence utilisée par l'équipe de Petersen, c'est-à-dire la condition d'écoute passive qui activait les aires de Broca et de Wernicke, ne permettant pas de les faire apparaître après la soustraction.

Une étude de Démonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, Rascol & Frackowiak (1992) a comparé l'écoute de sons purs, de phonèmes et de mots. Leur but était d'observer le traitement acoustique par les sons purs, le traitement phonologique grâce aux non-mots et le traitement sémantique *via* l'écoute des mots. La comparaison entre les traitements acoustique et phonologique active la partie supérieure du lobe temporal gauche, ainsi que le gyrus frontal inférieur. La comparaison entre les tâches phonologique et sémantique révèle des activations moins importantes du gyrus frontal inférieur, alors que les activations temporales étaient plus étendues par rapport à la comparaison précédente, allant jusqu'aux gyri angulaire (Aire de Brodmann 39) et supramarginal (Aire de Brodmann 40). L'élément central ressortant de cette étude est l'implication de l'aire de Broca dans des tâches de perception. Les auteurs expliquent cette activation comme reflétant l'accès à la « représentation articulatoire » des phonèmes, étape selon eux indispensable au décodage des sons langagiers.

Mazoyer, Tzourio, Frak, Syrota, Murayama, Levrier, Salamon, Dehaene, Cohen & Mehler (1993) mènent une étude, sur 16 hommes, contrastant l'écoute passive d'une histoire dans une langue étrangère (le tamoul) avec des mots français mais aussi des phrases constituées de logatomes, des phrases sémantiquement incongrues et à nouveau une histoire mais en français. Le gyrus temporal supérieur est recruté dans les deux hémisphères pour toutes ces conditions. L'écoute passive de mots français entraîne l'activation de l'aire de Broca, les

phrases avec les logatomes recrutent l'homologue droit de Broca, et la partie antérieure du gyrus temporal moyen de façon bilatérale (Aire de Brodmann 38). Cette dernière est aussi recrutée lors de l'écoute de phrases sémantiquement incongrues, alors que les phrases de l'histoire en français provoquent des activations à la fois dans l'aire de Brodmann 38 et dans les aires temporelles moyennes (Aires de Brodmann 21 et 37), mais aussi dans l'aire de Broca.

Douze personnes ont participé à l'étude de Price, Wise, Warburton, Moore, Howard, Patterson, Fracowiak & Friston (1996) qui se sont attachés à éclaircir les différentes implications des aires de Broca et de Wernicke, lors de la perception auditive passive de mots et durant leurs répétitions à voix haute. Ils relèvent des activations des aires de Brodmann 44 et 45 dans les deux tâches, mais ces activations sont supérieures lors de la répétition de mots et durant l'écoute passive à un faible débit. Ils constatent, d'autre part, que les aires temporelles supérieures sont recrutées aussi bien par la voix du locuteur que par celle d'une personne inconnue.

Ces quelques études mettent en lumière deux limites des données anatomocliniques, le recrutement exclusif de l'aire de Broca pour la production et de l'aire de Wernicke pour la perception de la parole, ainsi que la prépondérance de l'hémisphère gauche chez la plupart des individus. L'implication de l'hémisphère droit dans les traitements langagiers demeure encore fortement associée au domaine de la prosodie. Baum & Pell (1999) citent des chercheurs qui appuient cette idée depuis de nombreuses années (Weintraub, Mesulam & Kramer, 1981 ; Bradvik, Dravins, Holtas, Rosen, Ryding & Ingvar, 1991), alors que d'autres plaident pour un traitement des indices prosodiques répartis dans les deux hémisphères (Mayer *et al.* 2002 ; Warren & Griffiths, 2003 ; Gandour *et al.*, 2004 ; Zhao *et al.*, 2008), voire pour un traitement, en partie, subcortical (Bonder, Gur & Gur, 1989 ; Breitenstein, Daum & Ackerman, 1998). Zatorre, Evans, Meyer & Gjedde (1992) sont parmi les premiers à étudier spécifiquement les rôles de l'aire de Wernicke. En comparant l'écoute de syllabes à celle de bruits synthétiques, ces chercheurs mettent en exergue que le cortex auditif primaire est activé dans les deux conditions, alors que les parties postérieures des gyri temporaux moyen et supérieur répondent spécifiquement à l'écoute syllabique. Lorsqu'il était demandé aux sujets de discriminer les syllabes en fonction du lieu d'articulation, c'est l'hémisphère droit qui était le plus activé ; *a contrario* quand il était demandé aux sujets de discriminer les syllabes en fonction de leurs fréquences, c'est l'hémisphère gauche qui était le plus impliqué. Les auteurs poursuivent leurs investigations en se focalisant sur la prosodie affective et linguistique. À

leurs yeux, l'hémisphère gauche est impliqué dans le traitement des informations temporelles, et l'hémisphère droit dans celui des informations spectrales (Zatorre, Evans & Meyer, 1994 ; Zatorre, Bouffard, Ahad & Belin, 2002). Certaines études sur la perception des indices temporels chez des aphasiques (Robin, Tranel & Damasio, 1990), ou comparant de la musique à la parole, pourraient appuyer ce point de vue (Tervaniemi & Hugdhal, 2003). Boemio, Fromm, Braun & Poeppel (2005) cherchent à confirmer la répartition des informations spectrales et temporelles à travers les deux hémisphères en soumettant quinze sujets à une tâche d'écoute passive de stimuli non langagiers (des tons opposés à des fréquences modulées) de différentes durées : 12 ms, 25 ms, 45 ms, 85 ms, 160 ms et 300 ms. Les gyri et sillons temporaux supérieurs s'activent, à la fois, pour le traitement d'indices temporels et spectraux, exception faite des stimuli durant 12 ms pour lesquels aucune activation n'est observée. Les stimuli de 25 ms et de 45 ms activent davantage l'hémisphère gauche, alors que pour les stimuli de 85 ms, 160 ms et 300 ms, c'est l'hémisphère droit qui est le plus activé. Ces résultats tendent à réfuter l'hypothèse de Zatorre *et al.* (2002). *A contrario*, Schönwiesner, Rubsamen & Von Cramon (2005), comparent, dans une étude portant sur dix-neuf personnes, la perception de stimuli ayant des caractéristiques temporelles et spectrales équivalentes à celles de la parole et à celles de la musique, sans pour autant que les sujets puissent associer les stimuli à de la parole ou à de la musique. Les données IRMf montrent des activations qui plaident en faveur d'une répartition hémisphérique, telle que postulée par Zatorre. En effet, les stimuli « spectraux » activent les gyri de Heschl, et le gyrus temporal supérieur de l'hémisphère droit ; alors que les stimuli « temporels » activent le gyrus temporal supérieur de l'hémisphère gauche.

L'utilisation de la TEP afin d'étudier les processus engagés dans la parole articulée revêt deux avantages essentiels : cette technique offre la possibilité de réaliser des tâches de production de la parole, et dans un environnement sonore non perturbé. Si l'IRMf a largement supplanté la TEP, cela est dû à la nécessité d'injecter un produit de contraste radioactif en TEP (par conséquent avoir un appareil qui génère ledit produit), alors que les expériences en IRMf n'en ont pas besoin, ce qui les rend plus aisées à mettre en place. Nous allons aborder quelques études IRMf, qui se sont intéressées à la perception auditive de la parole, puis à la dimension audiovisuelle.

A.3.3. Les études en Imagerie par Résonance Magnétique fonctionnelle

Binder, Rao, Hammeke, Yetkin, Jesmanowicz, Bandettini, Wong, Estkowski, Goldstein, Haughton & Hyde (1994), dans l'une des premières expériences en IRMf sur la parole articulée, comparent la perception de mots et de logatomes monosyllabiques, chez cinq sujets. Les activations se chevauchent, dans les deux hémisphères, sur une large portion du gyrus temporal supérieur, jusqu'au sillon temporal supérieur. Binder, Frost, Hammeke, Rao & Cox (1996) comparent la perception passive de mots monosyllabiques anglais à celles de sons synthétiques en IRMf, chez dix participants. Les mots et les sons synthétiques activent les régions temporales supérieures de l'hémisphère gauche, comprenant le cortex auditif primaire (Aires de Brodmann 41 et 42). Les mots activent davantage la région ventrale du gyrus temporal supérieur, alors que l'activation la plus forte, pour les tons, se situe dans le tiers postérieur du sillon temporal supérieur, au niveau de la jonction temporo-pariéto-occipitale.

Benson, Richardson, Whalen & Lai (2006) conduisent une étude sur vingt-cinq locuteurs anglais. Ils cherchent à préciser les différences anatomo-fonctionnelles entre les niveaux acoustique et phonétique. Ils ont choisi d'utiliser des stimuli synthétiques de deux formes, la première, composée de cinq logatomes monosyllabiques (CVC), est opposée à la seconde composée d'items issus des précédents, mais manipulés informatiquement afin de les rendre méconnaissables (inversion du décours temporel et déplacement des formants). Les stimuli synthétiques de parole sont utilisés afin d'étudier le niveau phonétique, alors que les stimuli synthétiques non langagiers permettent d'accéder au traitement acoustique. Les résultats mettent en lumière des zones à travers les deux hémisphères. L'étude du niveau acoustique révèle que les activations de l'hémisphère droit sont situées dans le gyrus temporal inférieur, alors que c'est le gyrus angulaire qui est recruté dans l'hémisphère gauche (Aire de Brodmann 39). Le niveau phonétique révèle des activations bilatérales dans les sillons temporaux supérieurs, et des activations unilatérales, à droite dans le gyrus temporal inférieur, et à gauche dans le gyrus temporal moyen. Ces résultats suggèrent que les parties moyennes des sillons temporaux supérieurs des deux hémisphères sont impliquées dans la perception phonémique et non pas uniquement dans le traitement des informations acoustiques complexes. Néanmoins, les auteurs précisent qu'il existe une asymétrie dans les recrutements des deux hémisphères corticaux, l'hémisphère gauche étant préférentiellement impliqué dans les traitements fréquentiels. Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman &

Passing (2000) montrent, chez six hommes et quatre femmes, que le traitement auditif de sons purs, comparé à celui de bruits, active les parties supérieures du gyrus temporal supérieur dans chaque hémisphère, incluant les aires auditives associatives et le *planum temporale* (c'est-à-dire la partie plane non visible sur la face externe des aires de Brodmann 41 et 42). Lors d'un tâche impliquant des mots et des logatomes, Obleser, Zimmermann, Van Meter & Rauschecker (2006) n'observent pas de différences dans les activations du *planum temporale*, parmi les quinze participants. Au terme de leur méta-analyse, Griffiths & Warren (2002) pensent que le *planum temporale* n'est pas un site spécifique aux processus langagiers, mais qu'il est utile pour tous les types de sons nécessitant une analyse spectrotemporelle. Jäncke, Wüstenberg, Scheich & Heinze (2002) mènent une étude chez vingt et un participants. Ils constatent, pour leur part, des activations bilatérales du *planum temporale*, qui sont plus importantes pour la détection de syllabes CV débutant par une consonne sourde, par rapport à celles débutant par une sonore. Notons que la localisation et les limites du *planum temporale* sont encore sujettes à discussion (Wetsbury, Zatorre & Evans, 1999).

Blumstein, Myers & Rissman (2005) ont réalisé une étude portant sur la perception du délai d'établissement du voisement ou VOT, chez douze participants. Les stimuli étaient constitués de cinq syllabes synthétiques avec une variation de VOT de 0 à 40 millisecondes, reproduisant un passage de [da] à [ta]. Les résultats sont répartis à travers les deux hémisphères, et montrent une activation plus importante, dans l'hémisphère gauche, des gyri cingulaire et frontal inférieur lorsque les stimuli sont proches de la frontière phonémique (20 ms). En revanche, les activations les plus fortes du gyrus temporal moyen et du gyrus angulaire de l'hémisphère gauche ont lieu lors de la présentation du meilleur représentant de la catégorie. Les auteurs signalent aussi que le gyrus temporal supérieur montre une faible sensibilité aux différences de catégorie phonémique, ce dernier serait principalement impliqué dans l'analyse précoce des propriétés sensorielles de la parole. Rimol, Specht, Weis, Savoy & Hugdahl (2005) ont testé dix-sept sujets sains à l'aide de trois types de stimuli. Il s'agissait en premier lieu d'occlusives naturelles extraites de syllabes CV. Ces consonnes seules étaient contrastées, d'une part, avec les paires CV et, d'autre part, avec deux types de bruit de durée et d'intensité équivalant aux syllabes CV, ou aux consonnes seules. La comparaison « consonne seule / bruit équivalent » fait apparaître des activations unilatérales à gauche, dans les parties postérieures du gyrus temporal moyen et du sillon temporal supérieur. La comparaison « syllabe CV / bruit équivalent » recrute, de manière bilatérale, les mêmes

régions avec une asymétrie gauche. Les comparaisons inverses « bruit / consonne » et « bruit / syllabe » mettent en exergue des activations à droite du gyrus temporal supérieur et du gyrus supramarginal, ainsi que des zones préfrontales. Pour les auteurs, le gyrus temporal moyen et le sillon temporal supérieur seraient impliqués dans les traitements phonétique et phonologique. Liebenthal, Binder, Spitzer, Possing & Medler (2005) s'intéressent, eux aussi, à l'organisation des lobes temporaux en conduisant une étude sur vingt-cinq locuteurs de l'anglo-américain. Lors d'une tâche de discrimination de syllabes synthétiques CV du type « ABX » (le troisième stimulus « X » est comparé avec les deux premiers « A » et « B »), opposées à des patrons auditifs complexes non assimilables à de la parole, les auteurs observent une réponse plus forte du sillon temporal supérieur de l'hémisphère gauche, au niveau des aires de Brodmann 21 et 22. Alors que la partie dorsale supérieure du gyrus temporal répond de façon équivalente aux stimuli phonologiques et non phonologiques, et ce, de façon bilatérale. Pour les auteurs, le sillon temporal supérieur gauche joue un rôle crucial dans la perception de la parole, constituant un intermédiaire entre les gyri temporaux supérieurs, impliqués dans l'analyse acoustique de la parole et dans celle d'autres sons complexes, et les aires antérieures du sillon temporal supérieur gauche et du gyrus temporal moyen, recrutées pour les traitements linguistiques supérieurs. Notons que, dans cette étude de Liebenthal *et al.* (2005), la variation en âge des participants (de 19 à 50 ans) pourrait être discutée. Behne *et al.* (1998) comparent la perception audiovisuelle de syllabes CV s'opposant sur le voisement chez des adultes répartis en deux groupes d'âge : les sujets « âgés » (49 à 60 ans) utiliseraient davantage les indices visuels que les « jeunes » (19 à 30 ans). Déjà Desjardins, Rogers & Werker (1997), ou encore Wunderlich, Cone-Wesson & Shepherd (2006), observent des différences entre les enfants et les adultes, appuyant l'idée que la perception des indices visuels s'affinent avec l'expérience linguistique, particulièrement en condition bruitée (Wong, 2009). Notons pour finir que certains auteurs contestent l'existence d'aires corticales spécifiques aux traitements du langage articulé préférant raisonner sur un ensemble de structures recrutées en fonction du contexte (Price, Thierry & Griffiths, 2005), une idée qui avait déjà été défendue par Ojemann (1991).

A.3.3.1. La parole audiovisuelle et le cortex auditif

Nous présentons ici des études ayant pour but d'éclaircir la contribution des indices visuels pertinents dans la perception de la parole. Une des principales questions est de savoir

si les stimuli visuels congruents engendrent des activations du cortex auditif primaire, ou s'il existe une zone corticale spécialisée dans l'intégration audiovisuelle générale, voire spécifique à la parole bimodale.

Les régions temporales inférieures pouvant être impliquées dans la reconnaissance des visages (chez le macaque : Desimone, Albright, Gross & Bruce, 1984 ; Perrett *et al.* 1985 ; chez l'être humain : Haxby, Hoffman & Gobbini, 2000 ; Tsao, Freiwald, Tootell & Livingstone, 2006), alors que le rôle de l'aire MT/V5 (AB 19), dans la perception des mouvements visuels, semble désormais bien établi (Zeki *et al.* 1991 ; Smith, Greenlee, Singh, Kraemer & Hennig, 1998). Puce, Allison, Bentin, Gore & McCarthy (1998) précisent que l'aire MT/V5 semble subdivisée en différentes zones fonctionnelles. Puce *et al.* (1998) constatent qu'elle réagit davantage à la présentation de mouvements des yeux qu'à celle de mouvements de la bouche. L'une et l'autre de ces études observent, de surcroît, des activations du gyrus temporal supérieur.

À notre connaissance, Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen & David (1997) sont les premiers à étudier le rôle du cortex auditif primaire (Aires de Brodmann 41 et 42 ; Gyrus de Heschl) lors de la perception visuelle de mouvements langagiers en IRMf. S'inspirant d'une étude de Sams *et al.* (1991) menée avec la Magnétoencéphalographie (technique voisine de l'EEG, mais qui enregistre les champs magnétiques cérébraux), Calvert *et al.* (1997) ont mené une recherche sur deux panels de cinq participants, tous droitiers. Une première phase contrastait deux conditions : d'abord des mots entendus qu'il fallait répéter silencieusement, puis des périodes sans stimulus, durant lesquelles les sujets répétaient le mot « one ». La comparaison de ces deux conditions montre des activations bilatérales, mais prédominantes à gauche, des aires de Brodmann 41, 42 et 22. Durant la seconde phase, les mêmes participants voyaient soit un visage prononçant des nombres sans la source auditive, soit un visage fixe : là encore la condition contrôle consistait à répéter le nombre vu ou le chiffre « 1 ». Les régions activées par la soustraction sont le gyrus angulaire (Aire de Brodmann 39), le gyrus fusiforme (Aire de Brodmann 37) et le cortex visuel (Aire de Brodmann 19), mais aussi les aires auditives (Aires de Brodmann 22, 41 et 42). Au vu de ces résultats, Calvert *et al.* (1997) poursuivent l'étude avec cinq nouveaux participants. Reprenant la condition des nombres vus, ils proposent deux nouvelles conditions de contrôle, la vue d'un visage fixe est remplacée soit par la vue d'un visage grimaçant, soit par celle d'un visage prononçant un pseudo-mot. Les activations des pseudo-mots vus sont

similaires à celles des nombres vus, avec de surcroît des activations bilatérales de l'insula. Les grimaces activent des aires du cortex cingulaire (Aires de Brodmann 30 et 32), situées sur la face interne et considérées comme étant impliquées dans l'attention (Mellers *et al.*, 1995). Ces résultats indiquent, selon les auteurs, que les indices linguistiques pertinents modulent la perception de la parole au niveau prélexical, mais aussi probablement au niveau phonétique. Le réseau que constituent les aires activées au cours de cette étude est largement distribué. Pour la perception des mots entendus, les lobes temporaux sont activés (Aires de Brodmann 41, 42, 22 et 37), la perception visuelle des mots articulés recrute en plus les lobes pariétal (Aire de Brodmann 39) et occipital (Aire de Brodmann 19), alors que les logatomes induisent une activation bilatérale de l'insula. Les résultats de cette étude posent un débat qui demeure encore ouvert sur l'implication du cortex auditif primaire (Aires de Brodmann 41 et 42) dans la perception d'indices visibles linguistiquement pertinents.

Dans le même esprit, Pekkola *et al.* (2004) conduisent une étude à laquelle participaient dix sujets dont sept masculins. Tous étaient des locuteurs du finnois, avec une préférence manuelle droite. Les stimuli consistaient en la présentation du visage d'une femme, articulant silencieusement quatre voyelles du finnois : [a o i y]. La condition contrôle présentait le même visage, avec des cercles bleus sur la bouche, se déplaçant selon quatre orientations différentes : verticale, horizontale, oblique droite, oblique gauche. La présentation statique du même visage constituait la condition de référence. Les sujets connaissaient les voyelles du corpus et pouvaient s'entraîner en dehors du scanner. La consigne expérimentale était de cliquer lorsque deux voyelles identiques se suivaient ou lorsque deux cercles allaient dans la même direction. Les voyelles ont été correctement identifiées à 94 % ($\pm 7\%$) et les cercles à 91 % ($\pm 2.3\%$). Les auteurs ont utilisé l'étude de Penhune, Zatorre, MacDonald & Evans (1996), afin de localiser les limites du cortex auditif primaire. Pour neuf sujets sur dix, la perception visuelle des gestes articulatoires active le cortex auditif primaire davantage que la perception visuelle des cercles mouvants. Ces activations sont bilatérales avec une prédominance de l'hémisphère gauche. Notons qu'il ne s'agit pas d'analyse de groupe.

A contrario, Bernstein *et al.* (2002) contestent l'activation du cortex auditif primaire lors de la perception visuelle de la parole. Ils considèrent que les études, rapportant de telles activations, se trompent sur la localisation du cortex auditif primaire. Une expérience avec sept sujets droitiers a été mise en place. Les participants étaient de langue maternelle anglaise,

moyens ou bons lecteurs labiaux. Cette étude comprenait deux parties. La première proposait des tons purs de 1000 Hz, d'une durée de 100 millisecondes, cela ayant pour but de localiser précisément le cortex auditif primaire. Il y avait cinq tons par seconde durant 30 secondes, contrastés avec des périodes sans stimuli, où seul le bruit présent était celui des gradients de l'IRM (cf. B.1.2.). Il s'agissait d'écouter de façon passive les stimuli. La seconde partie concernait la perception visuelle de la parole ; une session de lecture labiale, présentant 120 monosyllabes contrastés avec 120 images du visage du locuteur, sur lequel étaient superposées des formes colorées (triangle vert, cercle jaune, étoile rouge, pentagone pourpre et carré bleu). La consigne expérimentale était alors de cliquer lorsque deux items identiques se suivaient. La stimulation par les tons purs active, de façon bilatérale, le *planum temporale* supérieur, le cortex auditif primaire et la part latérale du gyrus temporal supérieur. La session de lecture labiale ne semble pas activer des zones cérébrales supposées recouvrir les localisations du cortex auditif primaire, identifiées dans la phase précédente. Les activations de la session en lecture labiale étaient plus étendues, et incluaient des aires frontales, là encore de façon bilatérale. Les pics d'activation du lobe temporal étaient concentrés dans le sillon temporal supérieur, et le gyrus temporal moyen. Des pics additionnels ont été observés dans le gyrus frontal inférieur, moyen et supérieur. Les auteurs considèrent que ces résultats infirment le recrutement du cortex auditif primaire, lors de la perception visuelle de la parole, mais plaident en faveur d'un recrutement du sillon temporal supérieur dans l'intégration audiovisuelle.

Hall, Fussell & Summerfield (2005) reprennent cette hypothèse, avec un panel de trente-trois sujets, en présentant des phrases prononcées de façon naturelle qui mettent en exergue la richesse de la parole audiovisuelle. Quatre conditions différentes (écran blanc, visage statique, grimaces non linguistiques, parole auditive seule ou parole visuelle seule) permettent aux auteurs de hiérarchiser les traitements neuraux impliqués dans la lecture labiale, et de tester l'hypothèse selon laquelle les stimuli visuels accèdent à des « représentations sonores » dans le cortex auditif. En comparaison avec la vision passive d'un écran blanc, la condition « visage fixe » provoque des activations bilatérales autour de la limite du gyrus fusiforme (Aire de Brodmann 37), du cortex visuel (Aires de Brodmann 18 et 19) et du cervelet, ainsi que dans le milieu du gyrus frontal supérieur (Aire de Brodmann 6) et du gyrus précentral gauche (Aire de Brodmann 4). Avec le visage fixe comme comparaison, la « grimace » provoque des activations bilatérales de la jonction temporo-occipitale jusqu'à

l'aire MT/V5, sensibles aux mouvements (Aire de Brodmann 19). Pour la parole visuelle comparée aux grimaces, les auteurs observent des activations dans le gyrus temporal moyen (Aire de Brodmann 21), dans les gyri frontaux inférieur (Aires de Brodmann 44, 45 et 47) et moyen (Aire de Brodmann 9), et dans le lobe pariétal inférieur (Aire de Brodmann 7). La majorité des groupes de voxels ont des proportions deux fois supérieures dans l'hémisphère gauche. Ces régions pourraient être impliquées, selon Hall *et al.* (2005), dans le traitement lexical de la parole. Comparée aux périodes de repos, la parole audible génère des activations bilatérales à travers les gyri temporaux moyen (Aire de Brodmann 21) et supérieur (Aires de Brodmann 41, 42 et 22) ; les patrons d'activation de l'étude de groupe pour la lecture labiale n'activent pas les régions auditives le long du gyrus temporal supérieur, ce qui signifierait que la parole visuelle fluente ne recrute pas toujours le cortex auditif primaire. Un résultat important est que l'étendue des activations du gyrus temporal supérieur du côté gauche est fortement corrélée à la performance en labio-lecture. Il est possible que les labiolecteurs experts développent un accès facilitateur aux « représentations phonologiques ». Il serait intéressant d'associer les capacités en lecture labiale avec l'âge des participants. Le point important est que le recrutement du cortex auditif primaire ne semble pas essentiel.

A.3.3.2. Un site d'intégration audiovisuelle

Callan *et al.* (2003) défendent l'hypothèse inspirée, entre autres, de Scott, Blank, Rosen & Wise (2000), et considèrent que les gyrus et sillon temporaux supérieurs pourraient être les sites d'intégration des flux auditif et visuel. Six anglophones droitiers ont participé à cette expérience. Les stimuli étaient constitués de 96 monosyllabes anglais, du type : groupe consonantique + voyelle + groupe consonantique. Toutes les syllabes étaient prononcées par une unique locutrice de langue anglaise. Il y avait cinq conditions différentes : audio seule, audio seule bruitée, audiovisuelle, audiovisuelle bruitée et visuelle seule. La perception des indices visuels linguistiques en condition bruitée et en visuel seul augmente les performances d'identification, ainsi que l'activité neurale de zones corticales supposées accomplir l'intégration multimodale, c'est-à-dire le cortex auditif primaire, le gyrus temporal moyen, les gyrus et sillon temporaux supérieurs, ainsi que des zones considérées, par les auteurs, comme étant impliquées dans la planification et l'exécution de la parole, comme l'aire de Broca, le cortex prémoteur et l'aire motrice supplémentaire. Le gyrus temporal moyen semble être activé également en condition bruitée tout comme les cortex auditif gauche et droit. Malgré

cela, la comparaison audiovisuelle et audio seule ne fait pas apparaître de différences d'activations significatives hormis dans l'aire visuelle V5. Pour les auteurs, ces résultats semblent indiquer que les indices visuels pourraient être traités par un système additionnel dépendant de l'intégrité du signal auditif. Olson, Gatenby & Gore (2002) penchent pour un rôle plus large des gyrus et sillon temporaux supérieurs qui seraient davantage impliqués dans l'analyse des mouvements biologiques que dans l'intégration bimodale. Ces chercheurs ont mené une étude sur dix volontaires en IRMf, avec deux conditions uniquement visuelles et deux conditions AV non congruentes (type McGurk), dont l'une présentait un signal sonore désynchronisé. Les stimuli étaient prononcés par une locutrice, et ce dans les quatre conditions : visuelle statique, visuelle animée (présentant une vidéo avec un mot toutes les trois secondes) AV synchronisée et AV désynchronisée. La comparaison des deux conditions visuelles (animée vs. statique) active l'homologue droit de l'aire de Broca, le gyrus temporal supérieur gauche et les gyrus et sillon temporaux supérieurs droits. Les auteurs précisent qu'en élevant le seuil statistique ($p < 0.01$) seuls les gyri temporaux supérieurs droit et gauche restent activés. Le résultat le plus intéressant avancé par les auteurs est issu de la comparaison AV qui montre une activation de la région du *claustrum* gauche (fine couche de matière grise située aux abords de l'insula). Ces résultats, selon les auteurs, plaident en faveur de la théorie selon laquelle les sites de traitements unimodaux seraient aussi les sites d'intégrations bimodales *via* des relais subcorticaux. Miller & D'Esposito (2005) mènent une étude de perception de syllabes CVC en IRMf sur 11 femmes. Utilisant des stimuli concordants et de type « McGurk », ils constatent que le sillon temporal supérieur de l'hémisphère gauche est davantage activé par les stimuli concordants. Les stimuli de type « McGurk » recrutent aussi le sillon temporal supérieur, mais semblent de plus susciter une activation spécifique du sillon pariétal inférieur. Jones & Callan (2003) trouvent eux aussi des activations plus fortes pour des stimuli de type « McGurk » dans le lobe pariétal postérieur par rapport à des stimuli congruents.

Douze droitiers anglophones âgés ont participé à l'étude de Wright, Pelphrey, Allison, McKeown & McCarthy (2003) qui cherchent à spécifier les activations du sillon temporal supérieur pour trois modalités différentes (audiovisuelle, audio seule et visuelle seule). Un écran présentait un visage synthétisé, des épaules jusqu'au sommet du crâne, articulant des mots monosyllabiques. Les auteurs ont pratiqué une analyse à l'aide de huit régions d'intérêt, quatre par hémisphère (Sillon Intra Pariétal, Sillon Temporal Supérieur, Gyris Temporal

Supérieur et Gyrus Temporal Moyen). Ils constatent des activations bilatérales des gyri et sillons temporaux supérieurs pour les trois modalités. Ces activations sont supérieures en modalité audiovisuelle comparées aux présentations unimodales visuelle et auditive. En modalité visuelle, les activations les plus importantes se retrouvent dans la partie postérieure du sillon temporal supérieur de l'hémisphère droit, en dehors du cortex auditif primaire. Ils relèvent aussi une activation, en-dehors des régions d'intérêt, de l'aire MT/V5 (Aire de Brodmann 19) du lobe occipital qui est impliquée dans le traitement visuel du mouvement. Notons que l'utilisation d'un visage synthétique peut avoir des conséquences sur les tâches expérimentales, si les sujets ne sont pas informés de la nature linguistique des stimuli (Tuomainen, Andersen, Tiippana & Sams, 2005). Néanmoins, ces données indiquent un recrutement bilatéral et un chevauchement des aires corticales recrutées par la perception d'indices linguistiques audibles et visibles. De plus, elles sont en défaveur de l'hypothèse d'un recrutement du cortex auditif primaire, lors de la perception de la parole visuelle, celle-ci activant principalement la partie postérieure du sillon temporal supérieur. Beauchamp, Lee, Haxby & Martin (2003) ont comparé la perception visuelle de mouvements humains naturels à celle de mouvements équivalents mais présentés sous forme de points lumineux, ainsi que l'utilisation par un humain d'un outil (un marteau) et sa contrepartie sous forme lumineuse. Le sillon temporal supérieur semble répondre davantage aux stimuli « mouvements » naturels et lumineux. Alors que le gyrus temporal moyen et le sillon temporal inférieur seraient plus activés lors de la vision des objets. Le cortex temporal ventral semble subdivisé en deux : le gyrus fusiforme latéral serait recruté pour les stimuli « humains », alors que le fusiforme médian le serait pour les stimuli « objets ». Les faibles réponses observées avec les points lumineux seraient le signe, pour les auteurs, que le cortex temporal ventral est sensible aux indices de forme, de couleur et de texture. Le cortex temporal latéral, au niveau du gyrus temporal moyen, répond plus fortement aux points lumineux qu'aux vidéos suggérant que le mouvement est traité dans ce gyrus. Le sillon temporal supérieur répondant de façon significative aux deux types de stimulations, les auteurs pensent que ce sillon est responsable à la fois de l'intégration de la forme et la couleur, aussi bien que de celle des mouvements. Beauchamp, Lee, Argall & Martin (2004) précisent que des portions différentes (1 à 2 mm) du sillon temporal supérieur reçoivent des stimulations auditives et visuelles qui sont ensuite transmises vers une troisième partie « multisensorielle ».

Finalement, Calvert & Campbell (2003), dans une étude IRMf, se sont penchées sur les caractéristiques statiques et dynamiques du langage parlé. Certains gestes visibles peuvent être capturés dans des séquences d'images fixes. De précédentes études ont rapporté des activations des régions temporales supérieures chez des sujets voyant des stimuli visuels naturels de durées variables. Cette étude explore dans quelle mesure ces circuits sont activés par des stimuli visuels desquels les variables temporelles ont été évacuées. Dans le scanner, les participants ont reçu des instructions afin de se concentrer sur des cibles visuelles pré-spécifiées (« voo » & « ahv ») dans un corpus contenant d'autres syllabes. Dans la première condition, les séquences d'images comprenaient une série de plans fixes cibles montrant des gestes apicaux. Dans l'autre condition, des mouvements visibles de parole naturelle de durées équivalentes ont été proposés. En contraste, la condition de base présentait un visage au repos sur lequel la lettre « V » était superposée. Des activations ont été notées dans les régions postérieures associées à la perception de mouvements biologiques, et ce malgré le manque de mouvements apparents dans les séquences vidéo de parole. Des activations ont également été détectées dans les régions traditionnelles de traitement de la parole, incluant le cortex frontal inférieur gauche (Aire de Broca), le sillon temporal supérieur gauche et le gyrus supramarginal. Les images de parole « fixe » produisent des activations dans le cortex ventral prémoteur et le sillon pariétal inférieur antérieur, de façon bilatérale. Les visages mobiles engendrent significativement plus d'activations corticales que les séquences arrêtées, dans des régions similaires. Néanmoins, des différences ont été observées entre les deux types d'images. Dans le cortex visuel, les images fixes produisent relativement plus d'activations dans les régions visuelles primaires (V1/V2), pendant que les aires des mouvements visuels V5 sont préférentiellement recrutées pour les images animées. Dans les régions corticales activées davantage par les gestes visibles de parole naturelle, se trouvaient le cortex auditif (Aires de Brodmann 41 et 42), ainsi que le sillon temporal supérieur et le gyrus frontal inférieur gauches. La parole visible avec des variables de durée semble recruter préférentiellement les régions de « pur » traitement auditif spécialisées dans le langage, peut-être afin d'atteindre le sillon temporal supérieur, siège d'un possible mécanisme d'intégration audiovisuelle dynamique. Quand les variables de durée sont extraites de la parole naturelle, l'accès au système de traitement de la parole dans le lobe temporal gauche doit être réalisé *via* des « représentations » d' "*action-base*" de la parole, recrutant le cortex prémoteur ventral.

A.3.3.3. Perception de la parole et cortex moteur

Un nombre croissant d'études anatomo-fonctionnelles rendent compte, lors de tâches perceptives, d'activations d'aires cérébrales traditionnellement impliquées dans la production de la parole. Poldrack *et al.* (1999) observent en IRMf des activations des aires de Brodmann 44 et 45, lors de tâches de perception de mots et de pseudo-mots, alors que des tâches de compréhension activent davantage l'aire 47 et la part ventrale de l'aire 45. Paulesu *et al.* (2003) pensent, quant à eux, que la zone de Broca est recrutée pour le traitement de la lecture labiale. Ojanen *et al.* (2005) considèrent qu'elle serait responsable de l'intégration des sources auditive et visuelle, lors de la perception de la parole.

Pour Hsieh, Gandour, Wong & Hutchins (2001), qui mènent une étude sur différents indices linguistiques proposés à des sujets Chinois et Anglais, les activations de l'aire de Broca sont influencées par l'expérience linguistique des auditeurs. Friederici, Meyer & Von Cramon (2000), ou encore Grodzinsky & Friederici (2006), avancent qu'elle serait aussi recrutée dans des processus syntaxiques. Fiebach, Friederici, Müller & Von Cramon (2002) associent les activations de la partie supérieure de l'aire de Brodmann 44 à la conversion grapho-phonémique, et l'aire de Brodmann 45 à la sélection dans le lexique mental. Pour Scott & Johnsrude (2003), l'aire de Broca devrait englober un large réseau comprenant le cortex préfrontal et prémoteur, et seraient spécifiquement impliquée dans le traitement de la parole articulée, en particulier dans la segmentation en phonèmes.

Wilson, Pinar Saygin, Sereno & Iacoboni (2004) conduisent une étude IRMf sur le rôle du cortex moteur dans la perception de la parole, avec dix participants. Ils débutent par une tâche d'écoute passive de monosyllabes sans signification (/p a/ /g i/), contrastés avec du bruit blanc de même durée. Ensuite les sujets devaient soit bouger les doigts des deux mains, soit répéter silencieusement les monosyllabes précédemment entendus. Cette phase motrice a été utilisée pour localiser les aires impliquées dans la perception et dans la production de la parole, en vue d'établir des régions d'intérêt à partir de l'analyse individuelle de chacun des sujets. Les auteurs notent qu'ils n'observent pas de différence consistante entre les activations de deux monosyllabes (/p a/ /g i/). En plus, des aires habituellement impliquées dans la perception, la tâche d'écoute passive active un réseau bilatéral, comprenant la partie supérieure ventrale du cortex prémoteur, recouvrant la partie antérieure de l'aire de Brodmann 4, et la partie supérieure de l'aire de Brodmann 6.

Skipper, Nusbaum & Small (2005) ont eux aussi étudié les différences entre trois modalités, non pas en utilisant des syllabes ou des mots, mais des histoires d'une durée comprise entre 18 et 24 secondes. Sur les douze sujets que comprenait cette étude, seuls neuf ont été pris en compte pour les analyses. Dans la condition audiovisuelle, les participants voyaient et entendaient une personne raconter les histoires, puis ils l'entendaient sans la voir, et la voyaient ensuite privée de source auditive. En comparaison avec la ligne de base (événement nul), la perception audiovisuelle et la perception auditive seule engendrent des activations des aires temporales (gyrus et sillon temporaux supérieurs), mais aussi des aires frontales (gyrus frontal inférieur et cortex prémoteur). Les activations observées pour la perception audiovisuelle sont bilatérales, alors que seul le gyrus temporal supérieur est recruté par la perception auditive seule, dans l'hémisphère droit. La présentation visuelle seule active le gyrus frontal inférieur, de façon bilatérale, ainsi que le gyrus temporal inférieur gauche. Pour les auteurs, ces résultats appuient l'idée que la perception de la parole recrute des zones impliquées dans la production de la parole.

Pulvermüller, Huss, Kheriff, Moscoso Del Prado, Hauk & Shtyrov (2006) ont étudié, chez douze personnes, les activations de différentes parties du cortex moteur, lors de la perception de syllabes labiales ou coronales. Ils ont, tout d'abord, mené une tâche de localisation « motrice » durant laquelle les sujets devaient simplement bouger les lèvres, puis la langue. Ensuite, les participants devaient prononcer silencieusement des syllabes CV débutant, soit par le phonème [t], soit par le phonème [p], cela correspondant à la phase « articulatoire ». Les auteurs ont entrepris une comparaison avec une phase « perceptive », durant laquelle les mêmes sujets entendaient des syllabes de même type, ainsi que des bruits non langagiers. Les mouvements labiaux et linguaux activent fortement les cortex sensorimoteurs des deux hémisphères, tout comme la production silencieuse de syllabes qui génèrent en plus des activations des aires frontales inférieures. La perception syllabique active, pour sa part, les gyri temporaux supérieurs bilatéralement. La phase motrice ne présente pas d'asymétries significatives entre les deux hémisphères. Cependant, les activations des gyri pré- et post-centraux liées aux mouvements de la langue sont plus dorsales que celles liées aux mouvements des lèvres. Pour la phase « articulatoire », les activations de l'hémisphère gauche sont plus étendues que celles observées dans l'hémisphère droit. Là encore, les auteurs relèvent des activations pour le [t] dans les aires précentrales inférieures, alors que celles du [p] sont davantage situées dans le cortex moteur et prémoteur,

dorsalement à celles du [t]. Ces différences entre les activations des labiales, par rapport à celles de coronales, pourraient être liées aux indices visuels pertinents, ces articulations appartenant à des classes de visèmes distinctes (Mourand-Dornier, 1980 ; Gentil, 1981).

Fridriksson, Moss, Davis, Baylis, Bonilha & Rorden (2008) comparent la perception visuelle de dix syllabes CV à celle de dix items de mouvement orofaciaux non linguistiques, lors d'une tâche de discrimination à choix forcé, chez vingt participants. La comparaison des résultats pour les items non linguistiques par rapport aux syllabes révèle des activations des aires de Brodmann 21, 22, 42 de l'hémisphère droit, mais aussi de l'aire de Brodmann 18 de façon bilatérale, ainsi que des aires de Brodmann 22, 47 et 6 de l'hémisphère gauche. Les items non linguistiques activent l'aire de Brodmann 6, mais de façon significativement inférieure par rapport aux syllabes.

Des activations frontales sont régulièrement observées lors de tâches de discrimination langagière, pouvant refléter l'implication de la mémoire de travail dans ces processus (Doyon, Penhune & Ungerleider, 2003). Le recrutement de l'aire de Broca dans la production, mais aussi dans la perception de la parole, pourrait donc aussi être lié à des processus mémoriels ou attentionnels (Smith & Jonides, 1999 ; Giraud, Kell, Thierfelder, Sterzer, Russ, Preibisch & Kleinschmidt, 2004).

Aujourd'hui, il faut aborder le ou les rôles de l'aire de Broca avec une grande prudence ; il est envisageable qu'elle soit recrutée pour le contrôle et l'exécution de nombreuses tâches nécessitant une coordination motrice. La présence de neurones miroirs dans l'aire de Broca pourrait expliquer l'hétérogénéité des résultats expérimentaux concernant cette zone corticale. Nous suivons Iacobini & Wilson (2006) qui pensent qu'il est nécessaire de considérer l'aire de Broca comme faisant partie d'un réseau largement distribué, afin de pouvoir mieux cerner son ou ses rôles dans les processus langagiers et moteurs.

A.3.3.3.1. Le système des neurones miroirs : l'intrication de l'action et de la perception

L'idée selon laquelle les aires motrices corticales sont exclusivement recrutées à des fins exécutives, sans aucune fonction perceptive ou cognitive, a dominé jusqu'à récemment. L'homoncule moteur de Penfield, défini par les stimulations électriques des zones motrices, a longtemps mis en avant une organisation somatotopique des fonctions motrices (voir figure A9). Nous savons aujourd'hui que, bien que simple et intuitive, cette organisation somatotopique est exagérée. Il existe effectivement une séparation entre les régions corticales

contrôlant les pieds, la tête ou les bras, mais à l'intérieur de la région contrôlant par exemple les membres supérieurs, il n'y a pas de région dédiée exclusivement au contrôle des muscles de la main, du coude ou de l'épaule. Toutes les cellules corticales contrôlant un muscle en particulier sont entremêlées avec d'autres cellules ayant des fonctions différentes. Il existe un chevauchement très important entre les « représentations corticales » de différents muscles, mêmes pour des muscles agissant sur des articulations différentes. L'activation d'un petit volume de tissu cortical peut entraîner l'excitation de plusieurs muscles. Il n'existe donc pas d'organisation corticale simple, dans laquelle une région précise du cortex moteur contrôle un muscle ou un segment du corps. Postle, McMahon, Ashton, Meredith & de Zubicaray (2008) observent en IRMf un recouvrement des activations dans les aires de Brodmann 4 et 6 lors de tâches impliquant l'exécution, puis l'observation de mouvements des mains, des pieds et de la bouche. De plus, les différences interindividuelles du cortex moteur apparaissent comme non négligeables (Rademacher *et al.*, 2001).

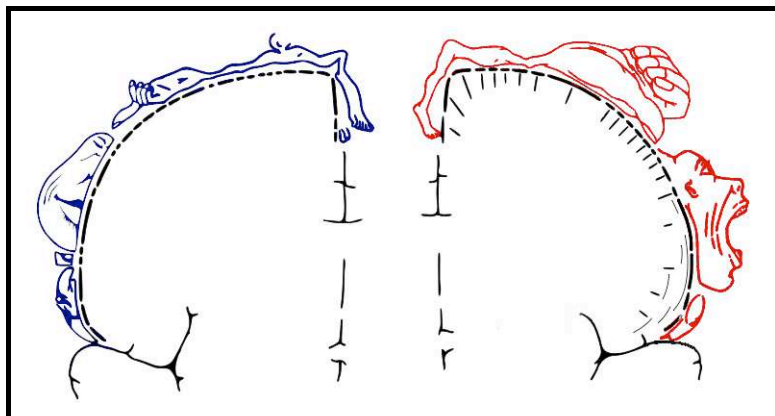


Figure A9 : Homoncles sensoriel (en bleu) et moteur (en rouge) de Penfield (Tiré de http://fr.wikipedia.org/wiki/Homoncule_moteur).

Fadiga, Fogassi, Pavesi & Rizzolatti (1995) signalent que les potentiels évoqués moteurs des bras augmentent lors de l'exécution d'une action mais aussi lors de la vision de cette action. La découverte d'une classe particulière de neurones dans le cortex moteur du macaque (Gallese *et al.*, 1996 ; Rizzolatti *et al.*, 1996) qui déchargent, à la fois, lorsqu'un individu effectue une action, mais aussi lorsque cet individu observe un congénère exécuter la même action, a ouvert de nouvelles perspectives pour la compréhension de l'organisation cognitive humaine, et a aussi permis de reconsidérer le système moteur non plus comme un bloc mais comme une mosaïque incluant des aires frontales et pariétales.

A.3.3.3.2. Les différents neurones miroirs

Dans les années 1930, il a été constaté que des neurones moteurs possédaient des propriétés visuelles : il s'agit des neurones « canoniques ». Ces neurones s'activent à la vue d'un objet sans que celui-ci soit utilisé. *A contrario*, les neurones miroirs ne s'activent que lors de l'interaction avec un certain type d'objet. Ils vont répondre sélectivement aux formes et aux dimensions des objets et ils sont associés à la fois à des mouvements transitifs (impliquant l'interaction avec un objet) et intransitifs (sans interaction avec un objet). Cela signifie que la perception d'un objet n'est pas uniquement passive, car elle induit l'activation d'un réseau du « possible ». L'apprentissage par imitation pourrait tirer profit de ce système (Iacobini *et al.*, 1999). En voyant un objet connu et qui fait partie de notre « patrimoine moteur », nous avons accès aux « potentialités » qu'il offre avant même de s'en servir !

Certains neurones répondent à un seul acte comme la saisie avec la main, d'autres répondront à deux actes comme saisir et tenir par exemple. Les interactions manuelles ne sont pas les seules à activer le système miroir, les mouvements de la bouche ou des pieds semblent eux aussi engendrer des réponses de neurones spécifiques (Buccino *et al.*, 2001 ; Buccino, Binkofski & Riggio, 2004a). En ce qui concerne la bouche, les neurones miroirs réagissent à des actes ingestifs et communicatifs. La plupart (environ 85 %) sera préférentiellement recrutée par des mouvements d'ingestion (actes transitifs) ; les neurones « communicatifs » vont réagir à des actes intransitifs. Ces neurones communicatifs s'activent lorsque le sujet voit l'examineur réaliser une protrusion ou un claquement avec ses lèvres mais aussi lors de la perception de ces mêmes gestes durant l'ingestion (Ferrari, Gallese, Rizzolatti & Fogassi, 2003). Ces activations seraient à mettre en relation avec un « patrimoine moteur » spécifique. En effet, la perception visuelle d'actes « ingestifs » ou « communicatifs » ne sera pas équivalente lorsqu'elle est réalisée par un autre être humain, par un singe ou par un chien. Les actes moteurs des chiens ne seraient pas interprétables par un humain en fonction de son patrimoine moteur. Par conséquent, ils n'activeront pas le réseau des neurones miroirs (Buccino, Binkofski & Riggio, 2004a).

Le cortex moteur n'est pas la seule région corticale qui semble pourvue de neurones miroirs, les lobes pariétal et frontal semblent aussi posséder des neurones de ce type. L'enregistrement de neurones isolés dans le lobe pariétal de primates montre que ces régions ne sont pas seulement sensibles aux stimuli visuels, mais aussi aux actions associées à ces

stimuli. Par exemple, un neurone peut répondre lorsqu'un objet est présenté, mais aussi quand le singe utilise cet objet, alors qu'il n'est pas dans son champ visuel (Rizzolatti *et al.*, 1996). De nombreuses aires visuelles semblent organisées autour de différents systèmes effecteurs « moteurs » (Rizzolatti *et al.*, 1996 ; Colby & Goldberg, 1999). Des neurones pariétaux réagissant à des stimuli audiovisuels, ou simplement auditifs, ont été aussi observés chez le singe (Kohler *et al.*, 2002). Ces neurones déchargent lors de l'observation de l'action, mais aussi lors de la perception auditive du bruit induit par cette action.

Les neurones miroirs du lobe frontal sont principalement associés aux aires de Brodmann 44 et 45 (Iacobini & Wilson, 2006). L'aire de Broca est considérée comme l'homologue de l'aire F5 du singe ; aussi Rizzolatti et ses collègues pensent qu'un système semblable soit probable chez l'être humain ; certains résultats relatifs à l'aire de Broca font écho à cette découverte (Binkofski *et al.*, 2000) ; (Grodzinsky, 2006). La figure A10 schématise le système des neurones miroirs chez l'être humain.

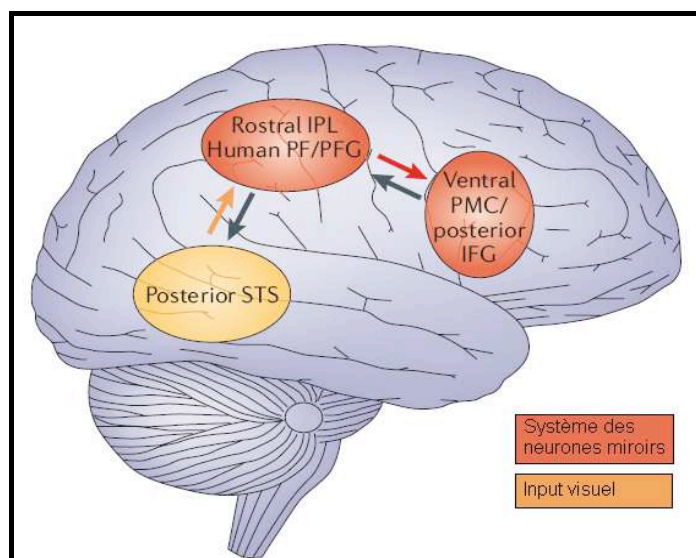


Figure A10 : Le système des neurones miroirs chez l'être humain (Tiré de Iacobini & Wilson, 2006).

A.3.3.3.3. L'intérêt des neurones miroirs pour les neurolinguistes

Dans le domaine de la parole, il est possible d'émettre plusieurs hypothèses quant aux différentes utilités du système des neurones miroirs, tant dans la phylogénèse (Meltzoff & Moore, 1983 ; Meltzoff & Moore, 1997) que dans l'ontogénèse (Rizzolatti & Sinigaglia, 2008).

La perception d'un congénère parlant pourrait induire l'activation des zones du cortex moteur contenant des neurones miroirs (Rizzolatti & Arbib, 1998). Cette idée découle d'une hypothèse selon laquelle la compréhension des actions d'autrui repose sur l'activation des « représentations motrices » de ces dernières :

“we understand action because the motor representation of that action is activated in our brain” (Rizzolatti, 2001)

« nous comprenons une action parce que la représentation motrice de cette action est activée dans notre cerveau. »

Rizzolatti et ses collègues défendent l'idée selon laquelle la récupération de patrons articulatoires lors de la perception de la parole, telle que supposée par la Théorie Motrice de Liberman & Mattingly (1985), pourrait avoir comme substrat le système des neurones miroirs. Une expérience de Fadiga, Craighero, Buccino & Rizzolatti (2002) révèle que les potentiels évoqués moteurs des muscles de la langue présentent une augmentation significative d'amplitude lors de l'écoute passive de mots comportant des articulations géminées pour lesquelles la langue entre en jeu. C'est le cas, par exemple, de mots contenant les géminées linguales « RR » par rapport à « FF » qui, elles, ne sollicitent pas la langue comme articulateur critique. Les potentiels évoqués moteurs des lèvres augmentent aussi lors de tâches de perception passive de la parole audible et visible par rapport à des stimuli auditifs et visuels non verbaux (Watkins, Strafella & Paus, 2003). Rizzolatti & Sinigaglia (2008) avancent l'idée que le système des neurones miroirs a subi des réorganisations durant l'évolution, afin de contribuer aux traitements de la parole, faisant ainsi émerger une classe spécifique de neurones : les « neurones miroirs échos ». Ces derniers pourraient constituer l'entrée dans le système moteur, permettant une correspondance entre les phonèmes perçus et le « patrimoine moteur phonologique ». Signalons, toutefois, que de récents travaux viennent nuancer ce point de vue.

Lotto, Hickok & Holt (2009) avancent que le système de neurones miroirs ne fournit pas d'explication au dilemme de la perception de la parole (récupération des invariants dans des flux sensoriels continus et sujets à une grande variabilité inter et intra-individuelle). La conversion d'un phonème en « actes moteurs » ou en « gestes articulatoires » ne saurait être efficace sans une « représentation » linguistiquement pertinente, et il y aurait, de surcroît, des contradictions entre la Théorie Motrice de la Perception et le système des neurones miroirs. Selon les auteurs, le système des neurones miroirs ne peut s'accommoder que d'une version

faible de la Théorie Motrice de Libermann ; en effet, le fait que des parties du cortex moteur soient activées durant certaines tâches de perception, ne signifie pas forcément que ce processus soit obligatoire et nécessaire pour la perception ou la compréhension de la parole articulée. L'argument le plus intéressant concerne les processus de compensation, étudiés grâce au paradigme du "*bite-block speech*" (paradigme dans lequel la mandibule est bloquée lors de la production de la parole). Si les gestes articulatoires interviennent directement dans le traitement de la parole, alors comment se fait-il que nous puissions percevoir des phonèmes, syllabes ou mots, produits en condition de perturbation de la parole et avec une action de compensation articulatoire ?

Scott, McGettigan & Eisner (2009) constatent, quant à eux, que les activations du cortex moteur dans des tâches de perception de la parole articulée, lors d'analyse en cerveau entier ("*whole brain analysis*"), sont très rares par rapport à celles observées dans les lobes temporaux ; cela suggère que le cortex moteur ne prend pas une place directe dans les réseaux de perception de la parole.

Rizzolatti et ses collègues concèdent que la compréhension des actions ne doit pas être sous la seule égide du système des neurones miroirs et que d'autres structures doivent pouvoir l'accomplir. Hickok (2009) met en lumière les difficultés existantes pour lier le système de neurones miroirs à la compréhension des actions chez le singe. De plus, l'auteur insiste sur les différences entre les rôles pressentis de ce système chez les humains et les simiens. La découverte du système des neurones miroirs a suscité un nouvel engouement pour la Théorie Motrice de Liberman & Mattingly (1985). Pour Hickok, la récupération des patrons moteurs par les neurones miroirs, lors de la perception de la parole, se heurte aux observations anatomocliniques des aphasiques de Broca. La destruction d'une large part des neurones miroirs devraient obligatoirement provoquer un déficit de compréhension chez les aphasiques de Broca.

Les tâches de perception de la parole impliquent à certains degrés un contrôle exécutif et le recrutement de la mémoire de travail (Blumstein *et al.*, 1977a ; Démonet *et al.*, 1992 ; Price *et al.*, 1996). Nous n'entrerons pas ici dans les dédales de ces processus, mais nous nous appuyerons sur le modèle de la mémoire de travail de Baddeley (1992, 1998, 2003, 2004). La mémoire de travail serait constituée d'un exécuteur central et de deux sous-systèmes : la tablette visuo-spatiale et la boucle phonologique. Le rôle de l'exécuteur central serait principalement de coordonner la mémoire de travail et la mémoire à long terme, et de

contrôler le traitement, l'encodage et la récupération des informations. La tablette visuo-spatiale est la composante qui permet de traiter les informations visuelles, gérées par le lobe occipital, et les informations spatiales qui seraient contrôlées par les lobes pariétaux. La boucle phonologique aurait deux composantes : un système de stockage phonologique et un système de répétition subvocale. Le stock phonologique aurait pour fonction de maintenir les informations sous forme « phonologique » durant quelques secondes ; la répétition subvocale aurait pour but de réactiver les informations du stock phonologique et d'y intégrer les informations issues de la perception visuelle. Paulesu, Frith & Frackowiak (1993) considèrent que le système de stockage phonologique correspond au gyrus supramarginal (Aire de Brodmann 40) alors que l'aire de Broca, particulièrement la *pars opercularis* (Aire de Brodmann 44), serait le siège du système de répétition subvocale.

Les questions soulevées par ces différentes études sont de deux ordres. La dichotomie, au niveau cérébral, entre la production et la perception de la parole est-elle anatomique, fonctionnelle ou anatomo-fonctionnelle ? Les indices visuels sont-ils traités par les zones assurant la perception langagière, ou par un réseau ad-hoc ?

Les études simultanées en IRMf et en EEG, comme la nôtre, sont rares. Nous disposons de ce fait de peu de points de comparaison. Malgré cela, certains travaux attestés dans la littérature peuvent nous fournir quelques pistes de réflexion pour l'exploitation de nos résultats.

A.3.4. Les études en potentiels évoqués

Les premières études en EEG ont apporté des informations sur les traitements auditifs (Davis, 1939). L'écoute passive de sons génère une onde négative de grande amplitude vers 100 millisecondes (N1 ou N100), avec une distribution maximum au vertex. Cette onde est générée au moins en partie dans le cortex auditif (Vaughan & Ritter, 1970). Cependant, cette onde sera divisée en plusieurs sous catégories en fonction des zones d'enregistrement. L'onde « N1b » est une négativité observée en regard des électrodes fronto-centrales : ce potentiel est souvent plus ample sur le vertex (Vaughan & Ritter, 1970). L'onde « Nd » ou “*Processing negativity*” est mise en évidence par la soustraction des réponses évoquées à un stimulus donné, se distinguant par son caractère attendu ou inattendu. La « Nd » indique que le stimulus était attendu. L'expérience typique suscitant une « Nd » est un paradigme d'attention

focale. Une série de stimuli auditifs est présentée, avec une périodicité soutenue, à une fréquence de 800 Hz pour l'oreille droite, et de 1500 Hz pour l'oreille gauche ; 10 % des stimuli ont une fréquence légèrement plus élevée (850 Hz et 1540 Hz). Il est demandé au sujet de les compter. Les potentiels évoqués, par les stimuli attendus, présentent une déflexion négative 100 millisecondes après la présentation du stimulus d'amplitude accrue, en comparaison avec la « N1 » évoquée aux stimuli standard (Näätänen, 1982 ; Näätänen, 1990 ; Martin, Tremblay & Korczak, 2008). L'attention portée aux stimuli rares module la composante « N1 ». En demandant au sujet d'être attentif à une oreille, il est possible de mesurer l'effet de l'attention et déterminer la « Nd » par soustraction entre les réponses aux stimuli attendus et non attendus. La composante « Nd » présente une amplitude plus large, dans le cortex auditif, lors des présentations auditives unimodales, par rapport à des présentations audiovisuelles (Besle, Fort, Delpuech & Giard, 2004). *A contrario*, l'étude de Ponton, Auer & Bernstein (2002) reporte un effet inverse, observant un renforcement lors des présentations audiovisuelles. L'étude de Giard & Peronnet (1999) met aussi en lumière un tel renforcement, précisons que cette expérience ne portait pas sur un matériau linguistique.

Une « N100 » peut aussi apparaître dans les contextes de traitement de l'indice de voisement (Trébuchon-Da Fonseca, Giraud, Badier, Chauvel & Liégeois-Chauvel, 2005 ; Hoonhorst, Colin, Markessis, Radeau, Deltenre & Serniclaes, 2007). De plus, pour les VOT positifs (supérieurs à + 30 ms), la « N100 » serait caractérisée par un double pic, le premier correspondant au relâchement articulaire et le second à la mise en action des plis vocaux. Les valeurs négatives (inférieures à -30 ms) mettent aussi en exergue ce double pic, le premier corrélé à la source glottique et le second au relâchement. Alors que pour les valeurs de VOT allant de -30 à +30 millisecondes, cette composante ne présenterait qu'un seul pic (Hoonhorst *et al.*, 2007). Trébuchon-Da Fonseca *et al.* (2005) relèvent, quant à eux, un double pic pour les valeurs négatives (-120 ms) et observent un pic unique pour des valeurs positives (-20 ms). Ils précisent que ces effets sont plus prononcés dans l'hémisphère gauche pour les sujets droitiers.

La négativité observée 200 ms après la présentation de divers types d'événements auditifs ou visuels est appelée « N2 » ou « N200 ». Cette composante est majoritairement observée sur les régions centrales du scalp, et apparaît dans les contextes où le sujet attend activement une stimulation déviante, qu'il doit repérer parmi d'autres stimuli. L'amplitude de la « N200 » augmente en fonction de l'attention portée à la stimulation (Novak, Ritter,

Vaughan & Witznitzer, 1990). Notons que les stimuli du type « McGurk » permettent de l'observer (pour les sonores : Colin, Radeau, Soquet, Demolin, Colin & Deltenre, 2002 ; pour les sourdes : Colin, Radeau, Soquet & Deltenre, 2004 ; Saint-Amour, De Sanctis, Molholm, Ritter & Foxe, 2007).

Giraud, Démonet, Habib, Marquis, Chauvel & Liégeois-Chauvel (2005) observent une corrélation entre le VOT et les ondes N1 et P2. Pour des VOT positifs, les auteurs reportent des ondes N1 et P2 suivies d'une onde négative culminant autour de 240 ms alors que seules les ondes N1 et P2 sont relevées pour les VOT négatifs. Les ondes N1 et P2 semblent être provoquées à la fois par des stimuli visibles et audibles (Fort, Delpuech, Pernier & Giard, 2002). Van Wassenhove *et al.* (2007) mettent en lumière des ondes N1 P2, dans une tâche de détection de cible syllabique, pour les items présentés en modalité audiovisuelle (congruente et non congruente « McGurk »), ainsi que pour l'auditif seul, mais pas pour la condition « visuelle seule ». Précédemment, Van Wassenhove, Grant & Poeppel (2005) avaient observé, pour les ondes N1 P2, des amplitudes significativement plus importantes en modalité auditive seule comparée à une présentation audiovisuelle. L'apport de la modalité audiovisuelle se traduirait par l'accélération des processus de traitement. Klucharev, Möttönen & Sams (2003) reportent eux aussi des ondes N1 P2 pour la détection de voyelles.

Alors que les composantes de la famille de la N200 sont suscitées par des événements physiquement « déviants » en regard du contexte, la composante N400 est sensible à une déviance dans un contexte plus abstrait, comme le sens (Astésano, Besson & Alter, 2004). La N4 ou N400 est enregistrée lors de la détection de stimuli déviants de nature linguistique (Kutas & Hillyard, 1989), phonologique, orthographique ou arithmétique, mais aussi lors de la présentation de visages et d'odeurs, ou dans le cadre de la parole audiovisuelle (Lebib, Papo, de Bode & Baudonniere, 2003). Le paradigme de "*Masked Priming*" est connu pour provoquer une N400 (Brown & Hagoort, 1993 ; Hagoort & Brown, 2000). Cette composante est principalement centro-pariétal mais les régions corticales frontales gauches pourraient aussi participer à l'évocation de la N400 comme le suggèrent les études de patients cérébrolésés (Friederici, Von Cramon & Kotz, 1999). Elle a été mise en évidence dans les réponses évoquées à des phrases écrites, dont le dernier mot est inattendu voire incongru. Par exemple « il étalait sur son toast de la chaussette », comparé à « c'était son premier jour de travail » ou encore « elle enfila ses chaussures à hauts TALONS ». Aucune composante particulière n'est observée lorsque le dernier mot est congruent sémantiquement et

physiquement. Lorsque le dernier mot d'une phrase est écrit en majuscule contrairement aux mots précédents écrits en minuscules, une P300 de latence plus tardive est enregistrée.

Selon Martin *et al.* (2008), la P3b pourrait refléter le traitement conscient de différences acoustiques présentes dans le signal acoustique comme une distinction entre les syllabes /ba/ et /da/. Ce potentiel nécessitant un type de paradigme qui requiert la présence d'événements rares, ne sera probablement pas observable dans nos données expérimentales.

Bernstein, Auer, Wagner & Ponton (2008) s'intéressent à la contribution relative des hémisphères corticaux en menant une expérience en EEG, mais s'appuyant sur de précédents résultats IRMf, en particulier ceux de Miller & D'Esposito (2005). Ces derniers mettent au jour la participation du cortex auditif primaire, du sillon temporal supérieur, de la partie moyenne du sillon intra-pariétal et du gyrus frontal inférieur. Bernstein *et al.* (2008) ont utilisé des syllabes CV naturelles ([g a b a d a ð a]), leur permettant de créer quatre modalités perceptives différentes (audio seule, visuelle seule, audiovisuelle congruente et audiovisuelle incongruente). La condition « audio seule » met en lumière une activité précoce (55-60 ms) dans les deux hémisphères cérébraux. Cette activité s'étend jusqu'à 180 millisecondes dans l'hémisphère droit, et seulement jusqu'à 120 millisecondes pour le gauche. Le lobe pariétal semble montrer une plus grande activité à droite, particulièrement dans l'intervalle de 55 à 90 millisecondes. Or c'est la *pars triangularis* du gyrus frontal inférieur qui semble être recrutée environ entre 65 et 70 millisecondes dans l'hémisphère gauche. En ce qui concerne, la modalité « visuelle seule », les premières activités sont enregistrées entre 65 et 80 millisecondes, mais les plus fortes variations sont relevées après 120 millisecondes, dans les régions occipitales. Les gyri pariétal et temporal inférieurs droits montrent, quant à eux, une activité entre 160 et 200 millisecondes. Le sillon pariétal inférieur et le gyrus supramarginal de l'hémisphère gauche produisent une activité précoce dès 60-65 millisecondes, pour la condition audiovisuelle congruente. Entre 60 et 80 millisecondes, les activités générées par cette condition sont similaires à celles observées en modalités « audio seule » pour la même période. Les auteurs relèvent ensuite une activité du cortex dorsolatéral préfrontal à partir de 85 millisecondes et cela jusqu'à 100 millisecondes. Une onde occipitale est aussi observée à partir de 90 millisecondes, et le sillon temporal supérieur semble être recruté entre 120 et 160 millisecondes. En comparaison, le cortex dorsolatéral préfrontal de l'hémisphère gauche est recruté plus précocement par la condition audiovisuelle non

congruente (à partir de 60 ms), alors que les activations de l'hémisphère droit sont sensiblement identiques quoique plus lentes.

Les potentiels évoqués ne permettent pas de préciser directement la chronologie des activations observées en IRMf, mais offre la possibilité d'identifier des réponses spécifiques, en fonction du type de stimulations. À l'instar des données IRMf, une grande variabilité est présente parmi les potentiels évoqués ; il apparaît clair que les différentes ondes doivent être replacées dans des réseaux distribués.

A.3.5. Les études couplant des données en EEG et en IRMf

Reprenant l'idée de Rauschecker (1998 ; cf. A.4.2.), Alain, Arnott, Hevenor, Graham & Grady (2001) ont tenté de mettre en lumière l'existence de deux réseaux dissociés dans le système auditif pour l'identification d'un son et pour sa localisation. Ils ont observé quinze sujets en IRMf, puis douze d'entre eux en EEG, lors d'une tâche d'identification portant sur des sons purs synthétiques variant en fréquence (S2 différent ou identique de S1). Ils ont aussi proposé une tâche de localisation consistant à dire si les deux sons venaient d'une même source, ou si le second venait de la gauche ou de la droite. L'augmentation de la réponse *BOLD* révèle des zones bilatérales communes aux deux tâches dans les gyri frontaux inférieur et supérieur, dans les cortex auditifs primaire et secondaire, ainsi que dans les gyri pariétaux supérieur et inférieur. Les deux tâches induisaient aussi une baisse du signal *BOLD* dans le cortex visuel. Afin de dégager les zones plus spécifiquement engagées dans chacune des tâches, les auteurs ont procédé à des soustractions par rapport à l'état de repos. La tâche de localisation impliquait plus d'activité dans le cortex auditif primaire. Ces activations s'étendaient antérieurement au cortex auditif associatif dans le pôle supratemporal, ainsi que dans le gyrus frontal inférieur, et ce dans l'hémisphère droit. La localisation activerait davantage, de façon bilatérale, les aires temporeles postérieures, ainsi que les gyri pariétaux inférieur et supérieur. Ils notent aussi une activation du sillon frontal supérieur droit, similaire à celle observée lors d'une étude de Courtney, Ungerleider, Keil & Haxby (1996) sur la localisation spatiale. Les résultats comportementaux en EEG sont similaires aux précédents, avec un score moyen de 82 %, alors qu'il y a une différence entre l'identification (1026 ms), significativement moins rapide que la localisation (954 ms). Les potentiels évoqués sont de deux ordres ; après S1, une onde positive apparaît entre 300 et 500 millisecondes dans les

régions fronto-temporales inférieures, pour l'identification, et dans les régions centro-pariétales pour la localisation. Les auteurs analysent cela comme l'effet du traitement et du maintien en mémoire échoïque pour la comparaison avec S2. La seconde modulation apparaît environ entre 300 et 400 millisecondes après S2, et est similaire à celle observée pour S1. Cette onde précède une P3b pariétale reflétant, selon les auteurs, la comparaison entre les deux stimuli. Ces résultats invitent les auteurs à penser que les réseaux d'identification et de localisation sont séparés anatomiquement et fonctionnellement. La différence fonctionnelle supposée repose sur les « degrés » d'activations dans les aires recrutées pour les deux tâches.

Dehaene-Lambertz *et al.* (2005) étudient dans quelles mesures l'assertion de Liberman "*Speech is special*" (1967) pourrait reposer sur des substrats neuraux. En effet, en utilisant des stimuli identiques, le comportement des auditeurs est sensiblement différent lors d'une tâche expérimentale selon qu'ils soient informés que ces stimuli sont de la parole ou non. Lorsque les auditeurs savent qu'ils sont confrontés à la parole, la capacité de discrimination augmentent entre deux catégories phonétiques et diminuent à l'intérieur d'une catégorie. Grâce à l'IRMf et aux potentiels évoqués, les auteurs voulaient vérifier si le changement de type de percepts engendre des différences significatives dans le cortex. Cette étude portait sur douze personnes en EEG, et sur dix-neuf autres en IRMf. Les stimuli étaient constitués de syllabes synthétiques. Le changement de type de percepts augmente la discrimination, présentant de meilleurs résultats pour le type « parole ». De plus, les auteurs observent que la "*MisMatch Negativity*" arrive plus tôt pour les stimuli assimilés à la parole. En IRMf, les résultats sont significativement asymétriques avec une prédominance de l'hémisphère gauche. Le changement de type de percepts augmente l'activité, dans l'hémisphère gauche, d'un groupe de voxels du gyrus supramarginal, ainsi que la partie postérieure du sillon temporal supérieur. Les auteurs pensent que les « représentations phonémiques et non phonémiques » sont traitées en parallèle. Le réseau phonémique pourrait avoir un effet inhibiteur sur le réseau non phonémique. L'identification et la localisation pourraient reposer sur une dissociation fonctionnelle générale du cortex.

Giraud, Kleinschmidt, Poeppel, Torben, Frackowiak & Laufs (2007) mènent une étude simultanément en IRMf et en EEG avec pour hypothèse principale qu'il existerait un lien entre le *timing* de la production / perception de la parole et les deux hémisphères cérébraux. Notons que cette étude n'utilise pas les potentiels évoqués, mais observent l'activité électrique spontanée du cerveau. Ils s'appuient sur la théorie de MacNeilage & Davis (1998)

pour avancer l'idée d'une corrélation entre le rythme des oscillations mandibulaires et les capacités de production et de perception des unités syllabiques. Reprenant la théorie "*Asymmetric Sampling in Time*" de Poeppel (cf. A.4.2.2.), ces auteurs supposent que les cortex auditifs des deux hémisphères ne réagissent pas aux mêmes propriétés intrinsèques de la parole. Les résultats semblent étayer cette idée, mettant en lumière une corrélation entre les ondes cérébrales « lentes » (onde thêta : 3 - 6 Hz) et l'hémisphère droit, alors que l'hémisphère gauche présente un rythme plus rapide (onde gamma : 28 - 40 Hz). Vouloumanos, Kiehl, Werker & Liddle (2001) réfutent cette hypothèse en s'appuyant sur le fait que des stimuli non langagiers équivalents en termes d'indices temporels n'activent pas l'hémisphère gauche.

Les premiers travaux anatomocliniques ont permis de préciser l'organisation fonctionnelle du cerveau humain. Insistant sur l'importance de certaines régions corticales dans les processus de traitements cognitifs, ces études ont ancré l'idée d'une dominance de l'hémisphère gauche pour la parole. Par la suite, ce concept de « dominance » s'est vu évincé et remplacé par celui de « latéralisation hémisphérique ». Les premières études en TEP appuient l'idée d'un recrutement de certaines structures, en particulier temporales, dans les deux hémisphères corticaux. Le nombre très important d'études en IRMf concernant la parole ne permet pas de faire une revue exhaustive. Néanmoins, nous retiendrons trois points essentiels :

Les deux gyri temporaux supérieurs semblent avoir un rôle prépondérant dans l'analyse des propriétés acoustiques et spectrales des sons langagiers.

Les zones frontales, traditionnellement impliquées dans la production de la parole, pourraient avoir un rôle non négligeable lors de la perception. Suivant les hypothèses de Liberman & Mattingly (1985) et de Rizzolatti & Craighero (2003), le recrutement de l'aire de Broca et du cortex prémoteur pourrait être incontournable. Mais il est possible qu'il soit facultatif, dépendant de l'environnement et des conditions perceptives. Il est aussi possible que ces aires corticales prennent place dans des mécanismes cognitifs de plus grande échelle, telle que la boucle phonologique de Baddeley (1992, 2003, 2004).

La dimension visuelle de la perception de la parole pourrait elle aussi prendre un chemin similaire. Les indices visuels pertinents pourraient être utilisés, dès qu'ils sont disponibles ou seulement lorsqu'il est besoin de désambiguïser le flux auditif. Une question sous-jacente est de savoir s'il existe une zone corticale réalisant l'intégration audiovisuelle, ou si celle-ci se produit au stade de l'accès aux « représentations » mémorisées.

Les études des potentiels évoqués par des stimuli langagiers, lors de session IRMf, sont encore très rares. La corrélation entre ces deux types de données, physiologiquement disjoints, est extrêmement difficile. L'avantage d'un recueil simultané est indéniable, mais celui-ci induit des contraintes supplémentaires, telle que la réduction des artefacts provoqués par les gradients de l'IRM.

L'idée principale ressortant des études en EEG est qu'il existerait une répartition hémisphérique des traitements du langage articulé. Il est possible que l'hémisphère gauche traite les informations « rapides », et que le droit soit responsable du traitement des informations « lentes ». L'hypothèse alternative serait que l'hémisphère gauche soit recruté pour les informations temporelles alors que le droit le serait pour les informations spectrales.

A.4. Modèles cognitif et anatomo-fonctionnel de la perception de la parole

Une manière usuelle de représenter un modèle cognitif consiste, d'une part à définir un certain nombre de modules, chacun respectivement responsable d'un aspect particulier du processus cognitif global, et d'autre part à relier ces modules par des voies de communication unilatérales ou bilatérales.

L'apparition des techniques d'imagerie cérébrale fonctionnelle a permis de corroborer certains modèles cognitifs déjà existants. L'exemple le plus frappant est la confirmation partielle, en imagerie fonctionnelle, du modèle de Morton qui, en 1969, postulait l'existence de deux voies distinctes pour la lecture. Parmi les modèles cognitifs de la perception, nous allons présenter brièvement le modèle de la perception à logique floue ("*Fuzzy Logical Model of Perception*" "*FLMP*" ; Massaro, 1988, 1990, 2001), avant d'aborder plus spécifiquement le modèle anatomo-fonctionnel de Hickok & Poeppel (2007), que nous avons pris comme cadre théorique pour l'analyse de nos résultats. Le choix d'un modèle, par essence réducteur, est toujours difficile. Il nous est apparu cependant nécessaire de replacer nos lectures et nos résultats expérimentaux dans un cadre nous permettant de les ordonner, mais aussi de les confronter à des études ne s'intéressant pas directement à la perception langagière.

A.4.1. Le modèle de perception à logique floue de Massaro

À nos yeux, le modèle de Massaro (1988, 1990, 2001) a le mérite de s'attacher particulièrement à la dimension multimodale de la perception de la parole. Une des assertions centrales de ce modèle est que les traitements perceptifs sont constitués par un ensemble de structures indépendamment du type de percepts. Cette hypothèse est en contradiction avec la Théorie Motrice (Liberman *et al.*, 1967 ; Liberman & Mattingly, 1985), qui présuppose un module de traitement spécifique à la parole articulée. Chez Massaro, les informations sensorielles, quelles qu'elles soient (les phonèmes, ou plutôt les sons, les lettres ou les mouvements manuels, *etc.*) sont censées être traitées de la même manière par le cerveau humain.

Le modèle de perception à logique floue avance que la correspondance entre un stimulus et sa catégorie perceptive s'opère en suivant trois étapes : l'étape d'évaluation des « traits », l'étape d'intégration des « traits » et l'étape de décision.

L'étape d'évaluation des « traits » consiste à convertir les sources disponibles d'informations en un ensemble de propriétés nommé « traits » (*“features”*). Chaque trait fournit une valeur continue et représente le degré de correspondance avec un patron prototypique interne, dans une dimension perceptive particulière. Ainsi, le degré d'aperture, par exemple, sera extrait et comparé avec le prototype.

L'étape d'intégration consiste à « intégrer » des traits issus de la première étape afin de déterminer la correspondance avec les prototypes disponibles (par exemple toutes les syllabes connues par l'auditeur). Finalement l'étape de décision consiste simplement à établir une corrélation entre l'entrée sensorielle et les prototypes.

Dans ce modèle, toutes les sources d'informations sont traitées afin d'être intégrées simultanément dans un patron de reconnaissance. Les flux audibles et visibles ont une influence réciproque sur la perception multimodale de la parole articulée. Les différentes sources sont évaluées indépendamment les unes des autres. Cela signifie que la parole visuelle n'aura pas d'effet sur la manière dont les informations auditives sont converties en valeur continue. Les deux sources seront combinées au niveau supérieur. Une assertion majeure de ce modèle est que l'influence d'une source d'information est maximale si l'autre source est neutre ou ambiguë :

“the influence of one source of information is greatest when the other source is neutral or ambiguous”.

Massaro (1999) pose une question des plus intéressantes concernant l'effet McGurk. Est-ce qu'il révèle quelque chose d'essentiel à propos de la parole ou concerne-t-il l'intégration multimodale dans son ensemble ? Pour l'auteur, la réponse est claire : l'effet McGurk n'est pas spécifique à la parole, mais procède de mécanismes ayant trait à nos modes de perception.

Pour Massaro, les informations des modalités auditive et visuelle sont traitées indépendamment lors de la première étape d'évaluation des traits. La deuxième étape prendrait en charge l'intégration des différentes modalités, pour réaliser la comparaison avec le stock phonologique qui contiendrait à la fois les informations auditive et visuelle. Le meilleur « candidat » sera ainsi sélectionné.

Ce bref aperçu du modèle de Massaro est utile, d'une part, afin de relativiser la portée de notre cadre théorique, et d'autre part, parce qu'il prend en compte la dimension visuelle, qui nous intéresse particulièrement.

A.4.2. Le modèle à deux voies de Hickok et Poeppel

En 2000, Hickok & Poeppel publient un premier article, se proposant de poser les jalons d'un modèle anatomo-fonctionnel de la perception de la parole, modèle qu'ils affinent au fil des parutions (Hickok & Poeppel, 2000 ; Hickok & Poeppel, 2004 ; Hickok & Poeppel, 2007). Le concept de traitement double remonte à Wernicke, qui distinguait deux réseaux dans le système auditif. Les modèles à deux voies introduisent l'idée que différents réseaux pourraient être impliqués dans divers aspects d'un même processus cognitif. Hickok & Poeppel (2007) reprennent cette idée, s'inspirant du modèle développé par Ungerleider & Mishkin (1982) et Mishkin, Ungerleider & Macko (1983) pour la perception visuelle, qui n'est pas sans rappeler celui de Morton (1969) pour la lecture, ou plus récemment dans le domaine somatosensoriel (Dijkerman & de Haan, 2007).

Dans le domaine visuel, les deux voies, l'une dite « ventrale » et l'autre dite « dorsale », assurent l'identification et la localisation spatiale (« quoi et où » “*What and Where*” ; Decety, et al., 1997). Pour la lecture, la voie ventrale est mise à contribution afin d'accéder au sens lexical de l'item lu, alors que la voie dorsale opère une analyse « pas à pas » des composants phonologiques. C'est la raison pour laquelle les appellations « voie lexicale » (ventrale) et « voie phonologique » (dorsale) coexistent. Rauschecker (1998) plaide pour une telle dichotomie “*What and Where*” dans le domaine de la parole.

Pour Hickok & Poeppel (2007), les processus langagiers sont répartis dans des réseaux distribués à travers les deux hémisphères corticaux, mais avec des spécialisations fonctionnelles distinctes. Les régions temporales antérieures ont été impliquées aussi bien dans les processus lexicaux et sémantiques que dans les processus phrastiques (processus d'intégration syntaxique et sémantique). Les patients souffrant de démence sémantique ont des atrophies des régions temporales antérieures bilatérales, associées à des déficits dans les tâches de dénomination, d'association sémantique et de compréhension d'un mot isolé (Gorno-Tempini *et al.*, 2004). Pour certains auteurs, ces données sont un argument en faveur de l'implication des régions temporales antérieures dans les fonctions sémantiques et lexicales (Scott *et al.*, 2000 ; Narain, Scott, Wise, Rosen, Leff, Iversen & Matthews, 2003 ; Spitsyna,

Warren, Scott, Turkheimer & Wise, 2006). Cependant, Hickok & Poeppel (2007) avancent l'idée que ces déficits pourraient être plus généraux, sachant que les atrophies impliquent d'autres régions, incluant les lobes temporaux inférieurs et moyens, les noyaux caudés bilatéraux, le thalamus postérieur droit, entre autres (Gorno-Tempini *et al.*, 2004). Friederici (1999) et Humphries, Love, Swinney & Hickok (2005) rapportent des activations plus fortes des régions temporales antérieures, lors de l'écoute ou de la lecture de phrases qu'avec des listes de mots ou de sons. Les lésions des régions temporales antérieures pourraient aussi engendrer des déficits dans la compréhension des structures syntaxiques complexes (Dronkers, Wilkins, Van Valin, Redfern & Jaeger, 2004). Cependant, les données issues de la démence sémantique sont contradictoires, ces patients conservant une bonne compréhension du niveau phrastique.

Pour Scott (2005) et Belin, Zatorre & Ahad (2002), la voie ventrale prendrait effectivement part aux mécanismes d'identification lexicale, la voie dorsale étant moins bien circonscrite. Alors que pour Warren, Wise & Warren (2005), la voie dorsale est fortement orientée vers la production de la parole. Le concept de modèle à deux voies pourrait correspondre à un schéma général entrant en jeu dans divers processus perceptifs et cognitifs.

A.4.2.1. Les premières étapes du modèle à deux voies de Hickok & Poeppel

Les deux premières phases qui entrent en jeu dans le modèle à deux voies de Hickok & Poeppel (2007) sont constituées des niveaux de traitements acoustique et phonétique ; les auteurs emploient les notions d'« analyse spectrotemporelle » (*"Spectrotemporal analysis"*) et de « réseau phonologique » (*"Phonological network"*). Ces deux premières étapes sont communes aux deux voies et seraient assurées par deux structures sises dans les lobes temporaux. Le modèle de Hickok & Poeppel (2007) diffère quelque peu de l'idée avancée par Zatorre et ses collègues (1992), la latéralisation hémisphérique étant régie par une différenciation d'ordre temporel, et non pas sur une opposition des informations spectrales vs. temporelles.

A.4.2.1.1. Traitement acoustique : « l'analyse spectrotemporelle »

Dans le modèle de Hickok & Poeppel (2007), l'analyse spectrotemporelle serait effectuée dans les cortex auditifs primaires des deux hémisphères cérébraux. Afin d'expliquer les contributions des deux hémisphères corticaux, le modèle de Hickok & Poeppel intègre la

théorie connue sous le sigle *AST* (“*Asymmetric Sampling in Time*”) de Poeppel (2003). L’auteur y avance l’idée d’une implication différente des deux hémisphères dans la perception de la parole, en fonction de fenêtres d’intégration temporelle distinctes. Partant du postulat que la parole contient des informations redondantes (Shannon, Zeng, Kamath, Wygonski & Ekelid, 1995) de différentes échelles temporelles (traits, segments, syllabes, etc.), il reprend l’idée d’une fenêtre « courte » allant de 20 à 40 millisecondes (niveau segmental), sous contrôle de l’hémisphère gauche, et d’une fenêtre « longue » allant de 150 à 250 millisecondes (niveau suprasegmental) régie par l’hémisphère droit. De plus, la fenêtre « courte » serait corrélée aux ondes gamma (28 - 40 Hz), et la fenêtre « longue » aux ondes thêta (3 - 6 Hz) et alpha (8 - 12 Hz). Samson, Belin, Thivard, Boddaert, Corzier & Zilbovicius, (2001) effectuent une méta-analyse d’études en TEP et en IRMf et citent un article de Tallal, Miller & Fitch (1993) qui avancent déjà l’idée d’une spécialisation du cortex temporal gauche pour les traitements temporels rapides.

A.4.2.1.2. Traitement phonétique : « le réseau phonologique »

Le réseau phonologique serait lui aussi assuré par les deux hémisphères cérébraux, dans les moitiés postérieures des sillons temporaux supérieurs, à la jonction des gyri temporaux moyen et supérieur. Les moitiés postérieures des sillons temporaux supérieurs des deux hémisphères seraient chargées d’assurer les traitements phonologiques et représentationnels. Ils auraient donc pour fonction principale de convertir les flux sensoriels en « représentation phonologique », expliquant ainsi pourquoi les sillons temporaux supérieurs sont aussi recrutés pour la production de la parole (Indefrey & Levelt, 2004 ; Wise, et al., 2001 ; Okada & Hickok, 2006), et durant le maintien actif d’informations phonémiques (Hickok, Buchsbaum, Humphries & Muftuler, 2003).

A.4.2.2. La séparation en deux voies

Pour Hickok & Poeppel (2007), les deux voies agissent en parallèle, et non pas de façon autonome et / ou alternative. La distinction opérée par les auteurs repose principalement sur la finalité des traitements : la voie ventrale serait orientée vers la lexicalité, alors que la voie dorsale serait une interface sensori-motrice.

A.4.2.2.1. La voie ventrale

La voie ventrale prend place dans les deux hémisphères et comprend deux subdivisions : une interface lexicale et un réseau combinatoire.

Les portions moyenne et supérieure des lobes temporaux moyens correspondent à l'interface lexicale qui relie les informations phonologiques et sémantiques sous la forme d'une interface sons / sens (Hickok & Poeppel, 2000 ; Hickok & Poeppel, 2004), avec une légère dominance de l'hémisphère gauche dans le gyrus temporal moyen et le sillon temporal inférieur postérieurs.

Le réseau combinatoire serait situé dans les portions antérieures du lobe temporal gauche, le gyrus temporal moyen antérieur et le sillon temporal inférieur antérieur. Des données IRMf ont révélé des activations des régions temporales moyennes dans le traitement sémantique lexical (Rodd, Davis & Johnsrude, 2005).

A.4.2.2.2. La voie dorsale

La voie dorsale, quant à elle, serait fortement latéralisée dans l'hémisphère gauche et comprend aussi deux subdivisions : une interface sensori-motrice et un réseau articulatoire.

Sise dans la scissure de Sylvius séparant les lobes pariétal et temporal, l'aire dénommée par Hickok & Poeppel « Spt », est considérée comme le siège de l'interface sensori-motrice. Cette aire rappelle l'aire « Tpt » chez le macaque localisée elle aussi entre le lobe pariétal et le lobe temporal, et pouvant jouer un rôle similaire (Aboitiz & Garcia, 1997 ; Aboitiz, Garcia, Bosman & Brunetti, 2006).

Le réseau articulatoire se trouverait dans le lobe frontal de l'hémisphère gauche, incluant le gyrus frontal inférieur ainsi qu'une partie plus dorsale des cortex moteur et prémoteur. Les auteurs pensent que la voie dorsale pourrait contribuer à la récupération des patrons articulatoires telle que supposée par la Théorie Motrice de la Parole (Lieberman *et al.* ; 1986), *via* le système des neurones miroirs (Rizzolatti & Arbib, 1998 ; Rizzolatti & Craighero, 2003). Mais pour Hickok & Poeppel (2007), cette récupération n'est pas essentielle à la compréhension de la parole : elle pourrait être utile dans des conditions de perception dégradée. Ils pensent que ce circuit audio-moteur fournit le mécanisme neural de la mémoire phonologique à court terme (Jacquemot & Scott, 2006). De plus, les interactions audio-motrices pourraient être cruciales dans l'apprentissage de nouveaux mots qui

impliquent de générer des « représentations sensorielles » de ces mots, codant les séquences de segments ou de syllabes. Ces « représentations sensorielles » pourront alors être utilisées, afin de produire les séquences motrices articuloaires.

Les figures A11 et A12 sont une illustration du modèle à deux voies de Hickok & Poeppel (2007).

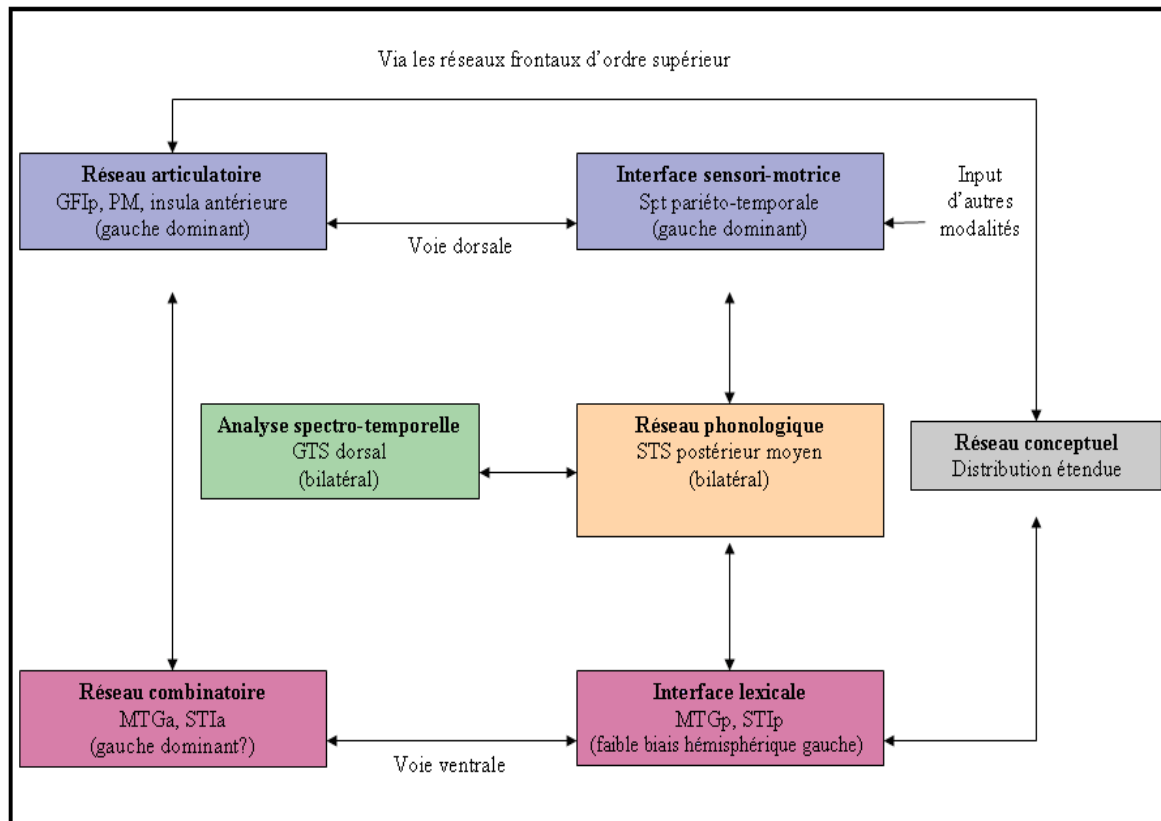


Figure A11 : Représentation schématique du modèle à deux voies de traitement du langage. (Tiré et adapté de Hickok & Poeppel, 2007).

- GFI : Gyrus Frontal Inférieur,
- GTM : Gyrus Temporal Moyen,
- GTS : Gyrus Temporal Supérieur,
- PM : Cortex Prémoteur,
- Spt : aire dans la fissure de Sylvius à la jonction pariéto-temporale,
- STI : Sillon Temporal Inférieur,
- STS : Sillon Temporal Supérieur.

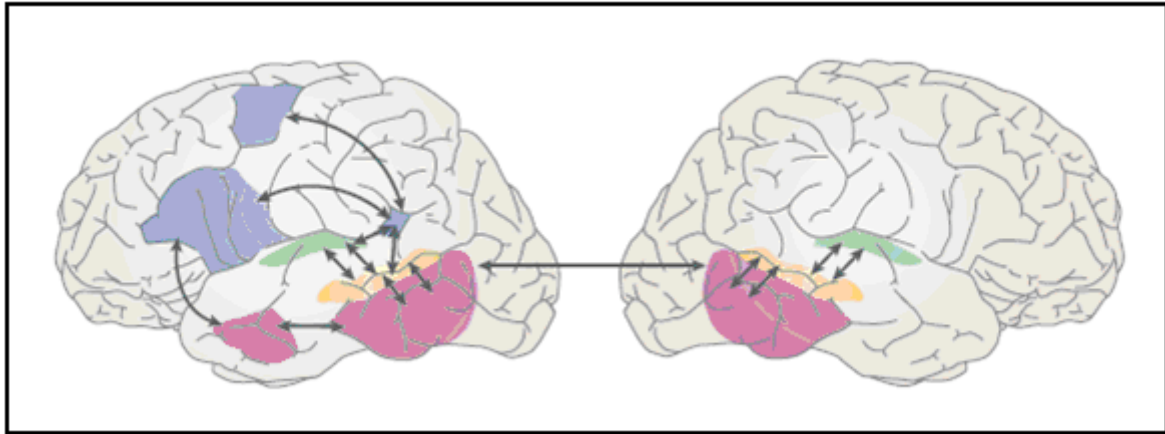


Figure A12 : Localisations anatomiques des composantes du modèle de deux voies de traitement du langage (Tiré de Hickok & Poeppel, 2007).

- En vert : gyrus temporal supérieur dorsal
- En jaune : sillon temporal supérieur postérieur
- En mauve : voie ventrale
- En bleu : voie dorsale

Nous retiendrons particulièrement dans le modèle à deux voies de Hickok & Poeppel (2007) que le système de perception de la parole englobe les deux lobes temporaux, en leur assignant des rôles fonctionnels distincts.

Les gyri et sillons temporaux supérieurs pourraient assurer le décodage acoustico-phonétique et l'accès aux « représentations phonologiques » correspondant aux deux premiers stades de l'évaluation et de l'intégration des traits dans le modèle de Massaro.

Certains auteurs pensent que ce sont les sillons temporaux supérieurs qui pourraient assurer les fonctions d'intégration audiovisuelle dans la perception du langage articulé (Wright *et al.*, 2003). Nous avons vu précédemment que l'étude de Calvert *et al.* (1997) soulève le débat du recrutement par le flux visuel des aires auditives primaires. Bernstein *et al.* (2002) avancent un argument de poids qui touche aux problèmes de localisation des cortex auditifs primaires. Pour Hickok & Poeppel (2007), c'est l'aire « Spt » qui serait une interface sensori-motrice et qui assurerait l'intégration des autres entrées sensorielles. Rappelons que dans le domaine visuel, la voie dorsale a été la première associée à la fonction « où » (Ungerleider & Haxby, 1994). Cependant, selon Hickok & Poeppel (2007), certains travaux suggèrent un rôle plus général dans l'intégration visuomotrice (Andersen, 1997 ; Milner & Goodale, 1995).

Malheureusement, les auteurs considèrent que la perception de la parole est fortement associée à une seule modalité (Hickok & Poeppel ; 2000). L'audition est bien sûr le vecteur préférentiel, mais nous savons qu'il ne faut pas mésestimer le rôle des autres modalités impliquées dans la perception langagière. L'idée d'une contribution adaptable des différentes modalités nous apparaît plus défendable, et nous suivons sur ce point le modèle de Massaro.

Il convient d'insister sur le fait que la dissociation anatomique en deux voies n'induit pas un clivage complet ; des différences fonctionnelles sont supposées, mais ces deux réseaux sont des canaux parallèles qui interagissent.

Le dernier point essentiel est l'implication des structures frontales dans la voie dorsale qui pourraient, grâce au système des neurones miroirs, contribuer à la production de la parole mais aussi faciliter sa perception.

A.5. Conclusion et hypothèses

Pendant longtemps, les noms de Broca et de Wernicke ont été associés à des zones cérébrales supposées être spécialisées respectivement pour la production et pour la perception de la parole. Si les travaux de ces deux pionniers nous indiquent que les processus de production et de perception, et les régions cérébrales qui les sous tendent, sont partiellement dissociés, il n'existe pas pour autant une ségrégation étanche entre eux. Les données expérimentales acquises en TEP et en IRMf montrent que les régions associées à l'Aire de Broca peuvent être recrutées, lors de la perception langagière. Plusieurs hypothèses ont été formulées, afin d'expliquer ce phénomène, dont l'une postule que l'Aire de Broca serait le siège des « représentations phonologiques ». L'accès au lexique mental se ferait aussi par le biais de telles « représentations ».

Une correspondance a longtemps été postulée entre les segments phonétiques présents dans le signal acoustique et les phonèmes d'une langue identifiés grâce aux paires minimales. En réalité, les signaux acoustiques sont plus proches de l'écriture cursive que d'un texte typographié. La variabilité est à la fois fréquentielle et temporelle. Du fait des différences dans la taille des cavités articulatoires des locuteurs (liées à leur âge, à leur sexe, *etc.*), leurs fréquences de résonance diffèrent d'un individu à l'autre. Chaque locuteur génère ainsi des signaux de parole avec une fréquence fondamentale différente ; ce phénomène induit aussi une répartition variable des formants sur l'échelle des fréquences. Par ailleurs, un locuteur peut parler plus ou moins rapidement en fonction des situations, et changer de vitesse d'élocution à l'intérieur d'un même énoncé. Cette variabilité dans l'organisation des sons de parole n'entrave cependant pas la stabilité perceptive, mais les difficultés posées par l'appariement entre les « représentations » proviennent aussi de ces contraintes inhérentes à la coproduction de segments proches. En effet, si les « représentations linguistiques » auxquelles devra être apparié le signal acoustique sont des unités segmentales discrètes, ce signal n'est pas, du fait de la coarticulation, constitué d'une séquence discrète de segments. Ainsi, chaque unité identifiée par un auditeur est extraite d'une portion de signal dont les caractéristiques dépendent aussi de la prononciation des segments qui l'environnent. Ce phénomène de coarticulation est lié au mode de production de la parole qui consiste à préparer les mouvements articulatoires correspondant à un segment, alors même que le segment qui précède est en cours de prononciation (effets d'anticipation). Les effets coarticulatoires se

manifestent aussi par persévérance ; ainsi, nous générons des mouvements destinés à produire un son alors que le son précédent n'est pas encore terminé. La relation entre les unités phonétiques et les signaux acoustiques et visuels n'est pas biunivoque (Nguyen, 2001).

Les connaissances lexicales pourraient faciliter la tâche du système perceptif en inhibant ou en renforçant certaines variables dans les situations problématiques. Afin de pouvoir identifier correctement les phonèmes constituant les mots, l'une des solutions qui ont été envisagées, pour faciliter l'appariement avec des « représentations phonémiques ou phonologiques », consiste à faire intervenir des procédures de rétroaction des niveaux lexicaux vers les niveaux prélexicaux. La majorité des indices qui ont été proposés comme marqueurs potentiels des frontières lexicales sont, en fait, de nature suprasegmentale. En effet, comment différencier deux suites identiques sur le plan segmental, comme « en terrain connu » et « en terre inconnue » ([ã t ε R ã k ɔ n y]), si ce n'est par le biais d'indices prosodiques ? Cette caractéristique implique la mise en relation des informations portées par plusieurs segments ou groupes de segments de la chaîne parlée. La comparaison des indices entre eux, ou les caractéristiques spécifiques d'une séquence de segments, fournit une information sur l'éventuelle présence ou absence d'une frontière lexicale.

La surdité verbale pure nous apprend que les niveaux prélexical et lexical bénéficient d'une relative indépendance. Mais les modèles d'accès au lexique mental postulent, pour la plupart, des mécanismes de rétroaction et de facilitation entre les différents niveaux représentationnels. L'étude bien connue de Mehler (1981) nous apprend que la perception d'une amorce syllabique facilite l'accès aux « représentations lexicales » des mots qui la contiennent. Les théories phonétiques cherchent à expliciter le fonctionnement du ou des niveaux prélexicaux. La Théorie Motrice (Liberman *et al.*, 1967 ; Liberman & Mattingly, 1985) a connu un regain d'intérêt suite à la découverte du système des neurones miroirs, qui pourrait être recruté afin d'accéder aux « représentations articulatoires ». Si, comme le disait Stetson (1951), la parole est davantage un ensemble de gestes rendu audible que des sons produits par des gestes, alors qu'en est-il des visèmes, ces gestes visibles ? L'étude de Sumbly & Pollack (1954) avait déjà mis en lumière l'importance de la source visuelle lorsque la source audible est perturbée par un environnement bruyant. Le système des neurones miroirs pourrait être recruté par la dimension visuelle de la parole, en déclenchant l'accès aux « représentations articulatoires ». Cet accès faciliterait à son tour l'accès aux « représentations sémantiques ». Massaro (1988, 1990, 2001) pense que les différentes sources perceptives sont

toujours intégrées. Une source, pour nous la vision, aura une influence maximale lorsqu'une autre est ambiguë, pour nous l'audition, et pourra le cas échéant désambiguïser la perception et par conséquent l'interprétation de la parole. La découverte de la classe spécifique baptisée les neurones miroirs échos relativise ce point de vue. Si des neurones miroirs de type « acoustique » existent, quelle est la nécessité d'avoir des neurones miroirs « visuels » ?

C'est que le système des neurones miroirs n'est pas destiné à la seule parole, mais bien plus probablement à toute interprétation de mouvements biologiques.

L'une des méthodes qui ont été proposées, afin de rendre compte de la capacité du système cognitif humain à gérer la variabilité consiste à effectuer une normalisation, permettant l'appariement entre un signal physique de forme variable et un référent. Dans notre domaine, cette normalisation pourrait être spectrale et / ou temporelle, mais aussi visuelle, voire articulaire (kinesthésique).

L'idée d'une prépondérance de la source acoustique se heurte principalement à l'hétérogénéité des traits présents dans les différentes langues étudiées. Pourquoi certains traits sont partagés par l'ensemble des langues, pourquoi certains sont sous-représentés ? Est-ce que tous les traits ont les mêmes propriétés ?

Les gestes articulatoires sont, à l'évidence, aussi variables que les signaux acoustiques qu'ils génèrent. Néanmoins, il est envisageable que des mécanismes de type catégoriels servent de « régulateurs » de la variabilité. Le phénomène de la perception catégorielle soulève toujours de nombreuses questions sur sa spécificité dans le domaine langagier (Cutting & Rosner, 1976 ; Jusczyk, Rosener, Cutting, Foard & Smith, 1977 ; Campbell, Woll, Benson & Wallace, 1999 ; Bimler & Kirkland, 2001 ; Bühlhoff & Newell, 2004) ou encore sur l'unicité de ce phénomène dans le règne animal (Fitch, Miller & Tallal, 1997 ; Wyttenbach, May & Hoy, 1996). Fry, Abramson, Eimas & Liberman, (1962) ont mené une étude sur la perception de voyelles synthétiques. Contrairement aux premiers résultats concernant des occlusives, les auditeurs étaient capables de distinguer deux stimuli associés à une même voyelle. Il se pourrait que la perception catégorielle n'intervienne que dans les traitements des consonnes (Schouten, Gerrits & Van Hesson, 2003) ou en fonction de la tâche (Gerrits & Schouten, 2004). S'il est envisageable que les consonnes et les voyelles ne soient pas traitées selon le même mode, il est aussi concevable que la perception catégorielle soit couplée à un système de perception continue. Une étude de Serniclaes, Ventura, Morais & Kolinsky (2005)

révèle que les personnes illettrées sont sensibles au phénomène de la perception catégorielle, au même titre que des adultes normo-lecteurs.

Les modèles d'accès au lexique mental diffèrent, mais ils se sont tous concentrés sur la façon dont les « représentations phonologiques » entrent en contact avec les « représentations lexicales ». Les traitements auditifs et / ou visuels sont réalisés par concordance partielle ou totale entre le stimulus et un prototype, jusqu'à ce qu'un item unique demeure. La majorité des modèles suit une analyse linguistique, comprenant un niveau acoustique, phonétique et / ou phonémique, et une analyse phonologique (tous n'ont pas chacune des étapes). Certains modèles d'accès au lexique se sont plutôt bien adaptés au problème de la segmentation lexicale d'un signal de parole continu. Le modèle COHORTE (Marslen-Wilson, 1987) propose un processus de sélection progressive, parmi un ensemble de candidats lexicaux possibles, afin de réaliser l'accès au lexique mental. Dans le modèle LAFS de Klatt (1979), le système de traitement de la parole comparerait chaque spectre à court-terme avec des *patterns* spectraux prototypiques, et rechercherait celui qui correspond le mieux au calcul effectué sur le signal. Chaque phonème de la langue serait lié à un nombre considérable de *patterns* prototypiques. L'appariement entre le produit des calculs et les « représentations spectrales prototypiques » stockées en mémoire permettrait alors d'avoir accès à des « représentations phonétiques ». TRACE (McClelland & Elman, 1986) adopte une approche dans laquelle les niveaux d'analyse sont les traits pertinents et les phonèmes, mais celui-ci ne présuppose pas de « représentations phonologiques » de la structure métrique, telles que la syllabe ou la more. SHORTLIST (Norris, 1990) fait intervenir ces structures suprasegmentales, mais ne précise pas un niveau explicite de traitement prélexical : le niveau de traitement phonémique constituerait la pierre de touche de l'édifice. Le modèle SARAH (Mehler *et al.*, 1990) défend la syllabe, qui serait le principal niveau de « représentation », et la propose comme unité élémentaire déclenchant l'appariement vers les « représentations lexicales ». Un problème demeure pourtant, le postulat consistant à considérer la parole comme un processus séquentiel est paradoxal. Au niveau phrastique ou propositionnel, les informations sont bien délivrées de façon séquentielle ; particulièrement pour la langue française, pour laquelle l'ordre des syntagmes (sujet / verbe / complément) est important pour la bonne compréhension. Néanmoins, la combinaison des différentes sources, auditive, visuelle et proprioceptive, peut aussi constituer un argument en défaveur des modèles séquentiels. Qui plus est, à un niveau moins élevé comme celui du mot, nous avons vu que le phénomène de coarticulation vient relativiser cette

séquentialité ; il est possible que les effets de coarticulation agissent sur l'accès au lexique mental (Nguyen, 2001).

Selon nous, les « représentations » « phonétiques », « phonologiques » ou « sémantiques » ne sont pas figées, le terme même de « représentation » s'avère assez contraignant.

Les langages codés (le louchébèm, le javanais ou le verlan, *etc.*) nous apprennent qu'un locuteur a une grande latitude sur ses configurations articulatoires. Les imitateurs, professionnels ou amateurs, mettent en exergue la capacité de reproduire la voix d'autrui, et par conséquent d'agir consciemment sur la sienne. *A contrario*, la perception de la parole tolère des variations beaucoup moins marquées. Un auditeur pourra éprouver de grandes difficultés, lors d'une conversation avec un locuteur ayant un accent régional très différent. La capacité cognitive de traitement de la variabilité, aussi grande soit-elle, a ses limites. Nous pensons que les « représentations phonologiques » contiennent deux sous-composantes, l'une pour la production, et l'autre pour la perception de la parole. C'est pour cette raison que le modèle à deux voies de Hickok & Poeppel (2007) s'est littéralement imposé à nous.

Les théories phonétiques cherchent à élucider quels indices présents dans les stimuli distaux sont utilisés pour l'appariement avec des « représentations phonologiques », ce qui correspond aux deux premiers stades du modèle à deux voies. Ensuite, ce sont les modèles d'accès au lexique mental qui s'attachent à éclaircir l'étape suivante, durant laquelle la voie ventrale interviendrait préférentiellement. La voie dorsale pourrait faciliter ces processus et pourrait préparer l'éventuelle réponse de l'auditeur, mais aussi faciliter le traitement des « représentations phonologiques ».

La variabilité présente dans les signaux linguistiques fait écho à celle observée dans les études anatomo-fonctionnelles. L'hétérogénéité des résultats issus des techniques d'imagerie pourrait être expliquée par une certaine confusion, dans les tâches expérimentales, ainsi que dans les niveaux de traitement et d'analyse que ces tâches expérimentales impliquent (Hickok & Poeppel, 2007). Un point essentiel avancé par Hickok & Poeppel est que le contraste entre des stimuli langagiers et non langagiers ("*speech vs. non speech*") ne peut pas exclure l'implication d'une région corticale. Hickok & Poeppel (2007) notent qu'il est difficile d'isoler des activations différentes pour des mots et des logatomes, de plus certaines aires seraient impliquées à la fois dans les traitements temporels rapides pour des signaux langagiers et non langagiers (Joanisse & Gati, 2003). Une région activée pour la

parole, mais aussi pour des stimuli non langagiers acoustiquement similaires, peut être impliquée dans le traitement de la parole. Il est possible d'opérer une distinction entre des études impliquant des tâches de “*speech processing*”, “*speech perception*”, et “*speech recognition*” (« traitement de la parole », « perception de la parole » et « reconnaissance de la parole » ; Poeppel, 2003). Pour l'auteur, le premier de ces termes recouvre les deux autres, il s'agit de toutes expériences utilisant la parole. Le terme de « perception de la parole » est utilisé pour les tâches sublexicales, alors que celui de « reconnaissance de la parole » est réservé aux études impliquant un accès au lexique mental.

Notre expérience portant sur la discrimination de syllabes naturelles, prononcées par une locutrice francophone, devrait s'inscrire dans l'étude de la « perception de la parole ». Il est difficile d'affirmer que la perception de syllabes faisant partie du système phonologique des sujets ne puisse pas déclencher un processus d'appariement lexical. La distinction entre la « perception de la parole » et la « reconnaissance de la parole » telle que postulée par Hickok & Poeppel (2007) n'est peut-être pas une réalité si tranchée.

Nos hypothèses :

L'interrogation liminaire et générale, soulevée dans notre travail, concerne le rôle et l'apport de la dimension visuelle dans les traitements perceptifs de la parole articulée. En effet, nous considérons que la parole est un phénomène multimodal, comme le suggèrent les études comportementales, d'une part, de la compréhension du langage oral (Sumby & Pollack, 1954), et d'autre part, de la perception de la parole au niveau prélexical (McGurk & MacDonald, 1976 ; Schwartz *et al.*, 2004).

L'intégration de la dimension visuelle dans les processus de discrimination phonologique est-elle sous-tendue par le recrutement de régions cérébrales dédiées, ou par une modulation de l'activité des réseaux impliqués dans le traitement phonologique ?

La réponse à cette question nécessite d'isoler le processus d'intérêt, à savoir l'intégration de la dimension visuelle. Pour ce faire, nous avons utilisé le paradigme de la soustraction cognitive (*cf.* A.1.4.3.2.1.), avec une tâche expérimentale et une tâche contrôle qui ne se différencient, lors du traitement d'un stimulus phonologique, que par l'apport de mouvements visuels pertinents de la parole.

Nous avons aussi introduit un second niveau de différenciation pour chacune des tâches en comparant trois paires syllabiques, faisant partie du système phonologique du français contemporain, entre elles, et par rapport à des stimuli non phonologiques.

Nous avons soumis les participants à deux tâches de discrimination de paires syllabiques s'opposant sur un trait pertinent. Pour la condition de contrôle, nous avons choisi des syllabes CV, que nous avons modifiées en inversant le décours temporel, et en manipulant leurs caractéristiques acoustiques par ordinateur (*cf.* B.1.2.2.). Nous considérons que ces « sons » demeurent prononçables par un locuteur, mais qu'ils n'ont plus de valeur phonologique.

Pour le traitement cérébral de la parole audiovisuelle, deux hypothèses concurrentes émergent de notre revue de la littérature. La première avance l'idée que les zones impliquées dans les traitements acoustico-phonétiques vont assurer le traitement de la dimension visuelle, ce qui pourrait se traduire par une augmentation significative de la réponse hémodynamique dans le cortex auditif primaire en modalité AV animée par rapport à l'AV fixe (Calvert *et al.*, 1997 ; Pekkola *et al.*, 2004). La deuxième hypothèse postule que l'intégration bimodale pourrait dépendre de régions plus spécialisées comme le Sillon Temporal Supérieur (Wright *et al.*, 2003 ; Beauchamp *et al.*, 2004 ; Beauchamp *et al.*, 2003 ; Callan *et al.*, 2003). Le modèle de Hickok & Poeppel (2007) propose, quant à lui, que l'aire qu'ils nomment « Spt » (sise dans la partie postérieure de la scissure de Sylvius au niveau de la jonction temporo-pariétale) soit impliquée dans l'intégration des sources d'informations supplémentaires.

La comparaison entre les deux modalités, AV fixe et AV animée, pourrait aussi révéler une activation bilatérale de l'aire MT/V5 (Aire de Brodmann 19), spécialisée dans la perception visuelle des mouvements (Tong, 2003 ; Zeki *et al.*, 1991 ; Smith *et al.*, 1998 ; Puce *et al.*, 1998). Selon Hall *et al.* (2005), les activations de cette région ne semblent pas liées au domaine langagier, mais à la perception du mouvement en général.

Dans le cadre théorique du modèle de Hickok & Poeppel (2007), nous nous attendons à ce que la comparaison des activations dans les deux modalités, AV fixe et AV animée, nous montre un recrutement des régions impliquées dans les deux premières étapes du modèle à deux voies, à savoir l'« analyse spectrotemporelle » et « le réseau phonologique », mais aussi un recrutement de la voie dorsale *via* le système des neurones miroirs (Rizzolatti & Craighero, 2003 ; Fadiga & Craighero, 2006 ; Buccino *et al.*, 2004a ; Buccino *et al.*, 2004b ; Ferrari *et al.*, 2003 ; Iacobini & Dapretto, 2006). Dans ce cas de figure, des activations du

lobe frontal sont à prévoir probablement dans l'hémisphère gauche, en particulier dans l'Aire de Brodmann 6 (Wilson *et al.*, 2004 ; Skipper *et al.*, 2005 ; Pulvermüller *et al.*, 2006). Une des questions principales est de savoir si l'aire de Broca peut être recrutée (Kohler *et al.*, 2002 ; Binkofski *et al.*, 2000 ; Grodzinsky, 2006). Les tâches de discrimination pouvant impliquer des structures frontales et pariétales (Caplan *et al.*, 1995 ; Doyon *et al.*, 2003), la question sous-jacente est de savoir si ces activations sont liées à la tâche ou aux traitements perceptifs de la parole articulée. *A priori*, un processus d'appariement lexical est envisageable à partir de la seule perception syllabique (Mehler *et al.*, 1990 ; Ferrand *et al.*, 1995). S'il a lieu, nous pourrions constater une implication de la voie ventrale, se traduisant par l'activation des régions temporales inférieures, pour les paires syllabiques, mais pas pour les stimuli non phonologiques.

Les potentiels évoqués, recueillis durant l'acquisition IRMf, pourraient nous éclairer sur la dynamique du recrutement des différentes régions cérébrales participant au traitement phonologique. La précision temporelle des potentiels évoqués pourraient, de plus, apporter des éléments nous permettant d'identifier des différences dans la chronométrie, l'amplitude, et la distribution des réponses suscitées par nos trois paires syllabiques (paires s'opposant sur la labialité, sur les lieux d'articulation et sur le voisement).

Suivant les résultats des études de Van Wassenhove *et al.* (2005), qui ont observé des potentiels évoqués plus précoces et moins amples en réponse aux stimuli en modalité audiovisuelle animée par rapport à la modalité auditive seule, un effet semblable dans nos résultats indiquerait que ce dernier soit dû à la présence de mouvements linguistiques pertinents.

L'environnement de l'IRM implique des conditions perceptives particulières. Le bruit des gradients de l'IRM, le système de transmission pneumatique, les électrodes de l'EEG, mais aussi la position allongée, ainsi que le confinement dans le tube de l'IRM sont des facteurs susceptibles d'influer sur les processus de perception de la parole, et partant sur les données comportementales recueillies durant l'expérience. Pour ces raisons, nous avons cherché à mesurer l'effet de l'environnement expérimental, en comparant les données comportementales de la discrimination dans deux autres environnements, l'un exempt de toutes perturbations, et l'autre avec une perturbation sonore : le bruit enregistré des gradients de l'IRM.

L'hypothèse principale est que l'impact des indices visuels est fonction du degré de perturbations de l'environnement. Dans les environnements perturbé et bruyant, l'influence des indices visuels pertinents devrait être plus important que dans l'environnement silencieux, pour lequel nous ne pensons pas observer de différences significatives des performances entre nos deux modalités. Les scores de discrimination devraient être supérieurs ou égaux mais pas statistiquement inférieurs pour la modalité audiovisuelle animée par rapport à la modalité audiovisuelle fixe.

L'hypothèse corollaire est que l'impact des indices visuels devrait être différent selon les types de paires. En introduisant trois paires syllabiques, qui ne sont pas équivalentes en termes d'indices visuels, nous pensons pouvoir évaluer l'apport de différents indices visuels. Dans les environnements bruyant et perturbé (IRMf/EEG), nous pensons que les paires s'opposant sur la labialisation vocalique (étirée vs. arrondie [i y]) va être influencée favorablement par la présence de mouvements visuels pertinents, tout comme les paires s'opposant sur les lieux d'articulation (extra vs. intra-buccales : [p t] et [b d]). En revanche, pour les paires s'opposant sur le voisement (sourdes vs sonores : [p b] et [t d]), faisant partie d'une classe unique de visème, nous pensons que l'influence des indices visuels sera nulle (Cathiard, 1994 ; Roy, 2005).

Nos hypothèses peuvent être récapitulées comme suit :

Première hypothèse :

La perception syllabique en modalité AV animée, par rapport à la modalité AV fixe, pourra :

- soit engendrer une activation plus forte dans le cortex auditif primaire,
- soit recruter une zone spécifiquement impliquées dans l'intégration audiovisuelle, comme le Sillon Temporal Supérieur.

La perception visuelle du mouvement peut aussi induire le recrutement bilatéral de l'aire MT/V5, dans la région occipito-temporale.

Sur le plan électrophysiologique, l'intégration de l'information AV peut se traduire par des réponses plus précoces en modalité AV animée, par rapport à la modalité AV fixe (Van

Wassenhove *et al.*, 2005). Dans la mesure où la perception audiovisuelle implique le traitement du mouvement contrairement à l'AV fixe, qui ne fournit pas d'indices visuels dynamiques, nous pouvons envisager l'apparition de réponses évoquées dans les régions participant au traitement des informations sensori-motrices, comme les régions perirolandiques.

Deuxième hypothèse :

En référence au modèle à deux voies, les structures qui sous-tendent l'analyse « spectrotemporelle » et le « réseau phonologique » devraient être impliquées dans la perception syllabique, et ce, dans les deux modalités AV fixe et AV animée.

En modalité AV animée, la perception syllabique pourrait aussi activer les régions frontales de la voie dorsale, *via* le système des neurones miroirs (Aires de Brodmann 6 et/ou 44 et 45).

Le recrutement de régions supposées correspondre à l' « interface lexicale » et au « réseau combinatoire » de la voie ventrale, lors de la discrimination des paires phonologiques pourrait refléter un processus d'appariement lexical.

Troisième hypothèse :

En ce qui concerne les données comportementales, nous pouvons émettre l'hypothèse que l'impact des indices visuels sera maximal dans l'environnement perturbé, et inexistant dans l'environnement silencieux. De plus, cet impact devrait être fonction de la catégorie visémique.

B. Partie expérimentale

B.1. Matériel et Méthode

Cette deuxième partie présente, d'une part, la méthode et les moyens expérimentaux employés pour l'acquisition des données, et d'autre part, les trois chapitres exposant nos résultats comportementaux, IRMf et EEG.

Le Comité Consultatif de Protection des Personnes Participant à une Recherche Biomédicale Alsace 1 a donné un avis favorable pour cette recherche, le 10/01/2006.

B.1.1 Les participants

Pour cette étude, nous avons recruté les participants parmi les étudiants de Licence de différentes UFR de l'Université de Strasbourg. Tous les sujets ont été informés des objectifs de la recherche avant d'y participer. Les personnes ayant pris part à l'expérience IRMf/EEG ont été informées des risques potentiels de la méthode et ont signé préalablement un consentement éclairé.

La première rencontre s'est effectuée à l'Institut de Phonétique de Strasbourg (IPS). Avant d'effectuer la tâche expérimentale, les sujets étaient informés par nos soins du but de l'étude et des critères d'exclusion. En l'occurrence, il ne fallait pas être gaucher ou porteur de lunettes, présenter de déficits langagiers quels qu'ils soient, incluant la dyslexie passée ou présente. Un questionnaire a été rempli par tous les sujets nous renseignant sur la date et le lieu de naissance, la préférence manuelle et d'éventuels antécédents de déficits auditifs ou visuels, de troubles de la parole ou du langage écrit ou oral. D'autres facteurs d'exclusion existent et sont liés à l'IRM comme la présence de prothèse métallique ou la claustrophobie. Deux sujets nous ont signalé qu'ils avaient été diagnostiqués dyslexiques durant leur enfance. Six gauchers se sont présentés, nous les avons informés qu'ils ne pourraient pas être recrutés pour cette étude. Malgré cela, ils ont tous souhaité participer aux tests comportementaux, dans l'environnement silencieux. Chaque sujet a effectué le test de l'Alouette (Lefavrais, 1967), et un test de dénomination rapide *RAN* ("*Rapid Automated Naming*"), afin de vérifier le niveau de lecture et la vitesse d'accès au lexique. Tous ont obtenu de bons résultats ne nécessitant l'exclusion d'aucun sujet.

Cinquante-cinq personnes ont pris part au test dans l'environnement silencieux (chambre sourde de l'IPS). Ce groupe était composé de vingt-huit femmes et de vingt-sept hommes (âge moyen : 22.6 ± 3.6) tous francophones nés en France métropolitaine de parents

eux aussi francophones. Vingt-neuf d’entre elles ont participé au test dans l’environnement perturbé, IRMf/EEG, parmi lesquelles un sujet pilote (une collègue phonéticienne), dont les résultats n’ont pas été analysés. Parmi ces vingt-huit sujets, l’un d’entre eux n’a pu mener à bien le protocole, et s’est extrait du scanner après dix minutes.

Après les premières analyses en IRMf, il est apparu que l’un des participants ne pouvait être pris en compte dans les analyses, du fait d’une incompatibilité avec le modèle de réponse *BOLD*. Les analyses portent par conséquent sur 26 sujets, 14 femmes et douze hommes (âge moyen : 22.6 ± 3.7). Finalement, 16 sujets ayant participé aux études dans les environnements silencieux et perturbé, 11 femmes et 5 hommes (âge moyen : 22.75 ± 4.1), ont aussi effectué les tâches de discrimination dans l’environnement bruyant.

Parmi les 26 sujets ayant pris part à l’acquisition en IRMf/EEG, seuls 11 d’entre eux ont pu être considérés lors de l’analyse des potentiels évoqués (8 femmes et 3 hommes ; âge moyen : 22.55 ± 3.1). Ce sont principalement des artefacts (mouvements oculaires et artefacts cardiaques) qui ont suscité l’exclusion des quinze autres participants, ainsi que des scores de discrimination inférieurs au hasard.

Tableau B1 : Âges et nombre de participants par environnement expérimental.

SUJETS	HORS SCANNER						SCANNER					
	Silencieux			Bruyant			IRMf			EEG		
Age	Moy.	E-T		Moy.	E-T		Moy.	E-T		Moy.	E-T	
		22,6	± 3,6		22,75	± 4,1		22,6	± 3,7		22,55	± 3,1
Nombre	Total	Fem	Masc	Total	Fem	Masc	Total	Fem	Masc	Total	Fem	Masc
	55	28	27	16	11	5	26	13	13	11	8	3

B.1.2. Paradigme de discrimination à choix forcé

Nous avons choisi d’étudier la perception syllabique en utilisant un paradigme de discrimination à choix forcé « AX ». Dans ce paradigme, deux stimuli sont présentés l’un après l’autre séparés par une pause. Au sein d’un essai composé de deux syllabes ou de deux stimuli non phonologiques, les participants devaient juger si le second stimulus était identique (AA) ou différent du premier (AB). Ce paradigme a été appliqué aux deux modalités audiovisuelle fixe (AVf) et audiovisuelle animée (AVa), dans nos trois environnements. Nous avons programmé les deux tâches de discrimination avec le logiciel Inquisit[®] de la firme Milliseconds[™].

Chaque tâche de discrimination comprenait huit catégories, constituées de 40 essais, soit un total de 320 paires. Les catégories sont scindées en deux groupes, l'un comprenant les paires appelant une réponse « identique » (AA), et l'autre comprenant les paires appelant une réponse « différente » (AB). Dans ces deux groupes, nous avons introduit des paires ne faisant pas partie du système phonologique du français contemporain. Ces paires non phonologiques sont utilisées afin de faire apparaître les zones cérébrales impliquées dans les processus de traitement de la parole, grâce à la méthode de la « soustraction cognitive » (*cf.* A.1.4.4.2.2.).

Nous avons *in fine* quatre catégories de paires syllabiques, une constituée de paires identiques, trois autres pour les paires syllabiques phonologiques, et deux catégories de paires non phonologiques : paires identiques et différentes. Dans le contexte particulier de l'expérience IRMf/EEG, des paires identiques phonologiques et différentes non phonologiques ont été placées durant les périodes de gradients de l'IRM, afin d'éviter l'habituation des participants à une présentation uniquement hors gradients. Rappelons que les gradients de l'IRM (voir figure B1) sont les périodes d'acquisition d'un volume cérébral durant lesquelles un bruit important est généré par le scanner. Ce bruit est dû à l'alternance des champs magnétiques intenses (*cf.* Introduction).

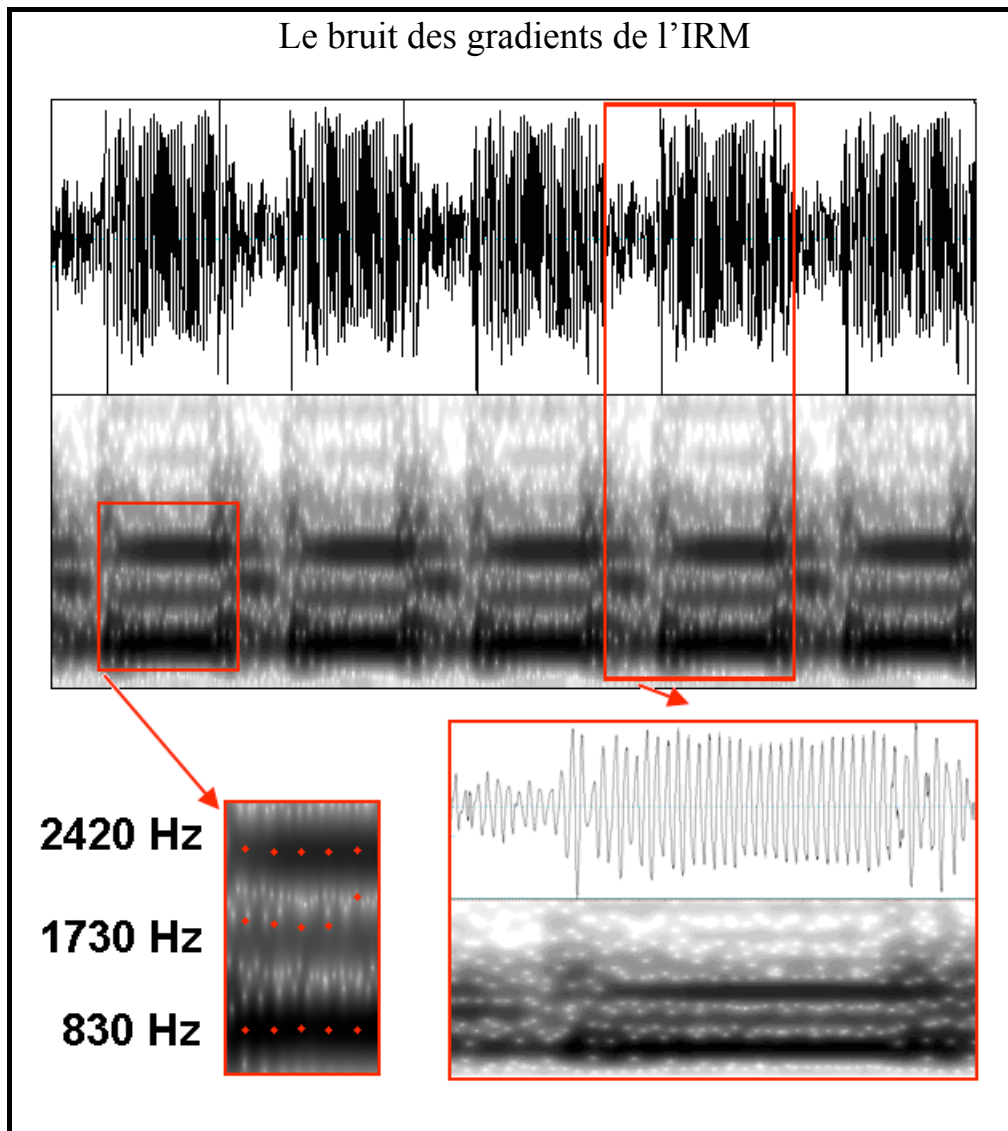


Figure B1 : Oscillogramme et spectrogramme du bruit produit lors des gradients de l'IRM, avec l'agrandissement d'une période et les valeurs des trois premiers formants en Hz.

Les mêmes tâches de discrimination ont été effectuées dans les différents environnements. Au cours des trois phases expérimentales, les participants donnaient leur réponse grâce à une souris d'ordinateur. Le clic gauche signifiait que les stimuli étaient différents et le clic droit qu'ils étaient identiques. Les données comportementales que nous avons choisies d'observer sont les Temps de Réponse et les scores de réussite à la tâche de discrimination. Les sujets étaient invités à répondre le plus rapidement tout en privilégiant la réussite de la tâche.

B.1.2.1. Les stimuli

Nous avons filmé une locutrice francophone, sans accent identifiable, âgée de 23 ans prononçant les syllabes isolément. Les films ont été réalisés avec le concours du Service Commun de l'Audio Visuel de l'Université de Strasbourg. La caméra (Sony DXC D30-Pal) enregistre vingt-cinq images par seconde, chaque image durant ainsi 40 millisecondes. Chaque séquence vidéo était constituée de neuf images, soit une durée de 360 millisecondes par syllabe. Cette durée a été décidée en fonction de notre protocole EEG/IRMf. Pour la modalité auditive, nous avons utilisé le signal acoustique acquis durant l'enregistrement des séquences filmées.

L'étude de la perception syllabique dans un environnement expérimental comme un scanner IRMf est loin de ressembler à la perception quotidienne de la parole, mais ce type de problème se pose de façon incontournable à tout phonéticien expérimentaliste. Si nous avons choisi de ne pas étudier la perception syllabique au moyen d'une parole synthétique, c'est que nous voulions demeurer dans une perspective « naturelle ». L'environnement expérimental, composé d'un scanner IRM, n'est bien sûr pas une condition très « écologique ». Néanmoins, il est possible de rencontrer des milieux « naturels » aussi bruyant que l'IRM, comme par exemple sur un chantier de construction ou dans une usine. Nous l'avons dit à maintes reprises, la parole est un moyen de communication très robuste, et ce malgré la variabilité ; loin de nous l'idée de contrôler cette variabilité, nous avons néanmoins choisi des syllabes produites par une seule locutrice. Parmi les productions recueillies, l'« homogénéité » acoustique des différents stimuli n'a pas été un critère majeur de sélection. La distinctivité des indices visuels nous a permis de choisir les stimuli qui nous semblaient les plus appropriés, pour étudier l'apport des visèmes en milieu perturbé. Nous avons utilisé le début de la closure consonantique (contact bilabial) comme point de repère. Cela explique les différences relativement importantes entre les durées vocalique et consonantique, entre celles des différents VOT, ainsi que les durées des silences acoustiques durant la closure, pour les plosives sourdes.

Nous avons effectué un cadrage de sorte que n'apparaisse que le bas du visage de notre locutrice (voir figure B2), afin de limiter les distracteurs et de concentrer l'attention des sujets. Cette démarche est critiquable, surtout en considérant notre choix de la parole naturelle dans une perspective écologique. Une présentation du visage complet de la locutrice aurait été

plus en accord avec ce parti pris. Néanmoins, nous avons choisi ce type de présentation afin d'évacuer l'attention que les sujets pourraient porter sur les yeux.



Figure B2 : Cadrage sur le bas du visage de la locutrice articulant une syllabe.

B.1.2.2. Corpus

Nos stimuli sont constitués d'une part de huit syllabes naturelles (voir figure B3) du type Consonne – Voyelle (CV) : [p i b i t i d i p y b y t y d y] et, d'autre part, de huit syllabes naturelles modifiées à l'aide du logiciel Audacity[®] (paires non phonologiques).

Tableau B2 : Durées segmentales et de VOT en millisecondes et valeurs des trois premiers formants en hertz des syllabes phonologiques.

Syllabes	Durée Consonne (ms)	VOT Klatt (ms)	VOT Lisker & Abramson (ms)	Durée Voyelle (ms)	Valeurs formantiques (Hz)		
					F1	F2	F3
p i		31	18	205	325	3146	4693
b i	126	22	-99	234	337	3510	4622
t i		106	96	197	370	3373	4456
d i	171	78	-80	189	325	3328	4374
p y		54	44	223	370	2372	3828
b y	224	30	-189	136	360	2352	3881
t y		96	87	246	279	2281	3965
d y	190	56	-123	170	416	2418	3783

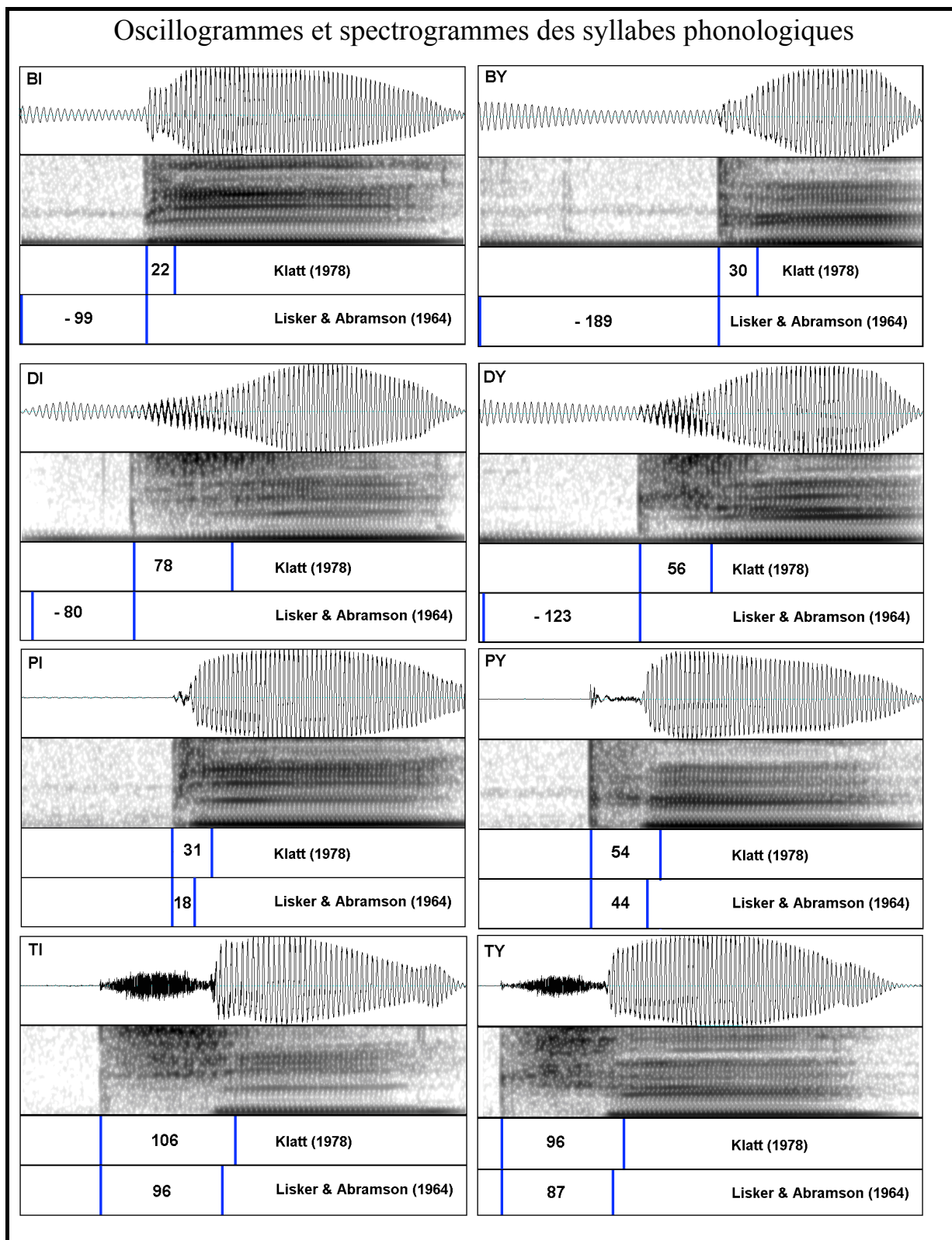


Figure B3 : Vues des oscillogrammes accompagnés des spectrogrammes des huit syllabes faisant partie du système phonologique du français contemporain et utilisées pour réaliser les paires minimales étudiées. Nous avons précisé les valeurs en millisecondes du VOT (indice de voisement), selon Klatt (1978) et selon Lisker et Abramson (1964).

Rappelons que le VOT de Lisker et Abramson (1964) est l'intervalle qui sépare l'explosion ou la détente acoustique (relâchement de l'occlusion) de l'apparition du premier signe périodique régulier sur le signal acoustique temporel (début de vibrations des plis vocaux). Pour ce que concerne le VOT de Klatt (1975), il s'agit de l'intervalle entre l'explosion et l'établissement d'une structure formantique clairement définie (l'apparition de zones de résonances vocaliques stables, le conduit vocal étant suffisamment dégagé).

Nous avons observé des valeurs relativement élevées des VOT pour l'occlusive apicale. Cela semble provenir de deux sources : une réalisation très bruitée (avec explosion/friction) et une production « appuyée », en position initiale absolue.

Les paires non phonologiques ont été créées à partir de syllabes naturelles prononcées par la même locutrice. Nous les avons modifiées afin de les rendre méconnaissables, tout d'abord en inversant le décours temporel des images et des sons, puis en appliquant une distorsion aux signaux acoustiques (voir figure B4). Cette distorsion comprenait une modification des basses fréquences (“*Low Frequency Oscillator*”) de 5 Hz avec une phase de départ de 180 degrés, une profondeur de 75 % avec une résonance de facteur 6, le tout pour une fréquence de l'effet à 85 %.

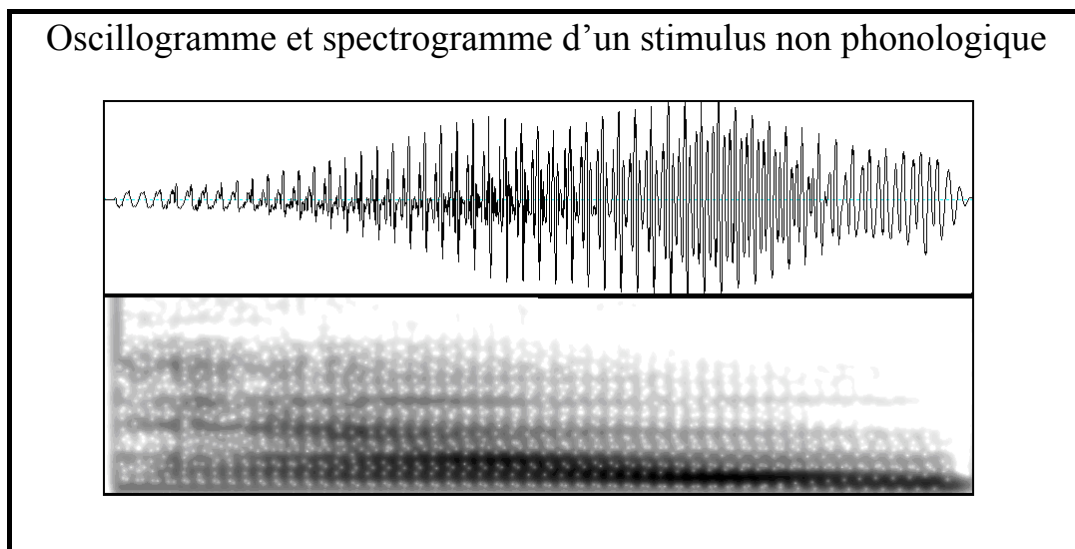


Figure B4 : Oscillogramme et spectrogramme d'une syllabe naturelle modifiée par ordinateur, d'une durée de 360 ms, utilisée afin de réaliser les paires de stimuli non phonologiques.

B.1.2.3. Les paires syllabiques

Les syllabes nous ont permis de réaliser des paires minimales. Les paires syllabiques s'opposant par un seul trait articulatoire pertinent ont été construites afin d'étudier trois contrastes opératoires en français contemporain. Ces paires minimales diffèrent en fonction de la labialisation (étirée *vs.* arrondie [i y]) pour l'opposition vocalique, en fonction du voisement (sourdes *vs.* sonores : [p b] et [t d]), et en fonction des lieux d'articulation (extra *vs.* intra-buccales : [p t] et [b d]).

Tableau B3 : Les paires syllabiques utilisées pour la tâche de discrimination à choix forcé en fonction du trait qui les opposent.

Labialité	Voisement	Lieux d'articulation
[p i p y]	[p i b i]	[t i p i]
[b i b y]	[p y b y]	[d i b i]
[t i t y]	[t i d i]	[t y p y]
[d i d y]	[t y d y]	[d y b y]
[p y p i]	[b i p i]	[t i p i]
[b y b i]	[b y p y]	[d i b i]
[t y t i]	[d i t i]	[t y p y]
[d y d i]	[d y t y]	[d y b y]

B.1.3. Les environnements expérimentaux

La spécificité des contraintes de l'environnement IRMf, *a fortiori* couplée à l'EEG, peut influencer sur les données comportementales. Cela nous a incité à mener deux tests complémentaires. Nous avons par conséquent répliqué l'expérience en EEG/IRMf dans la chambre sourde de l'Institut de Phonétique de Strasbourg. Le premier test visait à obtenir des données comportementales dans un milieu où les perturbations du scanner étaient absentes (environnement silencieux). Les sujets étaient installés confortablement, et il n'y avait aucun bruit de fond pouvant influencer sur la tâche de discrimination. Le deuxième test complémentaire introduisait une perturbation acoustique importante, qui consistait à entendre les gradients de l'IRM enregistrés durant la tâche de discrimination (environnement bruyant).

B.1.3.1. Environnement perturbé (IRMf/EEG)

L'acquisition des données anatomo-fonctionnelles a eu lieu au Laboratoire d'Imagerie et de Neurosciences Cognitives de Strasbourg (LINC, FRE 3289). Nous avons eu accès à un scanner corps entier (MRI S200 Bruker, Bruker Medical GmbH, Ettlingen, Allemagne), dans lequel les participants étaient allongés. La tête de la personne entre dans un cylindre (l'« antenne »), dédié aux acquisitions d'images fonctionnelles du cerveau (voir figure B5). Le diamètre du tube de l'IRM est de 90 cm, ce qui est réhibitoire pour les personnes souffrant de claustrophobie.

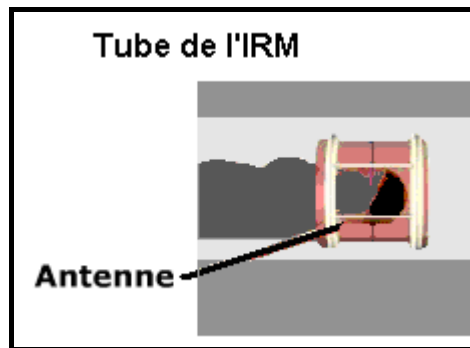


Figure B5 : Représentation schématique de la tête d'un sujet placée dans l'antenne de l'IRM.

Les sujets portaient des lunettes surmontées de prismes permettant de voir les images et les vidéos réfléchies par un miroir vers un écran en papier, placé dans l'axe du scanner. Un vidéo projecteur, à l'intérieur de la chambre contenant le scanner et protégé dans une cage de Faraday, était utilisé afin de projeter les stimuli visuels sur ce miroir (voir figure B6). Les participants entendaient les stimuli sonores, grâce à un casque dépourvu de matériau magnétique, et utilisant un système de transmission pneumatique AVOTEC® "*Silent Scan*" (SS3100 ; Stuart, FL).

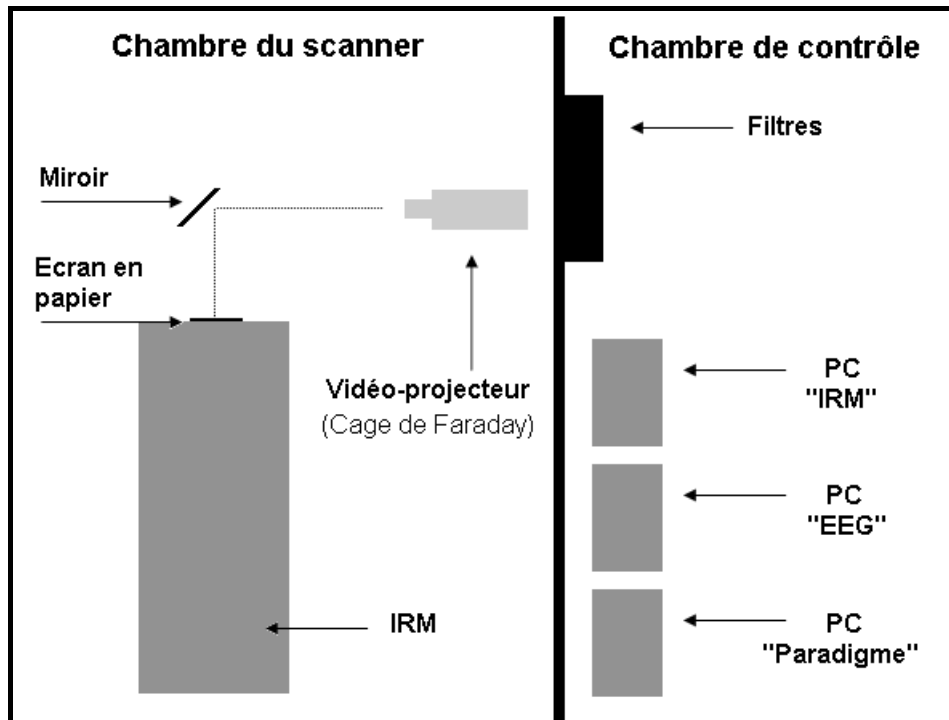


Figure B6 : Représentation schématique du dispositif d'acquisition des données.

Les principales perturbations présentes dans cet environnement sont les suivantes :

- Le bruit des gradients de l'IRM d'une intensité de 87 dB (SPL), mesure effectuée avec un sonomètre (Pulsar M30).
- Le système de transmission pneumatique AVOTEC®.
- Le format des fichiers (« .wav » et « .avi ») était imposé par le générateur de paradigme (Inquisit®), et devait être inférieur à 700 kilo-octets.

Ces perturbations, en particulier la transmission pneumatique, ont engendré des difficultés pour la discrimination des paires s'opposant sur les voyelles (labialité). La voyelle étirée [i] et la voyelle arrondie [y] faisant partie de deux groupes de visèmes distincts, nous avons choisi de maintenir cette catégorie de paires, afin d'optimiser l'apport des indices visuels pertinents. *A contrario*, une qualité sonore relativement acceptable a été préservée pour les paires consonantiques s'opposant sur les lieux d'articulation et sur le voisement.

Une perturbation supplémentaire est liée au port des électrodes nécessaires à l'enregistrement des données EEG. Les sensations douloureuses, provoquées par la pression

des électrodes sur le scalp, ont été limitées par la mise en place d'éponges sur le repose-tête de l'IRM.

Tous les sujets ont effectué d'abord la tâche de discrimination en modalité AV animée, puis la tâche en AV fixe, au cours d'une seule session IRMf/EEG. Nous avons choisi cet ordre de passage parce que les données en AV animée étaient prépondérantes pour notre étude. Deux modifications supplémentaires sont intervenues, lors de l'acquisition en IRMf/EEG. Les essais étaient présentés selon un mode pseudo-aléatoire (imposé par la réponse *BOLD* en IRMf), alors que lors des deux phases en dehors du scanner, nous avons utilisé un ordre totalement aléatoire.

De plus, les tâches de discrimination duraient 24 minutes, lors de l'acquisition anatomo-fonctionnelle, contrairement à celles effectuées dans les environnements silencieux et bruyant, qui n'avaient qu'une durée de 21 minutes. Cet écart est dû à la nécessité, lors d'un protocole événementiel en IRMf, d'inclure une condition de référence, constituée dans cette étude par des événements nuls, c'est-à-dire exempts de stimuli. Cette condition de référence est nécessaire pour l'exploitation des données anatomo-fonctionnelles.

B.1.3.2. Environnement silencieux

La chambre sourde de l'Institut de Phonétique de Strasbourg (IPS) a été utilisée lors des tests dans l'environnement silencieux. Les sujets étaient installés confortablement dans un siège réglable. Ils voyaient les stimuli visuels sur un écran d'ordinateur (Sensy M17GKA de 17"), et entendaient les paires de stimuli par le biais d'un haut-parleur (Sony SRS-200, 22 W) à une intensité sonore de 63 dB (SPL) mesurée avec un sonomètre (Pulsar M30). Avant de débiter les tâches de discrimination, les syllabes phonologiques et les stimuli non phonologiques ont été présentés aux sujets, qui notaient par écrit ce qu'ils avaient perçu. Cette procédure nous a permis de constater que les syllabes naturelles ont été identifiées à 100% ; alors que les stimuli non phonologiques n'étaient pas identifiables.

Contrairement à l'environnement perturbé, la moitié des participants a effectué le protocole audiovisuel animé en premier, puis le protocole audiovisuel fixe ; l'ordre inverse ayant été adopté pour l'autre moitié du groupe. Nous avons été ainsi en mesure de vérifier que l'ordre de passage n'influe pas de manière statistiquement significative sur les résultats.

B.1.3.3. Environnement bruyant

La chambre sourde de l'IPS a été de nouveau utilisée afin de réaliser ces tests. Cet environnement était en tous points identique à l'environnement silencieux, à l'exception de l'introduction de la perturbation sonore constituée par le bruit enregistré des gradients de l'IRM d'une intensité de 87 dB (SPL). Le bruit des gradients était émis par un second haut-parleur (Sony SRS-200, 22 W), contrôlé par un ordinateur portable (HP Compaq nc6120). Là encore, l'ordre de passage des deux tâches a été alterné.

Durant la préparation de notre étude, nous avons à plusieurs reprises testé le protocole expérimental en situation. Nous nous sommes aperçu que l'impact des conditions expérimentales, en particulier le bruit de gradients de l'IRM et le système de transmission pneumatique amagnétique, était plus important pour les paires s'opposant sur les voyelles (labialisée vs. non labialisée) par rapport aux deux autres, c'est-à-dire les paires s'opposant sur les lieux d'articulation et sur le voisement. Les paires s'opposant sur les voyelles ont été maintenues afin de pouvoir évaluer, plus précisément, l'apport des indices visuels combinés à une source sonore très dégradée, dans un environnement perturbé.

B.1.4. Recueil des données IRMf/EEG

Le couplage des deux techniques d'investigation neurophysiologique, que sont l'IRMf et l'EEG, est encore très rarement pratiqué. De nombreuses contraintes sont liées à l'utilisation de chacune des méthodes séparément, et le recueil simultané en induit de nouvelles.

B.1.4.1. Paradigme événementiel

Nous avons utilisé un paradigme d'IRMf événementiel (voir figure B7), car ce type de protocole permet de mesurer la réponse hémodynamique liée à un essai individuel, alors que les paradigmes en blocs ne le permettent pas. Les réponses hémodynamiques liées aux stimuli d'intérêt, dans une condition donnée, seront moyennées. Ensuite, la réponse moyenne, liée aux stimuli dits de « contrôle », sera soustraite à la réponse aux stimuli audiovisuels fixes ou animés. De la même façon, il est possible de contraster une condition expérimentale à une autre, afin de faire ressortir un processus particulier. L'intérêt du protocole événementiel est

de pouvoir isoler les réponses individuelles des stimuli, et de séparer par la suite les événements correspondant aux réponses justes de ceux des réponses fausses. Il présente aussi l'avantage d'être robuste, par rapport à l'habituation et à l'anticipation du sujet.

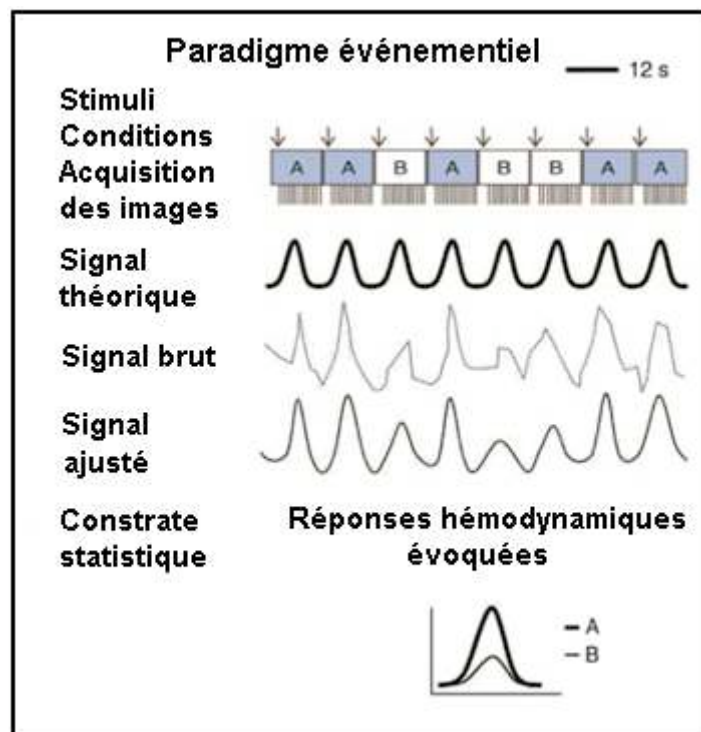


Figure B7 : Représentation schématique d'un paradigme événementiel

L'ordre de présentation des stimuli appartenant aux huit catégories a été pseudo-randomisé, à l'aide du logiciel « Optseq » (<http://surfer.nmr.mgh.harvard.edu/optseq>), afin de ne pas avoir de corrélation entre les événements des différentes conditions expérimentales, et permettre l'extraction de la réponse hémodynamique de chaque condition indépendamment, bien qu'elles se chevauchent. La séquence de stimuli ainsi générée est utilisée pour tous les sujets. Afin d'éviter que les sujets ne s'habituent à une présentation rythmée, deux catégories de stimuli ont été placées durant les gradients de l'IRM. Ces deux catégories n'ont pas été prises en considération dans nos analyses.

Durant l'acquisition d'un volume cérébral (4 sec.), 1800 ms étaient réservées pour la présentation des paires de stimuli. La période pré-stimuli durait 100 ms. Nos syllabes durant 360 ms et l'intervalle inter-stimuli étant de 400 ms, la durée totale de présentation est de 1120 ms (deux fois 360 ms plus 400 ms). L'analyse des potentiels évoqués a été réalisée à partir du début du second stimulus (760 ms). La période post-stimuli durait 580 ms. (Voir figure B8).

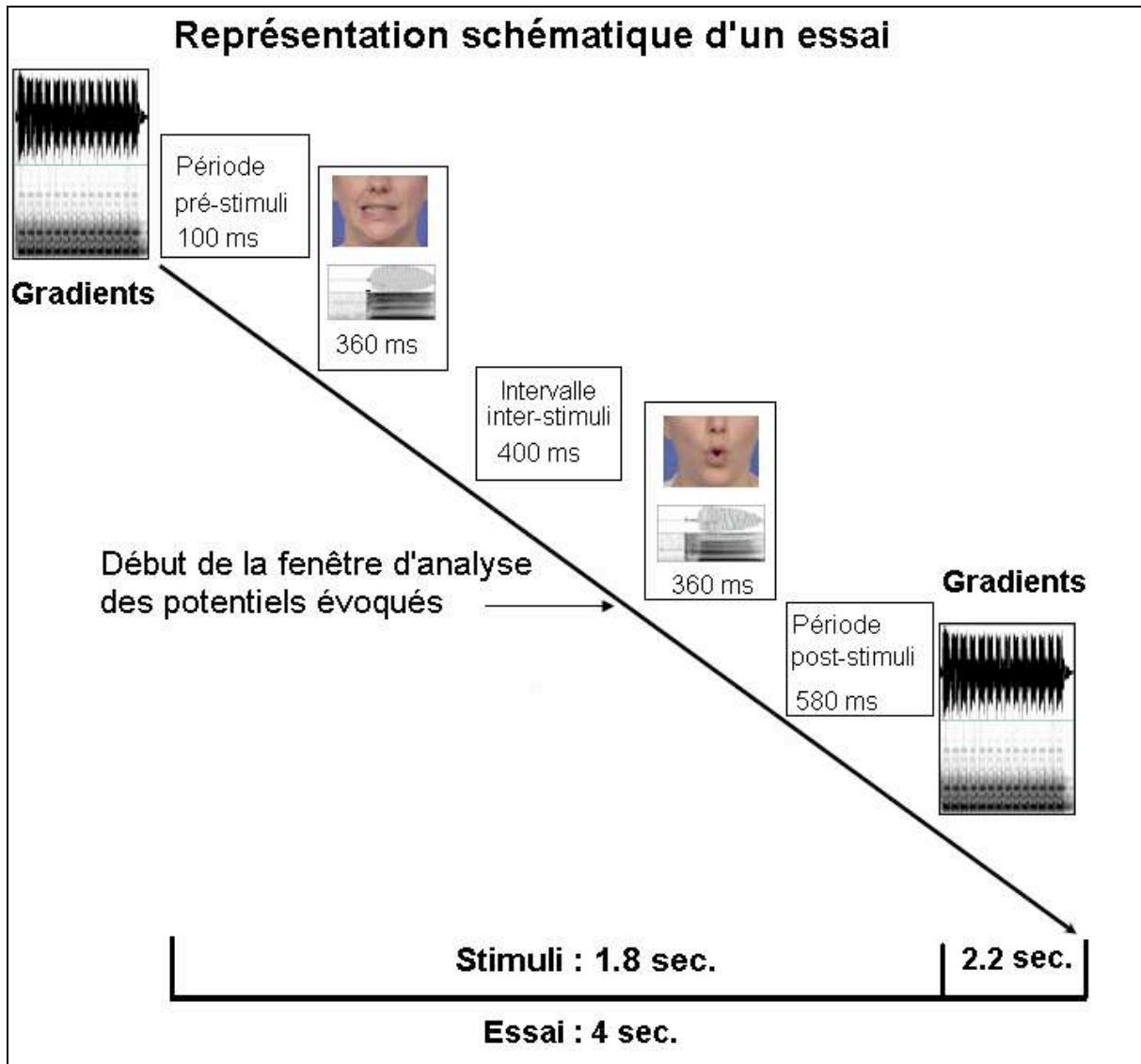


Figure B8 : Représentation schématique de la présentation d'un essai constitué d'une paire phonologique /pi py/ entre deux périodes de gradients de l'IRM.

B.1.4.2. Acquisition et prétraitement des images IRMf

Les images fonctionnelles ont été acquises dans un IRM 2 Tesla (MRI S200 Bruker, Bruker Medical GmbH, Ettlingen, Allemagne), équipé d'une bobine de gradients SK330. Les images ont été obtenues en imagerie écho-planaire *EPI* ("Echo-Planar Imaging") "single shot", qui est un mode d'acquisition rapide, avec un appareil 2 Tesla Bruker.

Le temps de répétition (« TR »), c'est-à-dire la durée d'acquisition d'un volume cérébral complet, était de 4 secondes. Par volume, 32 coupes de 4 mm d'épaisseur étaient acquises en orientation inter-commissurale AC-PC. La résolution spatiale, c'est-à-dire la taille

du voxel, était de 4 x 4 x 4 mm, pour une matrice de 64x64x32 voxels, et un temps d'écho (« TE ») de 43 millisecondes.

Les images morphologiques ont été acquises en imagerie *FSE* (“*Fast Spin Echo*”), qui permet d’avoir une bonne résolution spatiale (2x2x2 mm), avec un TR de 12 secondes, et un TE de 80 millisecondes. Elles sont composées de 80 coupes, de 2 mm d'épaisseur, résultant en une matrice de 128x128x80 voxels.

Nos images ont ensuite été traitées avec le logiciel *SPM5* (“*Statistical Parametrical Mapping*”, *Wellcome Department of Cognitive Neurology*, Londres, GB). Les opérations de prétraitements standards ont été effectuées. Cette étape est préalable à toute analyse, et est effectuée pour chacun des sujets. Les images ont d’abord été recalées dans le temps, puis dans l’espace sur la première image acquise, normalisées sur un atlas de référence (ici l’espace stéréotaxique *MNI* (“*Montreal Neurological Institute*”), et lissées enfin spatialement, avec un filtre gaussien de 6 mm de largeur.

Les coupes sont acquises séquentiellement, chacune est séparée de la suivante par un demi temps d’acquisition ; la première étape est donc celle du réalignement temporel, par rapport à une coupe de référence. Il est ensuite nécessaire de procéder à un réalignement spatial, afin de corriger les déplacements du crâne au cours de la session expérimentale. La réduction des artefacts s’opère volume par volume en prenant comme référence la moyenne de tous les volumes. Vient ensuite la normalisation, qui consiste à replacer tous les volumes acquis dans un gabarit commun, c’est-à-dire dans un atlas référentiel permettant de localiser les structures cérébrales, grâce à des coordonnées euclidiennes (x ; y ; z), et ainsi de comparer les analyses effectuées sur des sujets différents, en considérant les mêmes structures, repérées par leurs coordonnées, chez tous les sujets. Finalement, l’étape du lissage a pour but de réduire les petites différences anatomiques interindividuelles, et permet ainsi d’augmenter la cohérence spatiale (nécessaire pour l’analyse statistique suivante).

B.1.4.3. Acquisition et prétraitement du signal EEG recueilli pendant l’IRMf

Lors de la réalisation d’une tâche, des modifications de l’activité électrique engendrent sur le scalp des changements d’amplitude et de répartition topographique des potentiels évoqués. Pour enregistrer ces potentiels évoqués en présence de champs magnétiques intenses, sont utilisées des électrodes amagnétiques en or, en étain, en argent ou en chlorure d’argent, reliées à des amplificateurs différentiels à faible bruit. Les câbles blindés, reliant les

électrodes à l'amplificateur, sont fixés de telle sorte qu'ils n'entraient pas en contact avec la structure de l'IRM. L'amplificateur (Schwarzer EMR 32 Headbox, Munich Germany), placé derrière le scanner IRM est, quant à lui, relié par fibre optique à un ordinateur (PC), placé à l'extérieur de la chambre d'acquisition (cage de Faraday) réalisant les enregistrements des données EEG, grâce au logiciel Brainlab (Schwarzer, Munich Germany).

Les signaux EEG, enregistrés dans un scanner IRM, sont très perturbés par les champs magnétiques intenses et leurs commutations (gradients). Le système d'amplification dédié à l'acquisition dans l'IRM offre la possibilité d'enregistrer des signaux EEG non saturés (Hoffmann *et al.* 2000).

De plus, ces signaux sont parasités par l'activité cardiaque du sujet (voir figure B9). Par conséquent, il est nécessaire de réaliser une acquisition simultanée de l'électrocardiogramme (ECG), à l'aide d'un amplificateur (Physiogard, Bruker SARL, Wissembourg France), afin de pouvoir procéder au filtrage de cet artefact cardiaque (Otzenberger, Gounot, Marrer, Namer & Metz-Lutz, 2005). Cette étape est des plus importantes, un ECG exempt de perturbations est nécessaire pour l'exploitation subséquente. Les signaux EEG et ECG ont été enregistrés à la fréquence de 1000 Hz.

Deux voies d'acquisition supplémentaires ont été utilisées, l'une pour marquer la présentation des stimuli, l'autre afin de marquer le début et la fin des images IRMf.

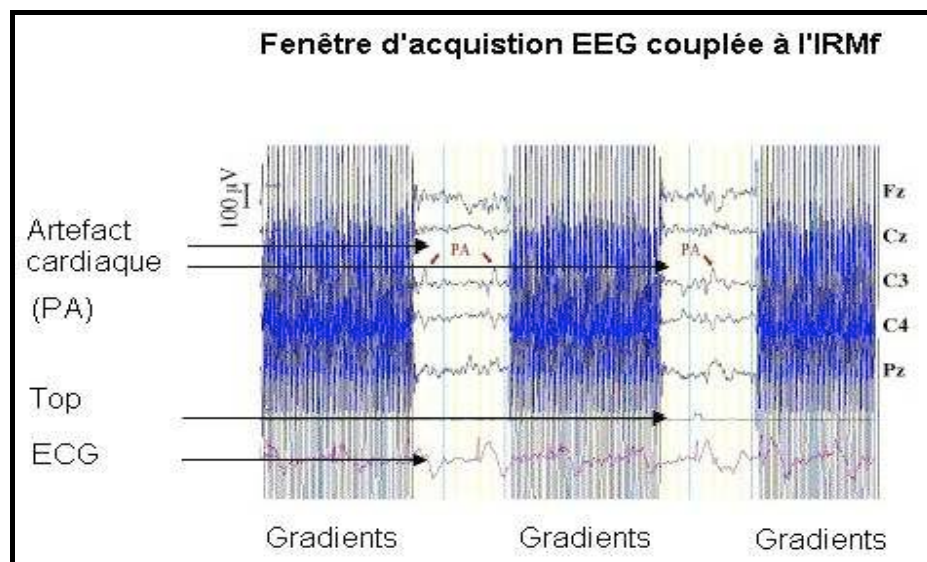


Figure B9 : Exemple pour cinq électrodes d'une fenêtre d'acquisition des signaux EEG, durant une passation IRMf, telle que proposée par le logiciel Brainlab.

Les électrodes en chlorure d'argent ont été fixées directement sur le scalp, décapé à l'éther, au moyen d'un gel susceptible d'améliorer la conduction (Elefix, Nikon Khoden). Les électrodes ont été disposées sur le crâne des sujets, par une infirmière spécialisée, selon le système « 10 - 20 ». L'infirmière commençait tout d'abord par mesurer les dimensions du crâne, afin de pouvoir adapter la disposition des électrodes à la morphologie individuelle.

Nous avons utilisé 23 électrodes en tout, une électrode dite « de terre » est placée sur le front (Nz) et trois autres sont nécessaires afin de mesurer les mouvements oculaires qui seront ensuite éliminés (voir figure B10). Parmi ces trois électrodes, deux sont placées sur les tempes, et la dernière, qui sert de référence, prend place sur la mastoïde gauche, c'est-à-dire sur la saillie de l'os temporal. Demeurent dix-neuf électrodes directement impliquées dans l'enregistrement des potentiels évoqués : trois sur le vertex Fz, Cz, Pz, et huit pour chaque hémisphère F3, F7, FC3, C3, CP5, T3, T5, P3 (à gauche) et F4, F8, FC4, C4, CP6, T4, T6, P4 (à droite). Les signaux EEG ont été acquis et visualisés grâce au logiciel Brainlab® (OSG, Rumst, Belgique) en utilisant un montage avec une référence calculée sur la moyenne des dix-neuf électrodes d'intérêt (Goldmann, G19 ; Schwarzer, Munich Germany).



Figure B10 : Photographies d'un sujet au cours et à la fin de la pose des électrodes.

L'impédance des électrodes doit nécessairement être contrôlée avant tout enregistrement ; elle est habituellement considérée comme acceptable à moins de 2 k Ω . Le signal électrique enregistré sur chaque électrode a une amplitude de quelques microvolts et il est amplifié afin de pouvoir être mesuré. Il est essentiel de s'assurer que toutes les électrodes répondent de façon satisfaisante après la pose. Une étape de vérification consiste à brancher chaque électrode sur l'amplificateur, avant que les sujets ne s'installent dans le tube de l'IRM.

Cela nous a permis, le cas échéant, de recoller une électrode, voire d'en remplacer certaines défectueuses.

Les signaux EEG ont été traités, *a posteriori*, à l'aide du logiciel Matlab 6.1 (Mathworks, INC.), utilisant un système de filtrage de l'artefact cardiaque ("*Pulse Artefact*", "*PA*"), développé spécifiquement au LINC (Otzenberger *et al.*, 2005). Les signaux ont été analysés, après filtrage des basses et hautes fréquences, dans la bande de 0.1 à 30 Hz. Un modèle de courbe de l'artefact cardiaque a été évalué pour chaque électrode chez chacun des sujets individuellement. Cette évaluation a été calculée sur une centaine d'ondes, prises en référence et enregistrées durant les premières secondes de l'acquisition juste avant que ne débute le protocole expérimental. Ce modèle d'artefact cardiaque sera ensuite « régressé » des données EEG (voir figure B11).

Méthode de filtrage de l'artefact cardiaque

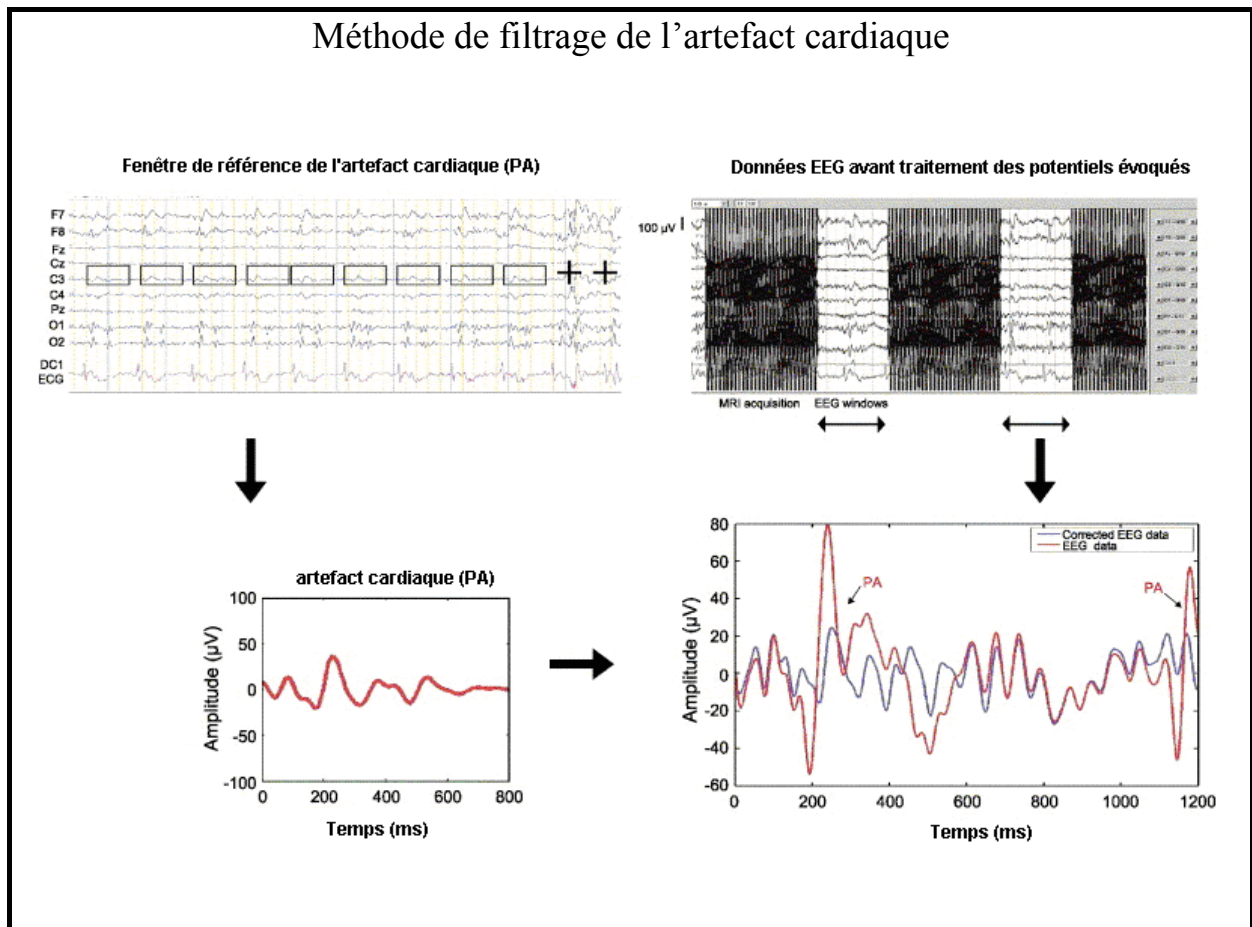


Figure B11 : Illustration de la méthode de filtrage de l'artefact cardiaque des données EEG acquises simultanément lors de la passation IRMf (Tiré et adapté de Otzenberger *et al.*, 2005).

En haut à gauche : Représentation des fenêtres de référence synchronisées sur l'ECG, et utilisées pour l'évaluation de l'artefact cardiaque.

En haut à droite : Capture de l'écran de l'ordinateur pendant l'acquisition des données EEG. Seules les fenêtres contenant des stimuli d'intérêt, présentés entre deux périodes de gradients de l'IRM, seront analysées.

En bas à gauche : Représentation d'un modèle d'artefact cardiaque évalué pour l'électrode C3 pour un sujet.

En bas à droite : Représentation des données d'une fenêtre entre deux périodes de gradients avant (en rouge) et après (en bleu) filtrage de l'artefact cardiaque.

B.1.4.4. Analyses des données IRMf et EEG

Afin de faire apparaître les aires cérébrales significativement activées par le paradigme, il est nécessaire d'opérer un traitement statistique sur tous les voxels du cerveau. Dans le même esprit, les données EEG nécessitent un traitement pour extraire les potentiels évoqués.

B.1.4.4.1. Analyses des images IRMf

Analyse individuelle

Dans une analyse individuelle (dite de premier niveau), une image de contraste, correspondant à la différence d'activation entre le signal *BOLD* induit par chaque tâche d'intérêt et celui induit par la condition de repos « événement nul », a été calculée pour chaque individu.

La première étape consiste à fabriquer un modèle prévisionnel des fluctuations hémodynamiques de la réponse *BOLD*, provoquées par les différents stimuli. Grâce au logiciel SPM5 fonctionnant sous l'environnement Matlab 7.2 (Mathworks, INC.), les signaux modèles induits par chaque type de tâche ont pu être calculés.

SPM5 détermine ensuite quelles sont les fluctuations du signal *BOLD*, dans chaque voxel, qui sont compatibles avec ledit modèle, et leurs probabilités. Une régression linéaire permet d'extraire la portion de signal correspondant au modèle pour chacun des stimuli dans chaque voxel. On opère une comparaison grâce à un test de *Student (T test)* entre l'amplitude des signaux et la portion de signal non expliquée par le modèle. En effet le reste de signal, non expliqué, correspond à la variance résiduelle ou « bruit ». Des cartes statistiques des probabilités d'activation, pour chacun des régresseurs, dans chaque voxel, sont calculées à l'aide du test T de Student, en comparant l'amplitude du signal au bruit.

Analyse de groupe

L'analyse de groupe en effet aléatoire permet, avec un seuil de probabilité statistique, de généraliser à l'ensemble de la population les résultats obtenus sur le groupe.

Il est nécessaire de réaliser des statistiques entre tous les sujets ayant participé à la session expérimentale, de façon à calculer des probabilités d'activations reproductibles dans l'ensemble d'une population. Pour chaque stimulus, on compare les cartes individuelles établies comme expliqué précédemment, voxel par voxel. La répartition statistique des activations des différents sujets est analysée, de façon à déterminer si une activation commune existe, ainsi que sa probabilité. Une carte de probabilité d'activation (spmT) est obtenue, permettant de visualiser les zones cérébrales qui sont significativement activées par les stimuli, ainsi que leur probabilité d'existence. Nous avons, ensuite, utilisé le logiciel

MRICron (www.mricro.com), afin de localiser les pics d'activation dans la cartographie de Brodmann.

Ce type d'analyse permet de faire, à partir du groupe de sujets, des inférences inductives. Lors de l'établissement de ces cartes de comparaison, il nous est offert la possibilité de déterminer deux paramètres cruciaux : la taille minimale à partir de laquelle nous considérons un groupe de voxels ("*cluster*") ; et le seuil statistique (« niveau p »).

Nous avons toujours choisi la taille de 25 voxels pour les *clusters*. En revanche, les comparaisons entre les paires phonologiques et la condition de référence ont un seuil statistique plus élevé (*FWE*, "*Family Wise Error*", $p < 0.01$) que les comparaisons entre les paires phonologiques et non phonologiques (non corrigé $p < 0.001$). Il en est de même pour les comparaisons sur deux échantillons (entre nos deux modalités), qui ont aussi un seuil non corrigé ($p < 0.001$).

B.1.4.4.2. Analyse en régions d'intérêt

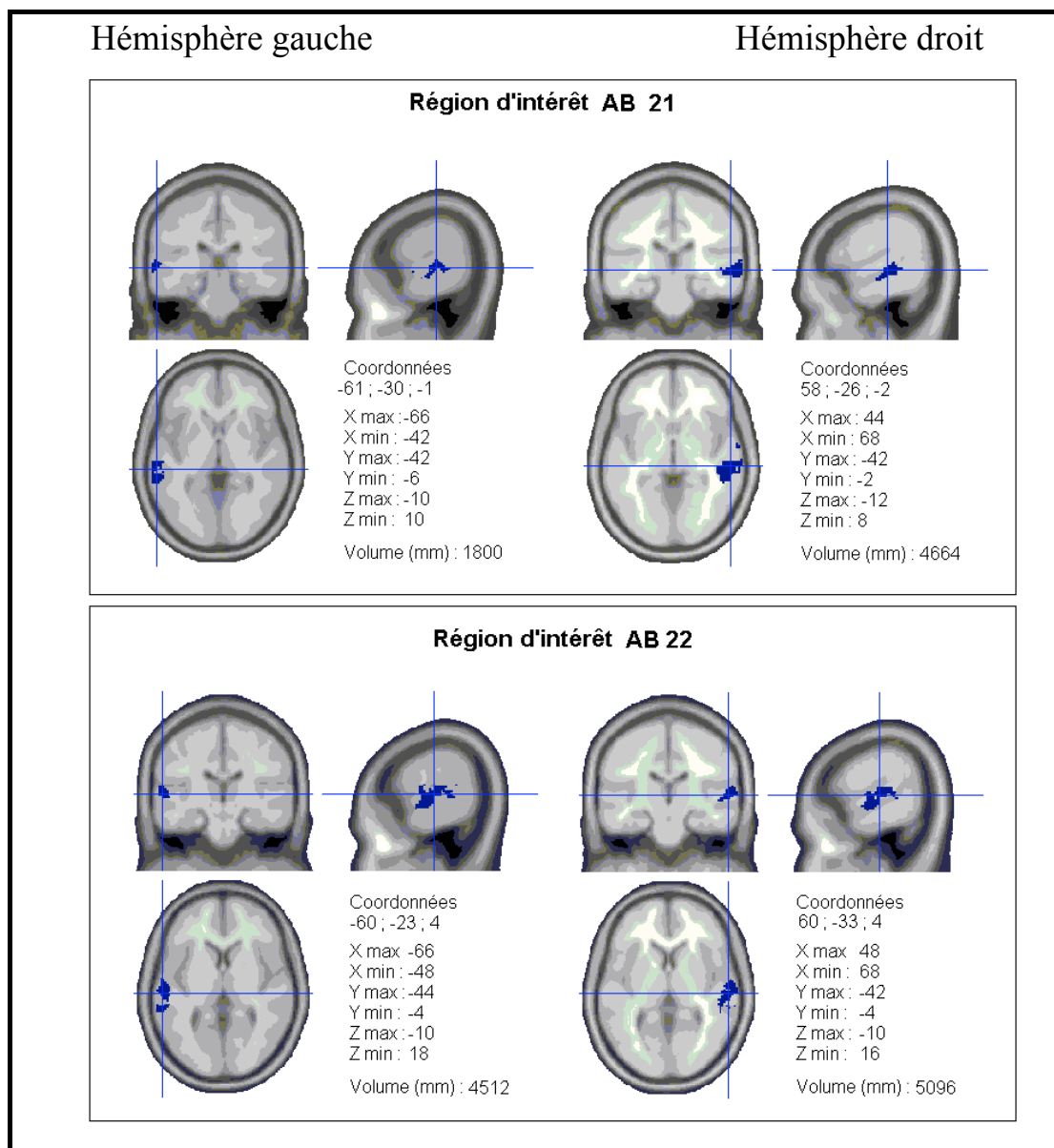
Afin de pouvoir compléter nos observations, nous avons défini des régions d'intérêt, grâce à la boîte à outils Marsbar de SPM. Les analyses en régions d'intérêt permettent d'extraire du signal *BOLD*, sur une portion de volume cérébral choisie, même si celui-ci est absent des cartes statistiques. Habituellement, les régions d'intérêt sont utilisées lorsqu'aucune activation n'est observée. Au contraire, si nous avons réalisé une analyse en régions d'intérêt, c'est afin de pouvoir effectuer des statistiques entre les pics d'activations des paires syllabiques et des stimuli non phonologiques.

Afin de restreindre l'étendue des régions d'intérêt aux zones impliquées dans nos tâches, nous avons recherché l'ensemble des zones d'activation calculées pour les trois types de paires syllabiques, en modalité AV animée. Nous avons ensuite calculé l'intersection de cette zone d'activation avec quatre aires de Brodmann dans chaque hémisphère. Par souci de simplicité, nous ferons référence aux régions d'intérêt des aires de Brodmann 21, 22, 41, 42, mais il ne s'agit que des portions de ces aires qui sont significativement activées, à l'issue des « tests T » de l'analyse de groupe (voir figure B12).

Au niveau du Gyrus Temporal Supérieur, les coordonnées de la « région d'intérêt 22 » sont pour l'hémisphère gauche : $x = -60$; $y = -23$; $z = 4$ (étendue de 4512 voxels), et pour l'hémisphère droit : $x = 60$; $y = -33$; $z = 4$ (étendue de 5096 voxels).

Au niveau du Gyrus Temporal Moyen, les coordonnées de la « région d'intérêt 21 » sont pour l'hémisphère gauche : $x = -61$; $y = -30$; $z = -1$ (étendue de 1800 voxels), et pour l'hémisphère droit : $x = 58$; $y = -26$; $z = -2$ (étendue de 4664 voxels).

Au niveau du cortex auditif primaire, les coordonnées centrales de la « région d'intérêt 41 » sont pour l'hémisphère gauche : $x = -43$; $y = -36$; $z = 13$ (étendue de 1680 voxels), et pour l'hémisphère droit : $x = 44$; $y = -34$; $z = 12$ (étendue de 1136 voxels) ; alors que les coordonnées de la « région d'intérêt 42 » sont pour l'hémisphère gauche : $x = -57$; $y = -34$; $z = 15$ (étendue de 1800 voxels), et pour l'hémisphère droit : $x = 58$; $y = -33$; $z = 13$ (étendue de 656 voxels).



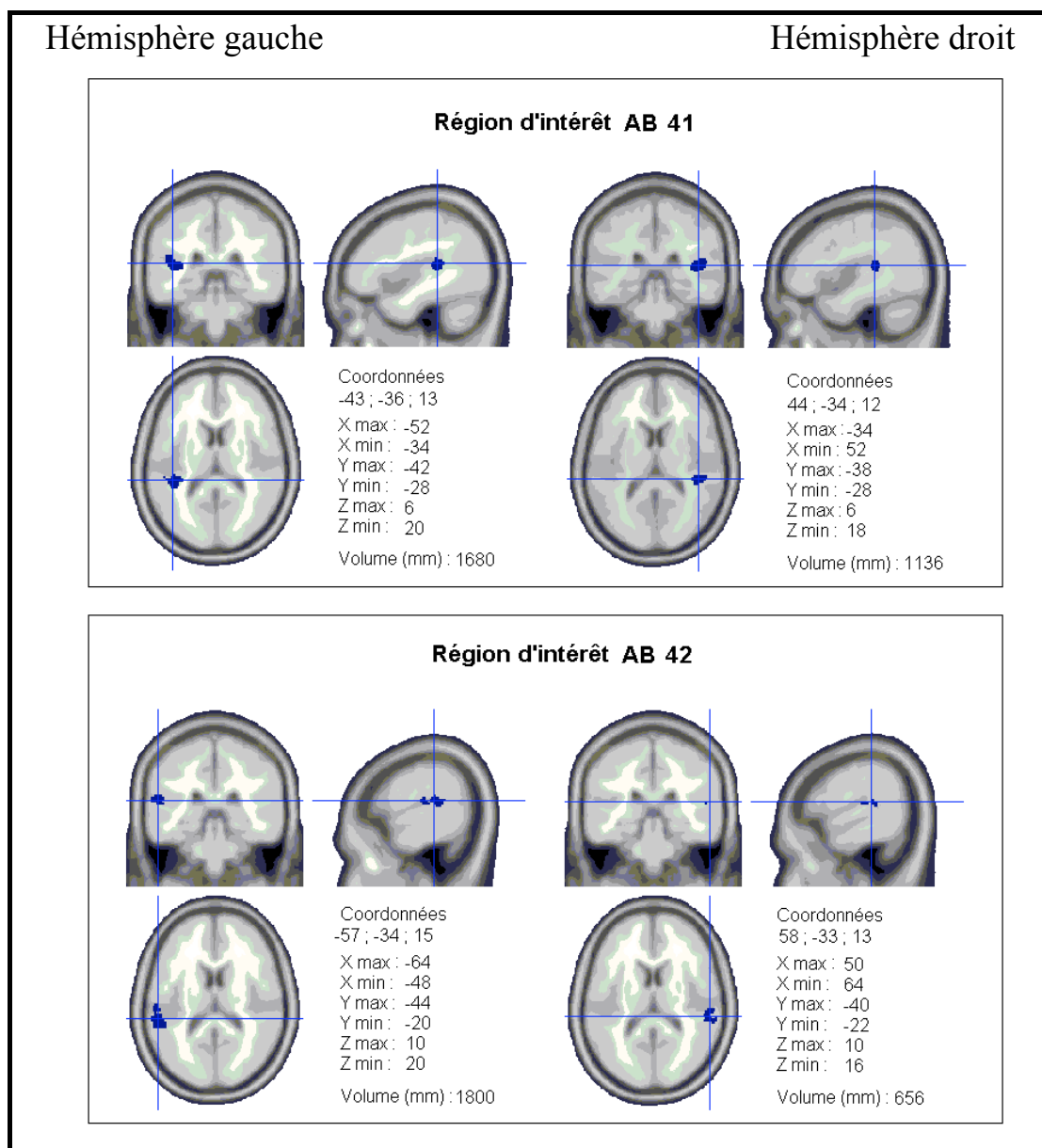


Figure B12 : Les huit régions d'intérêt accompagnées de leurs coordonnées (centre et étendue) et de la taille du volume en mm.

Dans chacune des huit régions d'intérêt, la valeur maximale d'activation, différence entre le signal *BOLD* entraîné par les tâches d'intérêt et la condition « nulle », a été calculée pour chaque sujet, et pour chaque région d'intérêt, à l'aide de la boîte à outils Marsbar. Nous avons ensuite pu effectuer, sur ces données, des analyses de variance (ANOVA : “*Analysis Of Variance*”) sur le groupe de 26 sujets.

B.1.4.4.3. Évaluation des potentiels évoqués

La période de temps de calcul des variations de potentiels électriques se situe entre – 100 ms et 800 ms ; le temps « 0 » correspond au début du second stimulus de la tâche de discrimination AX. Pour l'analyse des potentiels évoqués, nous avons éliminé les 200 dernières millisecondes (entre 600 ms et 800 ms), en raison de la présence des gradients de l'IRM pouvant perturber le recueil de l'activité électrique. Les potentiels évoqués ont été analysés sur 120 paires, en dehors des gradients de l'IRM (trois catégories de paires phonologiques répétées quarante fois, de manière pseudo-aléatoire).

Seules les réponses justes des sujets ont été prises en compte pour le traitement des potentiels évoqués. Parmi les 29 ayant participé à la passation IRMf/EEG, 16 ont obtenu des scores de discrimination supérieurs au hasard (> à 65 %) pour toutes les catégories de paires. Une analyse visuelle a permis d'éliminer les fenêtres perturbées par les mouvements oculaires. Huit sujets n'ont pu être pris en compte du fait de la présence de mouvements oculaires trop nombreux. Ainsi, 11 sujets sur les 26 ont été traités pour l'analyse des potentiels évoqués.

Les courbes individuelles de chaque sujet, pour toutes les électrodes, ont été analysées pour les trois catégories de paires syllabiques, pour nos deux modalités, AV fixe et AV animée (36 paires \pm 4). La technique du moyennage permet d'obtenir les différences de potentiel, reflétant indirectement les processus mentaux spécifiques des événements suscités. L'activité parasite indépendante de ces événements est éliminée par ce moyennage. Les courbes présentant les potentiels évoqués sont nommées « grande moyenne » ; elles correspondent à la moyenne des courbes individuelles de chacun des sujets.

B.1.4.4.4. Analyse des potentiels évoqués

Le logiciel ELAN-ERP (version 6.130), développé à Lyon (INSERM U280, France), nous a permis de traiter nos données de potentiels évoqués.

La première méthode d'analyse des potentiels évoqués est une analyse visuelle de la forme de la courbe représentant la grande moyenne. L'enregistrement de l'activité électrique du cerveau en présence de champs magnétiques intenses a une influence non négligeable sur les résultats. Nous avons eu des difficultés considérables pour analyser les potentiels évoqués, particulièrement en modalité AV fixe. Les moyennes obtenues présentent des écarts-types

conséquents. De ce fait, nous avons été contraint de procéder à des ajustements, afin d'être en mesure d'utiliser des calculs statistiques.

Le logiciel Matlab 6.1 (Mathworks, INC.) a été utilisé afin d'extraire la surface des amplitudes (aire entre la courbe et la ligne de base exprimée en $\text{Hz} * \mu\text{V}$), à l'intérieur de deux fenêtres temporelles. Une première analyse comparant les surfaces des ondes évoquées par la discrimination en modalités AV animée et AV fixe, par pas de 50 ms, réalisée sur l'électrode Fz toutes paires syllabiques confondues, nous a permis d'identifier deux fenêtres temporelles, dans lesquelles les réponses enregistrées dans les deux modalités étaient significativement différentes.

La première fenêtre s'étend de 110 ms à 220 ms, et la seconde de 220 ms à 300 ms, à partir du temps « 0 » qui correspond au début de la présentation du second stimulus (voir figure B13). Les différentes analyses de variance ont été réalisées sur ces fenêtres temporelles, afin de comparer les réponses évoquées, en fonction des deux modalités de présentation (AVf et AVa), pour les trois catégories phonologiques. Nous avons choisi d'observer plus particulièrement sept électrodes (Fz, CP5, CP6, T3, T4, T5, T6) sur les 19 dont nous disposions.

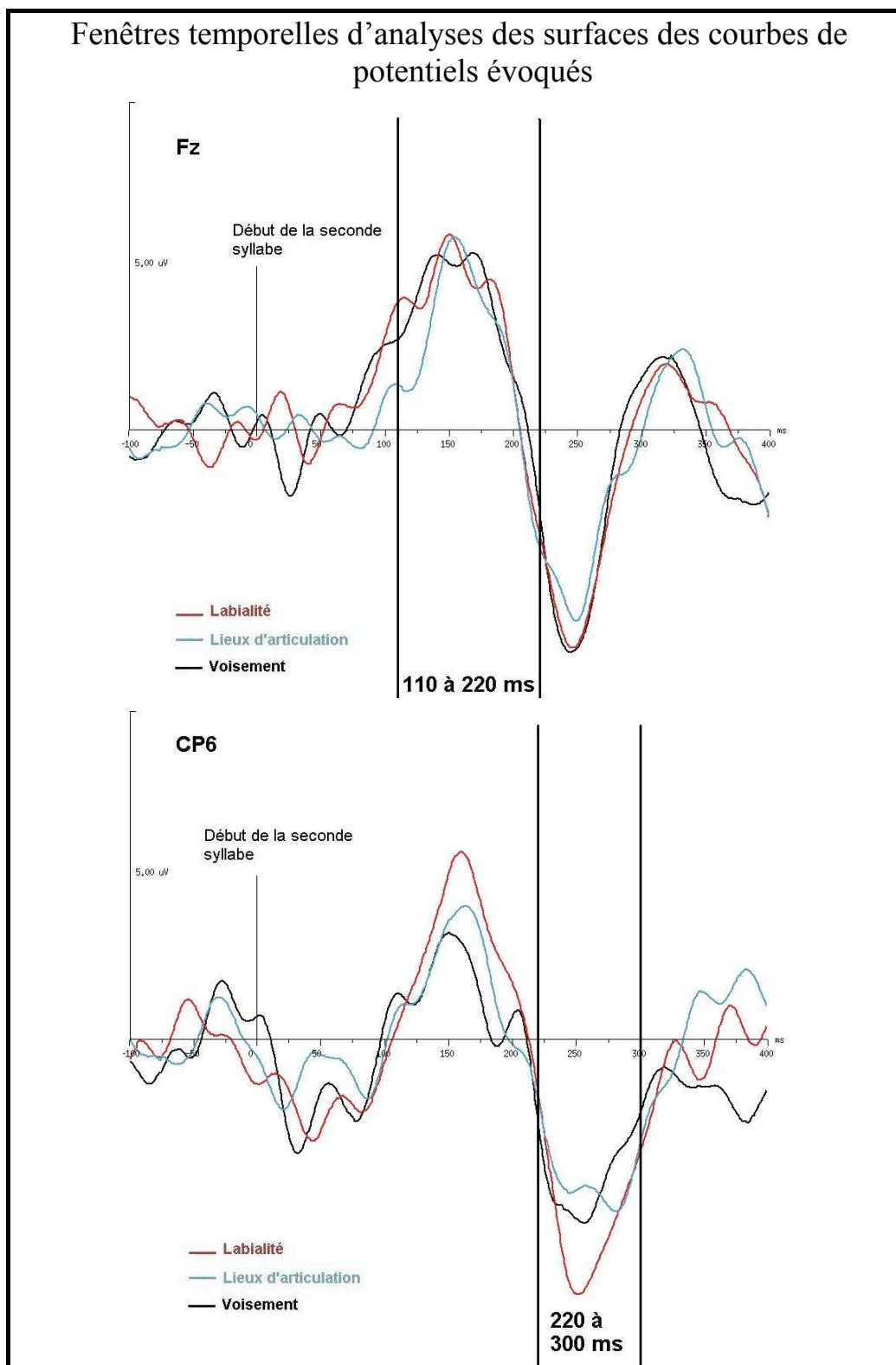


Figure B13 : Fenêtres d'analyse de surface de l'aire à partir des courbes des réponses électriques en AVa.. En haut : exemple pour l'électrode Fz de la fenêtre allant de 110 ms à 220 ms ; en bas : exemple pour l'électrode CP6 de la fenêtre allant de 220 ms à 300 ms.

B.2. Résultats : Données comportementales

L'analyse des données comportementales¹ obtenues pendant la réalisation de la tâche de discrimination, mise en place pour l'étude anatomo-fonctionnelle, nous offre la possibilité de vérifier l'impact de la présence des indices visuels pertinents sur la perception syllabique. Les deux modalités utilisées s'opposent entre elles par l'apport d'indices visuels, mais la source auditive est présente dans les deux cas. Nous débutons la présentation des résultats comportementaux par deux tableaux présentant les scores de discrimination et les temps de réponse moyens, pour tous les participants. Cinquante-cinq personnes ont pris part aux tests dans l'environnement silencieux, 16 dans l'environnement bruyant et 26 dans l'environnement perturbé, c'est-à-dire lors de l'acquisition en IRMf/EEG.

Tableau B4 : Scores de discrimination et écarts-types en pourcentage, pour les quatre types de paires, en fonction de la modalité dans les trois environnements.

Paires	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	AVf	AVa	AVf	AVa	AVf	AVa	AVf	AVa
Modalités								
Environ ^{mts}								
Silencieux (n = 55)	99,6 1.1	99,4 1.3	99,6 1.1	99,5 1.2	99,5 1.5	98,8 2.3	94 5	97 3.3
Perturbé (n = 26)	71,7 17	98,1 2.5	96,2 4.1	98,5 3.1	96,9 4.1	97,8 4.2	90,7 10.2	95,7 6.9
Bruyant (n = 16)	87,2 15.3	99,1 2.7	98,8 2.7	97,8 4.3	98,3 3.7	98 4.1	91,1 5.7	95,6 4

¹ Nous prions le lecteur de bien vouloir nous excuser pour la présentation des résultats, qui peut s'avérer pour le moins fastidieuse. Nous pensons, toutefois, que nous ne devons pas faire l'économie d'un exposé systématique de telles données.

Tableau B5 : Temps de réponse et écarts-types en millisecondes, pour les quatre types de paires, en fonction de la modalité dans les trois environnements.

Paires	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	AVf	AVa	AVf	AVa	AVf	AVa	AVf	AVa
Modalités								
Environ ^{mts}								
Silencieux	863 180	885 178	878 188	915 188	880 182	943 175	830 169	897 163
Perturbé	812 125	760 143	782 162	752 127	753 127	782 143	726 131	730 167
Bruyant	825 159	745 126	786 155	754 130	812 148	814 158	775 142	747 123

Des analyses de variance ont été effectuées, à l'aide du logiciel Statistica®, sur les moyennes individuelles obtenues dans les trois environnements. Les ANOVA ont pour but de révéler des différences significatives entre les moyennes pour les paires s'opposant sur la labialité, le voisement et le lieu d'articulation ainsi que pour les paires non phonologiques, en fonction des environnements expérimentaux et de la modalité de présentation. Les ANOVA mettent en évidence les différences significatives entre les groupes, mais ne stipulent pas ce sur quoi portent ces différences. Par conséquent, il faut pratiquer des tests *a posteriori* ou *post-hoc* qui peuvent nous indiquer entre quels groupes se situent la ou les différences significatives. Nous avons utilisé des tests *post-hoc* LSD ("*Least Significant Difference*").

Les catégories contenaient quarante items. Le seuil de hasard a été déterminé à 65 %. Tous les sujets ayant des résultats supérieurs ou égaux à 27 bonnes réponses, c'est-à-dire 67,5 %, ont été inclus dans les analyses statistiques. Notons qu'une réponse équivaut à 2,5 %.

Nous présentons tout d'abord les résultats des ANOVA à deux facteurs dans chaque environnement. Le premier facteur concerne les modalités et a, par conséquent, deux niveaux : AV fixe et AV animée. Le second facteur a trait aux différentes paires et contient quatre niveaux nommés en fonction du trait distinctif : labialité, lieux d'articulation et voisement avec de surcroît une catégorie de paires non phonologiques qui appelait la réponse « différente ».

B.2.1. Les performances dans l'environnement silencieux

Que ce soit en modalité AV fixe ou AV animée, les tâches expérimentales qui ont été mises en place sont d'une grande simplicité, en l'absence de perturbation, pour un adulte normo entendant. Les scores recueillis lors de l'acquisition en milieu silencieux sont tous excellents. Il n'y a aucun résultat inférieur au seuil de hasard dans le milieu silencieux. Les ANOVA ont donc été effectuées sur les scores de discrimination et les temps de réponse des 55 participants.

B.2.1.1. Les modalités dans l'environnement silencieux

Toutes paires confondues, il n'y a pas d'effet principal de la modalité pour les scores de discrimination, alors que les temps de réponse montrent un effet significatif ($F(1,54) = 5.6$ $p < 0,02$), reposant sur un avantage en AV fixe (862 ms), par rapport à l'AV animée (910 ms).

Tableau B6 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa (toutes paires confondues), pour les 55 participants, dans l'environnement silencieux.

Modalités	AV fixe		AV animée	
	M	E-T	M	E-T
Score en %	98,2	3,7	98,7	2,4
Temps de réponse en ms	862	179	910	177

B.2.1.2. Les types de paires dans l'environnement silencieux

Nos deux modalités confondues (voir figure B14), il y a un effet principal des paires pour les scores de discrimination, reposant sur un avantage des paires syllabiques, par rapport aux paires non phonologiques (95,5 % ; $F(3,362) = 86,26$ $p < 0.000001$). Les temps de réponses montrent un effet principal reposant sur une différence des paires non phonologiques et des paires s'opposant sur l'indice de labialité (non phonologique : 863 ms = labialité : 874 ms) par rapport aux deux autres paires syllabiques, qui sont à leur tour différentes l'une de l'autre, avec un avantage pour les paires s'opposant sur les lieux d'articulation (896 ms), par rapport à celle s'opposant sur le voisement (911 ms ; $F(3,362) = 18,87$ $p < 0.0000001$).

Tableau B7 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 55 participants, dans l'environnement silencieux.

Paires	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	99,5	2	99,6	1,1	99,2	1,2	95,5	4,58
Temps de réponse en ms	874	178	896	188	911	180	863	169

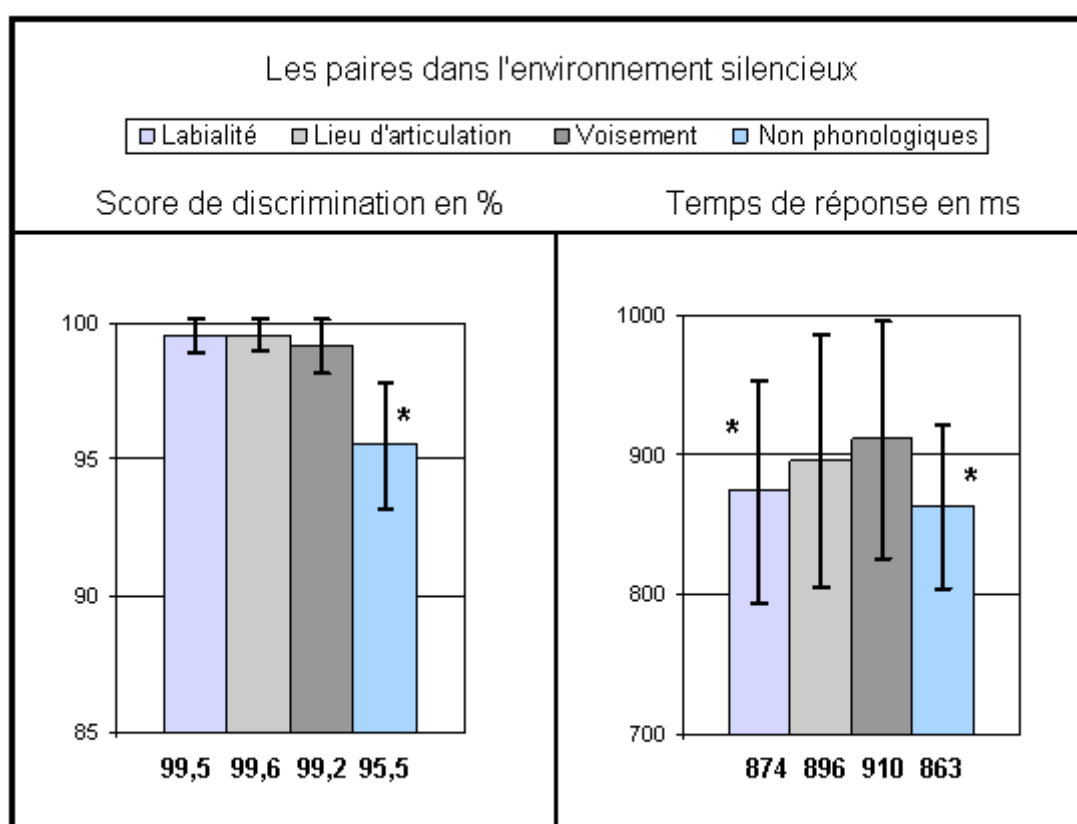


Figure B14 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, (les deux modalités confondues), pour les 55 participants, dans l'environnement silencieux ; les barres d'erreur représentent les écarts-types et les astérisques indiquent les différences significatives.

B.2.1.3. Les types de paires en fonction des modalités dans l'environnement silencieux

Les interactions entre les modalités et les paires sont significatives à la fois pour les scores de discrimination ($F(3,362) = 11,08$ $p < 0,00001$), et pour les temps de réponse ($F(3,362) = 7,84$ $p < 0,0001$).

Pour les scores de discrimination (voir figure B15), ce sont les paires non phonologiques en AV fixe (94,1 %) et en AV animée (97 %) qui sont significativement moins bien discriminées que toutes les autres. Pour les temps de réponse, les résultats sont moins tranchés. Les paires non phonologiques en AV fixe (829 ms) sont plus rapidement discriminées que toutes les autres. Les deux paires discriminées le plus lentement sont celles s'opposant sur les lieux d'articulation (915 ms) et le voisement, en AV animée (943 ms).

Tableau B8 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires dans les deux modalités, pour les 55 participants, dans l'environnement silencieux.

AV fixe	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	99,6	1,1	99,6	1,1	99,5	1,5	94,1	5
Temps de réponse en ms	863	179	878	188	879	182	829	169
AV animé	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	99,4	1,3	99,5	1,2	98,9	2,3	97	3,3
Temps de réponse en ms	884	178	915	188	943	175	897	163

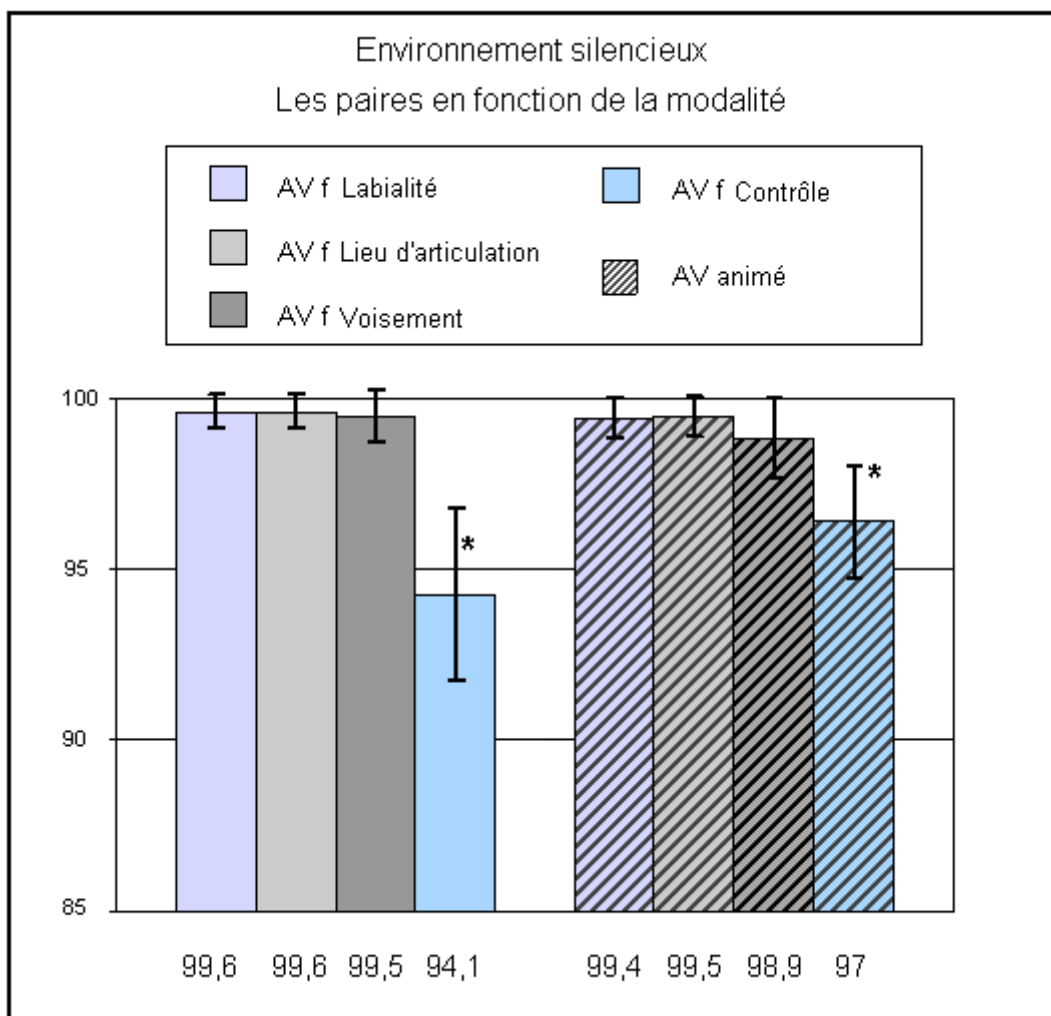


Figure B15 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 55 participants, dans l'environnement silencieux.

B.2.2. Les performances dans l'environnement perturbé (IRM)

Dans le scanner, milieu très perturbé, les stimuli s'opposant sur le trait de labialité ont été difficilement discriminés par 11 des 26 sujets ayant pris part à l'acquisition en IRMf/EEG. Les analyses portent par conséquent sur 16 sujets.

B.2.2.1. Les modalités dans l'environnement perturbé

Dans l'environnement le plus défavorable, la modalité AV animée (96,9 %) montre un effet significatif par rapport à la modalité AV fixe (92,1 %), toutes paires confondues ($F(1,15) = 11,56$ $p < 0,003$). Il n'y a pas d'effet de la modalité pour les temps de réponse.

Tableau B9 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa (toutes paires confondues), pour les 16 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement perturbé.

Modalités	AV fixe		AV animée	
	M	E-T	M	E-T
Score en %	92,1	8,9	96,9	5,4
Temps de réponse en ms	786	156	768	176

B.2.2.2. Les types de paires dans l'environnement perturbé

Pour les deux modalités confondues (voir figure B16), un effet principal est relevé ($F(3,45) = 9,74$ $p < 10^{-4}$) qui repose, d'une part, sur les paires s'opposant sur la labialité (90,3 %) qui sont strictement inférieures à toutes les autres et, d'autre part, sur les paires non phonologiques (93,7 %) qui sont inférieures aux paires s'opposant sur les lieux d'articulation (96,8 %) et sur le voisement (97,2 %). Les temps de réponse montrent aussi un effet significatif ($F(3,45) = 9,12$ $p < 10^{-4}$) ; les paires non phonologiques (740 ms) sont plus rapidement discriminées que les trois autres (Voisement : 782 ms ; Lieu d'articulation : 785 ms ; Labialité : 799 ms).

Tableau B10 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 16 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement perturbé.

Paires	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	90,3	11,2	96,8	4	97,2	4,3	93,7	7
Temps de réponse en ms	799	163	785	172	782	154	740	176

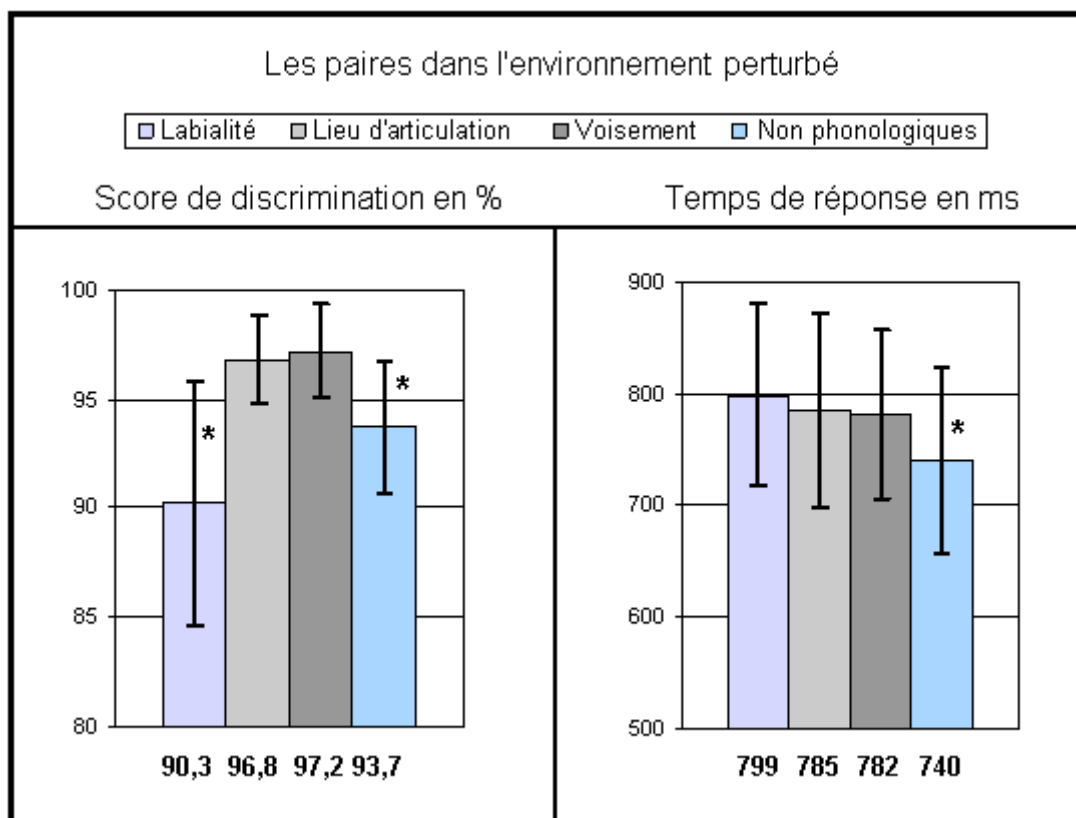


Figure B16 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 16 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement perturbé.

B.2.2.3. Les types de paires en fonction des modalités dans l'environnement perturbé

Une interaction significative est observée entre nos deux modalités et les différentes paires pour les scores de discrimination ($F(3,45) = 11,30$ $p < 10^{-4}$) et pour les temps de réponse ($F(3,45) = 3.05$ $p < 0.04$). Les paires syllabiques discriminées le moins efficacement (voir figure B17) sont celles mettant en jeu la labialité en AV fixe (82,8 %) ; c'est aussi l'indice le plus lentement identifié (827 ms). Les stimuli non phonologiques en AV fixe (92,5 %) et en AV animée (95 %) sont significativement moins bien discriminées que les autres, hormis les paires s'opposant sur le lieu d'articulation en AV fixe (95,8 %).

Tableau B11 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires en fonction des deux modalités, pour les 16 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement perturbé.

AV fixe	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	82,8	11,3	95,8	4,1	97,5	2,2	92,5	5,3
Temps de réponse en ms	827	147	805	191	777	137	734	144
AV animé	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	97,7	2,9	97,8	3,9	96,9	5,3	95	8,4
Temps de réponse en ms	772	178	766	154	788	174	746	208

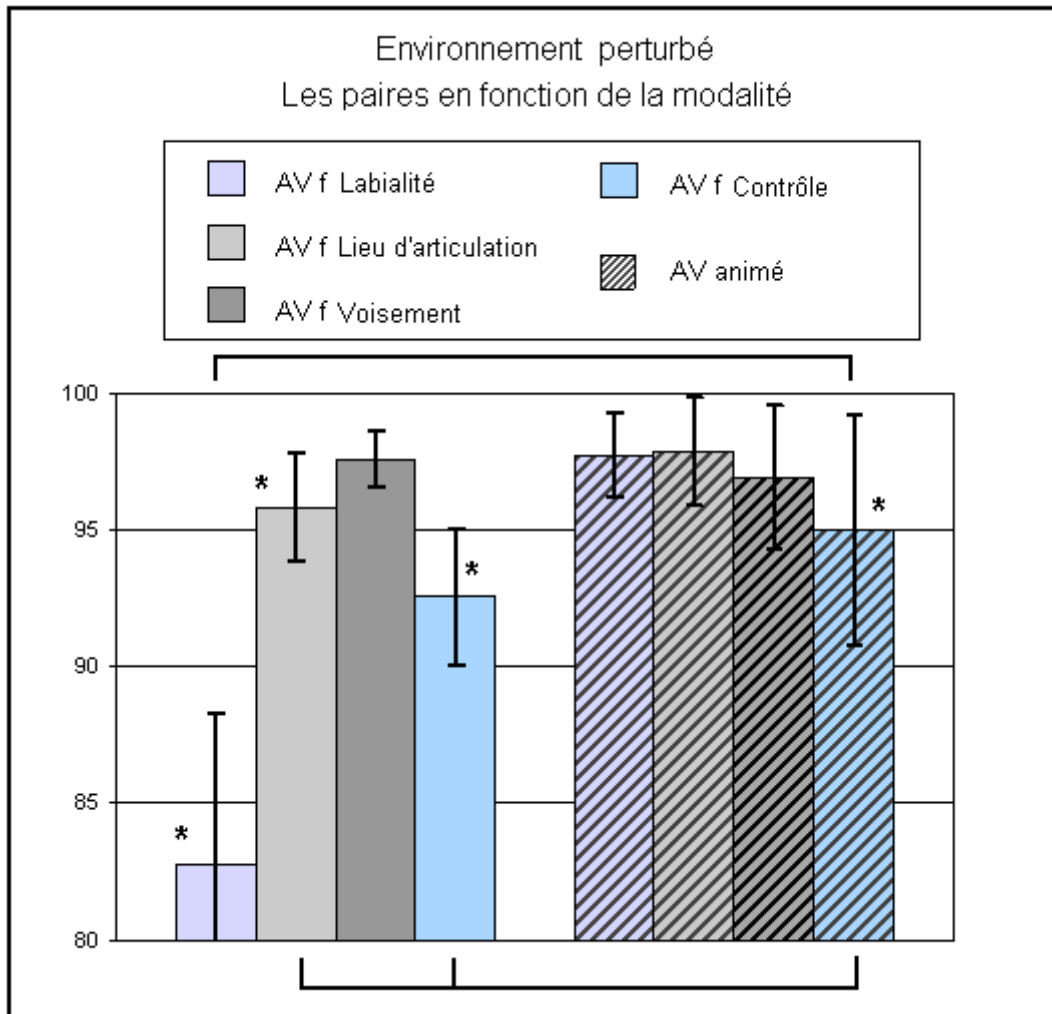


Figure B17 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 16 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement perturbé ; les barres d'erreur représentent les écarts-types.

B.2.3. Les performances dans l'environnement bruyant

Dans l'environnement bruyant, c'est-à-dire avec comme fond sonore les gradients de l'IRM, seuls 13 participants ont obtenu des scores supérieurs au hasard.

B.2.3.1. Les modalités dans l'environnement bruyant

Tous types de paires confondus, nos deux modalités ne montrent pas d'effet significatif, ni pour les scores de discrimination, ni pour les temps de réponse.

Tableau B12 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa (toutes paires confondues), pour les 13 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement bruyant.

Paires différentes	AV fixe		AV animé	
	M	E-T	M	E-T
Score en %	96,2	6	97,2	4
Temps de réponse en ms	773	145	754	128

B.2.3.2. Les paires syllabiques dans l'environnement bruyant

Les deux modalités confondues, les scores de discrimination montrent un effet significatif ($F(3,36) = 9,43$ $p < 10^{-4}$), reposant sur une différence entre les paires non phonologiques (94 %), par rapport aux trois paires syllabiques (Labialité : 96,6 % = Lieu d'articulation : 98,6 % = Voisement : 98,8 %). Les paires s'opposant sur la labialité ne sont pas différentes des paires s'opposant sur le lieu d'articulation, mais sont significativement moins discriminées que les paires s'opposant sur le voisement (voir figure B18). Les temps de réponse montrent un effet significatif ($F(3,36) = 5,19$ $p < 0,005$), induit par les paires s'opposant sur la labialité (788 ms), qui sont plus lentement discriminées que les trois autres (Non phonologique : 750 ms = Voisement : 756 ms = Lieu d'articulation : 763 ms).

Tableau B13 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 13 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement bruyant.

Paires différentes	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	96,6	7,8	98,6	1,6	98,8	1,1	94	5,3
Temps de réponse en ms	788	140	763	135	756	141	750	135

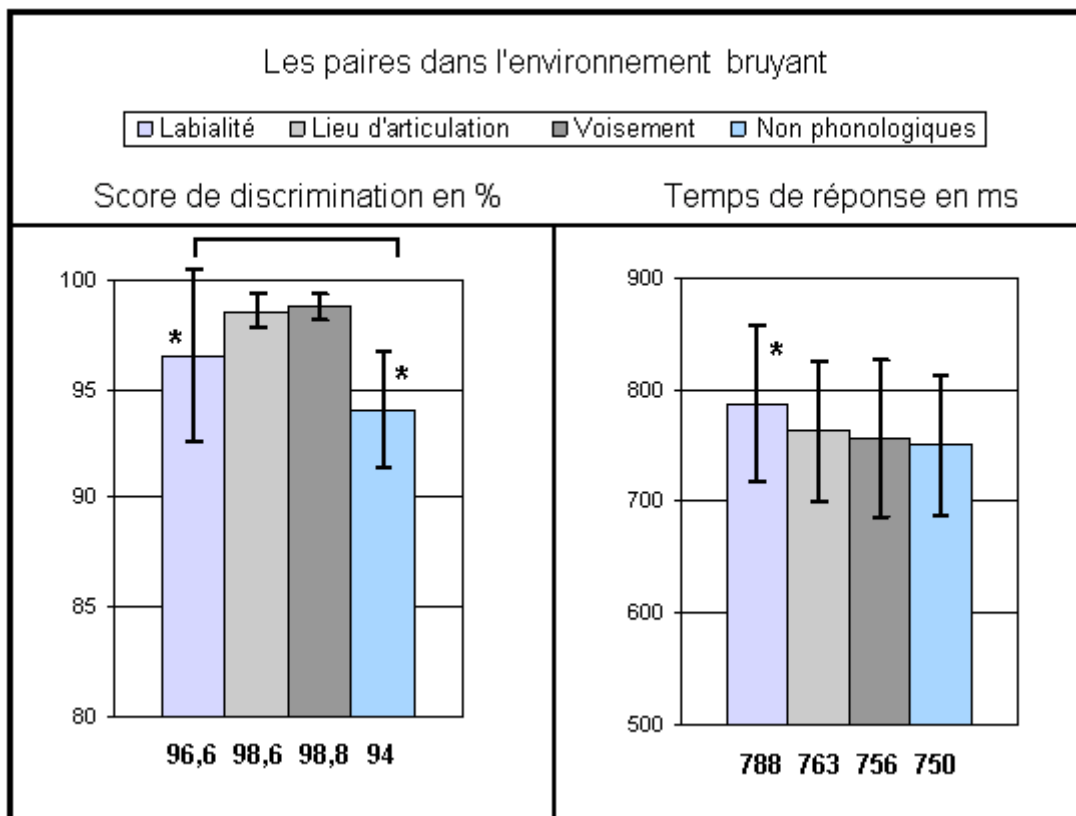


Figure B18 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 13 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement bruyant ; les barres d'erreur représentent les écarts-types.

B.2.3.3. Les paires syllabiques selon la modalité dans l'environnement bruyant

Pour les scores de discrimination (voir figure B19), un effet significatif de l'interaction entre modalités et paires est relevé ($F(3,36) = 5,16$ $p < 0,005$). Il repose sur une différence significative entre les paires non phonologiques en AV fixe (92,5 %), significativement moins bien discriminées que toutes les autres, à l'exception des paires s'opposant sur la labialité en AV fixe (93,5 %) et des paires non phonologiques en AV animée (95,6 %). La labialité en AV fixe est à son tour différente de toutes les autres paires, hormis les paires non phonologiques en AV animée. Notons que les paires s'opposant sur les lieux d'articulation et sur le voisement ne montrent pas de différences significatives entre les deux modalités. En ce qui concerne les temps de réponse, il n'y a pas d'interaction.

Tableau B14 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires en fonction des deux modalités, pour les 13 participants ayant obtenu un score supérieur au hasard (65 %), dans l'environnement bruyant.

AV fixe	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	93,5	7,8	99,3	1,6	99,5	1,1	92,5	5,3
Temps de réponse en ms	790	147	762	154	785	143	759	152
AV animé	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	99,6	1,4	97,9	4,5	98,1	4,3	95,6	4
Temps de réponse en ms	735	122	752	132	792	143	741	121

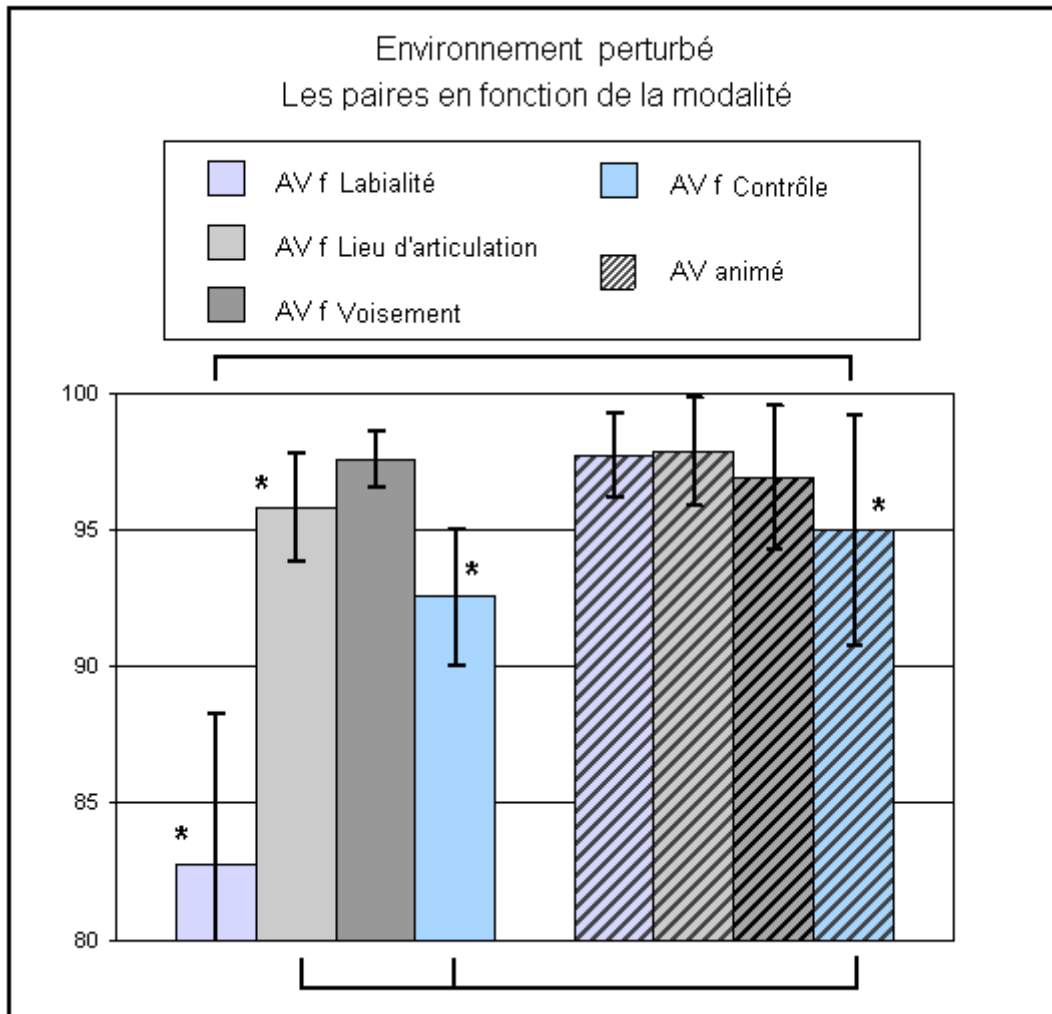


Figure B19 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 13 participants ayant obtenu un score supérieur au hasard (65 %) dans l'environnement bruyant ; les barres d'erreur représentent les écarts-types.

B.2.4. Comparaisons binaires entre les trois environnements

Pour ces comparaisons, nous avons utilisé des ANOVA à trois facteurs. Le premier facteur est celui observant les modalités avec deux niveaux : AV fixe et AV animée. Le deuxième facteur met en jeu les environnements perceptifs, il contient donc trois niveaux : silencieux, perturbé (IRMf/EEG) et bruyant. Le troisième facteur concerne les paires syllabiques, il compte quatre niveaux (syllabes s'opposant sur la labialité, sur le voisement, sur le lieu d'articulation et paires non phonologiques).

B.2.4.1. Comparaison entre les environnements silencieux et perturbé

B.2.4.1.1. Les modalités

Dans les deux environnements et toutes paires confondus, un effet principal de la modalité est observé ($F(1,15) = 9,80$ $p < 0,007$), qui repose sur un avantage significatif de l'AV animée (97,6 %) par rapport à l'AV fixe (95 %). Les temps de réponse ne sont pas significativement différents.

Tableau B15 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa (les deux environnements et toutes paires confondus), pour les 16 participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	AV fixe		AV animé	
	M	E-T	M	E-T
Score en %	95	4,4	97,6	7,7
Temps de réponse en ms	822	142	842	175

B.2.4.1.2. Les environnements silencieux et perturbé

Toutes paires et modalités confondues, un effet principal de l'environnement est constaté pour les scores de discrimination ($F(1,15) = 39,82$ $p < 0,00002$), il repose sur une différence dans l'environnement perturbé (94,5 %) qui est significativement moins bien discriminé que l'environnement silencieux (98 %) ; ainsi que pour les temps de réponse ($F(1,15) = 17,30$ $p < 0,0009$), qui repose sur une différence entre les environnements perturbé (777 ms) et silencieux (887 ms).

Tableau B16 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements silencieux et perturbé (toutes modalités et paires confondues), pour les 16 participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	Silencieux		Perturbé (IRM)	
	M	Et	M	Et
Score en %	98	3,6	94,5	7,7
Temps de réponse en ms	887	131	777	166

B.2.4.1.3. Les paires de stimuli dans les environnements silencieux et perturbé

Les paires de stimuli, tous environnements et modalités confondus, présentent un effet principal ($F(3,45) = 10,62$ $p < 0,00003$), reposant sur une différence significative entre les paires s’opposant sur le lieu d’articulation (98 %) et sur le voisement (97,9 %), qui sont significativement mieux discriminées que les paires non phonologiques (94,2 %), et celles s’opposant sur la labialité (94,9 %). Les temps de réponse montrent aussi un effet principal des paires ($F(3,45) = 7,42$ $p < 0,0004$; Non phonologique : 802 ms < Labialité : 836 ms = Lieu d’articulation : 841 ms = Voisement : 849 ms).

Tableau B17 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux environnements et les deux modalités confondus), pour les 16 participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	Labialité		Lieu d’articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	94,9	9,2	98	3,2	97,9	3,6	94,2	6,3
Temps de réponse en ms	836	151	841	164	849	155	802	165

B.2.4.2. Comparaison entre les environnements silencieux et bruyant

Seuls 13 participants ont obtenu des scores supérieurs au hasard dans l’environnement bruyant.

B.2.4.2.1. Les modalités

Dans les deux environnements et toutes paires confondus, un effet principal de la modalité est observé ($F(1,12) = 10,97$ $p < 0,007$), qui repose sur un avantage significatif de l’AV animée (98,5 %) par rapport à l’AV fixe (97,1 %). Les temps de réponse sont significativement différents ($F(1,12) = 13,31$ $p < 0,004$), avec un avantage pour l’AV fixe (776 ms) par rapport à l’AV animée (831 ms).

Tableau B18 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa (les deux environnements et toutes paires confondus), pour les 13 participants ayant obtenu des scores supérieurs au hasard (>65%).

Modalités	AV fixe		AV animée	
	M	E-T		M
Score en %	97,1	5,1	98,5	3,3
Temps de réponse en ms	776	126	831	140

B.2.4.2.2. Les environnements silencieux et bruyant

Toutes paires et modalités confondues, un effet principal de l'environnement est constaté pour les scores de discrimination ($F(1,12) = 5,06$ $p < 0,045$) ; il repose sur une différence dans l'environnement bruyant (97 %), dans lequel les sujets obtiennent des performances significativement inférieures à celles de l'environnement silencieux : 98,6 % ; ainsi que pour les temps de réponse, ($F(1,12) = 5,76$ $p < 0,04$), qui repose sur une différence entre les environnements silencieux (842 ms) et bruyant (764 ms).

Tableau B19 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements silencieux et bruyant (toutes modalités et paires confondues), pour les 13 participants ayant obtenu des scores supérieurs au hasard (>65%).

Environnements	Silencieux		Bruyant	
	M	ET	M	ET
Score en %	98,6	3,4	97	4,9
Temps de réponse en ms	842	124	764	136

B.2.4.2.3. Les paires syllabiques

Les paires de stimuli, tous environnements et modalités confondus, présentent aussi un effet principal ($F(3,36) = 24,16$ $p < 0,000001$), reposant sur une différence significative entre les paires s'opposant sur le lieu d'articulation (99,2 %), sur le voisement (99,1 %), et sur la

labialité (98,1 %), qui sont significativement mieux discriminées que les paires non phonologiques (94,7 %). Les temps de réponse montrent un effet principal des paires de stimuli ($F(3,36) = 8,50$ $p < 0,0003$; Non phonologique : 791 ms = Labialité : 793 ms = Lieu d'articulation : 802 ms < Voisement : 827 ms).

Tableau B20 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux environnements et les deux modalités confondus), pour les 13 participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	98,1	4,7	99,2	2,5	99,1	2,4	94,7	5,2
Temps de réponse en ms	793	126	802	141	827	135	791	142

B.2.4.3. Comparaison entre les environnements perturbé et bruyant

Seuls sept participants ont obtenu des scores supérieurs au hasard à la fois dans l'environnement bruyant et dans l'environnement IRM.

B.2.4.3.1. Les modalités

Dans les deux environnements et toutes paires confondus, un effet principal de la modalité est observé ($F(1,6) = 24,50$ $p < 0,003$), qui repose sur un avantage significatif de l'AV animée (97,4 %) par rapport à l'AV fixe (93,1 %). Les temps de réponse ne sont pas significativement différents.

Tableau B21 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, (les deux environnements et toutes paires confondus), pour les sept participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	AV fixe		AV animé	
	M	E-T		M
Score en %	93,1	8,7	97,4	4,1
Temps de réponse en ms	757	123	747	96

B.2.4.3.2. Les environnements perturbé et bruyant

Toutes paires et modalités confondus, il n'y a qu'un effet tendanciel de l'environnement pour les scores de discrimination ($F(1,6) = 5,42$ $p < 0,06$), les performances dans l'environnement bruyant (96,5 %) ne diffèrent pas statistiquement de celles observées dans l'environnement IRM (94 %) ; les temps de réponse ne sont pas significativement différents.

Tableau B22 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements perturbé et bruyant (toutes modalités et paires confondues), pour les sept participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	AV fixe		AV animé	
	M	E-T		M
Score en %	96,5	5,8	94	8,1
Temps de réponse en ms	726	76	778	131

B.2.4.3.3. Les paires syllabiques

Les paires de stimuli, tous environnements et modalités confondus, présentent aussi un effet principal ($F(3,18) = 8,15$ $p < 0,002$), reposant sur une différence significative entre les paires s'opposant sur le lieu d'articulation (97,8 %) et sur le voisement (97,6 %), qui sont

significativement mieux discriminées que les paires s'opposant sur la labialité (91,5 %) et les paires non phonologiques (94 %).

Tableau B23 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux environnements et les deux modalités confondus), pour les sept participants ayant obtenu des scores supérieurs au hasard (>65%).

Paires différentes	Labialité		Lieu d'articulation		Voisement		Non phonologique	
	M	E-T	M	E-T	M	E-T	M	E-T
Score en %	91,5	10,7	97,8	3,7	97,6	4	94	5,8
Temps de réponse en ms	766	118	748	104	773	121	721	92

B.2.5. Synthèse

Nous reportons ici les principaux résultats statistiques des performances de discrimination syllabique, en fonction des deux modalités, AV fixe et AV animée, et de l'environnement. Nous aurons l'occasion de revenir sur ces résultats lors de notre discussion générale.

Dans l'environnement très perturbé qu'est l'IRMf, les modalités AV fixe et AV animée sont significativement différentes (AVa : 96,9 % > AVf : 92,1 %) toutes paires confondues, alors qu'il n'y a d'apport significatif ni dans l'environnement bruyant (AVa : 97,2 % = AVf : 96,2 %), ni dans l'environnement silencieux (AVa : 98,7 % = AVf : 98,2 %).

Nous observons des différences significatives des environnements perturbé et bruyant par rapport à l'environnement silencieux exempt de toutes perturbations (silencieux : 98 % > perturbé : 94,5 % ; silencieux : 98,6 % > bruyant : 97 %), la différence étant plus importante entre les environnements silencieux et perturbé qu'entre les environnements silencieux et bruyant.

Dans l'environnement silencieux, toutes les paires phonologiques ont été discriminées sans difficulté particulière. L'interaction significative entre les modalités et les paires, pour les scores de discrimination, repose sur les scores enregistrés pour les paires non phonologiques, qui sont moins bien discriminées que les paires syllabiques (paires non phonologiques en AVf : 94,04 % et en AVa : 96,95 %).

Dans l'environnement perturbé, l'impact du système de transmission et le bruit des gradients a touché davantage les paires s'opposant sur la labialité. Une interaction significative est, à nouveau, observée entre les deux modalités et les différentes paires pour les scores de discrimination. Elle repose non seulement sur une différence entre les paires phonologiques et non phonologiques, mais, de surcroît, les paires s'opposant sur la labialité en AV fixe ont été très difficiles à discriminer. Nous avons recueilli les scores les plus faibles sur le contraste reposant sur la labialité en AV fixe (82,8 %). Les paires non phonologiques en AV fixe (92,4 %) et en AV animée (95 %) sont significativement moins bien discriminées que les autres hormis les paires s'opposant sur le lieu d'articulation en AV fixe (95,8 %).

Dans l'environnement bruyant, une interaction entre modalités et paires est relevée, pour les scores de discrimination. Les paires non phonologiques en AV fixe (92,5 %) sont significativement moins bien discriminées que toutes les autres à l'exception des paires

s'opposant sur la labialité en AV fixe (93,5 %), et des paires non phonologiques en AV animée (95,6 %). La labialité en AV fixe est à son tour différente de tous les autres contrastes hormis les paires non phonologiques en AV animée. Notons que les oppositions de lieu d'articulation et de voisement ne montrent pas de différences significatives entre nos deux modalités.

Ce tableau synoptique est destiné à donner une vue générale, mais ne présente qu'une partie des résultats exposés précédemment, à savoir les plus robustes.

Tableau B24 : Différences significatives lors de la comparaison des scores de discrimination (en pourcentage) dans les différents environnements. « LdA » : Paires s'opposant sur les Lieux d'Articulation. « VOT » : Paires s'opposant sur le voisement. « Lab » : Paires s'opposant sur la labialité.

	Silencieux vs Perturbé	Silencieux vs Bruyant	Bruyant vs Perturbé
Modalités	<p>AVa > AVf 97,6 > 95</p> <p>F(1,15) = 9,80 p < 0,007</p>	<p>AVa > AVf 98,5 > 97,1</p> <p>F(1,12) = 10,97 p < 0,007</p>	<p>AVa > AVf 93,1 > 97,4</p> <p>F(1,6) = 24,50 p < 0,003</p>
Environnements	<p>Silencieux > Perturbé 98 > 94,5</p> <p>F(1,15) = 39,82 p < 10⁻⁴</p>	<p>Silencieux > Bruyant 98,6 > 97</p> <p>F(1,12) = 5,06 p < 0,045</p>	<p>Bruyant = Perturbé 96,5 = 94</p> <p>F(1,6) = 5,42 p < 0,06</p>
Paires	<p>LdA = Vois > Lab = NP 98 = 97,9 > 94,9 = 94,2</p> <p>F(3,45) = 10,62 p < 10⁻⁴</p>	<p>LdA = Vois = Lab > NP 99,2 = 99,1 = 98,1 > 94,7</p> <p>F(3,36) = 24,16 p < 10⁻⁶</p>	<p>LdA = Vois > Lab = NP 97,8 = 97,6 > 91,5 = 94</p> <p>F(3,18) = 8,15 p < 0,002</p>

B.3. Résultats : IRMf

Deux types d'analyse sont présentés dans ce chapitre : les analyses de groupes et celles en régions d'intérêt (cf. B.1.).

Les analyses de groupes sont établies à partir de la moyenne des images anatomo-fonctionnelles de chaque sujet, et fournissent des données sur les zones cérébrales engagées dans le processus de discrimination syllabique.

Les analyses en régions d'intérêt nous permettent de comparer les pics d'activations de ces zones, en réalisant des comparaisons statistiques.

B.3.1. Résultats des analyses de groupe en effet aléatoire du cerveau entier

Les tableaux et les figures présentent les régions cérébrales significativement activées par la discrimination en modalités AV fixe et AV animée, par rapport à la condition de référence, c'est-à-dire, pour nous, les « événements nuls » (sans stimuli).

B.3.1.1. Activations cérébrales engendrées par la discrimination des paires syllabiques

Nous débutons en présentant les résultats pour les trois paires syllabiques et les paires non phonologiques.

B.3.1.1.1. Discrimination des paires s'opposant sur la labialité

Pour la modalité AV fixe, des activations du gyrus de Heschl de l'hémisphère gauche sont observées dans l'aire de Brodmann 42. Un groupe de voxels est activé dans le Gyrus Temporal Supérieur (Aire de Brodmann 22) de l'hémisphère gauche. Une partie de l'aire de Brodmann 48, sise dans la part mésiale du Gyrus Temporal Moyen, est aussi recrutée. Des activations de l'aire de Brodmann 21, dans le Gyrus Temporal Moyen, sont présentes dans l'hémisphère droit. Finalement, deux groupes de voxels sont activés dans le lobe occipital de l'hémisphère gauche (Aires de Brodmann 18 et 19).

Pour la modalité AV animée, une activation du gyrus de Heschl est observée dans l'hémisphère gauche, et s'étend sur les aires de Brodmann 41 et 42. Une activation du Gyrus Temporal Supérieur est aussi constatée dans l'aire de Brodmann 22, et s'étend jusqu'à la part mésiale du Gyrus Temporal Moyen (aire de Brodmann 48). Dans l'hémisphère droit, l'aire de Brodmann 21 du Gyrus Temporal Moyen présente un groupe de voxels activés, qui s'étend

jusqu'à l'Aire de Brodmann 22, dans le Gyrus Temporal Supérieur. Cela signifie qu'une partie du Sillon Temporal Supérieur est activée. Dans le gyrus fusiforme (Aire de Brodmann 37), nous observons trois pics d'activation dans l'hémisphère droit et un seul dans l'hémisphère gauche. Les aires de Brodmann 18 et 19 du lobe occipital sont aussi, en partie, recrutées dans les deux hémisphères. Finalement, des activations sont observées dans le gyrus postcentral de l'hémisphère gauche et dans l'Aire Motrice Supplémentaire. Voir figure B20.

Tableau B25 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur la labialité (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ FWE, et ayant une étendue spatiale supérieure à 25 voxels). AB : Aire de Brodmann ; H : hémisphère ; T : statistique de *Student*.

Régions activées par la discrimination de la labialité [y] / [i]

Régions	AB	H	AV fixe					AB	H	AV animé									
			Voxels	x	y	z	T			Voxels	x	y	z	T					
Gyrus de Heschl	42	G	57	-62	-38	14	9.02	42	G	160	-58	-40	14	9.68					
								41	G						-38	-30	10	9.57	
Gyrus Temporal Supérieur	22	G	179	-62	-18	6	9.64	41	G	324	-46	-38	10	8.01					
								22	G						-60	-10	0	10.03	
Part mésiale du Gyrus Temporal Moyen (GTM)	22	G	106	-58	-28	6	8.35	22	G	48	-62	-20	6	9.57					
								48	G						-56	-16	16	9.30	
Gyrus Temporal Moyen (GTM)	48	D	144	44	-32	6	9.02	21	D	534	64	-20	-6	10.71					
															48	D	34	-24	6
Gyrus Temporal Supérieur Gyrus fusiforme	21	D	106	58	-20	-8	9.42	21	D	22	64	-36	-10	9.59					
								21	D						60	-12	-6	9.40	
Lobe occipital	19	G	28	34	-70	-28	9.01	37	D	33	42	-42	28	10.75					
								37	D						96	-44	-54	-30	8.82
								37	G						-32	-60	-26	8.65	
								37	G						-36	-50	-32	7.73	
								37	D						53	34	-54	-30	8.74
								37	D						32	-58	-22	8.63	
								37	D						28	-50	-26	8.31	
								37	D						367	48	-68	-4	9.68
								19	D						46	-76	-4	9.31	
								19	D						42	-74	-20	8.67	
								19	G						74	-40	-86	-6	10.84
								19	G						-42	-74	-4	8.61	
Cortex moteur	18	D	354	10	-96	-6	11.80	18	D	18	2	-86	-12	11.32					
								18	D						-6	-82	-16	7.98	
								18	G						38	-8	-66	-22	9.11
								18	G						25	-22	-96	4	8.77
								2	D						145	42	-28	44	9.18
								3	D						44	-36	64	8.00	
Aire Motrice Supplémentaire	2	D	2	42	-36	54	7.73	6	G	99	-4	-2	60	8.31					
								6	G						12	0	62	7.84	
								6	D										

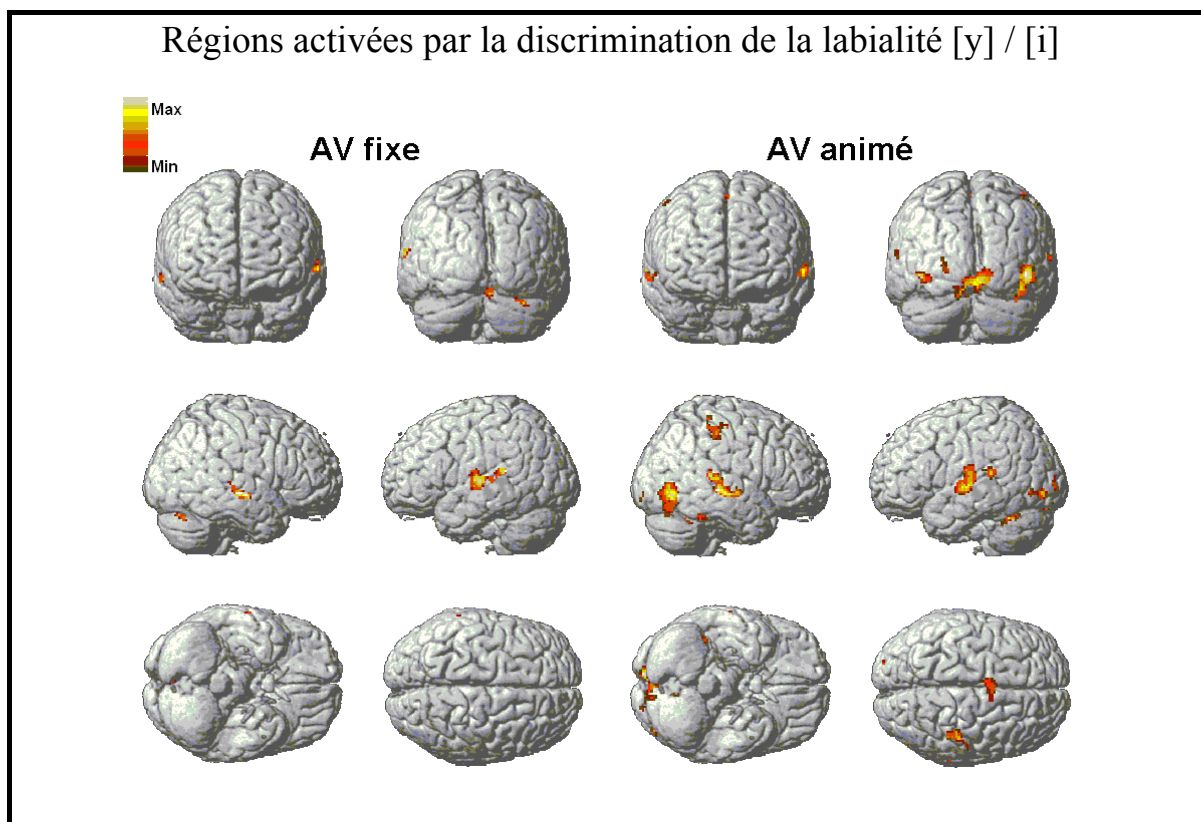


Figure B20 : Rendus de régions cérébrales significativement activées en AVf et AVa lors de la soustraction « stimuli-événement nul », pour les paires s’opposant sur la labialité (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ *FWE*, et ayant une étendue spatiale supérieure à 25 voxels).

B.3.1.1.2. Discrimination des paires s’opposant sur les lieux d’articulation

Pour la modalité AV fixe, une activation dans le gyrus de Heschl de l’hémisphère gauche est observée dans l’aire de Brodmann 42. Celle-ci fait partie d’un groupe de 463 voxels activés dans le Gyrus Temporal Supérieur (Aire de Brodmann 22). La part mésiale antérieure du Gyrus Temporal Moyen présente aussi une activation dans l’hémisphère droit (Aire de Brodmann 48). Des activations de l’aire de Brodmann 21, dans le Gyrus Temporal Moyen, sont observées dans l’hémisphère droit, et s’étendent jusqu’à l’aire de Brodmann 48, dans la part mésiale de ce gyrus. Un groupe de 89 voxels, qui s’étend sur les deux hémisphères, est activé dans le lobe occipital (Aire de Brodmann 18). Pour finir, une partie de l’Aire Motrice Supplémentaire est aussi recrutée, avec un pic situé dans l’hémisphère gauche.

Pour la modalité AV animée, deux activations dans le gyrus de Heschl, l’une dans l’aire de Brodmann 41, l’autre dans l’aire de Brodmann 42, sont observées dans les deux

hémisphères. L'activation de l'aire de Brodmann 42, dans le gyrus de Heschl, s'étend sur l'aire de Brodmann 22. Dans l'hémisphère gauche, une activation dans le Gyrus Temporal Supérieur est aussi constatée dans l'aire de Brodmann 22, et s'étend jusqu'à la part mésiale du Gyrus Temporal Moyen (aire de Brodmann 48). Dans l'hémisphère droit, une partie de l'aire de Brodmann 21 du Gyrus Temporal Moyen est aussi recrutée, cette activation s'étend jusqu'à la part mésiale du Gyrus Temporal Moyen (aire de Brodmann 48). Dans le gyrus fusiforme (Aire de Brodmann 37), nous observons un pic d'activation dans l'hémisphère droit, qui appartient à un groupe dont le maximum est situé sur la face interne de l'aire de Brodmann 18. Les aires de Brodmann 18 et 19, dans le lobe occipital, sont aussi, en partie, recrutées ; l'aire de Brodmann 18 de façon bilatérale, et l'aire de Brodmann 19 uniquement à gauche. Une activation est présente dans le cortex moteur de l'hémisphère droit. Finalement, un groupe de 25 voxels est activé dans le cortex prémoteur de l'hémisphère gauche (Aire de Brodmann 6). Voir figure B21.

Tableau B26 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur les lieux d'articulation (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ FWE, et ayant une étendue spatiale supérieure à 25 voxels).

Régions activées par la discrimination des lieux d'articulation [p] / [t] et [b] / [d]

Régions	AB	H	AV fixe					AB	H	AV animé				
			Voxels	x	y	z	T			Voxels	x	y	z	T
Gyrus Temporal Supérieur	22	G	463	-64	-20	4	11.08	22	G	428	-64	-18	4	11.28
	22	G		-64	-32	10	10.07	22	G		-62	-8	4	10.62
Part mésiale du GTM							48	G		-56	-16	14	8.83	
Gyrus de Heschl	41	G		-52	-34	14	8.76	41	G	111	-48	-38	18	8.89
							41	G		-38	-32	12	8.45	
							42	G	98	-60	-38	10	10.13	
Gyrus Temporal Supérieur							22	G		-60	-40	0	7.63	
Gyrus Temporal Moyen	21	D	508	64	-22	-4	12.27	21	D	742	64	-24	-2	14.24
	21	D		52	-26	-2	9.81	21	D		64	-16	10	9.78
Part mésiale du GTM	48	D		44	-32	6	9.09	48	D		44	-24	0	10.48
Part mésiale antérieure GTM	48	D	97	24	10	-4	10.15							
Lobe occipital	18	D	89	4	-82	-20	8.54	18	D	189	10	-94	-6	9.90
	18	G		-4	-80	-26	7.56	18	D		12	-86	-8	9.65
								18	G	76	-16	-62	-20	8.14
Gyrus fusiforme								37	G		-36	-50	-32	7.94
								19	G	91	48	-74	-12	8.72
Aire Motrice Supplémentaire	6	G	283	-2	8	56	11.35	6	G	303	-6	-2	60	12.30
								6	D		4	6	52	8.82
Cortex moteur								4	D		58	-14	50	8.03
								3	D	169	44	-28	62	10.31
								3	D		42	-32	52	8.08
								3	D		50	-20	56	7.73
Cortex prémoteur								6	G	25	-50	-4	52	10.59

Régions activées par la discrimination des lieux d'articulation [p] / [t] et [b] / [d]

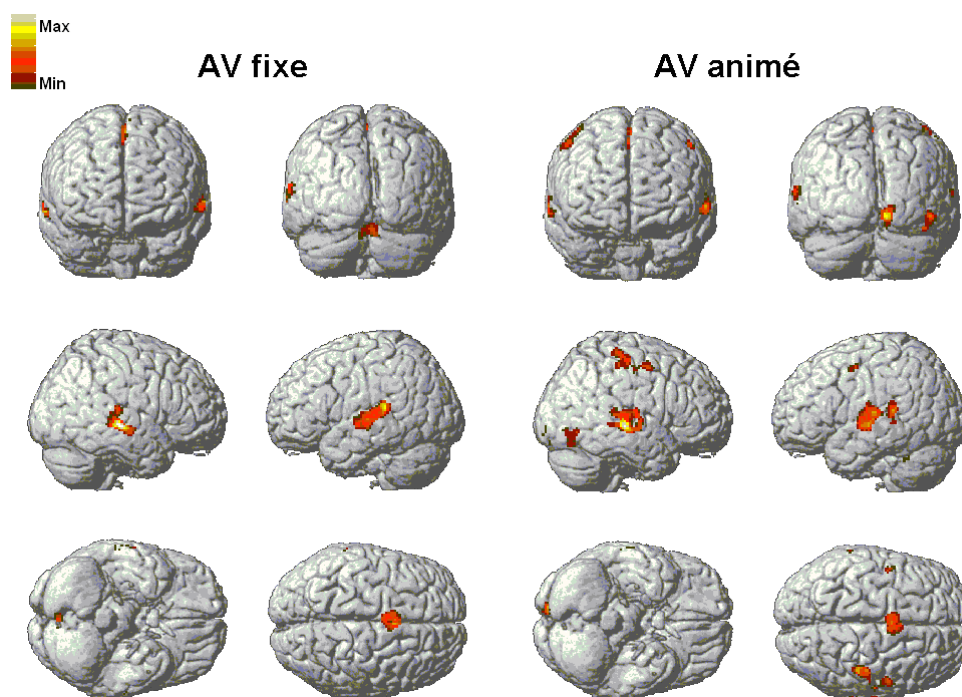


Figure B21 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « stimuli-événement nul » pour les paires s'opposant sur les lieux d'articulation (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ FWE, et ayant une étendue spatiale supérieure à 25 voxels).

B.3.1.1.3. Discrimination des paires s'opposant sur le voisement

Pour la modalité AV fixe, le Gyrus Temporal Supérieur (Aire de Brodmann 22) présente une activation dans l'hémisphère gauche, qui s'étend jusqu'au gyrus de Heschl (Aire de Brodmann 42). Dans l'hémisphère droit, nous pouvons observer, dans le Gyrus Temporal Moyen, une activation, qui s'étend jusqu'à la part mésiale antérieure (Aire de Brodmann 48). Finalement, l'Aire Motrice Supplémentaire est aussi recrutée, avec un pic situé dans l'hémisphère gauche.

Pour la modalité AV animée, un groupe de 47 voxels est activé dans le gyrus de Heschl de l'hémisphère gauche (Aire de Brodmann 41). Dans l'hémisphère gauche, des activations du Gyrus Temporal Supérieur sont aussi constatées dans l'aire de Brodmann 22, et s'étendent jusqu'à l'Aire de Brodmann 21, dans le Gyrus Temporal Moyen. Cela signifie

qu'une partie du STS est recrutée. Dans l'hémisphère droit, l'aire de Brodmann 21 dans le Gyrus Temporal Moyen est aussi recrutée. Dans le gyrus fusiforme (Aire de Brodmann 37), nous observons une activation dans l'hémisphère droit, qui s'étend jusqu'à l'aire de Brodmann 19. Les aires de Brodmann 18 et 19, dans le lobe occipital, sont aussi, en partie, recrutées dans l'hémisphère gauche. Des activations du cortex moteur, dans l'hémisphère droit, mais aussi de l'Aire Motrice Supplémentaire sont observées. À nouveau, un groupe de 27 voxels est activé dans le cortex prémoteur de l'hémisphère gauche (Aire de Brodmann 6). Voir figure B22.

Tableau B27 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur le voisement (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ *FWE*, et ayant une étendue spatiale supérieure à 25 voxels).

Régions activées par la discrimination du voisement [p] / [b] et [t] / [d]

Régions	AB	H	AV fixe					AB	H	AV animé				
			Voxels	x	y	z	T			Voxels	x	y	z	T
Gyrus Temporal Moyen							21	G	442	-64	-38	0	9.63	
Gyrus Temporal Supérieur	22	G	297	-62	-20	4	9.67	22	G		-62	-20	6	9.35
	22	G		-62	-10	0	8.19	22	G		-62	-32	12	9.22
Gyrus de Heschl	42	G		-62	-40	14	8.47	41	G	47	-42	-40	10	8.50
								41	G		-36	-36	6	7.88
Part mésiale du GTM	48	D	388	46	-22	2	9.79							
Gyrus Temporal Moyen	21	D		66	-26	2	10.58	21	D	409	64	-22	-4	12.79
	21	D		44	-30	6	9.09	21	D		64	-20	6	8.66
								21	D		44	-30	6	8.27
Gyrus fusiforme								37	D	261	48	-68	-4	8.36
Lobe occipital								19	D		44	-74	-20	8.35
								19	D		38	-76	-26	8.08
								17	D	682	10	-92	-8	10.77
								17	D		2	-88	-12	10.72
								17	D		-4	-86	-6	9.84
								18	G	26	-4	-66	-24	8.50
								19	G	38	-36	-82	-8	8.62
								19	G	80	-16	-56	-16	9.59
								19	G		-22	-60	-26	7.40
Aire Motrice Supplémentaire	6	G	102	-4	2	64	10.52	6	G	421	-4	0	60	12.41
Cortex moteur								2	D	637	44	-28	48	10.23
								4	D		56	-18	52	10.11
								4	D		42	-20	52	9.88
Cortex prémoteur								6	G	27	-50	-2	52	9.91

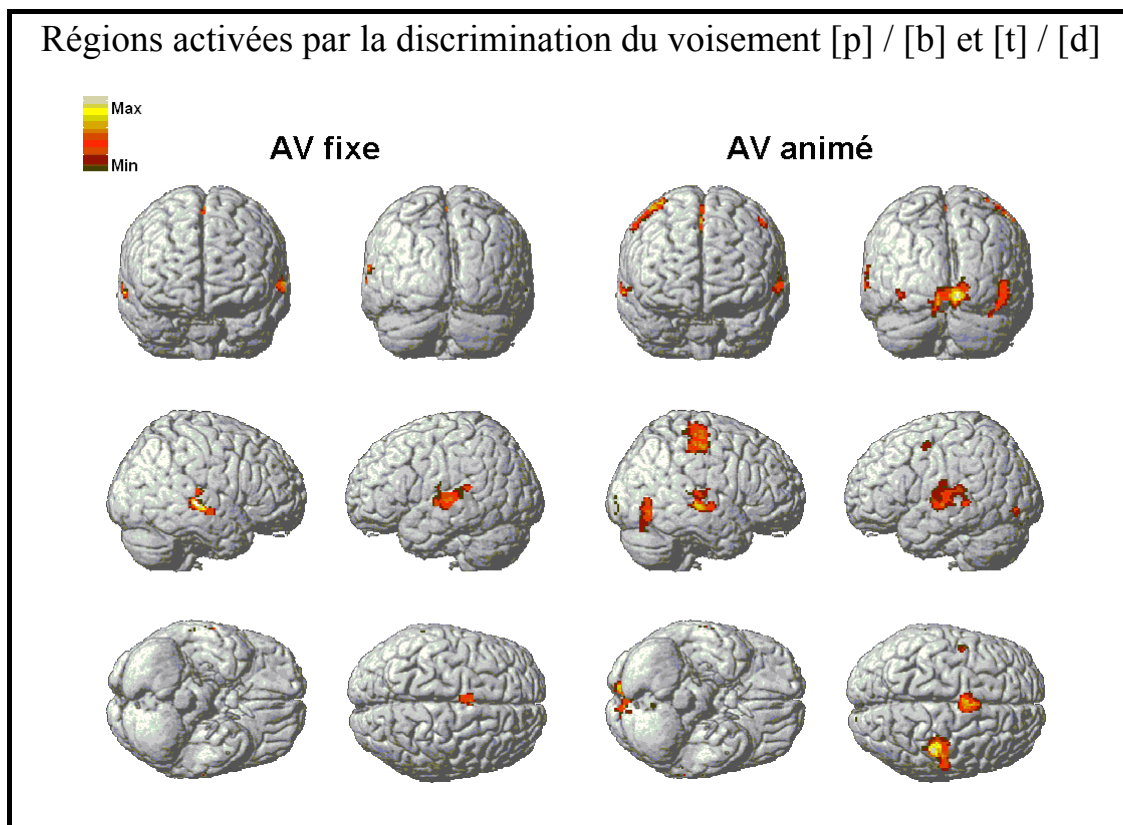


Figure B22 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s’opposant sur le voisement (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ *FWE*, et ayant une étendue spatiale supérieure à 25 voxels).

B.3.1.1.4. Discrimination des paires non phonologiques

Pour la modalité AV fixe, les Gyri Temporaux Moyen et Supérieur sont activés dans les deux hémisphères. Dans l’hémisphère droit, ces activations s’étendent sur l’aire de Brodmann 21, jusqu’à la part mésiale du Gyrus Temporal Moyen (Aire de Brodmann 48). Dans l’hémisphère gauche, le gyrus de Heschl est aussi recruté (Aire de Brodmann 42), tout comme le Gyrus Fusiforme (Aire de Brodmann 37), et le lobe occipital (Aire de Brodmann 19).

Pour la modalité AV animée, le Gyrus Temporal Supérieur (Aire de Brodmann 22) et le gyrus de Heschl sont activés dans l’hémisphère gauche, ainsi que le Gyrus Angulaire (Aire de Brodmann 39), et le lobe occipital (Aire de Brodmann 18). Dans l’hémisphère droit, les Gyri Temporal Moyen (Aire de Brodmann 21) et Fusiforme (Aire de Brodmann 37) présentent des activations, ainsi que le lobe occipital (Aires de

Brodmann 17 et 19). Nous pouvons aussi observer une activation bilatérale dans l'Aire Motrice Supplémentaire, avec un pic situé dans l'hémisphère gauche. Voir figure B23.

Tableau B28 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires non phonologiques (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ *FWE*, et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination des sons non phonologiques

Régions	AB	H	AV fixe				AB	H	AV animé					
			Voxels	x	y	z			T	Voxels	x	y	z	T
Gyrus Temporal Moyen	21	D	463	62	-22	-4	10.60	21	D	97	66	-24	0	9.56
Gyrus Temporal Supérieur	22	D		62	-14	-6	12.22	22	G	35	-62	-20	6	8.35
Part mésiale du GTM	48	D		46	-24	2	10.10							
	48	G	45	-36	-30	8	9.15							
Gyrus Temporal Supérieur	22	G	222	-64	-20	6	10.13							
	22	G		-60	-10	0	9.87							
Gyrus de Heschl	42			-62	-36	14	9.41	41	G	41	-44	-38	12	8.23
Gyrus fusiforme	37	G	30	-28	-56	-34	7.89	37	D		48	-68	-6	8.16
Lobe occipital	19	G	26	-32	-80	-24	9.46	19	D	136	42	-74	-20	9.21
								17	D	225	12	-92	-8	8.86
								18			18	-96	8	7.44
								18	G	32	-38	-84	10	7.77
Gyrus angulaire								39	G		-40	-82	18	7.59
Aire Motrice Supp	6		87	2	6	58	9.53	6	D	74	2	8	54	8.88
								6	G		-4	2	58	7.70

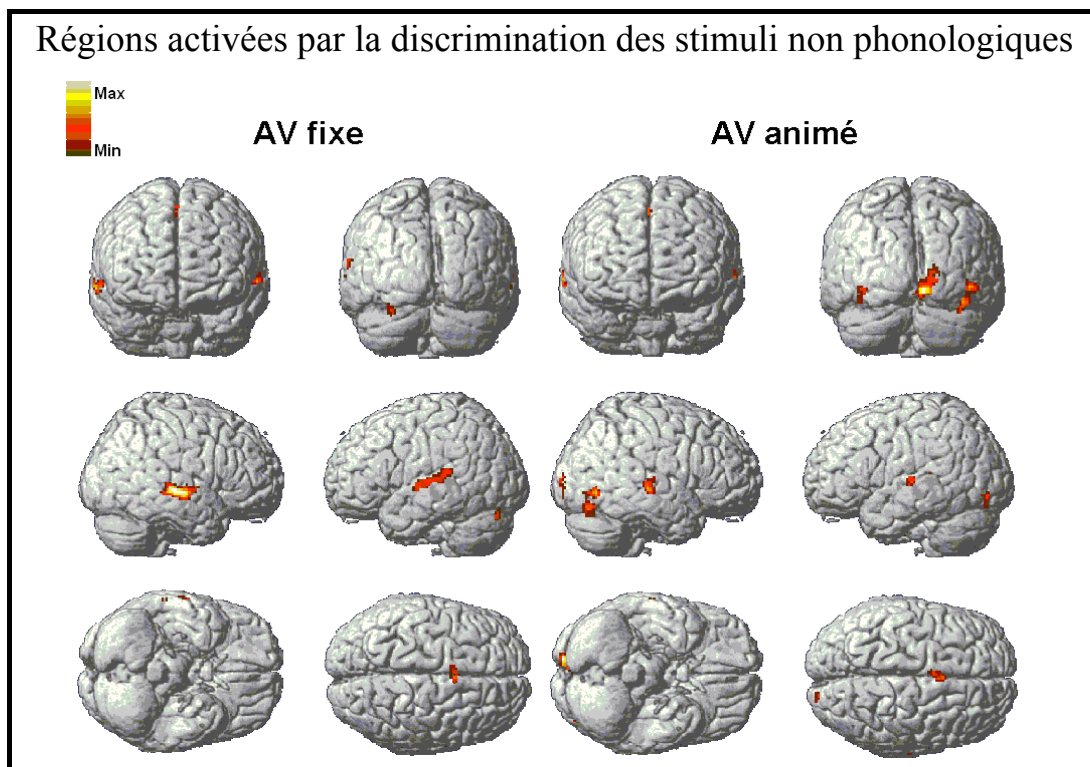


Figure B23 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires non phonologiques (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.01$ *FWE*, et ayant une étendue spatiale supérieure à 25 voxels).

B.3.2. Comparaison de la discrimination en présentation AV fixe et AV animée

Nous présentons ici des tests T de *Student* sur deux échantillons. Cette opération a pour but de révéler les groupes de voxels qui sont significativement plus activés en modalité AV animée par rapport à l'AV fixe. Comme nous l'avons signalé précédemment, le seuil statistique est non corrigé avec un « p » inférieur à 0.001.

B.3.2.1. Discrimination des paires s'opposant sur la labialité

Un groupe de 878 voxels de l'hémisphère droit est activé dans le gyrus fusiforme (Aire de Brodmann 37), et s'étend jusqu'à l'aire de Brodmann 19, dans le lobe occipital. Deux activations sont observées dans le lobe pariétal (Aires de Brodmann 7 et 40). Dans l'hémisphère gauche, seul le gyrus fusiforme (Aire de Brodmann 37) présente une activation. Voir figure B24.

Tableau B29 : Régions cérébrales significativement activées, pour les paires s'opposant sur la labialité, lors de la soustraction « AVa-AVf » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination de la labialité en AV animée, par rapport à l'AV fixe

Régions	AB	Hémis.	Voxels	x	y	z	T
Gyrus fusiforme	37	G	294	-44	-68	2	4.75
	37	G		-40	-96	-6	4.70
	37	D	878	42	-66	-6	6.33
Lobe occipital	19	D		48	-76	-4	5.46
Lobe pariétal	40	D	53	36	-40	56	3.80
	7	D	50	24	-64	60	3.85

B.3.2.2. Discrimination des paires s'opposant sur les lieux d'articulation

Un groupe de 522 voxels est activé dans le gyrus fusiforme de l'hémisphère droit. Dans l'hémisphère gauche, une activation est présente et s'étend de l'aire de Brodmann 19, dans le lobe occipital, jusqu'au gyrus fusiforme (Aire de Brodmann 37). Voir figure B24.

Tableau B30 : Régions cérébrales significativement activées, pour les paires s'opposant sur les lieux d'articulation, lors de la soustraction « AVa-AVf » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination des lieux d'articulation en AV animée, par rapport à l'AV fixe

Régions	AB	Hémis.	Voxels	x	y	z	T
Lobe occipital	19	G	59	-42	-74	-4	3.94
Gyrus fusiforme	37	G		-44	-68	4	3.73
	37	D	522	48	-66	-6	4.83
	37	D		54	-70	6	3.79

B.3.2.3. Discrimination des paires s'opposant sur le voisement

Dans l'hémisphère droit, un groupe de 687 voxels est activé dans le gyrus fusiforme (Aire de Brodmann 37). Un groupe de 146 voxels est activé dans l'hémisphère gauche à la jonction du gyrus fusiforme (Aire de Brodmann 37) et du lobe occipital (Aires de Brodmann 18 et 19) Voir figure B24.

Tableau B31 : Régions cérébrales significativement activées, pour les paires s'opposant sur le voisement, lors de la soustraction « AVa-AVf » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination du voisement en AV animée, par rapport à l'AV fixe

Régions	AB	Hémis.	Voxels	x	y	z	T
Gyrus fusiforme	37	D	687	42	-66	-6	5.79
Lobe occipital	19	G	146	-42	-72	-4	4.34
	18	G		-38	-92	-6	3.87
Gyrus fusiforme	37	G		-46	-66	-6	3.71

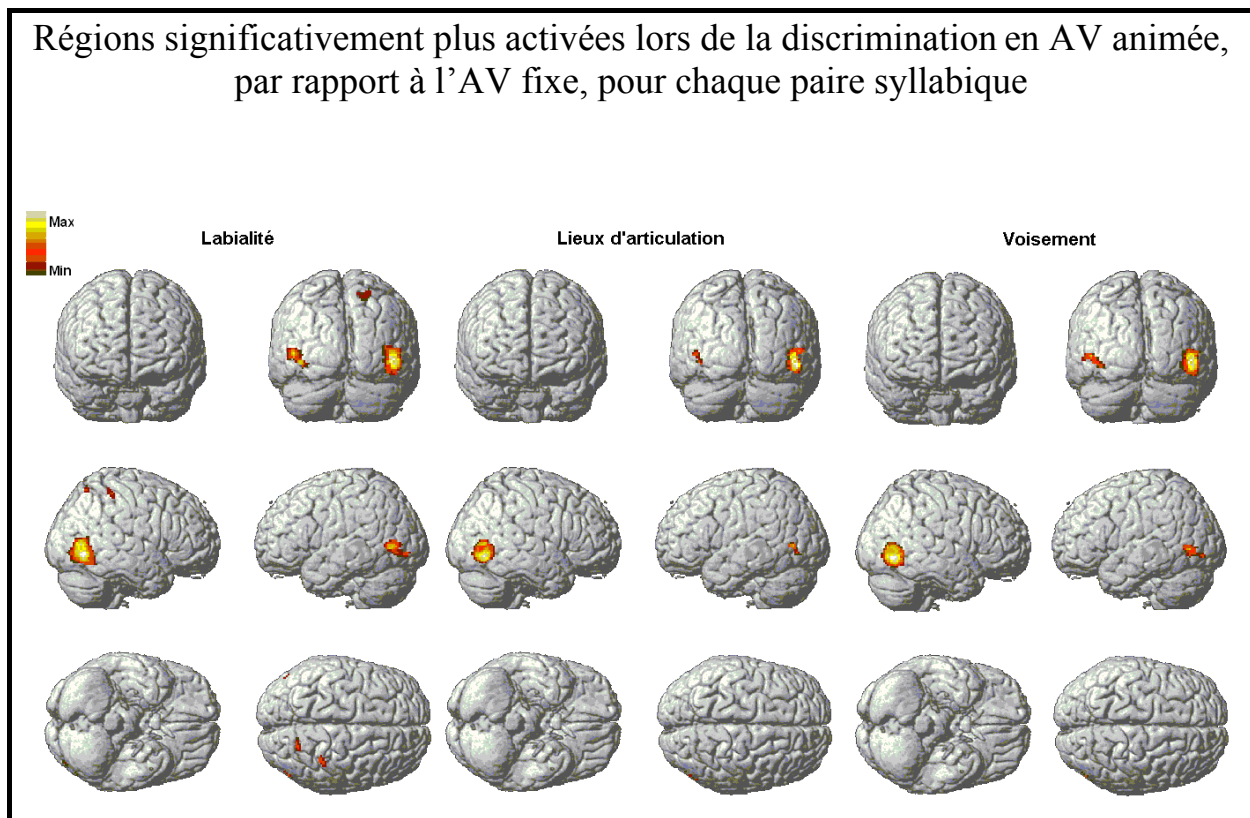


Figure B24 : Rendus des régions cérébrales significativement activées, lors de la soustraction « AVa-AVf », à gauche pour la labialité, au centre pour les lieux d'articulation et à droite pour le voisement, (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

B.3.3. Les activations engendrées par la discrimination des stimuli phonologiques par rapport aux non phonologiques

Lors des comparaisons entre les paires phonologiques et non phonologiques, nous avons été contraint de modifier le seuil statistique, afin de révéler des activations. Les tableaux et les figures présentent les régions cérébrales significativement plus activées par la discrimination phonologique en modalité AV fixe et AV animée, par rapport à la discrimination des paires non phonologiques.

B.3.3.1. Comparaison de la discrimination des paires s'opposant sur la labialité et des paires non phonologiques

Pour la modalité AV fixe, quatre groupes de voxels sont activés dans l'hémisphère gauche : deux le sont dans le lobe occipital (Aire de Brodmann 18) et deux dans les Gyri Frontaux Inférieurs (Aire de Brodmann 44) et supérieur (Aire de Brodmann 6).

Pour la modalité AV animée, six groupes de voxels sont activés à travers les deux hémisphères. Dans l'hémisphère gauche, l'opercule rolandique, (Aire de Brodmann 43) et la part mésiale du Gyrus Temporal Moyen (Aire de Brodmann 48) sont activés, tout comme le gyrus fusiforme (Aire de Brodmann 37). Ce dernier est aussi recruté dans l'hémisphère droit. Pour finir le gyrus angulaire (Aire de Brodmann 39) présente une activation dans les deux hémisphères. Voir figure B25.

Tableau B32 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur la labialité-paires non phonologiques » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination de la labialité, par rapport aux stimuli non phonologiques

Régions	AB	H	AV fixe					AB	H	AV animé				
			Voxels	x	y	z	T			Voxels	x	y	z	T
Gyrus Frontal Inférieur	44	G	190	-48	10	22	5.36							
Gyrus Frontal Supér	6	G	76	-46	-2	50	4.21							
Lobe occipital	17	G	61	-16	-98	10	4.33							
	17	G		-20	-94	16	4.18							
	17	G	37	-14	-68	4	4.17							
Part mésiale du GTM								48	G	60	-56	-26	26	4.06
											-56	-18	28	3.77
											-48	-30	24	3.61
Gyrus angulaire								39	D	327	44	-74	26	5.21
											54	-68	24	5.61
											58	-62	16	4.32
Gyrus fusiforme								39	G	47	-42	-72	16	4.79
								37	G	62	-42	-60	-24	4.85
								37	D	95	28	-38	-28	4.76
Opercule rolandique								43	G	66	-56	-6	26	5.22
											-52	4	-30	4.21

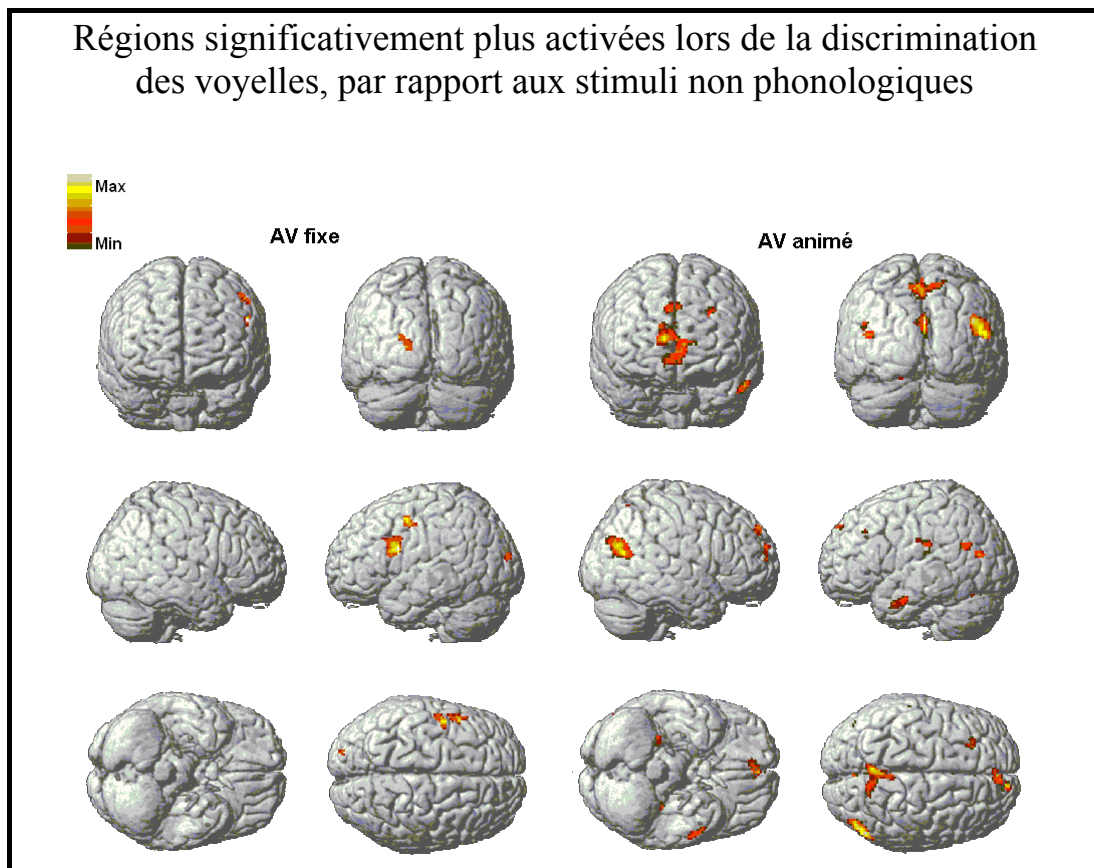


Figure B25 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « paires phonologiques s’opposant sur la labialité-paires non phonologiques » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

B.3.3.2. Comparaison de la discrimination des paires s’opposant sur les lieux d’articulation et des paires non phonologiques

Pour la modalité AV fixe, trois groupes de voxels sont activés dans l’hémisphère gauche, dont deux dans le Gyrus Temporal Moyen (Aire de Brodmann 21), et le troisième dans le Gyrus Frontal Supérieur (Aire de Brodmann 6).

Pour la modalité AV animée, une activation est observée dans le Gyrus Temporal Moyen (Aire de Brodmann 21), dans le gyrus de Heschl (Aire de Brodmann 42), et dans l’opercule rolandique (Aire de Brodmann 43) de l’hémisphère gauche. Le gyrus angulaire présente une activation dans les deux hémisphères. Voir figure B26.

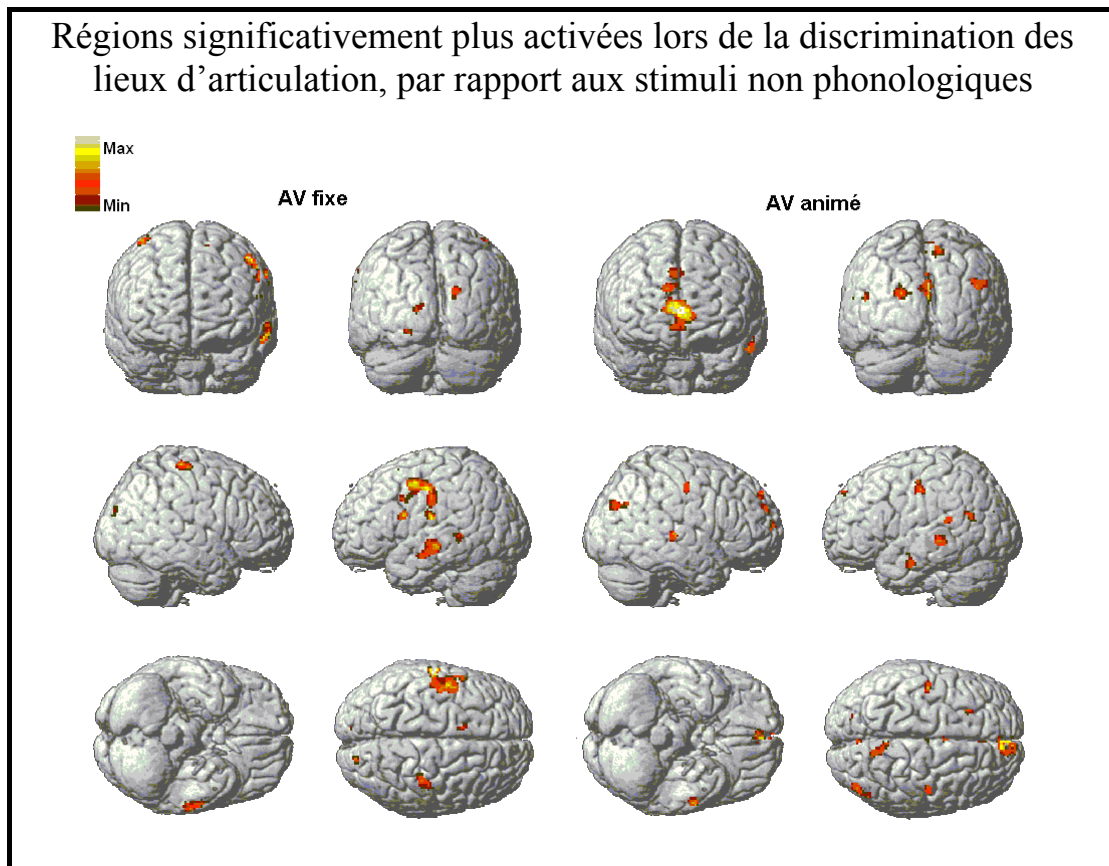


Figure B26 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur les lieux d'articulation-paires non phonologiques » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

B.3.3.3. Comparaison de la discrimination des paires s'opposant sur le voisement et des paires non phonologiques

Pour la modalité AV fixe, trois groupes de voxels sont activés dans l'hémisphère gauche ; l'un dans le Gyrus Temporal Moyen (Aire de Brodmann 21), l'autre dans le Gyrus Frontal Supérieur (Aire de Brodmann 6), et le dernier dans le lobe occipital (Aire de Brodmann 18).

Pour la modalité AV animée, deux groupes de voxels sont activés ; dans l'hémisphère droit, il s'agit de l'opercule rolandique (Aire de Brodmann 43), et du Gyrus Temporal Moyen (Aire de Brodmann 21), dans l'hémisphère gauche. Voir figure B27.

Tableau B34 : Régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur le voisement-paires non phonologiques » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

Régions significativement plus activées lors de la discrimination du voisement,
par rapport aux stimuli non phonologiques

Régions	AB	H	AV fixe					AB	H	AV animé				
			Voxels	x	y	z	T			Voxels	x	y	z	T
Gyrus Temporal Moyen	21	G	31	-62	-26	-6	4.38							
Gyrus Frontal Supérieur	6	G	31	-46	0	52	4.13							
Opercule rolandique							43	D	188	66	-12	28	5.04	
								D		66	-2	22	4.87	
								D		64	8	14	4.74	
Gyrus Temporal Moyen							21	G	211	-52	6	-24	4.95	
							21	G		-56	-2	30	4.94	
							21	G		-62	-8	-16	4.39	
Lobe occipital	18	G	163	-10	-92	16								

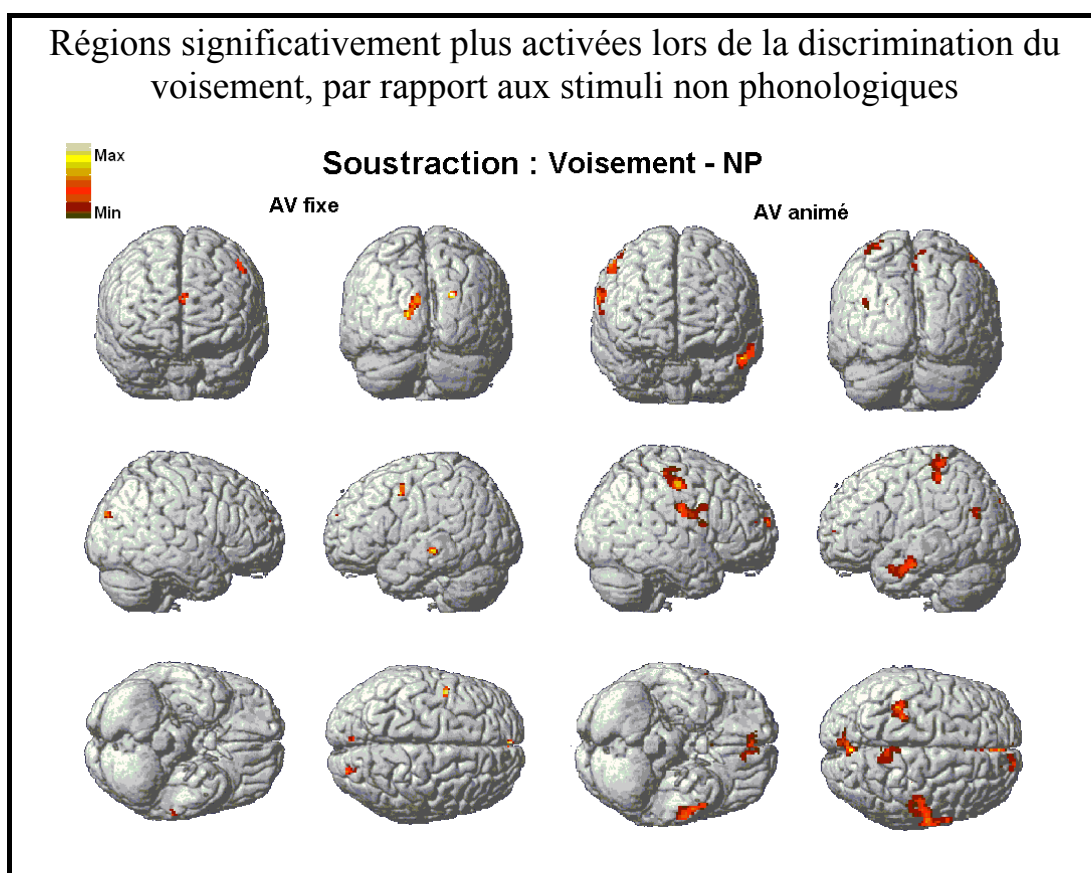


Figure B27 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur la labialité-paires non phonologiques » (test-t sur un échantillon en effet aléatoire, activations significatives au seuil $p < 0.001$ non corrigé et ayant une étendue spatiale supérieure à 25 voxels).

B.3.4. Tableaux récapitulatifs des activations en IRMf

Tableau B35 : Tableau synoptique des activations en modalité AV animée

Paires en modalité AV Fixe	Soustraction	Régions activées	Hémisphère	Aire de Brodmann
Labialité	AVf > Référence	Gyrus de Heschl	G	42
		Gyrus Temporal Supérieur	G	22
		Part mésiale du Gyrus Temporal Moyen	G	48
		Gyrus Temporal Moyen	D	21
		Lobe occipital	G	18 / 19
Lieux d'articulation	AVf > Référence	Gyrus de Heschl	G	42
		Gyrus Temporal Supérieur	G	22
		Part mésiale antérieure du Gyrus Temporal Moyen	D	48
		Gyrus Temporal Moyen	D	21
		Part mésiale du Gyrus Temporal Moyen	D	48
		Lobe occipital	G / D	18
		Aire Motrice Supplémentaire	G / D	6
Voisement	AVf > Référence	Gyrus Temporal Supérieur	G	22
		Gyrus de Heschl	G	42
		Gyrus Temporal Moyen	D	21
		Part mésiale antérieure du Gyrus Temporal Moyen	D	48
		Aire Motrice Supplémentaire	G / D	6
Non phonologique	AVf > Référence	Gyrus Temporal Supérieur	G / D	22
		Gyrus de Heschl	G	42
		Gyrus Temporal Moyen	G / D	21
		Gyrus Fusiforme	G	37
		Lobe occipital	G	19
Labialité	AVf > Non phonologique	Gyrus Frontal Inférieur	G	44
		Gyrus Frontal Supérieur	G	6
		Lobe occipital	G	18
Lieux d'articulation	AVf > Non phonologique	Gyrus Temporal Moyen	G	21
		Gyrus Frontal Supérieur	G	6
Voisement	AVf > Non phonologique	Gyrus Temporal Moyen	G	21
		Gyrus Frontal Supérieur	G	6
		Lobe occipital	G	18

Tableau B36 : Tableau synoptique des activations en modalité AV animée

Paires en modalité AV animée	Soustraction	Régions activées	Hémisphère	Aire de Brodmann
Labialité	Labialité > Référence	Gyrus de Heschl	G	41 / 42
		Gyrus Temporal Supérieur	G	22
		Part mésiale du Gyrus Temporal Moyen	G	48
		Gyrus Temporal Moyen	D	21
		Gyrus Temporal Supérieur	D	22
		Sillon Temporal Supérieur	D	Jonction 21 / 22
		Gyrus Fusiforme	G / D	37
		Lobe occipital	G	18 / 19
		Aire Motrice	G / D	6
		Supplémentaire		
Lieux d'articulation	Lieux d'articulation > Référence	Gyrus de Heschl	G / D	41 / 42
		Gyrus Temporal Supérieur	G	22
		Part mésiale du Gyrus Temporal Moyen	G	48
		Gyrus Temporal Moyen	D	21
		Gyrus Fusiforme	G	37
		Lobe occipital	G	19
		Lobe occipital	G / D	18
		Aire Motrice	G / D	6
		Supplémentaire		
		Cortex Moteur	D	3 / 4
Cortex Prémoteur	G	6		
Voisement	Voisement > Référence	Gyrus de Heschl	G	41
		Gyrus Temporal Supérieur	G	22
		Gyrus Temporal Moyen	G	21
		Sillon Temporal Supérieur	G	Jonction 21 / 22
		Gyrus Temporal Moyen	D	21
		Aire Motrice	G / D	6
		Supplémentaire		
		Gyrus Fusiforme	D	37
		Lobe occipital	D	19
		Lobe occipital	G	18 / 19
Aire Motrice	G / D	6		
Supplémentaire				
Cortex Moteur	D	3 / 4		
Cortex Prémoteur	G	6		
Non phonologique	Non phonologique > Référence	Gyrus Temporal Supérieur	G	22
		Gyrus de Heschl	G	42
		Gyrus Angulaire	G	39
		Lobe occipital	G	18
		Gyrus Temporal Moyen	D	21
		Gyrus Fusiforme	D	37
		Lobe occipital	D	17 / 19
		Aire Motrice	G / D	6
Supplémentaire				

Paires en modalité AV animée	Soustraction	Régions activées	Hémisphère	Aire de Brodmann
Labialité	AVa	Part mésiale du Gyrus Temporal Moyen	G	48
	>	Opercule rolandique	G	43
	Non phonologique	Gyrus Fusiforme	G	37
		Gyrus Fusiforme	D	37
		Gyrus Angulaire	G/D	39
Lieux d'articulation	AVa	Gyrus Temporal Moyen	G	21
	>	Gyrus de Heschl	G	42
	Non phonologique	Opercule rolandique	G	43
		Gyrus Angulaire	G/D	39
Voisement	AVa	Opercule rolandique	D	43
	> Non phonologique	Gyrus Temporal Moyen	G	21
Labialité	AVa	Gyrus Fusiforme	D	37
	>	Lobe occipital	D	19
	AVf	Lobe pariétal	G	7 / 40
		Gyrus Fusiforme	G	37
Lieux d'articulation	AVa	Gyrus Fusiforme	D	37
	>	Gyrus Fusiforme	G	37
	AVf	Lobe occipital	G	19
Voisement	AVa	Gyrus Fusiforme	D	37
	>	Gyrus Fusiforme	G	37
	AVf	Lobe occipital	G	18 / 19

B.3.5. Analyses en régions d'intérêt

Les zones cérébrales significativement activées durant la tâche de discrimination syllabique diffèrent en localisation et en nombre de voxels. Pour prolonger notre analyse, nous avons cherché à savoir s'il existait des différences significatives du niveau d'activation, exprimé par la valeur maximale des pics d'activation (T max), pour chaque sujet, en fonction de la modalité de présentation AV fixe ou AV animée et du type de paires syllabiques.

Nous avons confronté les activations communes des deux modalités, dans quatre régions d'intérêt définies dans chaque hémisphère par l'intersection des groupes de voxels activés, lors de la discrimination des trois paires syllabiques, dans le gyrus de Heschl (AB 41 et 42), le Gyrus Temporal Moyen (AB21) et le Gyrus Temporal Supérieur (AB22).

Une première analyse de variance à trois facteurs (2 modalités x 2 hémisphères x 4 régions d'intérêt : AB 21, 22, 41, 42) a été effectuée sur le niveau d'activation obtenu pour chaque type de paires syllabiques.

Pour la discrimination des paires s'opposant sur la labialité (voyelles), seul un effet principal des régions d'intérêt est observé ($F(3,75) = 7.67$ $p < 0.0002$). Les résultats sont similaires, pour les paires différant sur le voisement, avec un effet principal des régions d'intérêt ($F(3,75) = 12.44$ $p < 0.000001$), et les paires s'opposant sur les lieux d'articulation ($F(3,75) = 15.55$ $p < 0.0000001$). Les analyses *a posteriori* montrent que pour chaque paire, l'effet repose sur une différence du niveau d'activation entre les régions d'intérêt AB 21 et AB 22 supérieures à celles des AB 41 et AB 42.

Une seconde ANOVA à trois facteurs (2 modalités x 2 hémisphères x 3 paires syllabiques) portant sur le niveau d'activation dans chacune des quatre régions d'intérêt (AB 21, 22, 41 et 42) a révélé un effet principal des paires syllabiques sur 3 des quatre régions d'intérêt.

Pour la région d'intérêt « AB 21 », l'effet principal des paires ($F(2,50) = 10,30$ $p < 0.0002$) repose sur une différence entre les trois catégories : les paires s'opposant sur la labialisation ont les seuils les plus bas (3,09), viennent ensuite les paires s'opposant sur le voisement (3,39), et pour finir, les paires s'opposant sur les lieux d'articulation (3,76).

Pour la région d'intérêt « AB 22 » cet effet ($F(2,50) = 8,15$ $p < 0.0009$) repose sur une différence des paires s'opposant sur les lieux d'articulation (3,57), significativement

supérieures aux paires s'opposant sur la labialité (3,14), et à celles s'opposant sur le voisement (3,18).

En ce qui concerne la région d'intérêt « AB 41 », l'effet ($F(2,50) = 3,05$ $p < 0,04$), repose sur une différence entre les paires s'opposant sur le voisement (2,29), significativement inférieures à celles s'opposant sur les lieux d'articulation (2,55), alors que les paires s'opposant sur la labialité ne sont pas différentes des deux autres (2,42).

Enfin, nous avons réalisé des ANOVA à deux facteurs (4 régions d'intérêt x 3 syllabes) pour chaque hémisphère.

Pour l'hémisphère gauche, un effet principal des régions d'intérêt est constaté, en modalité AV fixe ($F(3,75) = 5.66$ $p < 0.002$), et en AV animée ($F(3,75) = 4.36$ $p < 0.007$). Les analyses *a posteriori* montrent que pour, les deux modalités, les pics d'activation sont supérieurs dans les régions d'intérêt « AB 21 » et « AB 22 ».

Nous observons aussi un effet principal des syllabes, en AV fixe ($F(2,50) = 3.44$ $p < 0.041$) et en AV animée ($F(2,50) = 3.55$ $p < 0.037$), reposant sur une différence entre les syllabes qui s'opposent sur les lieux d'articulation significativement plus activées que celles s'opposant sur le voisement et que celles s'opposant sur la labialité.

Pour l'hémisphère droit, un effet principal des régions d'intérêt est constaté, en modalité AV fixe ($F(3,75) = 11.95$ $p < 0.000002$), et en AV animée ($F(3,75) = 10.43$ $p < 0.000008$). La encore, cet effet repose sur une différence entre les régions d'intérêt « AB 21 » et « AB 22 », dont les niveaux d'activation sont statistiquement supérieurs à ceux des régions d'intérêt « AB 41 » (2.21) et « AB 42 ».

En modalité AV fixe, un effet principal des syllabes ($F(2,50) = 7.66$ $p < 0.002$) repose sur une différence entre les syllabes s'opposant sur les lieux d'articulation (3.05), dont les pics d'activation sont supérieurs aux deux autres paires syllabiques (Labialité : 2.46 = Voisement : 2.63).

En modalité AV animée, nous observons une interaction ($F(6,150) = 3.27$ $p < 0.005$), tous les niveaux d'activation des paires syllabiques sont supérieurs dans les régions « AB 21 et 22 ».

B.3.6. Synthèse

Nous reprenons maintenant les résultats anatomo-fonctionnels de la discrimination syllabique, en fonction des deux modalités AV fixe et AV animée.

Les analyses, comparant la perception des paires syllabiques par rapport aux événements nuls, révèlent que les pics maximaux activations sont observés dans le gyrus de Heschl de l'hémisphère gauche, dans les deux modalités AV fixe et AV animée. Ceux-ci sont circonscrits, en modalité AV fixe, dans l'aire de Brodmann 42, exception faite des paires s'opposant sur les lieux d'articulation, qui activent uniquement l'aire de Brodmann 41. En modalité AV animée, seule l'aire de Brodmann 41 présente un pic d'activation pour les paires s'opposant sur la labialité, contrairement aux paires s'opposant sur les lieux d'articulation et sur le voisement, dont les pics d'activations sont répartis sur les aires de Brodmann 41 et 42. Pour les paires non phonologiques, les pics sont concentrés dans l'aire de Brodmann 42. Dans les deux modalités, nos résultats montrent que l'aire de Brodmann 22 dans le Gyrus Temporal Supérieur de l'hémisphère gauche est, en partie, activée pour tous les types de paires.

En ce qui concerne la comparaison entre les deux modalités, pour les trois paires syllabiques, nous pouvons constater que la soustraction des activations obtenues en AV fixe par rapport à celles obtenues en AV animée révèle un groupe de voxels, situé dans le cortex temporo-occipital, à la limite des aires de Brodmann 19 et 37, et ce dans les deux hémisphères. Néanmoins, l'étendue de ces activations diffère, étant plus large dans l'hémisphère droit. De plus, pour les paires s'opposant sur la labialité, la soustraction des activations en AV fixe, par rapport à celles obtenues en AV animée, révèle deux activations pariétales dans l'hémisphère droit, l'une dans le gyrus angulaire (Aire de Brodmann 39) et l'autre dans le lobule pariétal (Aire de Brodmann 40).

Finalement, la comparaison entre les paires syllabiques et les paires non phonologiques fait aussi apparaître des activations, quoiqu'à un seuil statistique plus faible.

La soustraction des activations entre les paires s'opposant sur la labialité et les paires non phonologiques révèle des activations frontales, en modalité AV fixe, et des activations pariétales, en modalité AV animée. Des activations significatives de l'opercule rolandique sont présentes dans l'hémisphère gauche, pour les paires s'opposant sur la labialité et sur les lieux d'articulation, et dans l'hémisphère droit, pour les paires s'opposant sur le voisement.

Les tableaux suivants sont des récapitulatifs des ANOVA effectuées sur les régions d'intérêt.

Tableau B37 : Résultats des ANOVA à trois facteurs (modalités : 2 niveaux ; hémisphères : 2 niveaux ; Régions d'intérêt : 4 niveaux).

	Labialité	Voisement	Lieux d'articulation
Modalités	Non significatif	Non significatif	Non significatif
Hémisphères	Non significatif	Non significatif	Non significatif
Régions d'intérêt	$AB_{21} = AB_{22}$ $>$ $AB_{41} = AB_{42}$ $F(3,75) = 7.67$ $p < 0.0002$	$AB_{21} = AB_{22}$ $>$ $AB_{41} = AB_{42}$ $F(3,75) = 12.44$ $p < 0.000001$	$AB_{21} = AB_{22}$ $>$ $AB_{41} = AB_{42}$ $F(3,75) = 15.55$ $p < 0.000001$

Tableau B38: Résultats des ANOVA à trois facteurs (modalités : 2 niveaux ; hémisphères : 2 niveaux ; Paires syllabiques : 3 niveaux). « LdA » : Paires s'opposant sur les Lieux d'Articulation. « VOT » : Paires s'opposant sur le voisement. « Lab » : Paires s'opposant sur la labialité.

	Région d'intérêt AB 21	Région d'intérêt AB 22	Région d'intérêt AB 41	Région d'intérêt AB 42
Modalités	Non significatif	Non significatif	Non significatif	Non significatif
Hémisphères	Non significatif	Non significatif	Non significatif	Non significatif
Paires	LdA > VOT > Lab F(2,50) = 10.30 p < 0.0002	LdA > VOT = Lab F(2,50) = 8.15 p < 0.0009	LdA > VOT LdA = Lab VOT = Lab F(2,50) = 3,05 p < 0,04	Non significatif F(2,50) = 2,93 p < 0.07

Tableau B39 : Résultats des ANOVA à deux facteurs, pour nous deux modalités (Régions d'intérêt : 4 niveaux ; Paires syllabiques : 3 niveaux).

	AV Fixe		AV Animée	
	HÉMISPHERE GAUCHE	HÉMISPHERE DROIT	HÉMISPHERE GAUCHE	HÉMISPHERE DROIT
Région d'intérêt	AB 21 = AB 22 > AB 41 = AB 42 F(3,75) = 5.66 p < 0.002	AB 21 = AB 22 > AB 41 = AB 42 F(3,75) = 11.95 p < 0.000002	AB 21 = AB 22 > AB 41 = AB 42 F(3,75) = 4.36 p < 0.007	AB 21 = AB 22 > AB 41 = AB 42 F(3,75) = 10.43 p < 0.000008
Paires	Lda > VOT = LAB F(2,50) = 3.44 p < 0.041	Lda > VOT = LAB F(2,50) = 7.66 p < 0.002	Lda > VOT = LAB F(2,50) = 3.55 p < 0.037	Non significatif F(2,50) = 0.16 p < 0.9

B.4. Résultats : Potentiels évoqués

Nous l'avons signalé précédemment (*cf.* B.1.3.1.), l'environnement IRMf est loin d'être idéal pour l'enregistrement de potentiels évoqués. Les variations moyennes de l'activité électrique sont de quelques microvolts avec des écarts-types élevés.

Avant d'aborder la présentation des résultats électrophysiologiques, nous exposons, brièvement, les données comportementales du groupe de 11 sujets. Rappelons que, parmi les 26 sujets ayant pris part à l'acquisition IRMf/EEG, seules les données électrophysiologiques de 11 d'entre eux ont pu être considérées dans l'analyse en potentiels évoqués (*cf.* B.1.1.).

Après avoir présenté les grandes moyennes des réponses évoquées de l'ensemble des 19 électrodes utilisées pour l'enregistrement des potentiels évoqués, nous exposerons les résultats des analyses de variance portant sur les potentiels évoqués recueillis sur les 7 électrodes présentant une variation significative des réponses évoquées en fonction de la modalité de présentation des paires syllabiques. Nous terminerons par les analyses de variance que nous avons effectuées sur les deux fenêtres temporelles.

B.4.1. Résultats comportementaux du groupe de onze sujets « EEG »

L'analyse de variance à deux facteurs (2 modalités x 3 paires syllabiques) révèle un effet principal des paires syllabiques ($F(2,20) = 5,92$ $p < 0,01$) sur les performances de discrimination, ainsi qu'une interaction entre les deux facteurs ($F(2,20) = 8,34$ $p < 0,003$). Les analyses *a posteriori* montrent que l'effet principal repose sur une différence entre les paires s'opposant sur la labialité (92,8 %), qui présentent des scores significativement plus faibles que ceux observés pour les paires s'opposant sur le voisement (96,7 %) et les lieux d'articulation (96,6 %).

L'analyse de l'interaction nous apprend que cet effet est dû aux paires s'opposant sur la labialité en AV fixe (88,6 %), elles sont significativement moins bien discriminées que toutes les autres (voir figure B28).

L'analyse portant sur les temps de réponse révèle que seule l'interaction est significative ($F(2,20) = 4,98$ $p < 0,02$). Celle-ci repose sur une différence significative des temps de réponse correctes, en fonction de la modalité pour les paires s'opposant sur les lieux d'articulation (AVf : 822 ms > AVa : 766 ms).

Tableau B40 : Moyennes et écarts-types des performances et temps de réponse de discrimination des 11 sujets.

AV fixe	Labialité		Lieu d'articulation		Voisement	
	M	E-T	M	E-T	M	E-T
Score en %	88,6	8,5	95,7	3,4	97,7	2,4
Temps de réponse en ms	813	140	822	213	779	147
AV animé	Labialité		Lieu d'articulation		Voisement	
	M	E-T	M	E-T	M	E-T
Score en %	97,1	3,1	97,1	4,3	95,7	5,8
Temps de réponse en ms	782	204	766	172	806	203

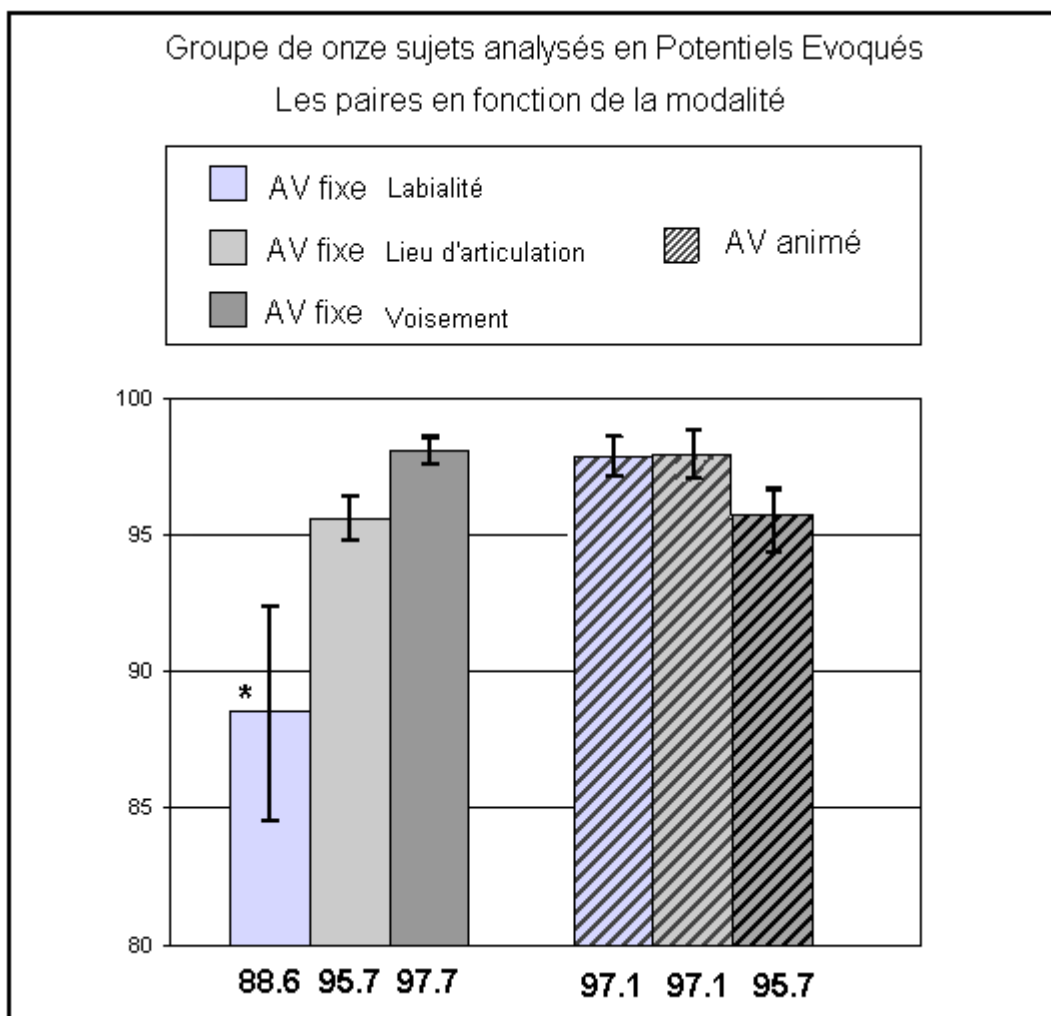


Figure B28 : Scores de discrimination en pourcentage pour les 11 sujets EEG

B.4.2. Les Grandes Moyennes issues des potentiels évoqués individuels

Les courbes électrophysiologiques sont présentées ici avec les différences de potentiels positives vers le haut et négatives vers la bas, afin de rendre la lecture des potentiels évoqués plus intuitive.

Nous débutons la présentation par les Grandes Moyennes des courbes de réponses évoquées ; viennent ensuite les cartographies (voir figure B29), montrant la répartition des potentiels évoqués, pour la moyenne des trois paires syllabiques, en fonction des deux modalités, AV fixe et AV animée.

L'analyse visuelle des courbes de réponses évoquées, entre 0 ms et 400 ms suivant la présentation de la seconde syllabe des paires à discriminer, nous permet de constater des

différences d'amplitude entre les modalités AV fixe et AV animée. Alors qu'entre 400 ms et 600 ms, les courbes sont relativement superposées pour les électrodes Fz, F4, FC4, F8, P3, Pz, P4, et en elles le sont aussi pour T3, T5, CP5 et CP6.

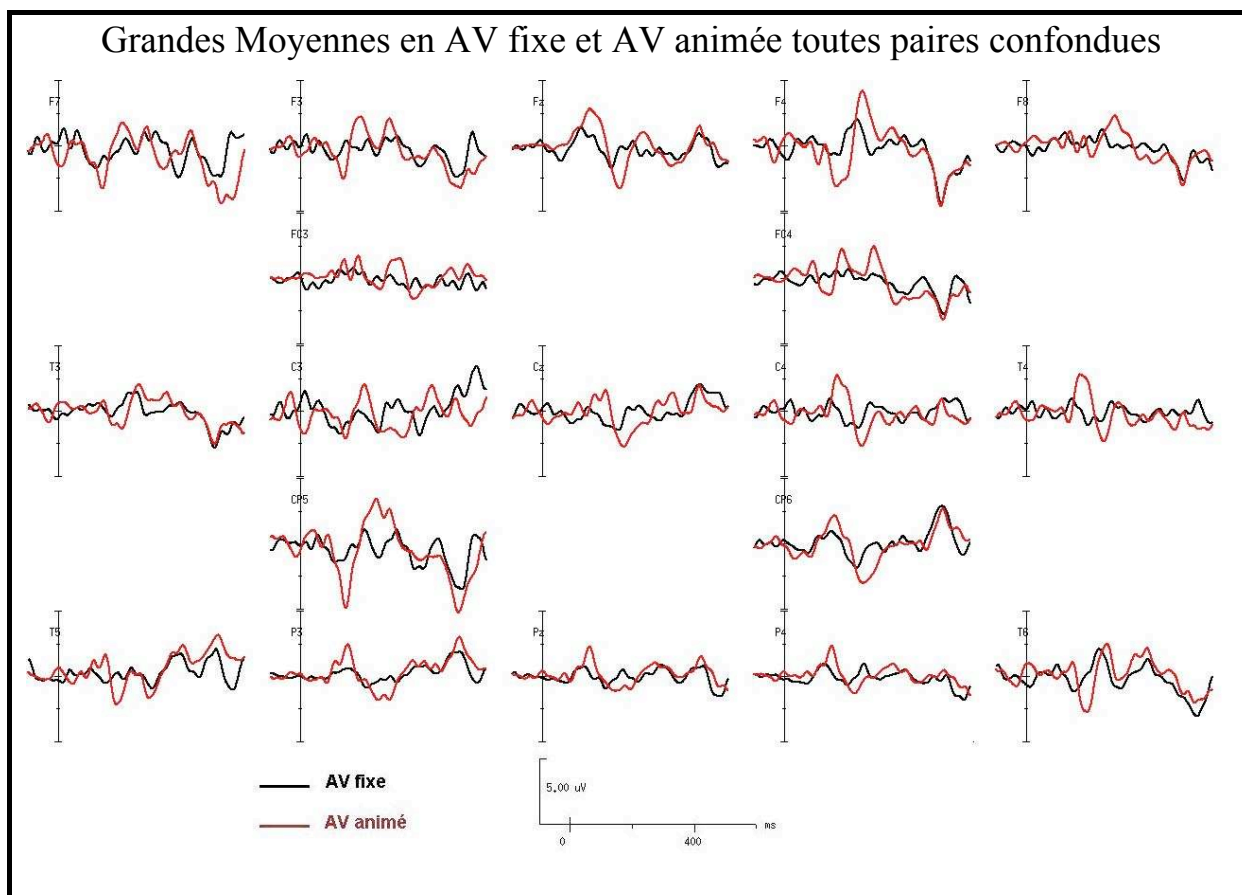


Figure B29 : Grandes Moyennes des potentiels évoqués par la discrimination syllabique, en modalités AVf et AVa, pour les trois paires phonologiques confondues.

Les cartographies d'amplitude rendent compte de l'évolution dans le temps de la distribution sur le scalp et de l'amplitude des ondes positives et négatives. Les cartes de la figure B30 ont été normalisées à $2,32 \mu\text{V}$, correspondant à la valeur maximale relevée dans les deux modalités, pour les trois catégories (types) de paires syllabiques.

Ces cartographies nous permettent d'observer que la répartition des différences de potentiel est relativement similaire pour les deux modalités ; les différences concernent principalement l'amplitude de ces ondes.

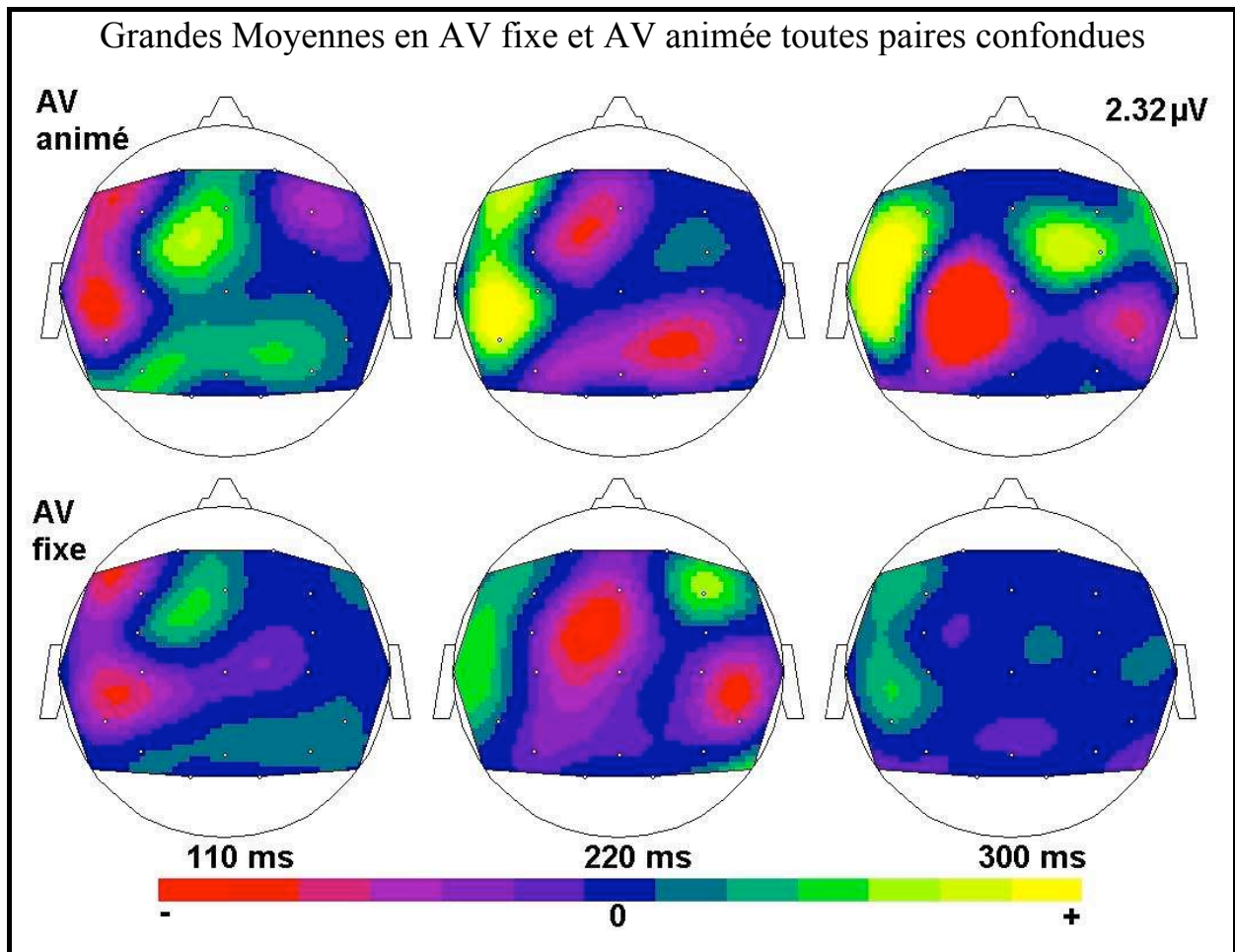


Figure B30 : Les cartes d'amplitude obtenues à 110 ms, 220 ms et 300 ms après le début de la seconde syllabe, en modalités AVf et AVa, pour les trois paires phonologiques confondues.

B.4.2.1. Les potentiels évoqués en modalité AV fixe

La Figure B31 représente les Grandes Moyennes des potentiels évoqués par la discrimination des trois paires syllabiques (en rouge : labialité ; en bleu : lieux d'articulation ; en noir : voisement) enregistrés sur les 19 électrodes.

L'analyse visuelle permet de constater que les différences entre les courbes, issues de la discrimination des paires s'opposant sur les lieux d'articulation et sur le voisement, sont minimales. Seule la courbe des potentiels évoqués par la discrimination de la labialité semble se différencier, sur certaines électrodes, des deux autres paires.

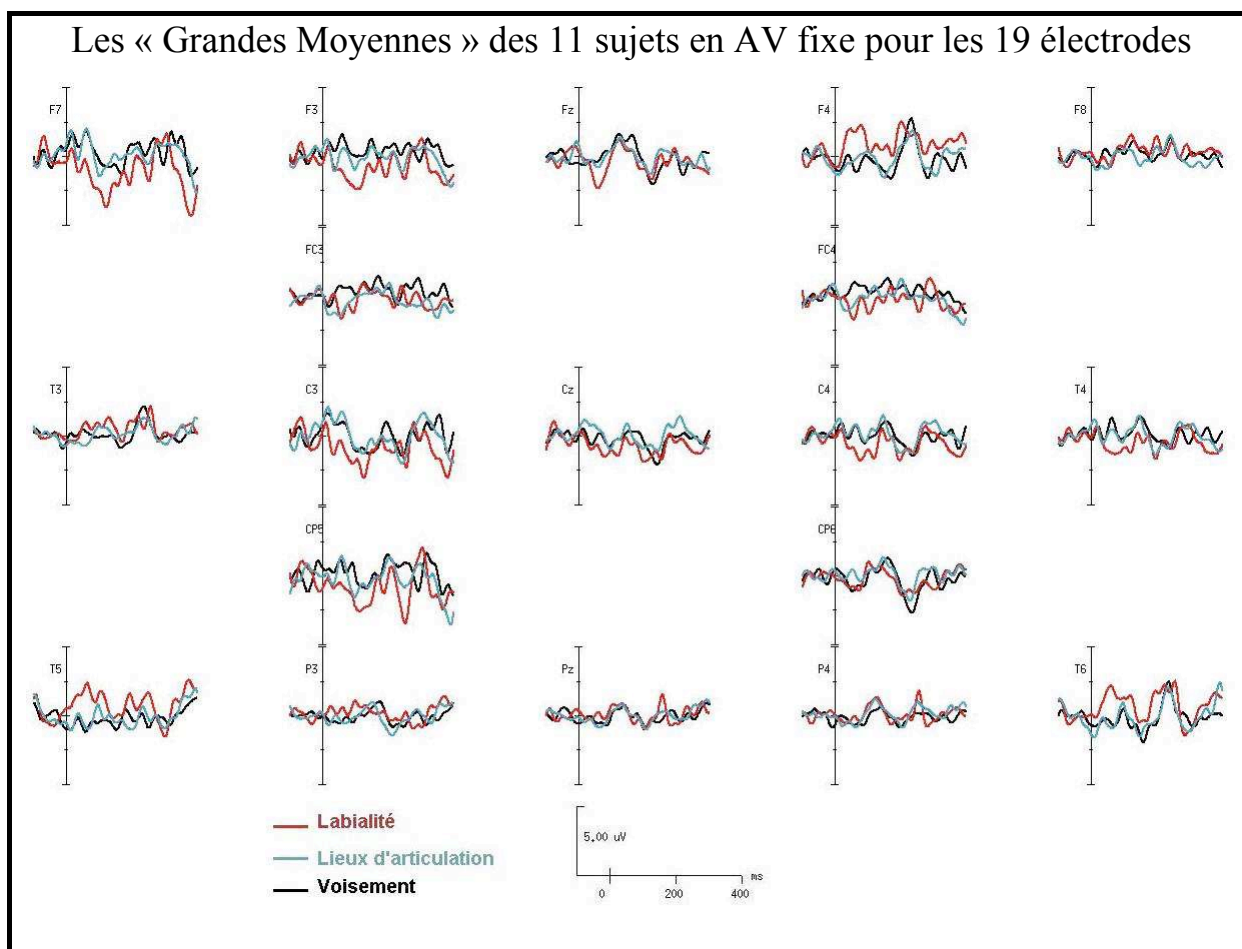


Figure B31 : Grandes moyennes pour les 11 sujets, sur les 19 électrodes, pour les trois paires phonologiques en modalité AV fixe.

B.4.2.2. Les potentiels évoqués en modalité AV animée

Les Grandes Moyennes des réponses évoquées par la discrimination des trois paires phonologiques sur les 19 électrodes sont présentées sur la figure B32.

En comparaison des potentiels évoqués obtenus lors de la présentation AV fixe, nous observons que les trois courbes, issues de la discrimination des différentes paires syllabiques, sont similaires sur la majorité des électrodes. Nous pouvons aussi constater que de nombreuses électrodes présentent des variations avant même le début de la seconde syllabe, sur laquelle porte la discrimination.

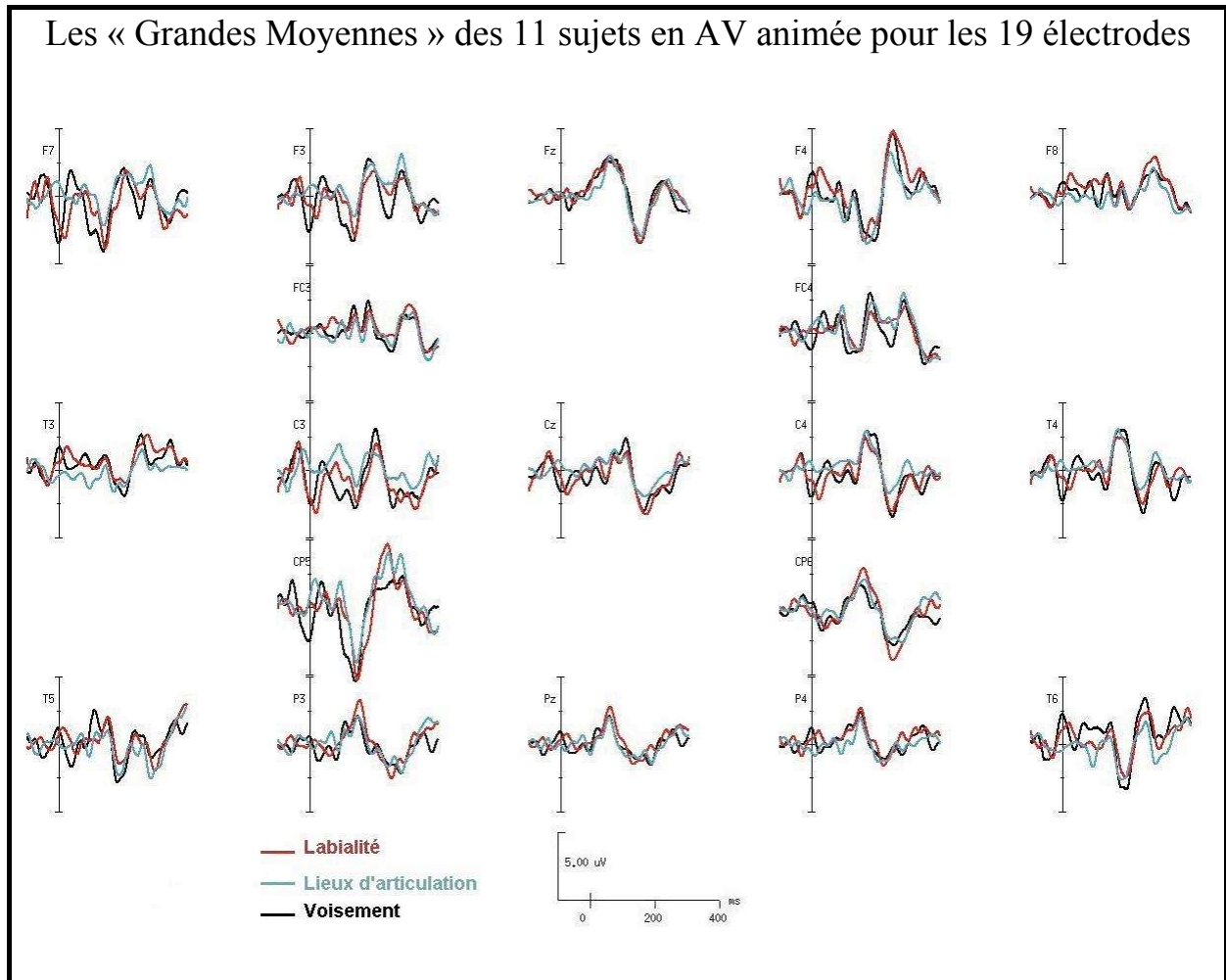


Figure B32 : Grandes moyennes, pour les 11 sujets, sur les 19 électrodes, pour les trois paires phonologiques en modalité AVa.

L'analyse comparative des réponses évoquées à la discrimination syllabique, toutes paires confondues, obtenues sur les 19 électrodes a permis d'identifier un ensemble de sept électrodes présentant une variation significative d'amplitude des réponses évoquées en fonction de la modalité de présentation AV fixe ou AV animée. Ces électrodes situées sur le vertex (Fz), en regard du cortex centro-pariétal gauche (CP5) et droit (CP6) et des régions temporales gauches (T3, T5) et droites (T4, T6) ont été retenues pour les analyses statistiques ultérieures.

B.4.2.2.1. Discrimination de la labialité

La figure B33 présente la moyenne des potentiels évoqués à la discrimination de la labialité chez les 11 sujets dans la modalité AV fixe et AV animée. À l'analyse visuelle, les électrodes Fz, CP5, CP6 et T4 semblent présenter des amplitudes supérieures en modalité AV animée par rapport à la modalité AV fixe. Nous pouvons aussi constater que les courbes des électrodes Fz, CP6 et T4 présentent une polarité opposée à celle de CP5, en modalité AV animée.

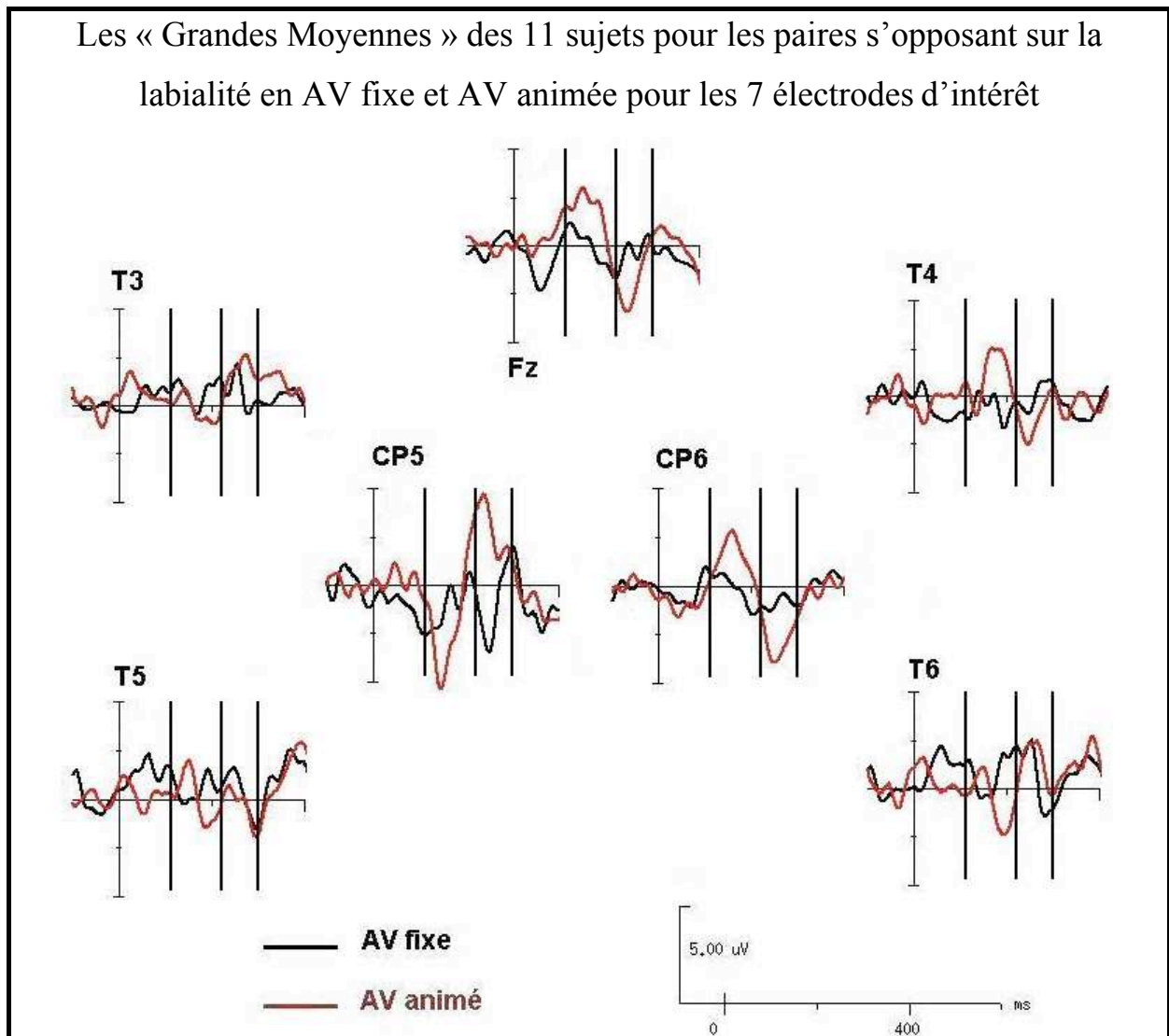


Figure B33 : Grandes moyennes des potentiels évoqués par la discrimination du trait de labialité chez les 11 sujets en fonction des modalités AVf et AVa.

B.4.2.2.2. Discrimination des lieux d'articulation

Les courbes de réponses évoquées par la discrimination des paires s'opposant sur les lieux d'articulation, présentées dans la figure B34, se caractérisent elles aussi par l'augmentation d'amplitude des réponses autour de 150 ms sur les électrodes Fz, CP5, CP6 et T4 en modalité AV animée, par rapport à la modalité AV fixe. L'analyse visuelle nous permet aussi de constater que les courbes des électrodes Fz, CP6 et T4 ont une polarité opposée à celle de CP5, en modalité AV animée.

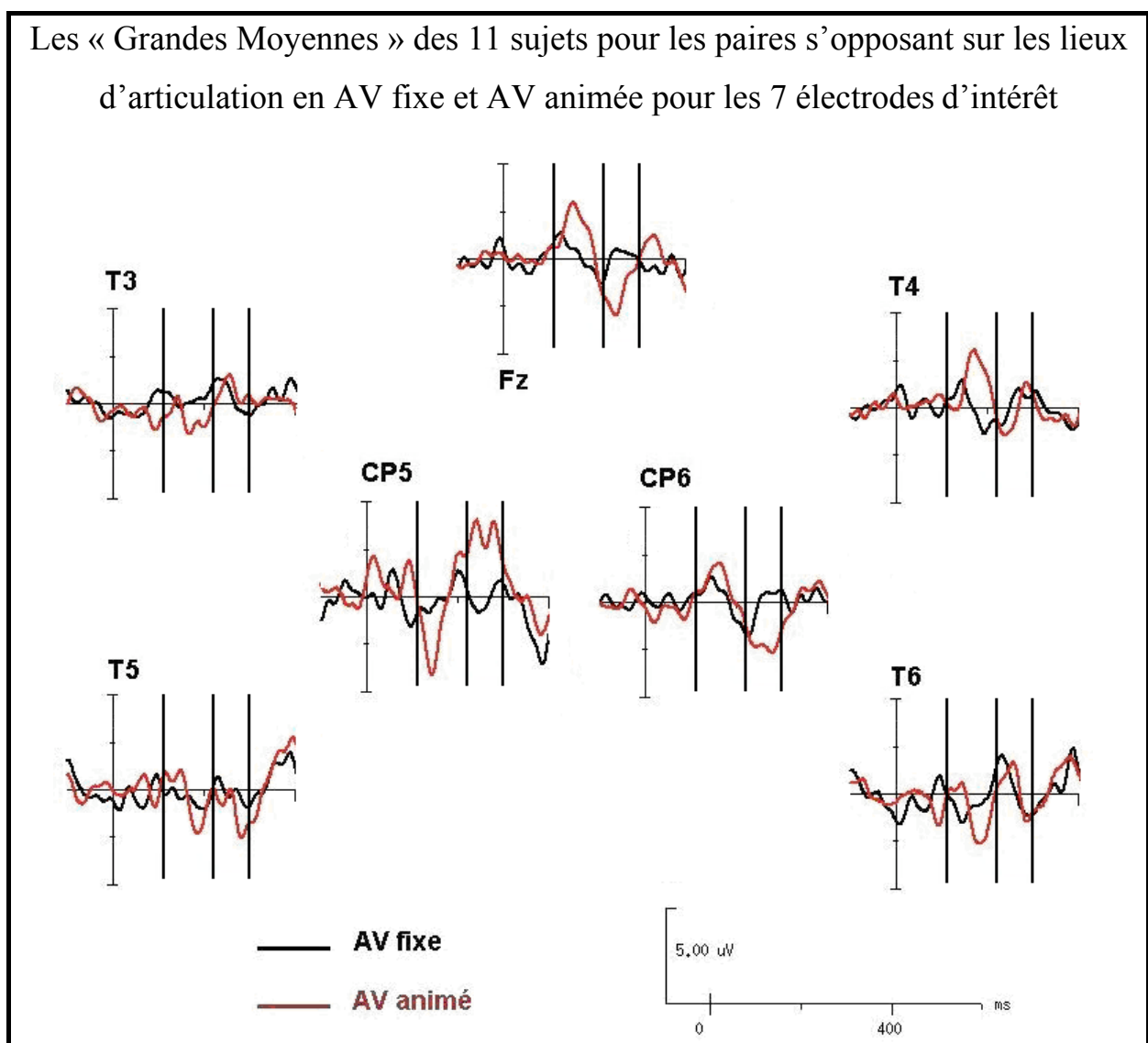


Figure B34 : Grandes moyennes des potentiels évoqués par la discrimination des lieux d'articulation chez des 11 sujets en fonction des modalités AVf et AVa.

B.4.2.2.3. Discrimination du voisement

Nous abordons, pour finir, les courbes de réponses évoquées par la discrimination des paires s'opposant sur le voisement. Les électrodes Fz, CP5 et T4, mais aussi T6 semblent présenter des amplitudes supérieures en modalité AV animée par rapport à la modalité AV fixe. Avant 200 ms, l'analyse visuelle nous permet aussi de constater que les courbes des électrodes Fz, CP6 et T4 met en lumière une polarité opposée à celle de CP5, en modalité AV animée. Les différences, après 200 ms, entre les amplitudes des réponses évoquées par les deux modalités sont moins marquées que pour les paires de la labialité et de lieux d'articulation. Voir figure B35.

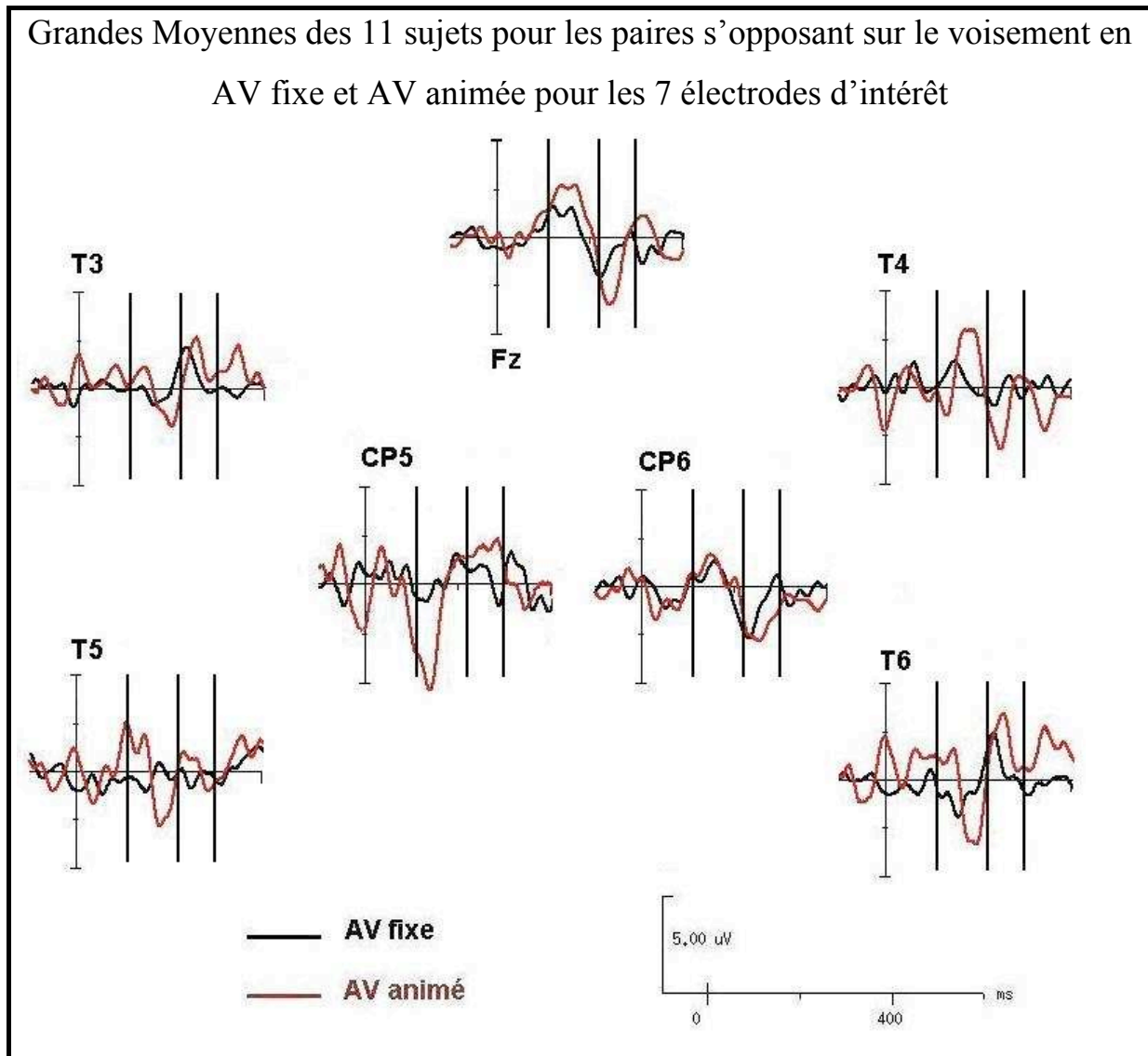


Figure B35 : Grandes moyennes des potentiels évoqués par la discrimination du voisement chez les 11 sujets en fonction des modalités AVf et AVa.

L'amplitude des réponses obtenues sur les électrodes Fz, CP5 et T4, mais aussi T6 semble supérieure en modalité AV animée par rapport à la modalité AV fixe. Avant 200 ms, l'analyse visuelle nous permet aussi de constater que les courbes des électrodes Fz, CP6 et T4 met en lumière une polarité opposée à celle de CP5, en modalité AV animée. Les différences, après 200 ms, entre les amplitudes des réponses évoquées par les deux modalités sont moins marquées que pour les paires de la labialité et de lieux d'articulation.

B.4.3. Résultats des analyses statistiques sur les potentiels évoqués

Nous avons effectué des analyses statistiques sur les surfaces des courbes de réponses évoquées par la discrimination syllabique à partir des deux fenêtres (110 ms à 220 ms ; 220 ms à 300 ms ; cf. B.1.4.4.4.).

B.4.3.1. Analyses de la fenêtre de 110 ms à 220 ms

L'ANOVA à trois facteurs (2 modalités x 3 paires syllabiques x 7 électrodes) ne montre qu'un effet principal des électrodes ($F(6,60) = 8.53$ $p < 0.000001$) accompagné d'une interaction (électrodes x modalités ; $F(6,60) = 7.88$ $p < 0.000003$) sur la surface des courbes, dans la fenêtre allant de 110 ms à 220 ms. Les analyses post-hoc (*LSD*) indiquent que l'effet principal repose sur une différence significative entre les électrodes Fz, T4 et CP6 et les électrodes CP5 et T6.

L'interaction est due à un effet des électrodes pour la modalité AV animée par rapport à la modalité AV fixe.

L'ANOVA à deux facteurs (2 modalités x 3 paires syllabiques), conduite sur chacune des sept électrodes d'intérêt, montre un effet principal de la modalité sur quatre électrodes (Fz : ($F(1,10) = 32.96$, $p < 0.0002$), CP5 : ($F(1,10) = 9.03$, $p < 0.02$, CP6 : ($F(1,10) = 8.93$, $p < 0.013$ et T4 : ($F(1,10) = 9.03$, $p < 0.02$). Cet effet principal est dû à une réponse plus ample dans la modalité AV animée.

Un effet principal des paires syllabiques n'est observé que pour l'électrode Fz ($F(2,20) = 5.07$, $p < 0.02$) reposant sur une différence significative des paires s'opposant sur les lieux

d'articulation, qui présentent des réponses moins amples que les paires s'opposant sur la labialité et celles s'opposant sur le voisement.

B.4.3.1.1. Analyses de chaque paire syllabique

Les ANOVA à deux facteurs (2 modalités x 7 électrodes), conduites sur chacune des paires syllabiques, révèlent un effet principal des électrodes (paires s'opposant sur la labialité : $F(6,60) = 4,37$ $p < 0,002$; sur le voisement : $F(6,60) = 6,65$ $p < 0,00002$; sur les lieux d'articulation : $F(6,60) = 2,92$ $p < 0,02$), ainsi que des interactions entre les deux facteurs (modalités et électrodes) que nous allons détailler dans ce qui suit.

Pour la labialité, les analyses post-hoc de l'interaction « modalités x électrodes » ($F(6,60) = 4,36$ $p < 0,002$) montrent des différences significatives pour CP5, CP6, T4 et Fz. Notons que les surfaces des courbes calculées sur l'électrode CP5, en modalité AV animée, sont strictement différentes de toutes les électrodes, hormis T6.

Pour les lieux d'articulation, les analyses post-hoc de l'interaction « modalités x électrodes » ($F(6,60) = 2,68$ $p < 0,03$) indiquent que seule l'électrode T4 présente une différence significative, en fonction de la modalité (AVa > AVf).

Pour le voisement, les analyses post-hoc de l'interaction « modalités x électrodes » ($F(6,60) = 5,23$ $p < 0,0003$) mettent en lumière des différences significatives pour CP5, pour T6, pour T4 et pour Fz. Notons que les surfaces des courbes calculées sur les électrodes Fz et CP5 en modalité AV animée sont strictement différentes l'une de l'autre mais aussi de toutes les électrodes en AV fixe.

B.4.3.2. Analyses de la fenêtre de 220 ms à 300 ms

L'ANOVA à trois facteurs (2 modalités x 3 paires syllabiques x 7 électrodes), sur la surface des courbes montre un effet principal de la modalité ($F(1,10) = 7.08$, $p < 0.03$) et des électrodes ($F(6,60) = 15.61$, $p < 0.0000001$). L'effet de la modalité repose sur une réponse significativement plus ample en modalité AV animée que AV fixe. L'effet principal des électrodes repose sur une différence significative entre les électrodes CP5, T3 et T6 et les électrodes CP6, Fz et T4, liée à une inversion de polarité entre ces deux groupes d'électrodes.

Une interaction significative est observée entre les facteurs « modalités » et « électrodes » ($F(6,60) = 7.97$, $p < 0.000001$) ; elle repose sur une différence entre les électrodes CP5, T3 et T6, par rapport aux électrodes CP6, Fz et T4 en AV animée. Signalons

que les électrodes T3 et T6, en AV fixe, ne sont pas statistiquement différentes des autres électrodes.

Une interaction significative est observée entre les facteurs « modalités » et « paires syllabiques » ($F(2,20) = 4.66, p < 0.021$). L'analyse post-hoc révèle une différence significative entre nos deux modalités pour les paires s'opposant sur les lieux d'articulation.

L'ANOVA à deux facteurs (2 modalités x 3 paires syllabiques), conduite sur chacune des sept électrodes d'intérêt, montre un effet principal de la modalité sur trois électrodes (Fz : ($F(1,10) = 19,43 p < 0.002$), CP5 : ($F(1,10) = 12,2 p < 0.006$, CP6 : ($F(1,10) = 15,15 p < 0.03$). Comme dans la fenêtre précédente, cet effet principal est dû à une réponse plus ample dans la modalité AV animée.

L'ANOVA à un seul facteur (paires syllabiques) pour chacune des électrodes ne montre d'effet principal que pour l'électrode T6 ($F(2,20) = 3.64, p < 0.044$) reposant sur les paires s'opposant sur la labialité significativement moins ample que les deux autres paires syllabiques (point d'articulation et voisement).

B.4.3.2.1. Analyses de chaque paire syllabique

Les ANOVA à deux facteurs (2 modalités x 7 électrodes) conduites sur chacune des paires syllabiques révèlent un effet principal des électrodes (paires s'opposant sur la labialité : $F(6,60) = 8,08 p < 0,000002$; sur le voisement : $F(6,60) = 10,57 p < 0,000001$; sur les lieux d'articulation : $F(6,60) = 9,70 p < 0,000001$), ainsi que des interactions entre les deux facteurs (modalités et électrodes) que nous allons détailler ci-dessous.

Pour la labialité, les analyses *post-hoc* de l'interaction « modalités x électrodes » ($F(6,60) = 6,36 p < 0,00004$) mettent en lumière des différences significatives pour CP5, CP6, T4 et Fz.

Notons que les surfaces des courbes, calculées sur l'électrode CP5 en modalité AV animée, sont strictement différentes de toutes les électrodes, hormis T3 en AV animée ; les courbes calculées sur l'électrode CP6, en modalité AV animée, sont strictement différentes de toutes les électrodes, hormis Fz en AV animée.

Pour les lieux d'articulation, les analyses post-hoc de l'interaction « modalités x électrodes » ($F(6,60) = 4,71 p < 0,0006$) indiquent des différences significatives pour CP5, CP6, Fz et T5.

Pour le voisement, les analyses post-hoc de l'interaction « modalités x électrodes » ($F(6,60) = 2,41$ $p < 0,04$) mettent en lumière des différences significatives pour CP5.

B.4.4. Synthèse

L'exploitation de potentiels évoqués enregistrés dans l'IRMf n'est pas chose aisée, nous avons choisi de nous focaliser sur les différences d'amplitude des courbes évoquées par la discrimination syllabique de chacune des deux modalités AV fixe et AV animée. Les deux fenêtres temporelles d'analyse, la première (de 110 ms à 220 ms) et la seconde (de 220 ms à 300 ms), présentent des modulations significatives en fonction de la modalité (AVf et AVa), et en fonction des électrodes d'intérêt.

Dans la fenêtre allant de 110 ms à 220 ms, les surfaces des courbes de potentiels évoqués par la discrimination syllabique sont statistiquement plus élevées en modalité AV animée, par rapport à l'AV fixe, pour quatre électrodes (Fz : vertex frontal, CP5 : centro-pariétale gauche, CP6 : centro-pariétale droite, T4 : temporale droite).

Les surfaces des courbes de la discrimination sont supérieures en AV animée, pour les paires s'opposant sur la labialité sur les électrodes Fz, CP5, CP6 et T4, et pour les paires s'opposant sur le voisement sur les électrodes Fz, CP5, T4 et T6. Alors que pour les paires s'opposant sur les lieux d'articulation, seule la surface de la courbe calculée sur l'électrode T4 est supérieure en AV animée par rapport à l'AV fixe.

Dans la fenêtre allant de 220 ms à 300 ms, l'amplitude des surfaces des courbes de potentiels évoqués en modalité AV animée sont statistiquement plus élevées, pour trois électrodes (Fz : vertex frontal, CP5 : centro-pariétale gauche, CP6 : centro-pariétale droite), par rapport à la modalité AV fixe. Les surfaces des courbes de la discrimination sont supérieures en AV animée, pour les paires s'opposant sur la labialité sur les électrodes Fz, CP5, CP6 et T4, comme dans la fenêtre précédente. *A contrario*, les surfaces des courbes pour les paires s'opposant sur les lieux d'articulation sont supérieures en AV animée sur les Fz, CP5, CP6 et T5. Alors que pour les paires s'opposant sur le voisement, seule la surface de la courbe calculée sur l'électrode T4 est supérieure en AV animée par rapport à l'AV fixe.

Tableau B41 : Récapitulatif des résultats des ANOVA à deux facteurs (modalités : 2 niveaux ; électrodes : 7 niveaux), effectuées sur les valeurs absolues de surfaces des courbes de réponses évoquées.

Modalités x Électrodes	Labialité	Lieux d'articulation	Voisement
110 à 220 ms	AV f < AV a CP5 CP6 T4 Fz $F(6,60) = 4,36 p < 0,002$	AV f < AV a T4 $F(6,60) = 2,68 p < 0,03$	AV f < AV a CP5 T6 T4 Fz $F(6,60) = 5,23 p < 0,0003$
220 à 300 ms	AV f < AV a CP5 CP6 T4 Fz $F(6,60) = 6,36 p < 0,00004$	AV f < AV a CP5 CP6 T5 Fz $F(6,60) = 4,71 p < 0,0006$	AV f < AV a CP5 $F(6,60) = 2,41 p < 0,04$

C. Discussion, Conclusion et Perspectives

C.1. Discussion : *Des gestes à la parole*

L'objectif principal de notre recherche était d'étudier les mécanismes cognitifs de la perception et de l'intégration de la dimension visuelle de la parole articulée, pour la discrimination de traits phonétiques. Pour ce faire, nous avons choisi de mettre en place un paradigme d'activation en IRMf couplée à l'acquisition de signaux EEG. Les études en IRMf, consacrées à la perception ou à la discrimination audiovisuelle de la parole, ont mis en lumière, l'implication des régions visuelles et des régions temporales spécialisées dans le traitement auditif de la parole, mais aussi le rôle possible de structures frontales. Certaines études apportent des données illustrant une capacité d'analyse de la dimension visuelle, par les aires temporales classiquement impliquées dans les traitements acoustiques de la parole articulée. La discussion de nos résultats portera tout d'abord sur les activations liées à la perception du mouvement durant la production de la parole, mises en évidence par la comparaison des modalités AV animée et AV fixe. Avant de replacer ces données dans le cadre théorique que nous avons choisi, c'est-à-dire le modèle à deux voies de Hickok & Poeppel (2007), nous discuterons de l'apport des données électrophysiologiques, à l'analyse de mécanismes cérébraux qui sous-tendent la discrimination audiovisuelle de la parole. Pour finir, nous discuterons de l'apport relatif des visèmes, en fonction des contraintes expérimentales, à partir des résultats des tests comportementaux.

C.1.1. La perception des mouvements orofaciaux phonologiques

La soustraction des activations suscitées par la modalité AV animée, par rapport à la modalité AV fixe, met en évidence, dans les deux hémisphères, des activations significativement plus élevées dans la partie inférieure du cortex occipito-temporal, et localisées dans les aires de Brodmann 19 et 37. Même en appliquant un seuil statistique non corrigé, cette comparaison ne révèle pas d'activation significative dans les régions du cortex auditif primaire, et du Sillon Temporal Supérieur, associées aux gestes phonologiques orofaciaux dans les études de Calvert *et al.* (1997) et de Pekkola *et al.* (2004). Des différences dans les paradigmes d'activation et dans les méthodes d'analyse des données IRMf peuvent expliquer cette discordance. Dans l'étude de Pekkola *et al.* (2004), l'analyse en régions d'intérêt, définies par les connaissances *a priori* issues, en partie, de l'étude Calvert *et al.*

(1997) a pu faire apparaître des activations dans le cortex auditif primaire qui n'atteignent pas le seuil de signification statistique, dans une analyse en cerveau entier. De plus, comme le suggèrent Bernstein *et al.* (2002), la région correspondant au gyrus de Heschl, dans l'étude de Calvert *et al.* (1997), pourrait en réalité appartenir au cortex auditif associatif (GTS).

Quant à l'activation occipito-temporale que cette soustraction met en évidence dans les deux hémisphères (Aire de Brodmann 19 et 37), elle peut être attribuée au traitement perceptif du visage (Denkova, Botzung & Manning, 2006 ; Hall *et al.*, 2005), mais aussi à la perception du mouvement visuel (Tong, 2003 ; Smith *et al.*, 1998), voire du mouvement orofacial en général (Puce *et al.*, 1998).

De nombreuses études portant sur les mouvements orofaciaux phonologiques ont aussi rapporté des activations de ces régions temporo-occipitales (Calvert *et al.*, 1997 ; Bernstein *et al.*, 2002 ; Calvert & Campbell, 2003 ; Callan *et al.*, 2003 ; Hall *et al.*, 2005). Au vu des données de la littérature et de nos conditions expérimentales, il convient de distinguer les activations de l'aire MT/V5 de celles du gyrus fusiforme (AB 37). Si les premières peuvent être liées à la perception visuelle du mouvement, *per se*, les secondes sont plus probablement à mettre en relation avec la présentation du visage de notre locutrice, même si celui-ci n'était que partiellement visible. L'activation simultanée de ces deux régions pourrait être spécifique au traitement du mouvement orofacial.

C.1.1.1. Perception visuelle et cortex auditif primaire

En accord avec les données de Bernstein, mais contrairement à celles de Calvert *et al.* (1997) et de Pekkola *et al.* (2004), les résultats de la soustraction entre les cartes d'activation obtenues, lors de la présentation AV animée par rapport à l'AV fixe, ne montrent pas d'activation plus marquée du cortex auditif primaire (gyrus de Heschl / Aires de Brodmann 41 et 42), en présence de mouvements orofaciaux phonologiques.

L'étude menée par Pekkola *et al.* (2004) contraste un visage articulante des voyelles avec un visage fixe, sur la bouche duquel ont été superposés des cercles en mouvement, ce qui masque partiellement la configuration du visage fixe. Dans notre expérience, nous dissociions les indices visuels phonologiquement pertinents fixes et animés. Cette caractéristique, propre à notre étude, peut expliquer la discordance entre les résultats observés. Quant à l'étude de Calvert *et al.* (1997), elle ne compare pas directement les mouvements du locuteur à une configuration fixe, mais au même visage grimaçant. Cette comparaison d'un même visage dont les mouvements orofaciaux phonologiques leur permet d'isoler les activations propres

aux mouvements linguistiques, avec notamment une activation du gyrus fusiforme gauche et des Gyri Temporaux Supérieurs bilatéraux.

Une définition préalable à notre étude, des aires auditives primaires de chacun des participants, nous aurait permis une meilleure caractérisation de l'influence de l'information visuelle sur les premières étapes du traitement auditif de la parole. De même, l'absence de condition expérimentale, offrant la possibilité d'étudier la seule modalité visuelle, ne nous permet pas de conclure sur ce point.

C.1.2. Intégration audiovisuelle et le sillon temporal supérieur

Il est probable, comme le suggèrent de nombreuses études (Beauchamp *et al.*, 2003 ; Beauchamp *et al.*, 2004 ; Van Atteveldt, Formisano, Goebel & Blomert, 2004), que le Sillon Temporal Supérieur prenne part dans les mécanismes d'intégration multimodale. La question est maintenant de déterminer, d'une part, si sa contribution est spécifique à l'intégration langagière et, d'autre part, si la nature de sa contribution diffère selon l'hémisphère cérébral impliqué.

Beauchamp *et al.* (2003, 2004) observent des activations du STS, lors de la perception passive d'objets et de stimuli mouvants, et précisent que des subdivisions fonctionnelles pourraient expliquer les différents rôles du STS. Van Atteveldt *et al.* (2004), dans une étude sur l'intégration bimodale de lettres et des phonèmes, observent, en comparaison avec les événements nuls, des activations du STS pour les présentations unimodale et bimodale. Binder *et al.* (1994, 1996) observent des activations du tiers postérieur du STS, lors de la perception passive de tons, alors que dans l'étude de Rimol *et al.* (2005), ce sont les soustractions entre des consonnes et des bruits, mais aussi entre des syllabes CV et des bruits, qui activent le STS. Liebenthal *et al.* (2005) rapportent, quant à eux, des activations plus importantes dans le STS gauche, pour la discrimination de syllabes CV. Ces exemples arguent en défaveur d'une implication spécifique du STS pour le traitement de la parole audiovisuelle. *A contrario*, certaines études constatent des activations du STS en présentation visuelle seule. Wright *et al.* (2003) rapportent des activations bilatérales du STS, lors de la perception de mots associés à un visage synthétique, et présentés selon trois modalités : auditive, visuelle et audiovisuelle. La présentation auditive et audiovisuelle recrute le STS de façon bilatérale, alors que la présentation visuelle semble préférentiellement recruter le STS de l'hémisphère droit. Pour leur part, Bernstein *et al.* (2002) observent des activations du STS, lors de sessions de lecture labiale, alors que Miller & D'Esposito (2005) constatent que les activations du STS

sont moindres, lors de la présentation de syllabes audiovisuelles non congruentes, de type « McGurk », par rapport à des syllabes congruentes.

Nos données, en accord avec ces résultats, indiquent que le STS est activé dans l'hémisphère droit, par la discrimination syllabique en modalités AV aussi bien fixe qu'animée. Ces activations sont plus étendues pour la modalité AV animée. Dans l'hémisphère gauche, elles sont circonscrites au tiers postérieur du STS (établissant la jonction avec la scissure de Sylvius).

Toutefois, l'analyse comparative du niveau d'activation des régions d'intérêt englobant le STS (qui correspondent aux activations observées dans les aires de Brodmann 21 et 22) ne montrent pas de variation significative de la réponse hémodynamique à la discrimination de syllabes audiovisuelles, selon la présence ou non de mouvements, ou en fonction de l'hémisphère cérébral.

Dans notre étude, les activations associées à la discrimination des paires non phonologiques sont similaires, bien que les régions temporales recrutées, en modalité AV animée, soient moins étendues que celles observées pour les paires syllabiques. Il n'est pas exclu que ces paires non phonologiques, doublées par des mouvements articulatoires pertinents dont le décours a été inversé, aient perturbé l'intégration bimodale.

L'ensemble des résultats, que nous venons de discuter, indiquent que la modalité AV animée recrute davantage les tiers postérieurs du STS que la modalité AV fixe. Cela suggère que l'intégration multimodale dépend de la partie postérieure du STS, alors que les traitements « acoustico-phonétiques » impliqueraient davantage la partie moyenne.

C.1.3. Le timing de la perception audiovisuelle de la parole

En nous focalisant sur les différences entre les deux modalités perceptives, nous espérons préciser le décours de l'intégration du mouvement visuel, dans la discrimination syllabique, par l'étude électrophysiologique concomitante de l'acquisition IRMf. Les potentiels évoqués, recueillis dans ce contexte particulier, nous montrent, pour chacune des paires syllabiques, une modulation significative des réponses évoquées dans deux fenêtres temporelles : la fenêtre de 110 ms à 220 ms et la fenêtre de 220 ms à 300 ms.

Aux alentours de 150 ms après le début du second stimuli, nous observons une augmentation d'amplitude significative en regard de l'électrode T4 (région temporale moyenne droite) en AV animée, par rapport à l'AV fixe, quelle que soit la paire syllabique ; ainsi que sur les électrodes Fz (située sur le vertex) et CP5 (centro-pariétale gauche) pour les

paires s'opposant sur la labialité et celles s'opposant sur le voisement. Dans la seconde fenêtre, une augmentation d'amplitude significative en AV animée, par rapport à l'AV fixe, est observée sur l'électrode CP5, pour chacune des paires syllabiques.

Rappelons que ces potentiels évoqués, acquis durant l'IRMf, reflètent les processus qui sous tendent, non seulement, la perception syllabique, mais aussi la discrimination de deux exemplaires à comparer, ce qui requiert la mémoire à court terme, et cela, dans des conditions très perturbées. En raison du dessein expérimental adopté pour l'acquisition de nos données électrophysiologiques, nos résultats ne peuvent être comparés à ceux de la littérature, qui ont observé principalement des modulations de l'amplitude et de la latence des réponses N1 P2. Contrairement à l'étude de Van Wassenhove *et al.* (2005), les potentiels évoqués par la discrimination AV animée ont une amplitude supérieure à ceux observés lors de la perception AV fixe. L'étude de Ponton *et al.* (2002) rapporte un effet inverse, observant un renforcement lors des présentations audiovisuelles de l'onde, concordant avec nos propres résultats.

C.1.4. La parole audiovisuelle et les deux voies du modèle de Hickok & Poeppel

L'idée d'une implication des aires temporales de l'hémisphère droit, dans les processus de perception de la parole, n'est pas nouvelle. Corroborée par des données anatomocliniques (Metz-Lutz & Dahl, 1984), mais aussi en neuro-imagerie (pour la TEP : Mazoyer *et al.*, 1993 ; Zatorre *et al.*, 1994 ; pour l'IRMf : Binder *et al.*, 1994 ; Binder *et al.*, 2000 ; Benson *et al.*, 2006 ; Jäncke *et al.*, 2002 ; Rimol *et al.*, 2005 ; Liebenthal *et al.*, 2005), il s'agit à présent de préciser quel(s) rôle(s) elles assurent dans le traitement de la parole.

Selon le modèle de Hickok & Poeppel (2007), les parties dorsales du Gyrus Temporal Supérieur des deux hémisphères sont recrutées pour l'analyse spectrotemporelle, alors que les tiers postérieurs des deux Sillons Temporaux Supérieurs sont impliqués dans ce qu'ils nomment le « réseau phonologique ». Les cartes d'activation issues de nos analyses de groupe, en effet aléatoire sur le cerveau entier, révèlent des activations significatives du STS et du GTS. Toutefois, nous n'observons pas la répartition symétrique des activations dans l'hémisphère droit et dans l'hémisphère gauche, postulée par le modèle de Hickok & Poeppel (2007).

En effet, lors de la présentation en AV fixe, la discrimination de syllabes recrute dans l'hémisphère gauche le gyrus de Heschl, dont l'activation s'étend postérieurement vers le Gyrus Temporal Supérieur (Aire de Brodmann 22), où elle est limitée par le bord inférieur de

la scissure de Sylvius. Dans l'hémisphère droit, l'activation s'étend du Sillon Temporal Supérieur jusqu'au tiers postérieur du gyrus de Heschl. Un *pattern* d'activation similaire, quoique plus étendu, est observé lors de la présentation en AV animée. À cette différence s'ajoute une activation significative de l'opercule rolandique gauche, lors de la présentation de syllabes, contrastée avec la présentation des stimuli non phonologiques en AV animée. Cette région du cortex sensori-moteur du conduit bucco-phonatoire, dont la contribution n'est pas envisagée dans le modèle anatomo-fonctionnel du traitement auditif de la parole de Hickok & Poeppel (2007), pourrait être mise en relation avec la mobilisation de « représentations motrices » ou de « schémas moteurs », pour la discrimination de syllabes présentées en modalité AV animée. Il convient toutefois de rappeler que la comparaison avec la modalité AV fixe ne révèle pas d'activations significativement plus élevées de cette région. Par conséquent, nous ne pouvons établir une relation directe entre l'activation de cette région et la perception des mouvements de la parole.

Les régions temporales moyennes n'apparaissent pas significativement activées par la discrimination des paires syllabiques, que ce soit en modalité AV fixe ou AV animée. En revanche, la comparaison de la discrimination entre les paires syllabiques et les paires non phonologiques, en modalité AV animée, révèle des activations dans les régions temporales moyennes antérieures de l'hémisphère gauche. Selon Hickok & Poeppel (2007), ces régions feraient partie du « réseau combinatoire » de la voie ventrale, qui participe au processus d'appariement lexical. Nos résultats indiquent que la discrimination audiovisuelle de paires syllabiques active de façon statistiquement supérieure le « réseau combinatoire », par rapport aux paires non phonologiques.

C.1.5. Perception audiovisuelle et réseau des neurones miroirs

Les premières questions relatives au rôle et l'implication des aires motrices dans la perception verbale ont été soulevées par la présence d'activation de l'aire de Broca, dans des tâches n'impliquant aucune production articulée (Mazoyer *et al.*, 1993 ; Price *et al.*, 1996).

La découverte du système des neurones miroirs a multiplié les interrogations sur l'implication des régions motrices dans la perception de la parole (Rizzolatti *et al.*, 1996 ; Gallese *et al.*, 1996). Le modèle à deux voies de Hickok & Poeppel (2007) postule que l'aire de Broca et la partie supérieure de l'aire prémotrice (AB 6) font partie de la voie dorsale.

Dans notre étude, la discrimination, en modalité AV animée, des syllabes s'opposant sur le voisement et sur les lieux d'articulation est associée à l'activation significative d'une

région du cortex prémoteur (AB 6). Les coordonnées des pics observés (voisement : $x = -50$; $y = -2$; $z = 52$; lieux d'articulation : $x = -50$; $y = -4$; $z = 52$) sont à mettre en parallèle avec celles rapportées ($x = -50$; $y = -6$; $z = 47$) dans l'étude de Wilson *et al.* (2004). Ces auteurs ont constaté un recouvrement des activations liées à la perception passive de syllabes CV, avec celles liées à la production orale des mêmes syllabes. Contrairement à ces travaux, nos analyses en cerveau entier révèlent des activations du cortex prémoteur, alors que Wilson *et al.* (2004) ont procédé à une analyse en régions d'intérêt. Néanmoins, n'ayant pas mené de phase de localisation motrice, nous pouvons seulement constater la similarité de nos pics avec ceux rapportés par Wilson.

C.1.6. L'impact des indices visuels sur la discrimination auditive des traits phonétiques

Il est important de remarquer que parmi les participants ayant terminé les deux tâches de discrimination dans l'environnement IRMf/EEG, seuls 16 d'entre eux ont obtenu des scores supérieurs au hasard, pour toutes les paires syllabiques. Cela est un premier signe de l'impact des conditions environnementales, mais pose aussi un problème quant à la généralisation de nos résultats à l'ensemble de la population, mais aussi quant à l'interprétation de nos données anatomo-fonctionnelles. Le fait que les analyses statistiques portent sur un groupe inférieur à 25 personnes, dans les milieux perturbé et bruyant, nous invite à relativiser la portée de nos résultats neurophysiologiques.

L'analyse statistique des données comportementales, obtenues dans les trois environnements expérimentaux, montre que les indices visuels ont une influence relative au degré de perturbations présentes dans l'environnement. La présence de mouvements visuels contribue à la désambiguïsation des signaux acoustiques dégradés, dans l'environnement très perturbé qu'est l'IRM, améliorant les scores de discrimination toutes paires confondues. Dans l'environnement silencieux et bruyant (perturbé par le seul bruit des gradients), il n'y a pas d'apport significatif du mouvement visuel par rapport à la configuration fixe. L'influence significative de l'environnement est expliquée par les meilleures performances dans l'environnement silencieux comparées à celles de l'environnement perturbé des tests en IRMf. L'absence de différences significatives des performances obtenues dans les environnements perturbé et bruyant, témoigne du rôle prépondérant du bruit des gradients sur la discrimination syllabique. Ces données sont en accord avec l'hypothèse d'un gain progressif, fonction du degré de perturbations (Sumbly & Pollack, 1954).

Nous avons comparé les performances de discrimination de trois traits distinctifs, le premier (le voisement) ne comportait pas de mouvements pertinents visibles, alors que les deux autres (la labialité et les lieux d'articulation) appartiennent à des classes de visèmes différentes.

Dans l'environnement silencieux, toutes les paires syllabiques ont été significativement mieux discriminées que les paires non phonologiques, aussi bien en présentation AV fixe qu'en présentation AV animée. L'absence de différences significatives des performances de discrimination entre les différentes paires, hormis l'allongement des temps de réponse pour la discrimination du voisement et des lieux d'articulation en modalité AV animée, indique que, dans cet environnement, l'apport ne dépend pas de la pertinence phonologique du mouvement visuel.

Dans l'environnement perturbé, le bruit des gradients a davantage altéré la discrimination des paires s'opposant sur la labialité et des paires non phonologiques. Comme le suggère l'interaction significative entre les deux modalités et les différentes paires, la moindre augmentation des performances en présence du mouvement visuel, pour les paires non phonologiques, comparée à celle des paires syllabiques, indique que la contribution du mouvement visuel dépend de sa pertinence phonologique.

Dans l'environnement bruyant, l'interaction entre modalités AV fixe et AV animée et les différentes paires, nous apprend qu'il n'y a pas de différence significative entre les deux modalités. Exception faite des paires s'opposant sur la labialité en AV fixe qui sont significativement moins bien discriminées que toutes les autres paires syllabiques. Finalement, le mouvement visible inhérent à la labialité a permis de désambiguïser les signaux auditifs dégradés dans l'IRMf, mais aussi dans l'environnement bruyant.

Ces résultats appuient, selon nous, deux idées présentes dans la littérature. L'apport favorable des indices visuels augmente en fonction du rapport signal sur bruit, et il semble agir davantage sur les structures du système phonologique.

C.2. Conclusion et Perspectives

Les mécanismes cognitifs de la perception et de la compréhension de la parole demeurent méconnus et sujets à discussion. La source auditive, qui est prépondérante, est caractérisée par une grande variabilité, intra- et interlocuteurs ; les théories phonétiques achoppent sur la définition d'invariants, certaines contestent même leur existence. Les modèles d'accès au lexique mental font abstraction de cette variabilité, et se concentrent principalement sur le « format » des « représentations » dites « lexicales ». Les oreilles et les yeux sont des systèmes sensoriels complexes qui traitent quantité d'informations « continues » qu'ils vont « convertir » en influx nerveux. Sont-ce là les premières étapes de la segmentation en éléments discrets ? Est-ce que le système perceptif a un besoin critique d'éléments discrets pour opérer une catégorisation, puis un appariement avec une ou des représentations ?

Il apparaît que les représentations lexicales sont constituées de multiples niveaux intégrant les concepts de phonèmes et de graphèmes, de syllabes et de morphèmes, mais aussi de mots, voire de groupes de mots. Des représentations orthographique, sémantique et phonologique coexistent probablement dans le lexique mental. Il est possible de connaître un mot, dans sa forme écrite ou orale, sans en posséder le sens. De même, la représentation « orthographique » d'un mot peut être ignorée sans pour autant que cela affecte le sens ou la dimension orale de celui-ci. Le sens, quant à lui, n'est pas toujours maîtrisé, alors que l'orthographe et la prononciation seront correctes.

Nous savons que la prise en compte des informations visuelles, inhérentes à la production de la parole articulée (les visèmes), peut soutenir l'interprétation des signaux acoustiques dans un environnement bruyant (Sumbly & Pollack, 1954). Dans le même temps, nous savons que l'intégration des sources audible et visible est très précoce, et agit au niveau syllabique (McGurk & MacDonald, 1976). La question sous jacente est par conséquent de savoir si cette influence est décisive ou accessoire. Autrement dit, à quels niveaux représentationnels la dimension visuelle prend-elle place ?

Hormis les interrogations concernant la manière dont les « représentations » sont enregistrées, la principale question portant sur ce concept concerne le degré d'abstraction ; une représentation est-elle symbolique ou conserve-t-elle de nombreux détails de surface ?

Nous penchons pour la seconde solution. Le trait de voisement, par exemple, nécessite une capacité d'analyse temporelle fine, preuve en est les différences de VOT entre deux langues pour un seul phonème, ce qui pourrait plaider en faveur de l'existence de « fenêtre(s) temporelle(s) ».

Dans le domaine visuel, la distinction entre la vision fovéale et la vision périphérique nous apprend que le cerveau n'a pas besoin d'une acuité parfaite de tout le champ visuel pour en analyser les composantes. Dans le domaine de la perception de la parole articulée, le système perceptif fait abstraction du bruit ambiant en se concentrant sur les signaux linguistiques. Cela pourrait expliquer la capacité à sélectionner une source audible parmi d'autres (effet "*cocktail party*"), mais aussi une source visuelle parmi d'autres (ségrégation « fond-forme »). Lorsqu'un mot subit une altération de l'un de ses phonèmes, due au bruit ou encore à un lapsus, les effets *top-down* permettent d'accéder à la « bonne » cible, mais n'empêchent pas d'identifier ladite altération. Si comme l'avance Serniclaes, Van Heghe, Mousty, Carré & Sprenger-Charolles (2004), la dyslexie développementale, trouble du langage écrit, est provoquée par un déficit du langage oral, en l'occurrence la perception allophonique, alors il est envisageable que l'écrit ait une influence sur l'oral.

Nous pensons, comme (MacNeilage et Davies (1998), que la syllabe est une « forme naturelle », émergeant des cycles d'ingestion, dans laquelle des « cibles » (Schwartz *et al.*, 2002) vont être réalisées et vont s'enchaîner avec les syllabes précédente et subséquente. Nous envisageons le concept de « propriétés émergentes » comme une possibilité d'ajouter, au fur et à mesure des expériences linguistiques, des "*specifiers*" (Fowler, 1986, 1996) supplémentaires et améliorer les rapports de « rigidité » entre les « gestes » (Browman & Goldstein, 1990).

Suivant le modèle de perception à logique floue de Massaro (1990), mais aussi s'appuyant sur l'effet « McGurk » (McGurk & MacDonalds, 1976), l'intégration des différentes sources est incoercible, et pourrait prendre place dans une « diagonale idéale » (Lindblom, 1987) de négociation entre le locuteur et l'auditeur.

Une « banque d'analyseurs syllabiques » (modèle SARAH : Mehler *et al.*, 1990) pourrait utiliser la fenêtre d'analyse temporelle précoce (Hickok & Poeppel, 2007), afin de réduire le nombre de « représentations » en concurrence. Le point d'unicité (COHORTE : Marslen-Wilson *et al.* 1978, 1987) pourrait s'appliquer dans ce « cadre syllabique ».

En admettant que le cortex auditif primaire de l'hémisphère gauche soit impliqué dans la perception acoustique, mais aussi dans la perception phonétique, il est envisageable que le processus de perception catégorielle entre en jeu, afin de compléter une perception continue des indices suprasegmentaux. Le Sillon Temporal Supérieur pourrait réaliser un appariement entre des « représentations phonologiques multimodales » et les différentes sources perceptives. Le concept de « représentation » peut fort bien recouvrir un réseau de régions corticales, dont chacune prend une part dans la mise en mémoire et dans la restitution d'un ou de plusieurs invariants, afin de produire, mais aussi de percevoir la parole. La découverte des neurones miroirs (Rizzolatti, 1996 ; Gallese, 1996) nous indique que le cerveau dispose de mécanismes neuronaux, susceptibles de mettre à profit toutes les dimensions perceptives. Contrairement aux idées avancées par la Théorie Motrice de Liberman & Mattingly (1985), nous pensons que l'accès à la dimension motrice des « représentations phonologiques multimodales » ne serait qu'un moyen subsidiaire utilisé pour optimiser la perception de la parole. Même s'il est vrai que la dimension visuelle engage les composantes motrices de la parole, la dimension auditive demeure prépondérante.

Il est important de noter que parmi les 26 personnes ayant terminé les deux tâches de discrimination dans l'environnement IRMf/EEG, seules 16 d'entre elles ont obtenu des scores supérieurs au hasard, pour toutes les paires syllabiques. Cela nous invite à la prudence, quant à l'interprétation de nos données anatomo-fonctionnelles. De plus, si les sciences expérimentales progressent en considérant que, lors de l'observation d'une variable d'intérêt, toutes choses sont égales par ailleurs, rien n'est moins sûr dans le domaine cognitif. L'organisation cérébrale individuelle est sujette à des influences, dont nous ignorons la plupart des effets.

Les perspectives, à court terme, sont de continuer l'analyse de nos données expérimentales, en particulier celle des potentiels évoqués pour lesquels nous avons éprouvé de fortes difficultés, du fait de l'absence presque complète de point de comparaison. Nous aimerions aussi pouvoir développer une corrélation statistique des données acquises en IRMf et en EEG. Nos tentatives, en ce sens, ce sont jusqu'à maintenant toutes soldées par un échec.

À moyen terme, nous souhaiterions avoir la possibilité de mener une nouvelle acquisition simultanée, en utilisant des mots monosyllabiques en condition d'intérêt, et la

parole dite « délexicalisée » en condition de contrôle. Les mots offrent la possibilité d'étudier l'accès au lexique mental, alors que la parole délexicalisée pourrait fournir un accès aux représentations « phonologiques ». La soustraction des activations liées à la perception des mots par rapport à celles liées à la perception de la parole délexicalisée devrait nous permettre d'observer le niveau sémantique, alors que la comparaison inverse pourrait nous permettre d'accéder au niveau phonologique. Il nous faudrait faire en sorte de répéter un plus grand nombre d'essais, de manière à pouvoir analyser les potentiels évoqués dans les 100 premières millisecondes après le début du stimulus.

À long terme, nos perspectives demeurent inchangées depuis notre maîtrise, et sont de mieux connaître le fonctionnement cérébral du langage articulé, et par là même, d'entrouvrir, comme le dit Steven Pinker, une fenêtre sur la cognition humaine.

Bibliographie

Aboitiz F. & Garcia V. R. (1997).

The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research*, 25 (3). 171-183.

Aboitiz F., Garcia R.R., Bosman C. & Brunetti E. (2006).

Cortical memory mechanisms and language origins. *Brain and Language*, 98 (1). 40-56.

Abry C., Stefanuto M., Vilain A. & Laboissière R. (2002).

Que nous apprennent les "tan, tan" de Tan de Broca sur l'hypothèse d'une syllabe émergeant du babbillage ? in *Phonetics, Phonology, and Cognition*, Durand, J. Laks, B. Oxford University Press.

Alain C., Arnott S.R., Hevenor S., Graham S. & Grady C.L. (2001).

"What" and "where" in the human auditory system. *Proceedings of the National Academy of Science*, 98. 12301-12306.

Amunts K. & Von Cramon D. Y. (2006).

The anatomical segregation of the frontal cortex: what does it mean for function? *Cortex*, 42 (4). 525-528.

Andersen R. A. (1997).

Multimodal integration for the representation of space in the posterior parietal cortex *Philosophical Transactions of the Royal Society London Series B Biological Sciences*, 352. 1421-1428.

Anderson J.M., Gilmore R., Roper S., Crosson B., Bauer R.M., Nadeau S., Berversdorf D.Q., Cibula J., Rogish M., Kortencamp S., Hughes J.D., Gonzalez Rothi L.J. & Heilman K.M. (1999).

Conduction aphasia and the arcuate Fasciculus: A Reexamination of the Wernicke-Geschwind Model. *Brain and Language*, 70. 1-12.

Astésano C., Besson M. & Alter K. (2004).

Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18. 172-184.

- Aubin J.-P. (1991).
Viability Theory. Birkhäuser. 543 p.
- Baddeley A. (1998).
Working Memory. *Life Sciences*, 321. 167-173.
- Baddeley A. (2003).
Working memory and language: an overview. *Journal of Communication Disorders*, 36 (3). 189-208.
- Baddeley A. (2004).
The Psychology of Memory. in *The Essential Handbook of Memory Disorders for Clinicians*, Baddeley, A.D. Kopelman, M.D. Wilson, B.A. eds. John Wiley & Sons. 1-14.
- Baker E., Blumstein S.E. & Goodglass H. (1981).
Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, 19. 1-15. Cité dans Hickok & Poeppel, 2007.
- Basso A., Casati G. & Vignolo L.A. (1977).
Phonemic identification defect in aphasia. *Cortex*, 13 (1). 84-95. Cité dans Hickok & Poeppel, 2007.
- Bates E., Wulfeck B., Dronkers N., Opie M., Fenson J., Kriz S., Jeffries R., Miller L. & Herbst K. (2001).
Differential Effects of Unilateral Lesions on Language Production in Children and Adults. *Brain and Language*, 79 (2). 223-265.
- Bates E., Wilson S.M., Saygin A.P., Dick F., Sereno M.I., Knight R.T. & Dronkers N.F. (2003).
Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6 (5). 448-450.
- Baum S. & Pell M. (1999).
The neural bases of prosody: Insights from lesion studies and Neuroimaging. *Aphasiology*, 13. 581-608.
- Beauchamp M.S., Lee K.E., Haxby J.V. & Martin A. (2003).
fMRI Responses to Video and Point-Light Displays of Moving Humans and Manipulable Objects. *Journal of Cognitive Neuroscience*, 15 (7). 991-1001.

- Beauchamp M.S., Lee K.E., Argall B.D. & Martin A. (2004).
Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41 (5). 809-823.
- Behne D., Wang Y., Alm M., Arntsen I., Eg R. & Valso A. (1998).
Changes in audio-visual speech perception during adulthood. *Australian Workshop on Auditory-Visual Speech Processing*.
- Belin P., Zatorre R. J. & Ahad P. (2002).
Human temporal response to vocal sounds. *Cognitive Brain Research*, 13. 17-26
- Benoît C., Mohamadi T. & Kandel S. (1995).
Effects of phonetic context on audio-visual intelligibility of French. *Rapport de Recherche de l'ICP*, 4. 1195-1203.
- Benson R.R., Richardson M., Whalen D.H. & Lai S. (2006).
Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *NeuroImage*, 31. 342-353.
- Berger K.W. (1972).
Speechreading, Principles and methods. Baltimore: National Educational Press. Cité dans Cathiard, 1994.
- Bernstein L.E., Auer E.T.Jr, Moore J.K., Ponton C.W., Don M. & Singh M. (2002).
Visual speech perception without primary auditory cortex activation. *Neuroreport*, 13 (3). 311-315.
- Bernstein L.E., Auer E.T. Jr, Wagner M. & Ponton C.W. (2008).
Spatiotemporal dynamics of audiovisual speech processing. *NeuroImage*, 39 (1). 423-435.
- Besle J., Fort A., Delpuech C. & Giard M-H. (2004).
Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20. 2225-2234.
- Bimler D. & Kirkland J. (2001).
Categorical perception of facial expressions of emotion: Evidence from multidimensional scaling. *Cognition and Emotion*, 15 (5). 633-658.

- Binder J.R., Rao S.M., Hammeke T.A., Yetkin F.Z., Jesmanowicz A., Bandettini P.A., Wong E.C., Estkowski L.D., Goldstein M.D., Houghton V.M. & Hyde J.S. (1994).
Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, 35 (6). 662-672.
- Binder J.R., Frost J.A., Hammeke T.A., Rao S.M. & Cox R.W. (1996).
Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119 (4). 1239-1247.
- Binder J.R., Frost J.A., Hammeke T.A., Bellgowan P.S., Springer J.A., Kaufman J.N. & Possing E.T. (2000).
Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10 (5). 512-528.
- Binkofski F., Amunts K., Stephan K.M., Posse S., Schormann T., Freund H.-J., Zilles K. & Seitz R.J. (2000).
Broca's region Subserves Imagery of Motion: A Combined Cytoarchitectonic and fMRI Study. *Human Brain mapping*, 11. 273-285.
- Bledowski C., Prvulovic D., Goebel R., Zanella F.E. & Linden D.E.J. (2004).
Attentional systems in target and distractor processing: a combined ERP and fMRI study. *NeuroImage*, 22. 530-540.
- Blumstein S. E., Baker E. & Goodglass H. (1977 a).
Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15. 19-30. Cité dans Hickok & Poeppel, 2007.
- Blumstein S. E., Cooper W. E., Zurif E. B. & Caramazza A. (1977 b).
The perception and production of voice onset time in aphasia. *Neuropsychologia*, 15 371-383. Cité dans Hickok & Poeppel, 2007.
- Blumstein S. E., Myers E. B. & Rissman J. (2005).
The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17 (9). 1353-1366.
- Boemio A., Fromm S., Braun A. & Poeppel D. (2005).
Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8 (3). 389-395.

- Bonder L.X., Gur R.E. & Gur R. (1989).
The effects of right and left hemiparkinsonism on prosody. *Brain and Language*, 36. 192-207.
- Bonmassar G., Schwartz D.P., Liu A.K., Kwong K.K., Dale A.M. & Belliveau J.W. (2001).
Spatiotemporal Brain Imaging of Visual-Evoked Activity Using Interleaved EEG and fMRI Recordings. *NeuroImage*, 13. 1035-1043.
- Bradvik B., Dravins C., Holtas S., Rosen I., Ryding E. & Ingvar D. (1991).
Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica*, 11. 114-126. Cité dans Baum & Pell, 1999.
- Breitenstein C., Daum I. & Ackerman H. (1998).
Emotional processing following cortical and subcortical brain damage: contribution of the fronto-striatal circuitry. *Behavioral Neurology*, 11. 29-42.
- Broca P. (1865).
Sur le siège du langage articulé. *Bulletin de la Société d'Anthropologie de Paris*, 6. 377-393.
- Browman C.P. & Goldstein L. (1990).
Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18. 299-320.
- Browman C.P. & Goldstein L. (2000).
Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5. 25-34.
- Brown C. & Hagoort P. (1993).
The processing nature of the N40: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, 5 (1). 34-44.
- Buccino G., Binkofski F. & Riggio L. (2004 a).
The mirror neuron system and action recognition. *Brain and Language*, 89 (2). 370-376.
- Buccino G., Lui F., Canessa N., Patteri I., Lagravinese G., Benuzzi F., Porro C. A. & Rizzolatti G. (2004 b).
Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *Journal of Cognitive Neuroscience*, 16 (1). 114-126.

Bülthoff I. & Newell F.N. (2004).

Categorical perception of sex occurs in familiar but not unfamiliar faces. *Visual Cognition*, 11 (7). 823-855.

Callan D. E., Jones J. A., Munhall K., Callan A. M., Kroos C. & Vatikiotis-Bateson E. (2003).

Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*, 14 (17). 2213-2218.

Calliope. (1988).

La parole et son traitement automatique. CNET, Masson. 714 p.

Calvert G.A., Bullmore E.T., Brammer M.J., Campbell R., Williams S.C.R., McGuire P.K., Woodruff P.W.R., Iversen S.D. & David A.S. (1997).

Activation of auditory cortex during silent lipreading. *Science*, 276. 593-596.

Calvert G. A. & Campbell R. (2003).

Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15 (1). 57-70.

Campbell R., Woll B., Benson P.J. & Wallace S.B. (1999).

Categorical Perception of Face Actions: Their Role in Sign Language and in Communicative Facial Displays. *The Quarterly Journal of Experimental Psychology*, 152 (A1). 67-95.

Caplan D., Gow D. & Makris N. (1995).

Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits, *Neurology*, 45. 293-298.

Castaigne P., Lhermitte F., Signoret J.-L. & Abelanet R. (1980).

Description et étude scannographique du cerveau de Leborgne: la découverte de Broca. *Revue de neurologie*, 136 (10). 563-583. Cité dans Abry *et al.* 2002.

Cathiard M.-A., Tiberghien G. & Abry C. (1992).

Face and Profile identification skills for lip-rounding in normal-hearing French subjects. *Bulletin de la Communication Parlée*, 2. 43-58.

Cathiard M.-A. (1994).

La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de parole. Thèse, Université Pierre Mendès-France, Grenoble. 335 p.

- Chomsky N. & Halle M. (1968).
 Sound Pattern of English. New York: Harper & Row. 484 p.
- Colby C.L. & Goldberg M.E. (1999).
 Space and Attention in Parietal Cortex. *Annual Review Neuroscience*, 22. 319-349.
- Colin C., Radeau M., Soquet A., Demolin D., Colin F. & Deltenre P. (2002).
 Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*, 113 (4). 495-506.
- Colin C. & Radeau M. (2003).
 Les illusions McGurk dans la parole : 25 ans de recherches. *L'Année Psychologique*, 103 (3). 497-542.
- Colin C., Radeau M., Soquet A. & Deltenre P. (2004).
 Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. *Clinical Neurophysiology*, 115 (9). 1989-2000.
- Content A., Kearns R.K. & Frauenfelder U.H. (2001).
 Boundaries versus Onsets in Syllabic Segmentation. *Journal of Memory and Language*, 45. 177-199.
- Courtney S.M., Ungerleider L.G., Keil K. & Haxby J.V. (1996).
 Object and Spatial Visual Working Memory Activate Separate neural Systems in Human Cortex. *Cerebral Cortex*, 6. 39-49.
- Cutler A., Mehler J., Norris D. & Segui J. (1987).
 Phoneme identification and the lexicon. *Cognitive Psychology*, 19 (2). 141-177.
- Cutler A. & Mehler J. (1993).
 The periodicity bias. *Journal of Phonetics*, 21. 103-108.
- Cutting J.E. & Rosner B.S. (1976).
 Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics*, 20 (1). 87-88.
- Damasio H. & Damasio A.R. (1980).
 The Anatomical Basis of Conduction Aphasia. *Brain*, 103. 337-350.
- Damasio H. (1991).
 Cerebral localization of the aphasias. in *Acquired aphasia* Sarno, M. ed. Academic, San Diego. 45-71. Cité dans Hickok & Poeppel, 2007.

- Dawson G. D. (1951).
 A summation technique for detecting small signals in large irregular background. *Journal of Physiology*, *115*. 2-3.
- Debener S., Ullsperger M., Siegel M., Fiehler K., Von Cramon Y.D. & Engel A.K. (2005).
 Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *The Journal of Neuroscience*, *25* (50). 11730-11737.
- Decety J., Grèzes J., Costes N., Perani D., Jeannerod M., Procyk E., Grassi F. & Fazio F. (1997).
 Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, *120*. 1763-1777.
- Dehaene-Lambertz G., Pallier C., Serniclaes W., Sprenger-Charolles L., Jobert A. & Dehaene S. (2005).
 Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*. 21-33.
- Dekle D.J., Fowler C.A. & Funnel M.G. (1992).
 Audiovisual integration in perception of real words. *Perception & Psychophysics*, *51* (4). 355-362.
- Delmas A. (1981).
 Voies et centres nerveux: 10ème édition. Masson, Paris. 296 p.
- Démonet J.F., Chollet F., Ramsay S., Cardebat D., Nespoulous J.-L., Wise R.J.S., Rascol A. & Frackowiak R.S.J. (1992).
 The Anatomy of Phonological and Semantic Processing in Normal Subject. *Brain*, *115* (6). 1753-1768.
- Démonet J.F., Thierry G. & Cardebat D. (2005).
 Renewal of the Neurophysiology of Language : Functional Neuroimaging. *Physiological Review*, *85*. 2612-2625.
- Denes G. & Semenza C. (1975).
 Auditory modality-specific anomia: evidence from a case of pure word deafness. *Cortex*, *11* (4). 401-411.

- Denkova E., Botzung A. & Manning L. (2006).
Neural correlates of remembering/knowing famous people: An event-related fMRI study. *Neuropsychologia*, 44 (14). 2783-2791.
- Desimone R., Albright T.D., Gross C.G. & Bruce G. (1984).
Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience*, 4. 2051-2062.
- Desjardins R., Rogers J. & Werker J.F. (1997).
An Exploration of Why Preschoolers Perform Differently than Do Adults in Audiovisual Speech Perception Tasks. *Journal of Experimental Child Psychology*, 66. 85-110.
- Dijkerman C.H. & De Haan E.H.F. (2007).
Somatosensory processes subserving perception and action. *Behavioral and Brain Science*, 30. 189-239.
- Dubois C. (2004).
La parole délexicalisée et réitérée : Perturber le sémantisme afin de mettre en exergue les traits prosodiques. *Travaux de l'Institut de Phonétique de Strasbourg*, 34. 35-67.
- Donders, F.C. (1868).
La vitesse des actes psychiques. *Archives Néerlandaises des Sciences Exactes et Naturelles*, 3. Cité dans Steinberg, 1969.
- Doyon J., Penhune V. & Ungerleider L.G. (2003).
Distinct Contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, 41. 252-262.
- Dronkers N.F. (1996).
A new brain region for coordinating speech articulation. *Nature*, 384. 159-161.
- Dronkers N.F., Redfern B. B. & Knight R. T. (2000).
The neural architecture of language. in *The New Cognitive Neurosciences*, Gazzaniga, M. S. ed. MIT, Cambridge, MA. 949-958.
- Dronkers N.F., Wilkins D.P., Van Valin R.D. Jr, Redfern B.D. & Jaeger J.J. (2004).
Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92. 145-177.

- Dumay N., Frauenfelder U.H. & Content A. (2002).
The Role of the Syllable in Lexical Segmentation in French: Word-Spotting Data. *Brain and Language*, 81. 144-161.
- Dupoux E. (1989).
Identification des mots parlés, détection de phonèmes et unité prélexicale. École des Hautes Études en Sciences Sociales, Paris.
- Edelman R., Wielopolski P. & Schmitt F. (1994).
Echo-planar imaging. *Radiology*, 192. 600-612.
- Eimas P.D. (1963).
The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 6. 206-217.
- Fadiga L., Fogassi L., Pavesi G. & Rizzolatti G. (1995).
Motor Facilitation During Action Observation: A Magnetic Stimulation Study. *Journal of Neurophysiology*, 73 (6). 2608-2611.
- Fadiga L., Craighero L., Buccino G. & Rizzolatti G. (2002).
Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15. 399-402.
- Fadiga L. & Craighero L. (2006).
Hand actions and speech representation in Broca's area. *Cortex*, 42. 486-490.
- Fayol M., Gombert J.-E., Lecocq P., Sprenger-Charolles L. & Zagar D. (1992).
Psychologie cognitive de la lecture, PUF. 288 p.
- Ferrand L., Grainger J. & Segui J. (1995).
Amorçage phonologique masqué et dénomination. *L'Année Psychologique*, 95. 645-659.
- Ferrari P.F., Gallese V., Rizzolatti G. & Fogassi L. (2003).
Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17. 1703-1714.
- Fiebach C. J., Friederici A. D., Müller K. & Von Cramon Y. D. (2002).
fMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. *Journal of Cognitive Neuroscience*, 14 (1). 11-23.

- Fisher C.G. (1968).
Confusions among visually perceived consonants. *Journal of Speech and Hearing Research, 11*. 796-804.
- Fitch R. H., Miller S. & Tallal P. (1997).
Neurobiology of speech perception. *Annual Review of Neuroscience, 20*. 331-353.
- Fort A., Delpuech C., Pernier J. & Giard M.-H. (2002).
Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14*. 20-30.
- Fowler C. (1986).
An event approach of the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*. 3-28.
- Fowler C. (1996).
Listeners do hear sounds, not tongues. *Journal of Acoustical Society of America, 99*, (3). 1730-1741.
- Fridriksson J., Moss J., Davis B., Baylis G.C., Bonilha L. & Rorden C. (2008).
Motor speech perception modulates the cortical language areas. *NeuroImage, 41*. 605-613.
- Friederici A.D. (1999).
Lexical integration sequential effects of syntactic and semantic integration difficulty. *Memory cognitive processes, 27*.
- Friederici A.D., Von Cramon D.Y. & Kotz S.A. (1999).
Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain, 122*. 1033-1047.
- Friederici A. D., Meyer M. & Von Cramon Y. (2000).
Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language, 74*. 698-705.
- Frost R. (1998).
Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological Bulletin, 123* (1). 71-99.
- Fry D.B., Abramson A.S., Eimas P.D. & Liberman A.M. (1962).
The identification and discrimination of synthetic vowels. *Language and Speech, 5* (4). 171-189.

- Gallese V., Fadiga L., Fogassi L. & Rizzolatti G. (1996).
Action recognition in the premotor cortex. *Brain*, *119*. 593-609.
- Gandour J., Tong Y., Wong D., Talavage T., Dziedzic M., Xu Y., Li X. & Lowe M. (2004).
Hemispheric roles in the perception of speech prosody. *NeuroImage*, *23*. 344-357.
- Gentil M. (1981).
Étude de la perception de la parole : lecture labiale et sosies labiaux. IBM, France.
Cité dans Cathiard, 1994..
- Gerrits E. & Schouten M.E.H. (2004).
Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66* (3). 363-376.
- Giard M.H. & Peronnet F. (1999).
Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Journal of Cognitive Neuroscience*, *11*. 473-490.
- Giraud A-L., Kell C., Thierfelder C., Sterzer P., Russ M.O., Preibisch C. & Kleinschmidt A. (2004).
Contributions of Sensory Input, Auditory Search and Verbal Comprehension to Cortical Activity during Speech Processing. *Cerebral Cortex*, *14* (3). 247-255.
- Giraud A-L., Démonet J.F., Habib M., Marquis P., Chauvel P. & Liégeois-Chauvel C. (2005).
Auditory Evoked Potential Patterns to Voiced and Voiceless Speech Sounds in Adult Developmental Dyslexics with Persistent Deficits. *Cerebral Cortex*, *10*. 1524-1534.
- Giraud A-L., Kleinschmidt A., Poeppel D., Torben L.E., Frackowiak R.S.J. & Laufs H. (2007).
Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron*, *56*. 1127-1134.
- Goodale M.A. & Humphrey G.K. (1998).
The objects of action and perception. *Cognition*, *67*. 181-207.
- Gorno-Tempini M.L., Dronkers N.F., Rankin K.P., Ogar J.M., La Phengrasamy B.A., Rosen H.J., Johnson J.K., Weiner M.W. & Miller B.L. (2004).
Cognition and Anatomy in Three Variants of Primary Progressive Aphasia. *Annals of Neurology*, *55*. 335-346.

Griffiths T.D. & Warren J.D. (2002).

The planum temporale as a computational hub. *Trends in Neurosciences*, 25 (7). 348-353.

Grodzinsky Y. (2006).

The Language Faculty, Broca's Region, and the Mirror System. *Cortex*, 42. 464-468.

Grodzinsky Y. & Friederici A.D. (2006).

Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, 16. 240-246.

Hagoort P. & Brown C.M. (2000).

ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, 38. 1518-1530.

Hall D.A., Fussell C. & Summerfield Q. (2005).

Reading Fluent Speech from Talking Faces: Typical Brain Networks And Individual Differences. *Journal of Cognitive Neuroscience*, 17 (6). 939-953.

Haxby J.V., Hoffman E.A. & Gobbini M.I. (2000).

The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4 (6).

Hecker V., Vaxelaire B., Sock R. & Cathiard M.-A. (1999).

Les effets perceptifs de gestes anticipatoires en français. *Travaux de l'Institut de Phonétique de Strasbourg*, 29. 1-29.

Heider F.K. & Heider G.M. (1940).

An experimental investigation of lipreading. *Psychological Monographs*, 52. 124-153.

Hertz-Pannier L., Lehericy S., Cordoliani Y.S., Le Bihan D., Marsault C. & Brunelle F. (2000).

IRM fonctionnelle cérébrale : bases physiologiques, techniques et méthodologiques, et applications cliniques. *Journal de Radiologie*, 81. 717-730.

Hickok G. & Poeppel D. (2000).

Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4 (4). 131-138.

Hickok G., Buchsbaum B., Humphries C. & Muftuler T. (2003).

Auditory-motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, 15 (5). 673-682.

Hickok G. & Poeppel D. (2004).

Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92. 67-99.

Hickok G. & Poeppel D. (2007).

The cortical organization of speech processing. *Nature Review Neuroscience*, 8. 393-402.

Hickok G. (2009).

Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience*, 21 (7). 1229-1243.

Hietanen J.K., Manninen P., Sams M. & Surrakala V. (2001).

Does audiovisual speech perception use information about facial configuration? *European Journal of Neuroscience*, 13 (3). 395-407.

Holender D. (1988).

Représentations phonologiques dans la compréhension et dans la prononciation des mots écrits. *Cahiers du Département des Langues et des Sciences du Langage*, 6. 31-84.

Hoonhorst I., Colin C., Markessis E., Radeau M., Deltenre P. & Serniclaes W. (2007).

N100 component: an electrophysiological cue of voicing perception. *Clinical Neurophysiology*, 120 (5). 897-903.

Houdé O., Mazoyer B. & Tzourio-Mazoyer N. (2002).

Cerveau et psychologie, introduction à l'imagerie cérébrale anatomique fonctionnelle. PUF. 609 p.

Hsieh L., Gandour J., Wong D. & Hutchins G.D. (2001).

Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, 76. 227-252.

Humphries C., Love T., Swinney D. & Hickok G. (2005).

Response of Anterior Temporal Cortex to Syntactic and Prosodic Manipulations During Sentence Processing. *Human Brain mapping*, 26. 128-138.

Iacobini M., Woods R.P., Brass M., Bekkering H., Mazziotta J.C. & Rizzolatti G. (1999).

Cortical Mechanisms of Human Imitation. *Science*, 86. 2526-2528.

- Iacobini M. & Wilson S.M. (2006).
Beyond a single area: Motor control and language within a neural architecture encompassing Broca's area. *Cortex*, 42. 503-506.
- Iacobini M. & Dapretto M. (2006).
The mirror neuron system and the consequences of its dysfunction. *Nature*, 942 (7). 942-951.
- Indefrey P. & Levelt W.J.M. (2004).
The spatial and temporal signatures of word production components. *Cognition*, 92. 101-144.
- Jacquemot C., Dupoux E., Pallier C. & Bachoud-Lévi A.-M. (2002).
Comprehending spoken words without hearing phonemes: A case study. *Cortex*, 38. 869-873.
- Jacquemot C. & Scott S.K. (2006).
What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10 (11). 480-491.
- Jakobson R., Fant G. & Halle M. (1952).
Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge: MIT Press. 115 p.
- Jäncke L., Wüstenberg T., Scheich H. & Heinze H. J. (2002).
Phonetic perception and the temporal cortex. *NeuroImage*, 15. 733-746.
- Joanisse M.F. & Gati J.S. (2003).
Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, 19. 64-79.
- Jonas S. (1981).
The supplementary motor region and speech emission. *Journal of Communication Disorders*, 14. 349-373.
- Jones J. A. & Callan D. E. (2003).
Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14 (8). 1129-1133.
- Jun S.A. & Fougeron C. (2000).
A phonological model of French intonation. Intonation: analysis, modelling and technology. A. Botinis, Dordrecht. 396 p.

- Jun S.A. & Fougeron C. (2002).
Realizations of Accentual Phrases in French Intonation. *Probus*, 14. 147-172.
- Jusczyk P.W., Rosener B.S., Cutting J.E., Foard C.F. & Smith L.B. (1977).
Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 21. 50-54.
- Kandel S. & Boë J.-L. (1996).
Traitement phonétique et représentation lexicale dans la reconnaissance des mots. *Bulletin de la Communication Parlée*, 3. 61-72.
- Kavanagh R., Darcey T. & Lehmann D. (1978).
Evaluation of methods for three dimensional localization of electrical sources in the human brain. *IEEE Transactions on Biomedical Engineering*, 25. 421-429.
- Klatt D.H. (1975)
Voice onset time frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18. 686-706.
- Klatt D.H. (1979).
Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7. 279-312.
- Klucharev V., Möttönen R. & Sams M. (2003).
Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, 18 (1). 65-75.
- Kohler E., Keysers C., Umiltà M.A., Fogassi L., Gallese V. & Rizzolatti G. (2002).
Hearing Sounds, Understanding Actions: Action representation in Mirror Neurons. *Science*, 297 (5582). 846-848.
- Kolinsky R., Morais J. & Segui J. (1992).
La reconnaissance des mots dans les différentes modalités sensorielles : étude psycholinguistique cognitive, PUF. 320 p.
- Krakow K., Woermann F.G., Symms M.R., Allen P.J., Lemieux L., Barker G.J., Duncan J.S. & Fish D.R. (1999).
EEG-triggered functional MRI of interictal epileptiform activity in patients with partial seizures. *Brain*, 122. 1679-1688.

Kuhl P.K. (1991).

Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories. *Perception & Psychophysics*, 50 (2). 93-107.

Kutas M. & Hillyard S.A. (1989).

An Electrophysiological Probe of Incidental Semantic Association. *Journal of Cognitive Neuroscience*, 1 (1). 38-49.

Lebib R., Papo D., De Bode S. & Baudonniere P. M. (2003).

Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neuroscience Letters*, 341 (3). 185-188.

Lefavrais P. (1967).

Test de l'Alouette. *Édition du Centre de Psychologie Appliquée, Paris*.

Liberman A.M., Harris K.S., Hoffman H.S. & Griffith B.C. (1957).

The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 54 (5). 358-368.

Liberman A.M., Cooper F.S., Shankweiler D.P. & Studdert-Kennedy M. (1967).

Perception of the speech code. *Psychological review*, 74. 431-461.

Liberman A.M. & Mattingly I.G. (1985).

The motor theory of speech perception revised. *Cognition*, 21. 1-36.

Lichtheim L. (1885).

On aphasia. *Brain*, 7. 433-484. Cité dans Caplan, 1987.

Liebenthal E., Binder J.R., Spitzer S.M., Possing E.T. & Medler D.A. (2005).

Neural substrates of phonemic perception. *Cerebral Cortex*, 10. 1621-1631.

Lindblom B. (1987).

Adaptative variability and absolute constancy in speech signals: Two themes in the quest for phonetic invariance. *Proceedings of the XIth ICPHS, Talinn* (3). 9-18.

Lisker L. & Abramson A.S. (1964).

A cross language study of voicing in initial stops: Acoustical measurements. *Journal of Acoustical Society of America*, 49. 389-391.

Lisker L. & Rossi M. (1992).

Auditory and visual cueing of the rounded feature of vowels. *Language and Speech*, 35 (4). 391-417.

- Lotto A.J., Hickok G. & Holt L.L. (2009).
 Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13 (3). 110-114.
- Luce P.A., Goldinger S.D., Auer E.T. Jr & Vitevitch M.S. (2000).
 Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62 (3). 615-625.
- MacNeilage P. & Davis B. (1998).
 The frame and content theory of evolution of speech production. *Behavioral and Brain Science*, 21. 499-511.
- Maddieson I. (1984).
 The effects of F0 on a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics*, 12. 9-15.
- Marslen-Wilson W.D., William D. & Welsh A. (1978).
 Processing Interactions and Lexical Access during Word Recognition in Continuous Speech. *Cognitive Psychology*, 10 (1). 29-63.
- Marslen-Wilson W.D. (1987).
 Functional parallelism in spoken word-recognition. *Cognition*, 25 (1-2). 71-102.
- Martin B.A., Tremblay K.L. & Korczak P. (2008).
 Speech Evoked Potentials: From the laboratory to the clinic. *Ear & Hearing*, 29 (3). 285-313.
- Martinet A. (1955).
 L'Économie des changements linguistiques. Berne. 290 p.
- Martinet A. (1967).
 Éléments de linguistique générale. *Armand Colin, Paris*. 217 p.
- Massaro D.W. (1988).
 Ambiguity in Perception and Experimentation. *Journal of Experimental Psychology: Human Perception and Performance*, 117 (4). 417-421.
- Massaro D.W. (1990).
 A Fuzzy Logical Model of Speech Perception. *Proceedings of the XXIV International Congress of Psychology, in Human Information Processing: Measures, Mechanisms, and Models*, Vickers D. & Smith P. L. eds. North-Holland, Amsterdam.

- Massaro D. W. (1999).
 Speechreading: illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3 (8).
- Massaro D. W. (2001).
 Speech Perception. in *International Encyclopedia of Social and Behavioral Sciences*, Smelser N.M. & Baltes P.B. eds. Amsterdam. 14870-14875.
- Mayer J., Wildgruber D., Riecker A., Dogil G., Ackermann H. & Grodd W. (2002).
 Prosody Production and Perception: Converging Evidence from fMRI Studies. *Speech Prosody*, Aix-en-Provence, France. 635-638.
- Mazoyer B.M., Tzourio N., Frak V., Syrota A., Murayama N., Levrier O., Salamon G., Dehaene S., Cohen L. & Mehler J. (1993).
 The Cortical Representation of Speech. *Journal of Cognitive Neuroscience*, 5 (4). 467-479.
- McClelland J.M. & Elman J.L. (1986).
 The TRACE Model of Speech Perception. *Cognitive Psychology*, 18. 1-86.
- McGlone J. (1984).
 Speech comprehension after unilateral injection of sodium amytal. *Brain and Language*, 22. 150-157. Cité dans Hickok & Poeppel, 2007.
- McGurk H. & MacDonald J. (1976).
 Hearing lips and seeing voices. *Nature*, 264 (5588). 746-748.
- McGurk H. (1981).
 Listening with eye and ear. in *The Cognitive Representation of Speech*, Myers T. Laver, J. Anderson, J. eds. North Holland. 336-337. Cité dans Colin & Radeau, 2003.
- Mehler J. (1981).
 The Role of Syllables in Speech Processing: Infant and Adult Data. *Philosophical Transactions of the Royal Society London Series B Biological Sciences*, 295. 333-352.
- Mehler J., Dupoux E. & Segui J. (1990).
 Constraining models of lexical access: the onset of word recognition. in *Cognitive Models of Speech Processing*, Altmann, G. ed. MIT Cambridge, MA.

- Mellers J.D.C., Bullmore E.T., Brammer M.J., Williams S.C.R., Andrew C., Sachs N., Andrews C., Cox T.S., Simmons A., Woodruff P.W.R., David A.S. & Howard R. (1995).
Neural correlates of working memory in a visual letter monitoring task: an fMRI study. *Neuroreport*, 7. 109-112. Cité dans Calvert *et al.*, 1997.
- Meltzoff A.N. & Moore M.K. (1983).
Newborn Infants Imitate Adult Facial Gestures. *Child Development*, 54. 701-709.
- Meltzoff A.N. & Moore M.K. (1997).
Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198 (4312). 75-78.
- Metz-Lutz M.-N. & Dahl E. (1984).
Analysis of word comprehension in a case of pure word deafness. *Brain and Language*, 23 (1). 13-25.
- Miceli G., Gainotti G., Caltagirone C. & Masullo C. (1980).
Some aspects of phonological impairment in aphasia. *Brain and Language*, 11. 159-169. Cité dans Hickok & Poeppel, 2007.
- Miglioretti D.L. & Boatman D. (2003).
Modeling variability in cortical representations of human complex sound perception. *Experimental Brain Research*, 153 (3). 382-387.
- Miller L.M. & D'esposito M.E. (2005).
Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech. *The Journal of Neuroscience*, 25. 5884-5893.
- Milner A.D. & Goodale M.A. (1995).
The Visual Brain in Action. *Oxford University Press*. 405-428. Cité dans Goodale & Humphrey, 1998.
- Mishkin M., Ungerleider L.G. & Macko C.A. (1983).
Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6. 414-417.
- Morton J. (1969).
Interaction of information in word recognition. *Psychological review*, 76. 165-178. Cité dans Seidenberg & McClelland, 1989.

Mourand-Dornier L. (1980).

Le rôle de la lecture labiale dans la reconnaissance de la parole. Thèse, Université de Franche Comté, Besançon. Cité dans Cathiard, 1994. 127 p.

Mulert C., Jäger L., Schmitt R., Bussfeld P., Pogarell O., Mölle H-J., Juckel G. & Hegerl U. (2004).

Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage*, 22. 83-94.

Näätänen R. (1982).

Processing negativity: An evoked potential reflection of selective attention. *Psychological Bulletin*, 3. 53-98.

Näätänen R. (1990).

The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Science*, 13. 201-288.

Narain C., Scott S.K., Wise R.J.S., Rosen S., Leff A., Iversen S.D. & Matthews P.M. (2003).

Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI. *Cerebral Cortex*, 13. 1362-1368.

Nguyen N. (2001).

Rôle de la coarticulation dans la reconnaissance des mots. *L'Année Psychologique*, 101. 125-154.

Norris D. (1994).

Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52 (3). 189-234.

Novak J.D., Ritter W., Vaughan H.G. Jr & Witznitzer M.L. (1990).

Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography Clinical Neurophysiology*, 75. 255-275.

Obleser J., Zimmermann J., Van Meter J. & Rauschecker J.P. (2006).

Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. *Cerebral Cortex*, 10. 562-568.

- Ogawa S., Lee T.M., Kay A.R. & Tank D.W. (1990).
Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Science*, 87. 9868-9872.
- Ojanen V., Möttönen R., Pekkola J., Jääskeläinen I.P., Joensuu R., Autti T. & Sams M. (2005).
Processing of audiovisual speech in Broca's area. *NeuroImage*, 25. 333-338.
- Ojemann G.A. (1991).
Cortical Organization of Language. *The Journal of Neuroscience*, 11 (8). 2281-2287.
- Okada K. & Hickok G. (2006).
Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, 98. 112-117.
- Olson I.R., Gatenby J.C. & Gore J.C. (2002).
A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research*, 14 (1). 129-138.
- Otzenberger H., Gounot D., Marrer C., Namer I.J. & Metz-Lutz M.-N. (2005).
Reliability of Individual Functional MRI Brain Mapping of Language. *Neuropsychology*, 19 (4). 484-493.
- Pallier C. (1994).
Rôle de la syllabe dans la perception de la parole : études attentionnelles. École des Hautes Études en Sciences Sociales, Paris. 151 p.
- Paulesu E., Frith C.D. & Frackowiak R.S.J. (1993).
The neural correlates of the verbal component of working memory. *Nature*, 362. 342-345.
- Paulesu E., Perani D., Blasi V., Silani G., Borghese N.A., De Giovanni U., Sensolo S. & Fazio F. (2003).
A Functional-Anatomical Model for Lipreading. *Journal of Neurophysiology*, 90. 2005-2013.
- Pekkola J., Ojanen V., Autti T., Jääskeläinen I.P., Möttönen R., Tarkiainen A. & Sams M. (2004).
Primary auditory cortex activation by visual speech: an fMRI study at 3T. *Neuroreport*, 16 (2). 125-128.

Penhune V.B., Zatorre R.J., MacDonald J.D. & Evans A.C. (1996).

Interhemispheric Anatomical Differences in Human Primary Auditory Cortex: Probabilistic Mapping and Volume Measurement from Magnetic Resonance Scans. *Cerebral Cortex*, 6. 661-672.

Perrett D.I., Smith P.A. J., Potter D.D., Mistlin A.J., Head A.S., Milner A.D. & Jeeves M.A. (1985).

Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. *Proceedings of the Royal Society of London*, 223 (1232). 293-317.

Petersen S.E., Fox P.T., Posner M.I., Mintun M. & Raichle M.E. (1988).

Positron Emission Tomographic studies of the cortical anatomy of single word processing. *Nature*, 331. 585-589. Cité dans Houdé *et al.*, 2002.

Pitt M.A. & McQueen J.M. (1998).

Is compensation for Coarticulation Mediated by the Lexicon? *Journal of Memory and Language*, 39. 347-370.

Poeppel D. (2001).

Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25. 679-693.

Poeppel D. (2003).

The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Communication*, 41. 245-255.

Poldrack R.A., Wagner A.D., Prull M.W., Desmond J.E., Glover G.H. & Gabrieli J.D.E. (1999).

Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *NeuroImage*, 10. 15-35.

Ponton C.W., Auer E.T. & Bernstein L.E. (2002).

Neurocognitive basis for audio-visual speech perception: evidence from event-related potentials. *7th International Conference on Spoken Language Processing* Denver, USA. 1697-1700.

Postle N., McMahon K.L., Ashton R., Meredith M. & De Zubicaray G.I. (2008).

Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *NeuroImage*, 43. 634-644.

- Price C.J., Wise R.J.S., Warburton E.A., Moore C.J., Howard D., Patterson K., Fracowiak R.S.J. & Friston K.J. (1996).
Hearing and saying the functional neuro-anatomy of auditory word processing. *Brain*, 119. 919-931.
- Price C.J. (2000).
The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy*, 197. 335-359.
- Price C.J., Thierry G. & Griffiths T. (2005).
Speech-specific auditory processing: where is it ? *Trends in Cognitive Sciences*, 9 (6). 271-276.
- Puce A., Allison T., Bentin S., Gore J.C. & McCarthy G. (1998).
Temporal Cortex Activation in Humans Viewing Eye and Mouth Movements. *The Journal of Neuroscience*, 15. 2188-2199.
- Pulvermüller F., Huss M., Kheriff F., Moscoso Del Prado M. F., Hauk O. & Shtyrov Y. (2006).
Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Science*, 103 (20). 7865-7870.
- Rademacher J., Bürgel U., Geyer S., Schirrmann T., Schleicher A., Freund H.-J. & Zilles K. (2001).
Variability and asymmetry in the human precentral motor system. A cyto-architectonic and myelo-architectonic brain mapping study. *Brain*, 124. 2232-2258.
- Rauschecker J.P. (1998).
Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8. 516-521.
- Rimol L.M., Specht K., Weis S., Savoy R. & Hugdahl K. (2005).
Processing of sub-syllabic speech units in the posterior temporal lobe: An fMRI study. *NeuroImage*, 26. 1059-1067.
- Rizzolatti G., Fadiga L., Gallese V. & Fogassi L. (1996).
Premotor cortex and the recognition of motor actions. *Cognitive brain research* 3.131-141.
- Rizzolatti G. & Arbib M.A. (1998).
Language within our grasp. *Trends in Neurosciences*, 21 (5). 188-194.

- Rizzolatti G. & Craighero L. (2003).
The Mirror-Neuron System. *Annual Review Neuroscience*, 27. 169-192.
- Rizzolatti G. & Sinigaglia C. (2008).
Les neurones miroirs. Odile Jacob, Paris. 236 p.
- Robin D.A., Tranel D. & Damasio H. (1990).
Auditory Perception of Temporal and Spectra Events in patients with Focal Left and right Cerebral Lesions. *Brain and Language*, 39. 539-555.
- Rodd J.M., Davis M.H. & Johnsrude I.S. (2005).
The Neural Mechanisms of Speech Comprehension: fMRI studies of Semantic Ambiguity. *Cerebral Cortex*, 15. 1261-1269.
- Rosenblum L. D., Schmuckler M. A. & Johnson J. A. (1997).
The McGurk effect in infants. *Perception Psychophysics*, 59 (3). 347-357.
- Ross L.A., Saint-Amour D., Leavitt V.M., Javitt D.C. & Foxe J.J. (2006).
Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex*, 10. 1147-1153.
- Roy C. & Sherrington C. (1890).
On the regulation of the blood supply of the brain. *Journal of Physiology*, 11. 85-108.
- Roy J.P. (2005).
Visual perception of anticipatory rounding gestures in French. *Proceedings of Interspeech Eurospeech 9th European Conference on Speech Communication and Technology*. 2949-2952.
- Saffran E.M., Marin O.S. & Yeni-Komshian G.H. (1976).
An analysis of speech perception in world deafness. *Brain and Language*, 3 (2). 209-228. Cité dans Metz-Lutz & Dahl, 1984.
- Saint-Amour D., De Sanctis P., Molholm S., Ritter W. & Foxe J.J. (2007).
Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, 45 (3). 587-597.
- Sams M., Aulanko R., Hämäläinen M., Hari R., Lounasmaa O.V., Lu S.-T. & Siomla J. (1991).
Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127. 141-145.

- Samson Y., Belin P., Thivard L., Boddaert N., Corzier S. & Zilbovicius M. (2001).
 Perception auditive et langage : imagerie fonctionnelle du cortex auditif sensible au langage. *Revue de neurologie*, 157 (8-9). 837-846.
- Scarborough R., Keating P., Baroni M., Cho T., Mattys S., Alwan A., Auer E.T. Jr & Bernstein L. (2006).
 Optical Cues to the Visual Perception of Lexical and Phrasal Stress in English. *UCLA Working Papers in Phonetics*, 105. 118-124.
- Schönwiesner M., Rubsamen R. & Von Cramon Y.D. (2005).
 Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral belt cortex. *European Journal of Neuroscience*, 22. 1521-1528.
- Schouten B., Gerrits E. & Van Hessen A. (2003).
 The end of categorical perception as we know it. *Speech Communication*, 41. 71-80.
- Schwartz J.L., Abry C., Boë L.-J. & Cathiard M.-A. (2002).
 Phonology in a theory of Perception-for-Action-Control. In *Phonetics, Phonology, and Cognition*. Durand, J. Laks, B. eds. Oxford University Press. 240-280.
- Schwartz J.L., Berthommier F. & Savariaux C. (2004).
 Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93. 69-78.
- Scott S. K., Blank C.C., Rosen S. & Wise R.J.S. (2000).
 Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123. 2400-2406.
- Scott S. K. & Johnsrude I.S. (2003).
 The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26 (2). 100-107.
- Scott S. K. (2005).
 Auditory processing - speech, space and auditory objects. *Current Opinion in Neurobiology*, 15. 197-201.
- Scott S. K., McGettigan C. & Eisner F. (2009).
 A little more conversation, a little less action - candidate roles for the perception motor cortex in speech perception. *Nature Neuroscience*, 10. 295-302.

Seeck M., Lazeyras F., Michel C.M., Blanke O., Gericke C.A., Ives J., Delavelle J., Golay X., Haenggel I.C.A., De Tribolet N. & Landis T. (1998).

Non-invasive epileptic focus localization using EEG-triggerred functional MRI and electromagnetic tomography. *Electroencephalography and clinical Neurophysiology*, 106. 508-512.

Seidenberg M.S. & McClelland J.L. (1989).

A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, 96 (4). 523-568.

Sekiyama K., Kanno I., Miura S. & Sugita Y. (2003).

Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47 (3). 277-287.

Serniclaes W., Van Heghe S., Mousty P., Carré R. & Sprenger-Charolles L. (2004).

Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, 87 (4). 336-361.

Serniclaes W., Ventura P., Morais J. & Kolinsky R. (2005).

Categorical perception of speech sounds in illiterate adults. *Cognition*, 98. 1-10.

Shannon R.V., Zeng F.-G., Kamath V., Wygonski J. & Ekelid M. (1995).

Speech Recognition with Primarily Temporal Cues. *Science*, 270 (5234). 303-304.

Sherrington C.S. (1906).

The integrative action of the nervous system. *Scribner*. Consulté sur <http://books.google.fr/>.

Skipper J.L., Nusbaum H.C. & Small S.L. (2005).

Listening to talking faces: motor activation during speech perception. *NeuroImage*, 25.

Smith A.T., Greenlee M.W., Singh K.D., Kraemer F.M. & Hennig J. (1998). 76-89.

The Processing of First- and Second-Order Motion in Human Visual Cortex Assessed by Functional Magnetic Resonance Imaging (fMRI). *The Journal of Neuroscience*, 18 (10). 3816-3830.

Smith E.E. & Jonides J. (1999).

Storage and Executive Processes in the Frontal Lobes. *Science*, 283. 1657-1661.

Sock R. (2001).

Théorie de la Viabilité en production-perception de la parole. *In Percevoir : Monde et Langage*. Keller, D. Durafour, J.-P. Bonnot, J.-F. Sock, R. eds. Mardaga, Liège. 285-316.

Soto-Faraco S., Navarra J. & Alsius A. (2004).

Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition*, 92 (3). 13-23.

Spitsyna G., Warren J.E., Scott S.K., Turkheimer F.E. & Wise R.J.S. (2006).

Converging Language Streams in the Human Temporal Lobe. *The Journal of Neuroscience*, 26 (28). 7328-7336.

Steinberg S. (1969).

The Discovery of Processing Stages: Extension's of Donders' Method. *Acta Psychologica*, 30. 276-315.

Stetson R.H. (1951).

Motor phonetics: A study of speech movements in action. *North-Holland Pub. Co*, Cité dans Krakow, 1999.

Stevens K.N. (1972).

The quantal nature of speech. *in Human Communication, A Unified View*. Denes, P.B. & Davides, E.E. eds. New York. 51-66.

Stevens K.N. (1981).

Evidence for the role of acoustic boundaries in the perception of speech sounds. *Journal of Acoustical Society of America*, 69 (S1). 61-74.

Sumby W.H. & Pollack I. (1954).

Visual Contribution to Speech Intelligibility in Noise. *Journal of Acoustical Society of America*, 26 (2). 212-216.

Summerfield Q. (1992).

Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society London Series B Biological Sciences*, 335 (1273). 71-78.

Tallal P., Miller S. & Fitch R.H. (1993).

Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences*, 682. 27-47.

Tervaniemi M. & Hugdhal K. (2003).

Lateralization of auditory-cortex functions. *Brain Research Reviews*, 43 (3). 231-246.

Tiippana K., Andersen T.S. & Sams M. (2004).

Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 16 (3). 457-472.

Timsit M. (2001).

Les potentiels évoqués (PE) : une fenêtre sur la cognition. in *123 Atelier*. Paris. 1-11.

Tong F. (2003).

Primary visual cortex and visual awareness. *Cognitive Neuroscience*, 4. 219-229.

Trébuchon-Da Fonseca A., Giraud K., Badier J-M., Chauvel P. & Liégeois-Chauvel C. (2005).

Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *NeuroImage*, 27. 1-14.

Troille E. & Cathiard M.-A. (2006).

L'intégration bimodale de l'anticipation du flux vocalique dans le flux consonantique. *Actes des XXVI^{èmes} Journées d'Études sur la Parole*, Dinard. 367-370.

Troubetzkoy N. (1939).

Phonemes and how to determine them. *Travaux du Cercle Linguistique de Prague*.

Tsao D.Y., Freiwald W.A., Tootell R.B.H. & Livingstone M.S. (2006).

A Cortical Region Consisting Entirely of Face-Selective Cells. *Science*, 311. 670-674.

Tuomainen J., Andersen T.S., Tiippana K. & Sams M. (2005).

Audiovisual speech perception is special. *Cognition*, 96. 13-22.

Ungerleider L.G. & Haxby J.V. (1994).

What and Where in the human brain. *Current Opinion in Neurobiology*, 4. 157-165.

Ungerleider L.G. & Mishkin M. (1982).

Two cortical visual systems. in *Analysis of visual behavior*, Ingle, D.J. Goodale, M.A. & Mans, R.J.W. eds. Cambridge, MA. 549-586. Cité dans Mishkin et al., 1983.

Van Atteveldt N., Formisano E., Goebel R. & Blomert L. (2004).

Integration of Letters and Speech Sounds in the Human Brain. *Neuron*, 43 (2). 271-282.

- Van Wassenhove V., Grant K. W. & Poeppel D. (2005).
 Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, 102 (4). 1181-1186.
- Van Wassenhove V., Grant K.W. & Poeppel D. (2007).
 Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45 (3). 598-607.
- Vaughan H.G. Jr & Ritter W. (1970).
 The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography Clinical Neurophysiology*, 28 (4). 360-367. Cité dans Choppe *et al.*, 1994.
- Vaxelaire B., Sock R. & Hecker V. (1999).
 Perceiving anticipatory phonetic gestures in French. *Eurospeech 99*, Budapest. 165-168.
- Vitevitch M.S., Luce P.A., Pisoni D.B. & Auer E.T. Jr. (1999 a).
 Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. *Brain and Language*, 68. 306-311.
- Vitevitch M.S., Luce P.A., Pisoni D.B. & Auer E.T. Jr. (1999 b).
 Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language* (40). 374-408.
- Vouloumanos A., Kiehl K.A., Werker J.F. & Liddle P.F. (2001).
 Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience*, 13 (7). 1-12.
- Warren J.D. & Griffiths T.D. (2003).
 Distinct Mechanisms for processing Spatial Sequences and Pitch Sequences in the Human Auditory Brain. *The Journal of Neuroscience*, 23 (13). 5799-5804.
- Warren J.E., Wise R.J.S. & Warren J.D. (2005).
 Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends in Cognitive Sciences*, 28 (12). 636-643.
- Watkins K.E., Strafella A.P. & Paus T. (2003).
 Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41. 989-949.

- Weintraub S., Mesulam M.-M. & Kramer L. (1981).
Disturbances in prosody: a righthemisphere contribution to language. *Archives of Neurology*, 38. 742-744. Cité dans Baum & Pell, 1999.
- Wetsbury C.F., Zatorre R.J. & Evans A.C. (1999).
Quantifying variability in the Planum temporale: a Probability Map. *Cerebral Cortex*, 9. 392-405.
- Wilkinson D. & Halligan P. (2004).
The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Neuroscience Reviews*, 5. 67-73.
- Wilshire C.E. & McCarthy R.A. (1996).
Experimental Investigations of an Impairment in Phonological Encoding. *Cognitive Neuropsychology*, 13 (7). 1059-1098.
- Wilson S.M., Pinar Saygin A.P., Sereno M.I. & Iacoboni M. (2004).
Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7 (7). 701-702.
- Wise R.J.S., Chollet F., Hadar U., Friston K.J., Hoffner E. & Frackowiak R.S.J. (1991).
Distribution of Cortical Neural Networks Involved in Word Comprehension and Word Retrieval. *Brain*, 114 (4). 1803-1817. Cité dans Houdé *et al.* 2002.
- Wise R.J.S., Scott S.K., Blank C.S., Mummery C.J., Murphy K. & Warburton E. A. (2001).
Separate neural subsystem within Wernicke's area. *Brain*, 124. 83-95.
- Wise S.P., Boussaoud D., Johnson P.B. & Caminiti R. (1997).
Premotor and Parietal Cortex: Corticocortical Connectivity and Combinatorial Computations. *Annual Review in Neuroscience*, 20. 25-42.
- Wong P.C.M. (2009).
Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, 47. 693-703.
- Wood C.C. (1976).
Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of Acoustical Society of America*, 60 (6). 1381-1389.
- Wright T.R., Pelphrey K.A., Allison T., McKeown M.J. & McCarthy G. (2003). Polysensory Interactions along lateral Temporal regions Evoked by Audiovisual Speech. *Cerebral Cortex*, 13. 1034-1043.

- Wunderlich J.L., Cone-Wesson B.K. & Shepherd R. (2006).
Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, 212. 185-202.
- Wytenbach R.A., May M.L. & Hoy R.R. (1996).
Categorical perception of Sound frequency by crickets. *Science*, 273. 1542-1544.
- Zaidel E. (1985).
Language in the right hemisphere. in *The Dual Brain: Hemispheric Specialization in Humans*. Benson, D.F. & Zaidel, E. eds. Guilford, New York. 205-231. Cité dans Hickok et Poeppel, 2007.
- Zatorre R.J., Evans A.C., Meyer E. & Gjedde A. (1992).
Lateralization of phonetic and pitch processing in speech perception. *Science*, 256. 846-849.
- Zatorre R.J., Evans A.C. & Meyer E. (1994).
Neural Mechanisms Underlying Melodic Perception and Memory for Pitch. *The Journal of Neuroscience*, 14 (4). 1908-1919.
- Zatorre R. J., Bouffard M., Ahad P. & Belin P. (2002).
Where is “where” in the human auditory cortex? *Nature Neuroscience*, 5 (2). 905-909.
- Zeki S., Watson J.D.G., Lueck C.J., Friston K.J., Kennard C. & Frackowiak R.S.J. (1991).
A Direct Demonstration of Functional Specialization in Human Visual Cortex. *The Journal of Neuroscience*, 11 (3). 641-649.
- Zerling J.-P. (1992).
Frontal lip shape for French and English vowels. *Journal of Phonetics*, 20. 3-14.
- Zhao J., Shu H., Zhang L., Wang X., Gong Q. & Li P. (2008).
Cortical competition during language discrimination. *NeuroImage*, 43. 624-633

Index des auteurs

Aboitiz	94
Abry	25, 43, 58
Alain	84
Amunts	53
Andersen	24, 70, 97
Anderson	54
Astésano	82
Aubin	44
Baddeley	79, 87
Baker	56, 57
Basso	57
Bates	57, 58
Baum	60
Beauchamp	70, 105, 209
Behne	64
Belin	61, 92, 93
Benoît	25
Benson	62, 101, 211
Bernstein	66, 81, 83, 97, 208, 209
Besle	81
Bimler	101
Binder	62, 64, 209, 211
Binkofski	76, 77, 106
Bledowski	39
Blumstein	56, 57, 63, 79
Boemio	61
Bonder	60
Bonmassar	39
Bradvik	60
Breitenstein	60
Broca	53, 54, 55, 56, 58, 59, 60, 68, 71, 72, 74, 77, 79, 80, 87, 99, 106, 212
Browman	46, 216
Brown	82
Buccino	76, 78, 105
Callan	68, 105, 208
Calliope	19
Calvert	65, 71, 97, 105, 207, 208
Campbell	65, 71, 101, 208
Caplan	57, 106
Castaigne	58
Cathiard	25, 26, 43, 107
Chomsky	20
Colby	77
Colin	24, 81, 82
Content	21

Courtney	84
Cutler	47
Cutting	101
Damasio	54, 57, 61
Dawson	28
Debener	38
Decety	91
Dehaene-Lambertz	85
Dekle	23
Delmas	12
Démonet	53, 59, 79, 82
Denes	56
Denkova	208
Desimone	65
Desjardins	64
Dijkerman	91
Donders	27
Dronkers	57, 58, 92
Dumay	21
Dupoux	50, 57
Edelman	35
Eimas	41, 101
Fadiga	75, 78, 105
Fayol	41
Ferrand	21, 47, 106
Ferrari	76, 105
Fiebach	72
Fisher	25
Fitch	93, 101
Fort	81, 82
Fougeron	6, 23
Fowler	23, 45, 216
Fridriksson	74
Friederici	72, 82, 92
Frost	42, 62
Fry	101
Gallese	12, 75, 76, 212, 217
Gandour	60, 72
Gentil	25, 74
Gerrits	101
Giard	81, 82
Giraud	74, 81, 82, 85
Goldstein	46, 62, 216
Gorno-Tempini	91
Griffiths	60, 63, 64
Grodzinsky	72, 77, 106
Hagoort	82
Hall	67, 105, 208

Haxby	65, 70, 84, 97
Hecker	26
Heider	6, 25
Hertz-Pannier	38
Hickok	14, 56, 78, 79, 89, 91, 92, 93, 94, 95, 96, 97, 103, 104, 105, 207, 211, 212, 216
Hietanen	23
Holender	41
Hoonhorst	81
Houdé	58
Hsieh	72
Humphries	92, 93
Iacobini	74, 76, 77, 105
Indefrey	93
Jacquemot	57, 94
Jakobson	20
Jäncke	63, 211
Joanisse	103
Jonas	59
Jones	69
Jun	23
Juszyk	101
Kandel	25, 42
Kavanagh	28
Klatt	49, 51, 102, 116, 117, 118
Klucharev	82
Köhler	77, 106
Kolinsky	47, 101
Krakow	38
Kuhl	41
Kutas	82
Lebib	82
Lefavrais	111
Lieberman	41, 45, 46, 78, 79, 85, 87, 89, 94, 100, 101, 217
Lichteim	54, 55
Liebenthal	64, 209, 211
Lindblom	43, 51, 216
Lisker	25, 116, 117, 118
Luce	49
MacDonald	12, 23, 66, 104, 215
Maddieson	21
Marslen-Wilson	47, 51, 102, 216
Martin	6, 70, 81, 83
Martinet	23, 43
Massaro	89, 90, 91, 97, 100, 216
Mayer	60
Mazoyer	58, 59, 211, 212
McClelland	48, 51, 102
McGlone	57

McGurk	12, 23, 24, 69, 82, 90, 104, 210, 215, 216
Mehler	21, 47, 50, 51, 59, 100, 102, 106, 216
Meltzoff	77
Metz-Lutz	56, 127, 211
Miceli	56
Miglioretti	57
Miller	69, 83, 93, 101, 209
Milner	97
Mishkin	91
Morton	89, 91
Mourand-Dornier	25, 74
Mulert	39
Näätänen	81
Narain	91
Nguyen	100, 103
Norris	47, 50, 51, 102
Novak	81
Obleser	63
Ogawa	35
Ojanen	72
Ojemann	64
Okada	93
Olson	69
Otzenberger	127, 129, 130
Pallier	21, 57
Paulesu	72, 80
Pekkola	66, 105, 207, 208
Penhune	66, 74
Perrett	65
Petersen	58
Pitt	47
Poepfel	14, 23, 56, 61, 82, 85, 89, 91, 92, 93, 94, 95, 96, 97, 103, 104, 105, 207, 211, 212, 216
Poldrack	72
Ponton	81, 83, 211
Postle	75
Price	56, 60, 64, 79, 212
Puce	65, 105, 208
Pulvermüller	73, 106
Rademacher	75
Rimol	63, 209, 211
Rizzolatti	12, 75, 76, 77, 78, 79, 87, 94, 105, 212
Robin	61
Rodd	94
Rosenblum	23
Ross	24
Roy	26, 35, 107
Saffran	56

Saint-Amour	24, 82
Sams	23, 24, 65, 70, 82
Samson	93
Scarborough	24
Schönwiesner	61
Schwartz	24, 39, 43, 51, 104, 216
Scott	68, 72, 79, 91, 92, 94
Seeck	38
Seidenberg	48
Sekiyama	23
Serniclaes	81, 101, 216
Shannon	93
Sherrington	12, 35
Skipper	73, 106
Smith	65, 74, 101, 105, 208
Sock	26, 44, 51
Soto-Faraco	24
Spitsyna	91
Steinberg	28
Stetson	100
Stevens	42, 51
Sumby	12, 24, 100, 104, 213, 215
Summerfield	24, 67
Tallal	93, 101
Tervaniemi	61
Tiippana	24, 70
Timsit	32, 33
Tong	105, 208
Trébuchon-Da Fonseca	81
Troille	26
Troubetzkoy	19
Tsao	65
Tuomainen	70
Ungerleider	74, 84, 91, 97
Van Atteveldt	209
Van Wassenhove	23, 82, 106, 108, 211
Vaughan	80, 82
Vaxelaire	5, 26
Vitevitch	49
Vouloumanos	86
Warren	60, 63, 92
Watkins	78
Weintraub	60
Wernicke	53, 54, 55, 56, 57, 59, 60, 91, 99
Wetsbury	63
Wilkinson	40
Wilshire	54
Wilson	48, 72, 74, 77, 106, 213

Wise	57, 59, 60, 68, 91, 92, 93, 132
Wong	62, 64, 72
Wood	41
Wright	69, 97, 105, 209
Wunderlich	64
Wytttenbach	101
Zaidel	57
Zatorre	60, 63, 66, 92, 211
Zeki	65, 105
Zerling	13
Zhao	60

Aire de Broca

54, 56, 58, 59, 60, 68, 72, 74, 77, 80, 87, 106, 212

Aire de Brodmann

10, 53, 56, 58, 59, 60, 62, 65, 67, 70, 71, 80, 105, 106, 108, 161, 162, 163, 166, 168, 171, 172, 174, 175, 177, 180, 181, 182, 185, 208, 211

Aire de Wernicke

56, 60

Audiovisuelle

11, 24, 26, 61, 64, 65, 67, 68, 69, 71, 73, 82, 83, 87, 97, 105, 106, 107, 108, 112, 207, 209, 210, 211, 212, 265, 267, 268

AV animée

105, 107, 108, 122, 132, 135, 136, 140, 141, 143, 144, 145, 146, 150, 152, 153, 154, 155, 156, 159, 160, 161, 163, 166, 168, 170, 171, 172, 173, 174, 175, 177, 180, 181, 182, 183, 184, 185, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 207, 208, 210, 211, 212, 214, 261, 267

AV fixe

105, 107, 108, 122, 135, 136, 140, 141, 143, 144, 145, 146, 147, 149, 150, 151, 152, 153, 154, 155, 156, 157, 159, 161, 162, 163, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 183, 184, 185, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 207, 208, 210, 211, 212, 214, 264, 267

BOLD

10, 35, 36, 37, 84, 112, 122, 131, 132, 134, 265

Bruit

11, 13, 21, 23, 24, 31, 40, 63, 67, 72, 77, 106, 113, 114, 119, 121, 123, 126, 131, 159, 213, 214, 216, 262

Cognitif

12, 14, 15, 27, 40, 46, 89, 91, 101, 217, 266

Cortex

29, 56, 57, 58, 60, 62, 64, 65, 66, 67, 68, 70, 71, 72, 73, 75, 76, 78, 79, 80, 83, 84, 85, 86, 87, 92, 94, 97, 105, 107, 133, 164, 167, 185, 195, 207, 208, 212, 213, 217, 219, 221, 222, 224, 228, 229, 230, 233, 240, 242, 243, 244, 247, 250, 265, 267

Cortex Auditif Primaire

56, 58, 60, 62, 65, 66, 68, 70, 83, 84, 105, 107, 133, 207, 208, 217, 267

Cortex moteur

56, 72, 73, 75, 76, 78, 79, 94, 164, 167, 265

Cortex prémoteur

68, 71, 72, 73, 87, 164, 167, 213

Couplage

36, 38, 40, 43, 123, 265

Discrimination

12, 24, 41, 47, 48, 55, 56, 57, 64, 74, 85, 104, 105, 106, 107, 108, 112, 113, 114, 119, 121, 122, 135, 136, 139, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 183, 185, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 202, 207, 209, 210, 211, 212, 213, 214, 217, 228, 229, 230, 235, 239, 249, 250, 259, 260, 261, 263, 264, 266, 267, 268

EEG

10, 11, 12, 13, 15, 26, 28, 29, 31, 33, 38, 39, 40, 65, 80, 83, 84, 85, 87, 106, 107, 111, 112, 113, 115, 119, 120, 121, 122, 123, 126, 127, 128, 129, 130, 135, 139, 144, 152, 189, 191, 207, 213, 217, 223, 234, 239, 245, 247, 262, 263, 264, 265, 266, 267

Électrode

29, 30, 128, 129, 130, 136, 137, 199, 200, 201, 202, 210

Fenêtre

14, 93, 127, 130, 136, 137, 199, 200, 201, 202, 210, 211, 216, 218, 247, 262, 267

Gyrus de Heschl

161, 163, 166, 168, 175, 183, 185, 208, 211

Gyrus Frontal Inférieur

10, 59, 67, 71, 73, 83, 84, 94, 95

Gyrus Temporal Moyen

10, 57, 60, 62, 63, 67, 68, 70, 94, 95, 133, 162, 165, 167, 169, 174

Gyrus Temporal Supérieur

10, 56, 59, 62, 63, 67, 68, 69, 73, 95, 96, 132, 208, 211

Intégration

40, 65, 67, 68, 70, 71, 72, 87, 89, 90, 91, 93, 97, 104, 105, 107, 207, 209, 210, 215, 216, 247, 265

IRMF

10, 12, 13, 26, 35, 36, 38, 39, 40, 61, 62, 69, 72, 75, 80, 85, 87, 93, 94, 99, 111, 115, 119, 123, 130, 189, 202, 207, 211, 217, 265, 266

Labialité

10, 13, 25, 26, 106, 121, 140, 141, 144, 145, 146, 149, 150, 152, 154, 155, 158, 159, 160, 161, 162, 163, 171, 173, 174, 175, 179, 183, 184, 185, 187, 189, 193, 196, 198, 199, 200, 201, 202, 211, 214, 260, 261, 263, 264, 267

Lexique mental

14, 15, 41, 47, 48, 49, 50, 51, 57, 72, 99, 100, 102, 103, 104, 215, 218, 265

Lieux d'articulation

13, 106, 107, 119, 121, 123, 140, 141, 143, 145, 150, 163, 165, 166, 171, 172, 173, 175, 176, 177, 183, 184, 185, 189, 193, 197, 198, 199, 200, 201, 202, 212, 214, 260, 261, 263, 264, 267

Modalité

12, 23, 24, 25, 43, 70, 82, 83, 97, 105, 106, 107, 108, 115, 122, 132, 135, 139, 140, 141, 144, 150, 153, 154, 156, 161, 163, 166, 168, 170, 173, 174, 175, 177, 180, 181, 182, 183, 184, 185, 189, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 207, 209, 210, 212, 214, 259, 261, 264, 266, 267

Modèle

14, 37, 47, 48, 49, 50, 56, 79, 89, 90, 91, 92, 95, 96, 97, 102, 103, 105, 108, 112, 129, 130, 131, 207, 211, 212, 216, 262, 266, 268

Mouvement

74, 105, 107, 108, 207, 208, 210, 213, 214

Neurones miroirs

12, 74, 76, 77, 78, 79, 94, 97, 100, 101, 105, 108, 212, 217, 243, 262, 265, 266, 268

Perturbation

13, 79, 106, 119, 121, 123, 141

Potentiels évoqués

11, 15, 28, 29, 30, 32, 33, 38, 40, 75, 78, 80, 81, 84, 85, 87, 106, 112, 124, 126, 128, 130, 135, 137, 189, 191, 192, 193, 194, 196, 197, 198, 199, 202, 210, 211, 217, 218, 247, 262, 264, 265, 266, 267

Région

29, 30, 53, 62, 69, 75, 76, 103, 105, 107, 132, 133, 134, 183, 184, 208, 210, 212, 213

Région d'intérêt

132, 133, 134, 183, 184

Sillon Temporal Inférieur

10, 70, 94, 95

Sillon Temporal Supérieur

10, 62, 63, 67, 69, 71, 83, 85, 95, 96, 167, 207, 209, 210, 211, 217, 267

Spt

10, 94, 95, 97, 105

Syllabe

21, 22, 23, 40, 45, 49, 50, 51, 63, 102, 115, 116, 118, 191, 193, 194, 216, 219, 240, 262, 264, 265

Tâche

11, 27, 28, 34, 40, 53, 61, 63, 64, 72, 73, 74, 82, 84, 85, 100, 101, 104, 106, 111, 113, 114, 119, 122, 126, 131, 135, 139, 183, 259

TEP

10, 34, 40, 58, 61, 87, 93, 99, 211

Théorie

12, 14, 21, 42, 43, 44, 45, 51, 69, 78, 79, 85, 89, 93, 94, 100, 103, 215, 217

Trait

13, 19, 20, 21, 25, 26, 40, 42, 46, 48, 51, 89, 90, 93, 101, 105, 119, 139, 143, 196, 207, 213, 216,

Visème

23, 24, 25, 26, 40, 107, 115, 121, 207, 214

Voisement

13, 20, 23, 48, 63, 81, 106, 107, 117, 119, 121, 123, 140, 141, 143, 145, 149, 150, 152, 154, 155, 157, 160, 166, 167, 168, 172, 173, 177, 178, 179, 183, 184, 185, 187, 189, 193, 198, 200, 201, 202, 211, 212, 214, 216

VOT

10, 13, 63, 81, 82, 115, 116, 117, 118, 159, 187, 216

Index des tableaux

Tableau B1 : Âges et nombre de participants par environnement expérimental.	112
Tableau B2 : Durées segmentales et de VOT en millisecondes et valeurs des trois premiers formants en hertz des syllabes phonologiques.	116
Tableau B3 : Les paires syllabiques utilisées pour la tâche de discrimination à choix forcé en fonction du trait qui les oppose.	119
Tableau B4 : Scores de discrimination et écarts-types en pourcentage, pour les quatre types de paires, en fonction de la modalité dans les trois environnements.	139
Tableau B5 : Temps de réponse et écarts-types en millisecondes, pour les quatre types de paires, en fonction de la modalité dans les trois environnements.	140
Tableau B6 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les 55 participants, dans l'environnement silencieux.	141
Tableau B7 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 55 participants, dans l'environnement silencieux.	142
Tableau B8 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires dans les deux modalités, pour les 55 participants, dans l'environnement silencieux.	143
Tableau B9 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les 16 participants, dans l'environnement perturbé.	145
Tableau B10 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 16 participants, dans l'environnement perturbé.	145
Tableau B11 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires en fonction des deux modalités, chez les 16 participants, dans l'environnement perturbé.	147
Tableau B12 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les 13 participants, dans l'environnement bruyant.	149
Tableau B13 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 13 participants, dans l'environnement bruyant.	149

Tableau B14 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires en fonction des deux modalités, chez les 13 participants, dans l'environnement bruyant.	151
Tableau B15 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les 16 participants.	153
Tableau B16 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements silencieux et perturbé, chez les 16 participants.	153
Tableau B17 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 16 participants.	154
Tableau B18 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les 13 participants.	155
Tableau B19 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements silencieux et bruyant, chez les 13 participants.	155
Tableau B20 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 13 participants.	156
Tableau B21 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, en modalités AVf et AVa, chez les sept participants.	157
Tableau B22 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les environnements perturbé et bruyant, chez les sept participants.	157
Tableau B23 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les sept participants.	158
Tableau B24 : Différences significatives lors de la comparaison des scores de discrimination (en pourcentage) dans les différents environnements.	160
Tableau B25 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur la labialité.	162
Tableau B26 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur les lieux d'articulation.	165
Tableau B27 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur le voisement.	167
Tableau B28 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires non phonologiques.	169
Tableau B29 : Régions cérébrales significativement activées, pour les paires s'opposant sur la labialité, lors de la soustraction « AVa-AVf ».	171
Tableau B30 : Régions cérébrales significativement activées, pour les paires s'opposant sur les lieux d'articulation, lors de la soustraction « AVa - AVf ».	172

Tableau B31 : Régions cérébrales significativement activées, pour les paires s'opposant sur le voisement, lors de la soustraction « AVa-AVf ».	172
Tableau B32 : Régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur la labialité-paires non phonologiques ».	174
Tableau B33 : Régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur les lieux d'articulation-paires non phonologiques ».	176
Tableau B34 : Régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur le voisement-paires non phonologiques ».	178
Tableau B35 : Tableau synoptique des activations en modalité AV animée	180
Tableau B36 : Tableau synoptique des activations en modalité AV animée	181
Tableau B37 : Résultats des ANOVA à trois facteurs (modalités : 2 niveaux ; hémisphères : 2 niveaux ; Régions d'intérêt : 4 niveaux).	186
Tableau B38: Résultats des ANOVA à trois facteurs (modalités : 2 niveaux ; hémisphères : 2 niveaux ; Paires syllabiques : 3 niveaux).	187
Tableau B39 : Résultats des ANOVA à deux facteurs, pour nous deux modalités (Régions d'intérêt : 4 niveaux ; Paires syllabiques : 3 niveaux).	187
Tableau B40 : Moyennes et écarts-types des performances et temps de réponse de discrimination des 11 sujets.	190
Tableau B41 : Récapitulatif des résultats des ANOVA à deux facteurs (modalités : 2 niveaux ; électrodes : 7 niveaux), effectuées sur les valeurs absolues de surfaces des courbes de réponses évoquées.	203

Index des figures

Figure A1 : Représentation schématique de la structure syllabique.	22
Figure A2 : Représentation schématique de la propagation d'un potentiel d'action. La flèche horizontale représente le dipôle électrique.	30
Figure A3 : Représentation schématique de l'emplacement des électrodes sur le scalp dans le système « 10-20 ».	31
Figure A4 : Extraction des potentiels évoqués par la technique du moyennage.	32
Figure A5 : Ondes négatives et positives observées sur les « grandes moyennes » issues des données EEG.	33
Figure A6 : Illustration de la méthode de la soustraction cognitive.	37
Figure A7 : Représentation schématique d'un paradigme en bloc et d'un paradigme événementiel.	38
Figure A8 : Le schéma de Lichteim contenant les trois « modules » et leurs connexions.	55
Figure A9 : Homoncules sensoriel et moteur de Penfield.	75
Figure A10 : Le système des neurones miroirs chez l'être humain.	77
Figure A11 : Représentation schématique du modèle de deux voies de traitement du langage.	95
Figure A12 : Localisations anatomiques des composantes du modèle de deux voies de traitement du langage.	96
Figure B1 : Oscillogramme et spectrogramme du bruit produit lors des gradients de l'IRM, avec l'agrandissement d'une période et les valeurs des trois premiers formants en Hz.	114
Figure B2 : Cadrage sur le bas du visage de la locutrice articulant une syllabe.	116
Figure B3 : Vues des oscillogrammes accompagnés des spectrogrammes des huit syllabes faisant partie du système phonologique du français contemporain et utilisées pour réaliser les paires minimales étudiées.	117
Figure B4 : Oscillogramme et spectrogramme d'une syllabe naturelle modifiée par ordinateur, d'une durée de 360 ms, utilisée afin de réaliser les paires de stimuli non phonologiques.	118
Figure B5 : Représentation schématique de la tête d'un sujet placée dans l'antenne de l'IRM.	120
Figure B6 : Représentation schématique du dispositif d'acquisition des données.	121
Figure B7 : Représentation schématique d'un paradigme événementiel	124
Figure B8 : Représentation schématique de la présentation d'un essai constitué d'une paire phonologique /pi py/ entre deux périodes de gradients de l'IRM.	125
Figure B9 : Exemple pour cinq électrodes d'une fenêtre d'acquisition des signaux EEG, durant une passation IRMf.	127
Figure B10 : Photographies d'un sujet au cours et à la fin de la pose des électrodes.	128

Figure B11 : Illustration de la méthode de filtrage de l'artefact cardiaque des données EEG acquises simultanément lors de la passation IRMf.	130
Figure B12 : Les huit régions d'intérêt accompagnées de leurs coordonnées et de la taille du volume en mm.	134
Figure B13 : Fenêtres d'analyse de surface de l'aire à partir des courbes des réponses électriques en AVa.	137
Figure B14 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 55 participants, dans l'environnement silencieux.	142
Figure B15 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 55 participants, dans l'environnement silencieux.	144
Figure B16 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires, chez les 16 participants, dans l'environnement perturbé.	146
Figure B17 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 16 participants.	148
Figure B18 : Moyennes et écarts-types des scores de discrimination en pourcentage, et des temps de réponse en millisecondes, pour les paires (les deux modalités confondues), pour les 13 participants.	150
Figure B19 : Moyennes et écarts-types des scores de discrimination en pourcentage, pour les paires en fonction des deux modalités, pour les 13 participants.	152
Figure B20 : Rendus de régions cérébrales significativement activées en AVf et AVa lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur la labialité.	163
Figure B21 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « stimuli-événement nul » pour les paires s'opposant sur les lieux d'articulation.	166
Figure B22 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires s'opposant sur le voisement.	168
Figure B23 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « stimuli-événement nul », pour les paires non phonologiques.	170
Figure B24 : Rendus des régions cérébrales significativement activées, lors de la soustraction « AVa – AVf », à gauche pour la labialité, au centre pour les lieux d'articulation et à droite pour le voisement.	173
Figure B25 : Rendus des régions cérébrales significativement activées en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur la labialité-paires non phonologiques ».	175
Figure B26 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s'opposant sur les lieux d'articulation-paires non phonologiques ».	177

Figure B27 : Rendus des régions cérébrales significativement activées, en AVf et AVa, lors de la soustraction « paires phonologiques s’opposant sur la labialité-paires non phonologiques ».	179
Figure B28 : Scores de discrimination en pourcentage pour les 11 sujets EEG	191
Figure B29 : Grandes Moyennes des potentiels évoqués par la discrimination syllabique, en modalités AVf et AVa, pour les trois paires phonologiques confondues.	192
Figure B30 : Les cartes d’amplitude obtenues à 110 ms, 220 ms et 300 ms après le début de la seconde syllabe, en modalités AVf et AVa, pour les trois paires phonologiques confondues.	193
Figure B31 : Grandes moyennes pour les 11 sujets, sur les 19 électrodes, pour les trois paires phonologiques en modalité AV fixe.	194
Figure B32 : Grandes moyennes, pour les 11 sujets, sur les 19 électrodes, pour les trois paires phonologiques en modalité AVa.	195
Figure B33 : Grandes moyennes des potentiels évoqués par la discrimination du trait de labialité chez les 11 sujets en fonction des modalités AVf et AVa.	196
Figure B34 : Grandes moyennes des potentiels évoqués par la discrimination des lieux d’articulation chez des 11 sujets en fonction des modalités AVf et AVa.	197
Figure B35 : Grandes moyennes des potentiels évoqués par la discrimination du voisement chez les 11 sujets en fonction des modalités AVf et AVa.	198

Tables des matières

Sommaire	9
Introduction	11
A. Partie théorique	17
A.1. De la phonétique à la neuro-imagerie : unités de bases et outils expérimentaux	19
A.1.1. Le phonème et ses traits	19
A.1.2. La syllabe	21
A.1.3. La perception visuelle de la parole et les visèmes	23
A.1.4. Les méthodes et techniques d'investigation du fonctionnement cérébral	26
A.1.4.1. La méthode anatomoclinique	27
A.1.4.2. La chronométrie mentale : les temps de réaction ou de réponse	27
A.1.4.3. La méthode psychophysiological : les potentiels évoqués	28
A.1.4.3.1. Courants cellulaires et potentiels de surface	29
A.1.4.3.2. Les potentiels évoqués	30
A.1.4.4. L'imagerie cérébrale fonctionnelle	34
A.1.4.4.1. La Tomographie par Émission de Positons (TEP)	34
A.1.4.4.2. L'Imagerie par Résonance Magnétique fonctionnelle	35
A.1.4.4.2.1. L'effet <i>BOLD</i>	35
A.1.4.4.2.2. La soustraction cognitive	36
A.1.4.4.2.3. Paradigme en bloc et paradigme événementiel	37
A.1.4.4.5. Le couplage de l'IRMf et de l'EEG	38
A.2. Théories phonétiques de la perception de la parole et modèles d'accès au «lexique mental»	41
A.2.1. Les théories de la perception de la parole articulée	42
A.2.1.1. Théorie auditive : la Théorie Quantique	42
A.2.1.2. Les théories interactionnelles	43
A.2.1.2.1. La Théorie de la Variabilité Adaptative	43
A.2.1.2.2. La Théorie de la Perception pour le Contrôle de l'Action	43
A.2.1.2.3. La Théorie de la Viabilité en production-perception de la parole	44
A.2.1.3. Les théories motrices	44
A.2.1.3.1. La Théorie de la Perception Directe	45
A.2.1.3.2. La Théorie Motrice de la Perception de la parole	45
A.2.1.3.3. La Phonologie Articulatoire	46
A.2.2. Les principaux modèles d'accès au lexique mental	47
A.2.2.1. COHORTE	47
A.2.2.2. TRACE	48
A.2.2.3. NAM	49
A.2.2.4. LAFS	49
A.2.2.5. SARAH	49
A.2.2.6. SHORTLIST	50
A.3. Les mécanismes cérébraux de la perception de la parole	53
A.3.1. Les apports des précurseurs	53
A.3.1.1. L'apport de l'étude des cas pathologiques	56
A.3.2. Les études en Tomographie par Émission de Positons	58
A.3.3. Les études en Imagerie par Résonance Magnétique fonctionnelle	62
A.3.3.1. La parole audiovisuelle et le cortex auditif	64
A.3.3.2. Un site d'intégration audiovisuelle	68
A.3.3.3. Perception de la parole et cortex moteur	72
A.3.3.3.1. Le système des neurones miroirs : l'intrication de l'action et de la perception	74
A.3.3.3.2. Les différents neurones miroirs	76
A.3.3.3.3. L'intérêt des neurones miroirs pour les neurolinguistes	77

A.3.4. Les études en potentiels évoqués	80
A.3.5. Les études couplant des données en EEG et en IRMf	84
A.4. Modèles cognitif et anatomo-fonctionnel de la perception de la parole	89
A.4.1. Le modèle de perception à logique floue de Massaro	89
A.4.2. Le modèle à deux voies de Hickok et Poeppel	91
A.4.2.1. Les premières étapes du modèle à deux voies de Hickok & Poeppel	92
A.4.2.1.1. Traitement acoustique : « l'analyse spectrotemporelle »	92
A.4.2.1.2. Traitement phonétique : « le réseau phonologique »	93
A.4.2.2. La séparation en deux voies	93
A.4.2.2.1. La voie ventrale	94
A.4.2.2.2. La voie dorsale	94
A.5. Conclusion et hypothèses	99
B. Partie expérimentale	109
B.1. Matériel et Méthode	111
B.1.1 Les participants	111
B.1.2. Paradigme de discrimination à choix forcé	112
B.1.2.1. Les stimuli	115
B.1.2.2. Corpus	116
B.1.2.3. Les paires syllabiques	119
B.1.3. Les environnements expérimentaux	119
B.1.3.1. Environnement perturbé (IRMf/EEG)	120
B.1.3.2. Environnement silencieux	122
B.1.3.3. Environnement bruyant	123
B.1.4. Recueil des données IRMf/EEG	123
B.1.4.1. Paradigme événementiel	123
B.1.4.2. Acquisition et prétraitement des images IRMf	125
B.1.4.3. Acquisition et prétraitement du signal EEG recueilli pendant l'IRMf	126
B.1.4.4. Analyses des données IRMf et EEG	130
B.1.4.4.1. Analyses des images IRMf	131
B.1.4.4.2. Analyse en régions d'intérêt	132
B.1.4.4.3. Évaluation des potentiels évoqués	135
B.1.4.4.4. Analyse des potentiels évoqués	135
B.2. Résultats : Données comportementales	139
B.2.1. Les performances dans l'environnement silencieux	141
B.2.1.1. Les modalités dans l'environnement silencieux	141
B.2.1.2. Les types de paires dans l'environnement silencieux	141
B.2.1.3. Les types de paires en fonction des modalités dans l'environnement silencieux	143
B.2.2. Les performances dans l'environnement perturbé (IRM)	144
B.2.2.1. Les modalités dans l'environnement perturbé	144
B.2.2.2. Les types de paires dans l'environnement perturbé	145
B.2.2.3. Les types de paires en fonction des modalités dans l'environnement perturbé	146
B.2.3. Les performances dans l'environnement bruyant	148
B.2.3.1. Les modalités dans l'environnement bruyant	148
B.2.3.2. Les paires syllabiques dans l'environnement bruyant	149
B.2.3.3. Les paires syllabiques selon la modalité dans l'environnement bruyant	150
B.2.4. Comparaisons binaires entre les trois environnements	152
B.2.4.1. Comparaison entre les environnements silencieux et perturbé	153
B.2.4.1.1. Les modalités	153
B.2.4.1.2. Les environnements silencieux et perturbé	153
B.2.4.1.3. Les paires de stimuli dans les environnements silencieux et perturbé	154
B.2.4.2. Comparaison entre les environnements silencieux et bruyant	154

B.2.4.2.1. Les modalités	154
B.2.4.2.2. Les environnements silencieux et bruyant	155
B.2.4.2.3. Les paires syllabiques	155
B.2.4.3. Comparaison entre les environnements perturbé et bruyant	156
B.2.4.3.1. Les modalités	156
B.2.4.3.2. Les environnements perturbé et bruyant	157
B.2.4.3.3. Les paires syllabiques	157
B.2.5. Synthèse	159
B.3. Résultats : IRMf	161
B.3.1. Résultats des analyses de groupe en effet aléatoire du cerveau entier	161
B.3.1.1. Activations cérébrales engendrées par la discrimination des paires syllabiques	161
B.3.1.1.1. Discrimination des paires s'opposant sur la labialité	161
B.3.1.1.2. Discrimination des paires s'opposant sur les lieux d'articulation	163
B.3.1.1.3. Discrimination des paires s'opposant sur le voisement	166
B.3.1.1.4. Discrimination des paires non phonologiques	168
B.3.2. Comparaison de la discrimination en présentation AV fixe et AV animée	170
B.3.2.1. Discrimination des paires s'opposant sur la labialité	171
B.3.2.2. Discrimination des paires s'opposant sur les lieux d'articulation	171
B.3.2.3. Discrimination des paires s'opposant sur le voisement	172
B.3.3. Les activations engendrées par la discrimination des stimuli phonologiques par rapport aux non phonologiques	173
B.3.3.1. Comparaison de la discrimination des paires s'opposant sur la labialité et des paires non phonologiques	174
B.3.3.2. Comparaison de la discrimination des paires s'opposant sur les lieux d'articulation et des paires non phonologiques	175
B.3.3.3. Comparaison de la discrimination des paires s'opposant sur le voisement et des paires non phonologiques	177
B.3.4. Tableaux récapitulatifs des activations en IRMf	180
B.3.5. Analyses en régions d'intérêt	183
B.3.6. Synthèse	185
B.4. Résultats : Potentiels évoqués	189
B.4.1. Résultats comportementaux du groupe de onze sujets « EEG »	189
B.4.2. Les Grandes Moyennes issues des potentiels évoqués individuels	191
B.4.2.1. Les potentiels évoqués en modalité AV fixe	193
B.4.2.2. Les potentiels évoqués en modalité AV animée	194
B.4.2.2.1. Discrimination de la labialité	196
B.4.2.2.2. Discrimination des lieux d'articulation	197
B.4.2.2.3. Discrimination du voisement	198
B.4.3. Résultats des analyses statistiques sur les potentiels évoqués	199
B.4.3.1. Analyses de la fenêtre de 110 ms à 220 ms	199
B.4.3.1.1. Analyses de chaque paire syllabique	200
B.4.3.2. Analyses de la fenêtre de 220 ms à 300 ms	200
B.4.3.2.1. Analyses de chaque paire syllabique	201
B.4.4. Synthèse	202
C. Discussion, Conclusion et Perspectives	205
C.1. Discussion :	207
C.1.1. La perception des mouvements orofaciaux phonologiques	207
C.1.1.1. Perception visuelle et cortex auditif primaire	208
C.1.2. Intégration audiovisuelle et le sillon temporal supérieur	209
C.1.3. Le timing de la perception audiovisuelle de la parole	210
C.1.4. La parole audiovisuelle et les deux voies du modèle de Hickok & Poeppel	211
C.1.5. Perception audiovisuelle et réseau des neurones miroirs	212

C.1.6. L'impact des indices visuels sur la discrimination auditive des traits phonétiques	213
C.2. Conclusion et Perspectives	215
Bibliographie	219
Index des auteurs	251
Index des notions	257
Index des tableaux	261
Index des figures	265
Annexes	273

Annexes



Figure A. : En haut, face interne de l'hémisphère droit ; en bas, face externe de l'hémisphère gauche

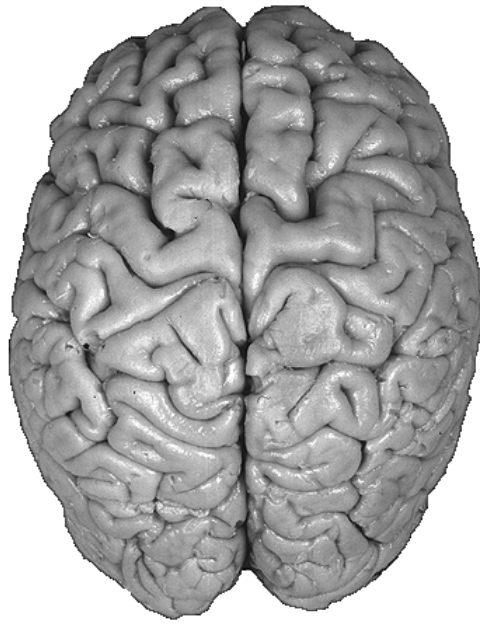


Figure B : Vue du dessus des hémisphères corticaux

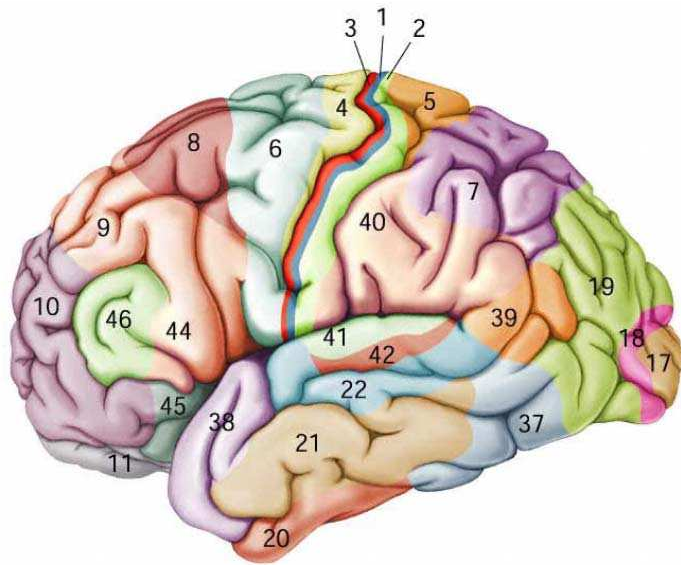


Figure C : Les zones de Brodmann

Résumé

Si, comme le disait Stetson (1951), la parole est davantage un ensemble de mouvements rendus audibles qu'un ensemble de sons produit par des mouvements, qu'en est-il de la dimension visible de ces mouvements ?

L'intelligibilité de la parole est améliorée par l'apport des visèmes, dans un environnement bruyant (Sumby & Pollack, 1954), ce qui nous indique un effet facilitateur sur l'accès au lexique mental. Selon Massaro (1990), l'influence d'une source est maximale si une autre est ambiguë. Pourtant dans l'effet « McGurk » (McGurk & MacDonald, 1976), les deux sources, audible et visible sont équipotentes. Le résultat est bien une perturbation, puisque la « cible » a été mal interprétée. Certains chercheurs ont montré que l'effet McGurk pouvait avoir lieu sur des mots et même sur des phrases. Cela corrobore l'idée que l'intégration de la source visuelle est précoce et prélexicale, mais aussi sujette à l'influence d'effet *top-down*.

Nous avons mené une étude anatomo-fonctionnelle simultanée en IRMf/EEG, en utilisant une tâche de discrimination à choix forcé, portant sur des syllabes CV, selon deux modalités perceptives : audiovisuelle animée et audiovisuelle fixe, afin de pouvoir observer les bases neurophysiologiques de la perception audiovisuelle syllabique. La tâche de discrimination portait sur des paires syllabiques s'opposant sur les trois traits suivants : la labialité vocalique, le lieu d'articulation et le voisement consonantiques.

Mots clés : Neurophysiologie, neurophonétique, parole, perception audiovisuelle, visèmes, syllabes, IRMf, EEG, labialité, lieu d'articulation, voisement.

Abstract

If, as Stetson said (1951), speech is rather a set of movements made audible than a set of sounds produced by movements, what about the visible aspects of these movements ?

In a noisy environment, speech intelligibility is improved by perceiving a speaker's face (Sumby & Pollack, 1954), a dimension which seemingly involves a facilitation effect in accessing the mental lexicon. Massaro (1990) assumes that the influence of one source of information is greatest when the other source is neutral or ambiguous. However, the McGurk effect suggests that audible and visible sources have an equal impact on the speech perception system (McGurk & MacDonald, 1976). The result is indeed a perturbation, in terms of misperception of the "target". Several studies claim that the McGurk effect operates on the lexical level as well as on word or phrasal levels. Taken together, previous studies indicate that the bimodal integration of the visual source is early and prelexical, moreover it could be influenced by a top-down effect.

We conducted a study with simultaneous recordings in fMRI/EEG, in a discrimination task, comprising CV syllables, in two perception modalities: audiovisual and audio only, in order to investigate the neural substrates of audiovisual syllabic perception. The discrimination task was based on syllable pairs, contrasting three features: vowel lip rounding, consonant place of articulation and voicing.

Keywords: Neurophysiology, neurophonetics, speech, audiovisual perception, visemes, syllables, fMRI, EEG, rounding, place of articulation, voicing.