

THÈSE DE DOCTORAT DE L'UNIVERSITÉ LUMIÈRE LYON 2

ECOLE DOCTORALE DE SCIENCES COGNITIVES

Présentée par Julien Besle

Pour obtenir le grade de Docteur de l'Université Lyon 2

Spécialité : Sciences Cognitives - Mention : Neurosciences

Interactions audiovisuelles dans le cortex auditif chez l'homme

Approches électrophysiologique et comportementale

Soutenance publique le 22 mai 2007 devant le jury composé de :

M^r Pascal Barone (Examineur)

M^{me} Nicole Bruneau (Rapporteur)

M^r Jean-Luc Schwartz (Rapporteur)

M^{me} Marie-Hélène Steiner-Giard (Directrice de thèse)

M^r Rémy Versace (Examineur)

Table des matières

I	Revue de la littérature	3
1	Convergence audiovisuelle en neurophysiologie	5
1.1	Aires associatives corticales	5
1.1.1	Études électrocorticographique (ECoG) de la convergence multisensorielle	5
1.1.2	Convergence audiovisuelle au niveau du neurone unitaire	8
1.1.3	Aires de convergence dans le cortex frontal	9
1.1.4	Effet de l'anesthésie sur les interactions multisensorielles	9
1.2	Convergence audiovisuelle dans le cortex visuel	10
1.3	Convergence corticale chez l'homme	11
1.4	Convergence sous-corticale	12
1.4.1	Colliculus Supérieur / Tectum optique	13
1.4.2	Autres structures sous-corticales	16
1.5	Études anatomiques de la convergence multisensorielle	17
1.6	Conclusion	19
2	Interactions Audiovisuelles en psychologie	21
2.1	Effets intersensoriels sur les capacités perceptives	22
2.1.1	Effets dynamogéniques	22
2.1.2	Modèles explicatifs de l'effet dynamogénique	22
2.1.3	Effet dynamogénique et théorie de la détection du signal	24
2.1.4	Modèles de détection d'un stimulus bimodal au seuil	24
2.2	Correspondance des dimensions synesthésiques	25
2.2.1	Établissement des dimensions synesthésiques	26
2.2.2	Réalité des correspondances synesthésiques	27
2.2.3	Correspondance des intensités	29
2.2.4	Résumé	30
2.3	Temps de réaction audiovisuels	31
2.3.1	Premières études	31
2.3.2	Paradigme du stimulus accessoire	33
2.3.3	Paradigme d'attention partagée	36
2.4	Conflit des indices spatiaux auditifs et visuels	42
2.4.1	Ventriloquie	43
2.4.2	Facteurs influençant l'effet de ventriloquie	45
2.4.3	Niveau des interactions dans l'effet de la ventriloquie	46

2.5	Conflit des indices temporels	47
2.6	Conclusion	48
3	Perception audiovisuelle de la parole	49
3.1	Contribution visuelle à l'intelligibilité	49
3.1.1	Complémentarité des informations auditives et visuelles de parole	50
3.1.2	Redondance des informations auditives et visuelles de parole	51
3.1.3	Facteurs liés à la connaissance de la langue	51
3.2	Effet McGurk	52
3.2.1	L'hypothèse VPAM	53
3.2.2	Intégration audiovisuelle pré-phonologique	54
3.2.3	Influence des facteurs linguistiques et cognitifs	55
3.3	Facteurs spatiaux et temporels	56
3.4	Modèles de perception de la parole audiovisuelle	58
3.4.1	Modèles post-catégoriels	58
3.4.2	Modèles pré-catégoriels	60
3.5	Conclusion	61
4	Intégration AV en neurosciences cognitives	63
4.1	Comportements d'orientation	63
4.1.1	Orientation vers un stimulus audiovisuel chez l'animal	64
4.1.2	Saccades oculaires vers un stimulus audiovisuel, chez l'homme	65
4.1.3	Expériences chez l'animal alerte et actif	66
4.2	Effet du stimulus redondant	67
4.2.1	Premières études	67
4.2.2	Tâches de discrimination	67
4.2.3	Tâche de détection	68
4.3	Perception des émotions	69
4.4	Objets écologiques audiovisuels	70
4.5	Conditions limites de l'intégration AV	71
4.6	Illusions audiovisuelles	72
4.6.1	Intégration audiovisuelle pré-attentive	72
4.6.2	Application du modèle additif	73
4.6.3	Activités corrélées à une illusion audiovisuelle	74
4.7	Perception audiovisuelle de la parole	74
4.8	Conclusion	77
5	Problématique générale	79
II	Méthodes	81
6	Approches électrophysiologiques	83
6.1	Bases physiologiques des mesures (s)EEG/MEG	83
6.2	ElectroEncéphaloGraphie (EEG)	84

6.2.1	Enregistrement	84
6.2.2	Analyse des potentiels évoqués (PE)	86
6.3	MagnétoEncéphaloGraphie (MEG)	90
6.3.1	Champs magnétiques cérébraux	90
6.3.2	Procédure d'enregistrement	91
6.4	StéréoElectroEncéphaloGraphie (sEEG)	92
6.4.1	Localisation des électrodes	92
6.4.2	Procédure d'enregistrement	93
6.4.3	Calcul du PE et rejet d'artéfacts	94
6.4.4	Résolution spatiale et représentation spatiotemporelle	94
6.4.5	Étude de groupe et normalisation anatomique	95
7	Approche méthodologique de l'intégration AV	99
7.1	Falsification de l'inégalité de Miller	99
7.1.1	Bases mathématiques et postulats	99
7.1.2	Application de l'inégalité	102
7.1.3	Biais potentiels	104
7.1.4	Analyse statistique de groupe	105
7.2	Modèle additif	106
7.2.1	Falsification du modèle additif en EEG/MEG	107
7.2.2	Interprétation des violations de l'additivité en EEG/MEG	109
7.2.3	Comparaison avec le critère d'additivité en IRM fonctionnelle	109
8	Méthodes statistiques en (s)EEG/MEG	111
8.1	Tests multiples	111
8.2	Tests Statistiques sur les données individuelles	113
8.2.1	Tests sur les essais élémentaires	113
8.2.2	Test du modèle additif par randomisation pour des données non ap- pariées	114
8.2.3	Remarques	115
III Interactions audiovisuelles dans la perception de la parole		117
9	Étude en EEG et comportement	119
9.1	Rappel de la problématique	119
9.2	Méthodes	120
9.2.1	Sujets	120
9.2.2	Stimuli	120
9.2.3	Procédure	121
9.2.4	Expérience comportementale complémentaire	122
9.2.5	Analyse des résultats	122
9.3	Résultats	123
9.3.1	Résultats comportementaux	123
9.3.2	Résultats électrophysiologiques	123

9.4	Discussion	125
9.4.1	Comportement	125
9.4.2	Résultats électrophysiologiques	127
10	Étude en sEEG	131
10.1	Introduction	131
10.2	Méthodes	134
10.2.1	Patients	134
10.2.2	Stimuli et procédure	134
10.2.3	Calcul des potentiels évoqués	134
10.2.4	Analyses statistiques	135
10.3	Résultats	136
10.3.1	Données comportementales	136
10.3.2	Réponses évoquées auditives	136
10.3.3	Réponses évoquées visuelles	138
10.3.4	Violations du modèle additif	141
10.3.5	Relations entre réponses auditives, visuelles et interactions audiovisuelles	144
10.4	Discussion	145
10.4.1	Activité du cortex auditif en réponse aux indices visuels de parole	146
10.4.2	Interactions audiovisuelles	149
10.4.3	Comparaison avec l'expérience EEG de surface	151
11	Effet d'indigage temporel	153
11.1	Introduction	153
11.2	Expérience comportementale 1	155
11.2.1	Méthodes	156
11.2.2	Résultats	159
11.2.3	Discussion	162
11.3	Expérience comportementale 2	163
11.3.1	Méthodes	164
11.3.2	Résultats	166
11.3.3	Discussion	169
11.4	Discussion générale	170
IV	Interactions audiovisuelles en mémoire sensorielle	173
12	Introduction générale	175
12.1	MMN Auditive	175
12.2	Rappel de la problématique	176
13	Étude comportementale	179
13.1	Introduction	179
13.2	Méthodes	180

13.2.1	Sujets	180
13.2.2	Stimuli	180
13.2.3	Procédure	181
13.2.4	Analyses	182
13.3	Résultats	182
13.4	Discussion	183
14	Additivité des MMNs auditives et visuelles	185
14.1	Introduction	185
14.2	Méthodes	187
14.2.1	Sujets	187
14.2.2	Stimuli	187
14.2.3	Procédure	187
14.2.4	Analyses	188
14.3	Résultats	188
14.4	Discussion	191
15	Représentation auditive d'une régularité AV	195
15.1	Introduction	195
15.2	Méthodes	196
15.2.1	Sujets	196
15.2.2	Stimuli	196
15.2.3	Procédure	197
15.2.4	Analyses	197
15.3	Résultats	198
15.4	Discussion	201
16	MMN à la conjonction audiovisuelle	205
16.1	Introduction	205
16.2	Méthodes	207
16.2.1	Sujets	207
16.2.2	Stimuli	207
16.2.3	Procédure	207
16.2.4	Analyses	208
16.3	Résultats	208
16.4	Expérience comportementale complémentaire	210
16.5	Discussion	211
V	Discussion générale	215
17	Discussion générale	217
17.1	Interactions audiovisuelles précoces dans la perception de la parole	217
17.2	Représentation d'un évènement audiovisuel en mémoire sensorielle auditive	218
17.3	Interactions audiovisuelles dans le cortex auditif	219

A Données individuelles des patients	223
B Articles	239
Bibliographie	287

Cinquième partie
Discussion générale

Chapitre 17

Discussion générale

17.1 Interactions audiovisuelles précoces dans la perception de la parole

Notre première expérience sur la perception audiovisuelle de la parole (chapitre 9 page 119) a montré, d'une part, que le temps de traitement de la parole auditive était diminué par la vision des mouvements articulatoires associés, sans que cette diminution ne puisse s'expliquer dans un modèle de traitement séparé des informations auditives et visuelles. D'autre part, ce gain comportemental semblait associé à une diminution de l'activité auditive entre 120 ms et 200 ms de traitement.

Notre seconde expérience chez le patient épileptique (chapitre 10 page 131) a montré que cet effet n'est pas le seul à prendre place dans les 200 premières millisecondes de traitement de la syllabe auditive. Les effets les plus reproductibles correspondent, d'une part, à une activation du cortex auditif par les indices visuels de parole et, d'autre part, à une modulation (essentiellement une diminution) du traitement des indices auditifs dans le cortex auditif, dont une partie pourrait correspondre aux effets observés en EEG de scalp. L'activation du cortex auditif par les mouvements articulatoires de la parole semble avoir lieu directement après le traitement de ces stimuli dans les aires visuelles (bien que la couverture spatiale éparse des électrodes intracérébrales ne permette pas de conclure définitivement sur ce point) et correspondrait donc à une activation *feedforward* (en termes temporels). Rappelons que l'information visuelle est disponible avant l'information auditive dans les stimuli de paroles. Cette activation visuelle du cortex auditif pourrait ensuite permettre la modulation du traitement phonétique des syllabes auditives dans le cortex auditif, et ce à des étapes de traitement relativement précoces (à partir de 50 ms). Le cortex auditif primaire semble relativement épargné par ces phénomènes.

Dans notre discussion de l'expérience en EEG de scalp (page 127), nous avons proposé plusieurs possibilités d'interprétation des violations de l'additivité :

- activation du cortex auditif par les indices visuels.
- effet d'indigence temporel intersensoriel.
- intégration des informations phonétiques auditives et visuelles à un stade pré-phonologique.
- amorçage phonologique intersensoriel.

Notre expérience en sEEG a bien montré que l'on pouvait dissocier les activations du cortex auditif par les stimuli visuels, des modulations de l'activité auditive sous influence visuelle, et donc que l'activation visuelle du cortex auditif ne pouvait expliquer la violation de l'additivité dans l'étude en EEG de scalp, ni a fortiori des autres violations du modèle additif en sEEG.

Nos expériences comportementales (chapitre 11 page 153) avaient pour but de tester l'existence d'un effet d'indigage temporel intersensoriel sur le TR, c'est-à-dire des interactions audiovisuelles dans la perception de la parole ne reposant pas sur une intégration phonétique ou phonologique. Les résultats montrent que, si un tel effet existe et peut être mis en évidence, il s'observe uniquement lorsque la performance est diminuée par la présence de bruit. Il est donc peu probable que l'indigage temporel intersensoriel explique à lui seul le gain de TR de la première expérience et l'ensemble des interactions audiovisuelles mises en évidence par nos mesures électrophysiologiques.

Il reste donc les deux dernières possibilités. Toutes deux impliquent l'existence d'une intégration des informations auditives et visuelles phonétiques. La première correspondrait plutôt à une intégration pré-phonologique, alors que la seconde (amorçage) préserve la possibilité d'une intégration post-phonologique. Nos données sEEG pointent plutôt vers l'hypothèse d'un amorçage, étant donné l'activation massive du cortex auditif par les indices visuels avant la présentation de la syllabe auditive. Bien entendu, le protocole utilisé dans les expériences électrophysiologiques ne permet pas de statuer sur le caractère pré-catégoriel ou catégoriel des représentations impliquées dans l'intégration. Pour exclure l'hypothèse d'amorçage, il faudrait utiliser des stimuli de parole dans lesquels les informations phonétiques visuelles et auditives sont synchrones. L'utilisation du modèle additif dans ce cas permettrait de caractériser la dynamique spatio-temporelle de véritables interactions phonétiques audiovisuelles.

17.2 Représentation d'un évènement audiovisuel en mémoire sensorielle auditive

Notre première expérience comportementale (chapitre 13 page 179) a montré que la détection d'une déviance audiovisuelle était plus rapide que prédit par un modèle d'indépendance des systèmes de détection auditifs et visuels. Notre seconde expérience, en EEG de scalp (chapitre 14 page 185), suggère qu'une partie des interactions audiovisuelles pouvant expliquer cette facilitation concerne l'accès à des représentations mnésiques sensorielles des évènements standards auditifs et/ou visuels (indexés par les MMN auditive et visuelle). Ces deux expériences montrent donc, a minima, l'existence d'interactions entre les systèmes auditif et visuel de détection de la déviance, eux-même basés sur l'existence de registres mnésiques spécifiques à chacune des deux modalités sensorielles.

D'un autre côté, la différence entre les MMN visuelles générées par des évènements purement visuels et audiovisuels (première expérience EEG, chapitre 14 page 185) et la différence entre les MMN auditives générées par des évènements purement auditifs et audiovisuels (seconde expérience EEG, chapitre 15 page 195), suggèrent que l'existence d'une association régulière entre stimuli auditifs et visuels peut modifier la construction de cha-

cune de ces deux traces mnésiques. La représentation d'une régularité audiovisuelle pourrait donc passer par l'inclusion réciproque d'informations auditives et visuelles dans les mémoires sensorielles sensori-spécifiques, comme le montre la topographie de la MMN à la double déviance auditive et visuelle dans la première expérience EEG. Le fait que cette modification de la MMN auditive n'ait pas lieu lorsque l'association entre le stimulus auditif et le stimulus visuel ne constitue pas elle-même un régularité (condition audiovisuelle équiprobable de la seconde expérience EEG, chapitre 15 page 195), suggère que cette inclusion réciproque correspond bien à la représentation de la régularité de l'association des deux traits auditif et visuel.

En revanche, nous ne sommes pas parvenu à montrer de façon convaincante que la violation d'une telle régularité audiovisuelle, sous la forme d'une violation de la conjonction des traits auditifs et visuels, suffit à générer activité de type MMN (chapitre 16 page 205). Il est donc possible que la représentation de la régularité audiovisuelle mise en évidence dans les trois premières expériences n'aboutisse pas à une véritable trace intégrée de la régularité audiovisuelle, permettant une détection automatique et rapide de sa violation. Une explication alternative serait que cette représentation existe mais que la force de la trace mnésique est trop faible dans notre protocole pour permettre une détection rapide de la déviance à la conjonction.

De manière générale, les interactions audiovisuelles mises en évidence dans cette série d'expériences étaient de faible amplitude et toujours à la limite de la significativité statistique, et donc probablement à la limite de la sensibilité de la technique d'enregistrement utilisée. En cela, elles s'opposent aux effets audiovisuels massifs provoqués par les syllabes McGurk, l'illusion de ventriloquie ou des stimuli audiovisuels entretenant des liens plus étroits ou plus écologiques. En utilisant de tels stimuli peut-être pourrait-on mettre en évidence de manière plus convaincante une MMN à la conjonction audiovisuelle.

17.3 Interactions audiovisuelles dans le cortex auditif

Nos deux séries d'expériences ont mis en évidence, entre autres, une influence des informations visuelles sur les traitements réalisés dans le cortex auditif. Cette influence pouvait se manifester de deux façons : soit par une activation des structures auditives en réponse à un stimulus visuel (cas de la parole), soit par une modulation de l'activité enregistrée en réponse à un stimulus auditif (cas de la parole et des représentations en mémoire sensorielle auditive).

Comment peut-on expliquer ces activations intersensorielles au regard de l'architecture connue des systèmes sensoriels auditif et visuel et en particulier de leurs interrelations ? Depuis la mise en évidence d'effets d'interaction multisensorielle dans les cortex sensori-spécifiques par les méthodes d'imagerie fonctionnelle chez l'homme (par exemple : Calvert et coll., 1997 ; Giard & Peronnet, 1999 ; Calvert et coll., 1999), principalement deux hypothèses anatomiques, pouvant expliquer ces effets, ont été proposées, l'une dans le cadre du modèle classique de convergence tardive, l'autre en opposition à ce modèle.

On trouve dans une revue de Mesulam (1998), une description détaillée de l'architecture du modèle classique de convergence tardive chez l'homme. Dans ce modèle, les aires

corticales auditives et visuelles sont totalement ségréguées dans le sens ascendant (*feedforward*) : il n'existe ni connexions latérales entre cortex de différentes modalités sensorielles et encore moins de projections sous-cortico-corticales intersensorielles (les projections sous-cortico-sous-corticales ne sont pas discutées). Les informations de différentes modalités sensorielles ne convergent qu'au niveau d'aires associatives hétéromodales, qui sont au nombre de quatre : le cortex pré-frontal, le cortex pariétal postérieur, le cortex temporal latéral (dont le STS) et le gyrus para-hippocampique, qui sont analogues à celles mises en évidence chez l'animal (voir la partie 1.5 page 17). Cependant, un aspect important de ce modèle, est que toutes les connexions sont bidirectionnelles. Ainsi, même en l'absence de connexions ascendantes ou latérales entre systèmes sensoriels, une influence intersensorielle dans les cortex modalité-spécifiques est possible par le biais de projections descendantes (*feedback*), depuis les aires associatives hétéromodales. Plusieurs auteurs ont proposé que ces voies descendantes soient à l'origine des effets intersensoriels dans les cortex auditifs ou visuels (par exemple : Calvert, 2001 ; Driver & Spence, 2000).

Pour d'autres auteurs, une partie de ces effets doit nécessairement s'expliquer en sortant de ce modèle (Schroeder et coll., 2003 ; Schroeder, Molholm, Lakatos, Ritter & Foxe, 2004 ; Bulkin & Groh, 2006 ; Ghazanfar & Schroeder, 2006). Ces auteurs se basent principalement sur deux arguments. D'une part, selon Foxe et Schroeder (2005), certains effets audiovisuels ont une latence trop courte pour être explicables par des projections descendantes (par exemple : Giard & Peronnet, 1999 ; Molholm, Ritter, Javitt & Foxe, 2004 ; Fort et coll., 2002a). D'autre part, des projections intersensorielles latérales et ascendantes entre cortex auditif et visuel existent : les premières ont déjà été mentionnées dans la partie 1.5 page 18 et ont été principalement observées du cortex auditif vers le cortex visuel ; selon Schroeder et coll. (2003), les secondes correspondraient au système de projection thalamo-cortical koniocellulaire (aussi appelé non spécifique), un système de projection diffus ne respectant pas la ségrégation des aires corticales (revue dans E. G. Jones, 2001). Bulkin et Groh (2006) mentionnent également l'existence de connexions sous-cortico-sous-corticales audiovisuelles, par exemple entre le colliculus supérieur et le colliculus inférieur (une structure sous-corticale auditive ; Doubell, Baron, Skaliora & King, 2000)

Les effets que nous avons mis en évidence dans le cortex auditif nécessitent une analyse détaillée des stimuli, afin, soit de discriminer les informations phonétiques, soit de distinguer les déformations de cercles dans différentes directions. Cela semble exclure les projections koniocellulaires qui ne possèdent pas une spécificité spatiale suffisante pour porter de telles informations (Schroeder et coll., 2003). De même le colliculus supérieur semble être impliqué dans des fonctions liées plus à la détection et la localisation du stimulus visuel qu'à son identification (voir la partie 1.4.1 page 13).

Les résultats concernant la mémoire sensorielle auditive donnent peu d'informations temporelles permettant de trancher entre projections latérales ou descendantes. En effet, ils ne donnent qu'une borne temporelle supérieure des interactions audiovisuelles nécessaires à l'inclusion d'informations visuelles dans la trace sensorielle auditive. Or ces effets étaient observés à la latence de la MMN, c'est-à-dire autour de 200 ms.

Dans le cas de la parole, nous avons proposé (Besle, Fort, Delpuech & Giard, 2004, annexe page 245) que la modulation de l'activation du cortex auditif pouvait être due à des

connections *feedback* entre le STS et le cortex auditif, le STS pouvant être activé par les informations visuelles présentées en avance. Cette proposition était basée sur le manque de données anatomiques montrant des projections d'aires sensorielles visuelles vers les aires sensorielles auditives, sur l'existence de projections d'aires polysensorielles homologues du STS vers le cortex auditif (Pandya, Hallett & Kmukherjee, 1969 ; Seltzer & Pandya, 1978), ainsi que des données électrophysiologiques chez le macaque, montrant que l'influence d'un stimulus visuel sur le traitement d'un son dans le cortex auditif possède un profil laminaire (profil spatial le long de différentes couches du cortex) de type descendant (Schroeder & Foxe, 2002). Depuis, des projections directes du cortex visuel vers le cortex auditif ont été mises en évidence (Hishida et coll., 2003 ; Cappe & Barone, 2005), mais il semble que, même dans ce cas, les aires visuelles d'origine possèdent déjà un caractère multisensoriel audiovisuel (voire la partie 1.5 page 18).

Les données en sEEG suggèrent que des composantes auditives relativement précoces (50 ms) sont modulées par les informations visuelles. Le problème est qu'il existe une asynchronie fondamentale entre les composantes visuelles et auditives de la parole, et que, dans nos stimuli, les informations visuelles étaient disponibles avant les informations auditives. De fait, les interactions audiovisuelles dans le cortex auditif étaient précédées d'une activation visuelle de ce même cortex, dès 120 millisecondes avant la présentation du son, soit 170 ms avant les premières modulations de l'activité auditive. Ce laps de temps est largement suffisant pour permettre à des effets descendants de se mettre en place.

En revanche, l'activation du cortex auditif par les stimuli visuels semble relativement précoce, puisqu'elle suivait immédiatement celle d'aires occipito-temporales et temporales postérieures, vraisemblablement visuelles. Les activations provenant du STS semblent, elles, se produire plus tardivement. Mais le STS est une vaste structure qui ne présente sans doute pas une unité fonctionnelle très prononcée et certaines parties du STS n'ont pas du tout été explorées chez nos patients. De même, les zones visuelles occipitales n'ont pas été explorées dans cette étude, si bien qu'on ignore à quelle latence avaient lieu les premières activations visuelles. Il est donc difficile de dire exactement à quel point les activations du cortex occipito-temporal et du cortex auditif étaient "précoces" et donc de se prononcer sur la nature *feedback* ou *feedforward* de ces activations.

