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 (Examinateur)
M^{me} Nicole Bruneau (Rapporteur)
M^{r} Jean-Luc Schwartz (Rapporteur)
M^{me} Marie-Hélène Steiner-Giard (Directrice de thèse)
M^{r} Rémy Versace (Examinateur)

Table des matières

Troisième partie

Interactions audiovisuelles dans la perception de la parole

Chapitre 9

Étude en EEG et comportement

Cette première étude a été réalisée lorsque j'étais en DEA à l'unité 280, sous la direction de Marie-Hélène Giard, et l'analyse des données s'est poursuivie au début de ma thèse. Cette étude ayant fait l'objet d'une publication (Besle, Fort, Delpuech & Giard, 2004), elle ne sera que brièvement présentée ici. Les détails en sont décrits dans la publication, intégrée au manuscrit en annexe (page 245).

9.1 Rappel de la problématique

Nombre de données comportementales ont montré que des indices visuels de parole (les mouvements des lèvres en particulier) pouvaient influencer la perception auditive de la parole. Une partie de ces interactions a vraisemblablement lieu, entre autres, à une étape précoce du traitement, avant la catégorisation phonologique des sons de parole (voir la partie 3.2.2 page 54). La plupart des études de neuroimagerie qui ont traité de la question de l'intégration des indices auditifs et visuels dans la perception de la parole ont cependant utilisé l'IRM fonctionnelle. Ces études ont montré l'implication de plusieurs aires corticales dans cette intégration mais ne pouvaient leur assigner de place dans la chaine des traitements, étant donné la faible résolution temporelle de la technique utilisée.

L'EEG est une technique d'enregistrement particulièrement adéquate pour tenter de mettre en évidence différentes étapes de traitement, où peut opérer l'intégration audiovisuelle. Au moment ou nous avons conduit cette expérience, les seuls résultats en EEG/MEG sur la perception de la parole avaient cependant soit montré des interactions audiovisuelles à des latences très tardives autour de 450 ms (en utilisant donc le modèle additif en dehors de son domaine d'application, Sams & Levänen, 1998), soit uniquement établi une borne temporelle supérieure pour les premières interactions audiovisuelles : c'est le cas des études ayant montré l'existence d'une MMN auditive, vers 200 ms, pour des syllabes déviant sur leur dimension visuelle (Colin, Radeau, Soquet, Demolin et coll., 2002 ; Möttönen et coll., 2002 ; Sams et coll., 1991, voir aussi la partie 4.6.1 page 72). L'utilisation du modèle additif dans les 200 premières millisecondes de traitement devrait donc permettre de mettre en évidence le décours temporel des interactions audiovisuelles ayant lieu en amont.

L'utilisation de techniques d'imagerie et du modèle additif permettent de plus d'étudier la perception de la parole naturelle, sans recourir à la présentation d'informations conflic-

tuelles ou bruitées comme cela a souvent été le cas dans les études comportementales de la perception de la parole bimodale. Nous avons donc enregistré les potentiels évoqués par des syllabes présentées soit uniquement dans la modalité auditive, soit uniquement dans la modalité visuelle, soit dans les deux modalités simultanément et avons comparé le potentiel évoqué audiovisuel à la somme des potentiels évoqués auditifs et visuels, de façon à déterminer les aires cérébrales et les étapes de traitement où ces interactions ont lieu.

9.2 Méthodes

9.2.1 Sujets

Seize sujets droitiers (dont 8 de sexe féminin), âgés en moyenne de 23 ans ont passé cette expérience. Aucun sujet ne souffrait de troubles neurologiques. Ils avaient tous une audition normale et une vision normale ou corrigée.

Treize autres sujets (dont 9 de sexe féminin) âgés en moyenne de 24,3 ans ont participé à l'étude comportementale.

9.2.2 Stimuli

Les stimuli étaient des syllabes /pa/, /pi/ , /po/ et /py/, prononcées par une locutrice de langue maternelle française et enregistrées à une fréquence d'échantillonnage de 25 images/s pour l'image et de 41 kHz pour le son. Trois exemplaires différents de chacune des syllabes ont été sélectionnés, sur un corpus d'une centaine de syllabes enregistrées, de manière à conserver une certaine variabilité naturelle de la parole, nécessaire pour que les sujets traitent les stimuli sur un plan linguistique et ne se contentent pas de discriminer les stimuli sur des traits de surface non pertinents, tels qu'une légère différence d'éclairage ou de position des lèvres au départ de la syllabe. Ces 12 syllabes ont été sélectionnées de sorte qu'elles aient approximativement toutes la même structure temporelle audiovisuelle et ont ensuite été légèrement modifiées de façon à présenter des caractéristiques temporelles véritablement identiques (temps séparant le début du mouvement des lèvres, l'ouverture de la bouche, l'explosion de la consonne et début du voisement) et ainsi minimiser la variabilité des réponses évoquées auditives et visuelles. La structure audiovisuelle des syllabes finales est décrite dans la figure 9.1 page ci-contre. Seule la partie inférieure du visage de la locutrice était présentée aux sujets, la bouche ayant une taille de 2,2° d'angle visuel. Le niveau sonore était confortable.

Dans ces stimuli, les informations visuelles commençaient 240 ms avant le début des informations auditives. Les mouvements des lèvres dans les 6 premières images avant l'ouverture de la bouche étaient toutefois de faible amplitude. Nous avons vérifié dans une pré-expérience sur un groupe de 7 sujets qu'ils ne pouvaient donner d'indice sur l'identité de la syllabe. Nous avons pour cela demandé au sujet de tenter d'identifier la syllabe visuelle, qui pouvait être tronquée à la 6ème, la 8ème ou la 14ème image. Les résultats (table 9.1 page suivante) montrent que les sujets répondent au hasard (26% de bonnes réponses en moyenne) lorsque la syllabe s'arrêtait à la 6ème trame. Dès la 8ème image cependant, les informations visuelles étaient suffisantes pour atteindre la performance ob-

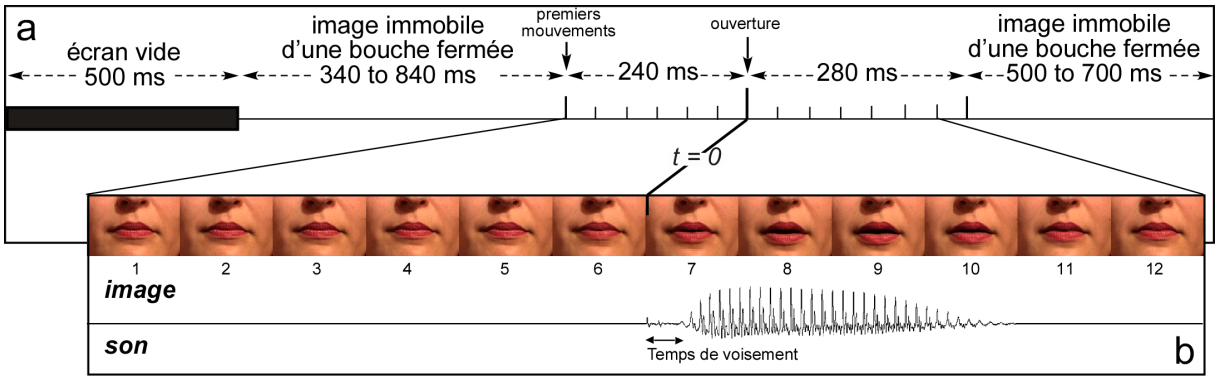

FIG. 9.1 – Structure temporelle d'un essai audiovisuel (a) et d'une syllabe audiovisuelle (b). chaque trame vidéo durait 40 ms. Le temps 0 pour le calcul des PE était pris au début du son.

	/pa/	/pi/	/po/	/pu/	moyenne
6 trames	21%	31%	10%	43%	26%
8 trames	45%	74%	23%	51%	50%
14 trames	81%	81%	37%	27%	53%

TAB. 9.1 – Résultats de la pré-expérience comportementale. Les valeurs indiquent le pourcentage de reconnaissance en fonction du type de syllabe et du nombre de trames présentées

servée lorsque la syllabe était présentée dans sa totalité. L'analyse des erreurs de cette expérience a, par ailleurs, montré que les syllabes /po/ et /py étaient souvent confondues.

9.2.3 Procédure

Les 3 exemplaires des 4 syllabes étaient présentées de façon auditive, visuelle ou audiovisuelle. Tous les essais étaient présentés aléatoirement dans un même bloc de stimuli. Au total, 1116 stimulations étaient présentées, réparties en 16 blocs d'une durée approximative de 2 min 30. Au début de chaque bloc, l'une des 4 syllabes était désignée comme cible (chaque syllabe pouvait donc être cible ou non-cible selon le bloc). Le sujet devait répondre en appuyant sur un bouton lorsqu'il entendait la syllabe cible (seulement pour les essais audiovisuels et auditifs).

Nous avons longuement hésité à demander aux sujets de détecter la cible quelle que soit sa modalité de présentation, y compris en condition visuelle seule, c'est-à-dire en lisant sur les lèvres. Dans ce cas, nous aurions pu lier plus directement les résultats des potentiels évoqués aux résultats comportementaux calculés à partir de l'inégalité de Miller, qui prend en compte les TR auditifs, visuels et audiovisuels et exclut un simple effet de facilitation statistique des TR. Cependant le fait de demander aux sujets une réponse dans les trois conditions aurait nécessité un effort attentionnel plus important en condition de lecture labiale que dans les deux autres modalités. Les effets de cette attention visuelle sur la réponse évoquée auraient pu se manifester de manière plus importante dans la condition visuelle seule, que dans la condition audiovisuelle et auraient donc pu apparaitre de manière

erronée comme des violations de l'additivité (voir aussi Besle, Fort & Giard, 2004, et la partie 7.2.1 page 108 pour une discussion plus détaillée). Nous avons donc demandé au sujet de ne répondre que sur la base des indices auditifs. Les sujet devaient cependant fixer la bouche durant toute l'expérience, et ceci était vérifié grâce à une caméra vidéo.

9.2.4 Expérience comportementale complémentaire

Pour appliquer l'inégalité de Miller et vérifier l'existence d'un gain comportemental audiovisuel, nous avons donc mené une expérience comportementale complémentaire avec un autre groupe de sujets. Les stimuli et les conditions de stimulation étaient identiques, excepté que les sujets devaient répondre dans les 3 conditions de stimulation. Dans cette expérience, seules les syllabes /pa/ et /pi/, plus faciles à discriminer visuellement, pouvaient être cible. Cette expérience complémentaire permettra a minima de conclure que les stimuli utilisés pour le calcul des interactions audiovisuelles, au moyen du modèle additif, sont susceptibles de donner lieu à un effet de facilitation audiovisuelle qui n'est pas dû à une facilitation statistique.

9.2.5 Analyse des résultats

Les TR auditifs et audiovisuels dans l'expérience d'EEG ont été comparés par un test de Student. Les TR auditifs, visuels et audiovisuels dans l'expérience comportementale complémentaire ont été analysés par application de l'inégalité de Miller et comparaison des fractiles de distribution des TR audiovisuels et de la somme des distributions des TR unimodaux (voir la partie 7.1.4 page 105).

Les PE n'ont été calculés que sur les essais non-cibles, pour exclure toute activité motrice dans les signaux analysés. Les essais où le sujet avait répondu par erreur ont également été exclus de l'analyse des PE. Le nombre moyen d'essais était de 160 par condition et par sujet. Le temps zéro utilisé pour le moyennage des PE et à partir duquel étaient mesurées les latences correspondait au début de la syllabe auditive. La ligne de base était prise entre -300 et -150 ms pré-stimulus. Cette fenêtre de latence est un compromis entre la nécessité de rapprocher le plus possible la ligne de base de la fenêtre d'analyse, et celle d'éviter l'inclusion de potentiels évoqués visuels dus au mouvement des lèvres qui commençait 240 ms avant la présentation du son (bien que ces mouvements soient de très faible amplitude).

La différence entre le PE audiovisuel et la somme des PE unimodaux statistiquement a été testée à chaque échantillon temporel et à chaque électrode par un test de Student apparié, dans les 200 premières millisecondes post-stimulus. Les tests multiples ont été pris en compte en exigeant, pour chaque électrode, un nombre minimal de 24 échantillons significatifs successifs, d'après la table proposée par Guthrie et Buchwald (1991, voir la partie 8.1 page 111).

9.3 Résultats

9.3.1 Résultats comportementaux

En ce qui concerne l'expérience en EEG, les sujets ont été plus rapides pour répondre aux cibles audiovisuelles (400 ms) qu'aux cibles auditives (423 ms). Bien que cette différence soit assez faible, elle était très significative ($t(15) = 4,33$; $p < 0,001$). Le pourcentage d'erreurs (oublis ou fausses alertes) était inférieur à 1% dans les deux conditions.

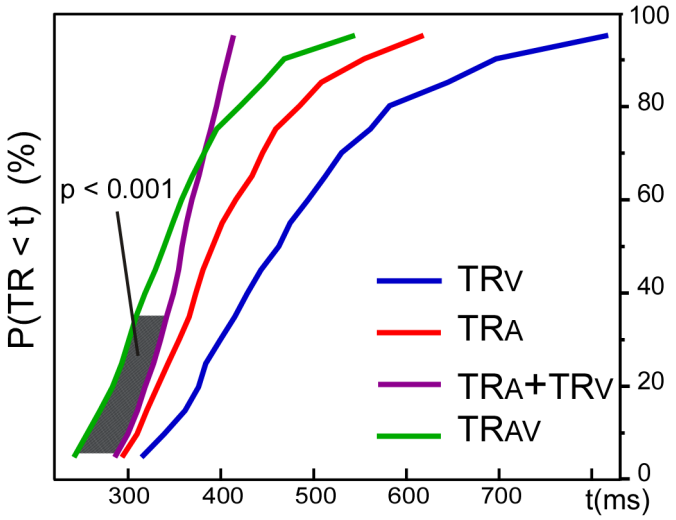

FIG. 9.2 – Application de l'inégalité de Miller. TR_V : fonction de répartition des temps de réaction visuels ; TR_A : fonction de répartition des temps de réaction auditifs ; $TR_A + TR_V$: somme des 2 fonctions de répartition unimodales ; TR_{AV} : fonction de répartition des temps de réaction audiovisuels. La partie hachurée désigne les zones où les fractiles correspondants des deux fonctions de répartition sont significativement différents.

Concernant l'expérience comportementale complémentaire, les TR dans les conditions visuelle, auditive et audiovisuelle étaient respectivement de 496, 418 et 356 ms. La figure 9.2 montre les fonctions de répartition (pour l'ensemble des sujets) des TR visuels, auditifs et audiovisuels, ainsi que la somme des fonctions de répartition unimodales. Pour les 9 premiers fractiles, les TR bimodaux sont significativement inférieurs à ceux prédits par les modèles d'activations séparées et représentés par la somme des fonctions de répartition unimodales.

9.3.2 Résultats électrophysiologiques

la figure 9.3.A (page suivante) montre les PEs obtenus dans chaque modalité dans les 300 premières millisecondes. La réponse visuelle unimodale (courbe bleue) montre principalement un pic négatif vers 40 ms, dont le maximum se situe sur les électrodes occipitales et forme une topographie occipitale bilatérale (non illustrée). La topographie et la latence de cette onde suggèrent qu'il pourraient s'agir de l'onde N1 visuelle dont le pic est habituellement observé vers 180 ms. Dans notre cas, ce pic suivait le début du mouvement des

lèvres de 280 ms, ce qui pourrait s'expliquer par le début très progressif des mouvements, et donc à la fois un temps de traitement plus lent à s'établir et une plus grande variabilité des réponses élémentaires, qui auraient pour effet d'étaler cette composante dans le temps. Il se peut aussi qu'il s'agisse d'une composante spécifique au traitement d'un mouvement. Cette réponse était suivie d'une deuxième composante visuelle négative dont le maximum se situait de façon bilatérale sur les électrodes pariéto-centrale, vers 160 ms post-stimulus.

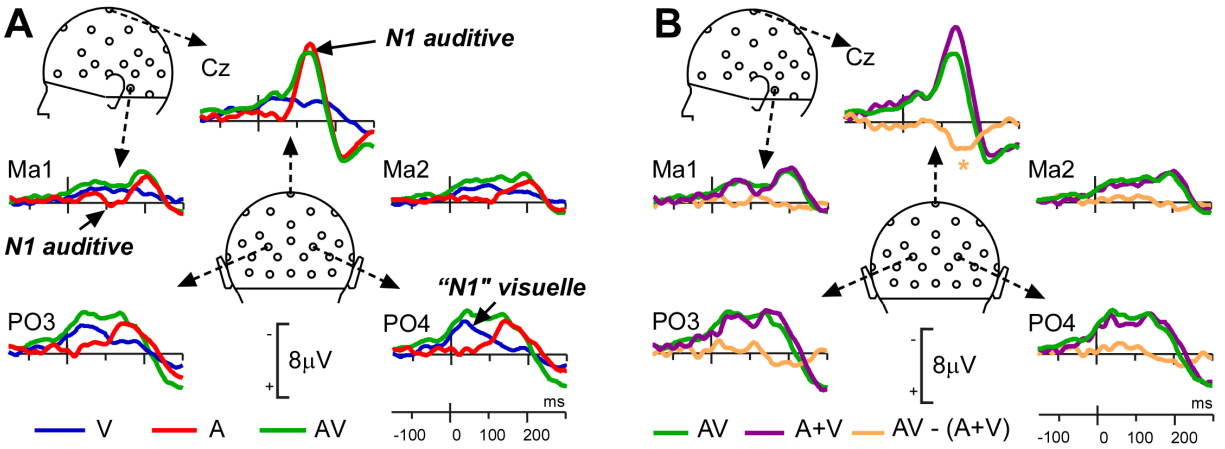

FIG. 9.3 – A. Réponses évoquées par les syllabes auditives (A), visuelles (V) et audiovisuelles (AV) entre -150 et 300 ms à un sous-ensemble d'électrodes. B. Application du modèle additif. A+V : somme des réponses auditives et visuelles. AV-A+V : violation du modèle additif. L'étoile indique les violations significatives au seuil corrigé.

La réponse auditive unimodale (courbe rouge) se caractérisait par une onde négative dont le pic maximum vers 135 ms était associé à une inversion de polarité sur les électrodes mastoïdes. La topographie de cette onde ainsi que celles des densités radiales de courant associées sont visibles sur la figure 9.5 (1re colonne, page 126). C'est une topographie typique d'activités prenant place dans le cortex auditif. Cette onde correspond sans nul doute à l'onde N1 auditive. L'onde N1 était suivie d'un onde de polarité inverse (l'onde P2) dont l'amplitude était maximale à 205 ms post-stimulus.

La figure 9.3.B compare les PE audiovisuels (courbe verte) à la somme des PE unimodaux (courbe mauve). Ces deux courbes sont globalement identiques excepté sur les électrodes fronto-centrales entre 100 et 200 ms, c'est-à-dire dans une fenêtre de temps correspondant à l'onde N1 auditive et à la deuxième composante visuelle. Les résultats détaillés du test statistique du modèle additif sont donnés dans la figure 9.4 page ci-contre. On peut constater que la différence entre la réponse bimodale et la somme des réponses unimodales est significative sur une grande partie des électrodes fronto-centrales d'environ 120 ms à 200 ms post-stimulus et que le nombre d'échantillons significatifs successif dépasse largement 24 ms sur ces électrodes, ce qui permet d'exclure un effet dû au nombre important de tests réalisés. La topographie des interactions audiovisuelles était à peu près stable entre 120 et 190 ms post-stimulus.

Pour tenter de comprendre la nature de ces interactions audiovisuelles, nous avons comparé (figure 9.5 page 126) leur topographie à la topographie des réponses unimodales, à la latence où la violation du modèle additif était la plus significative, ce qui correspond

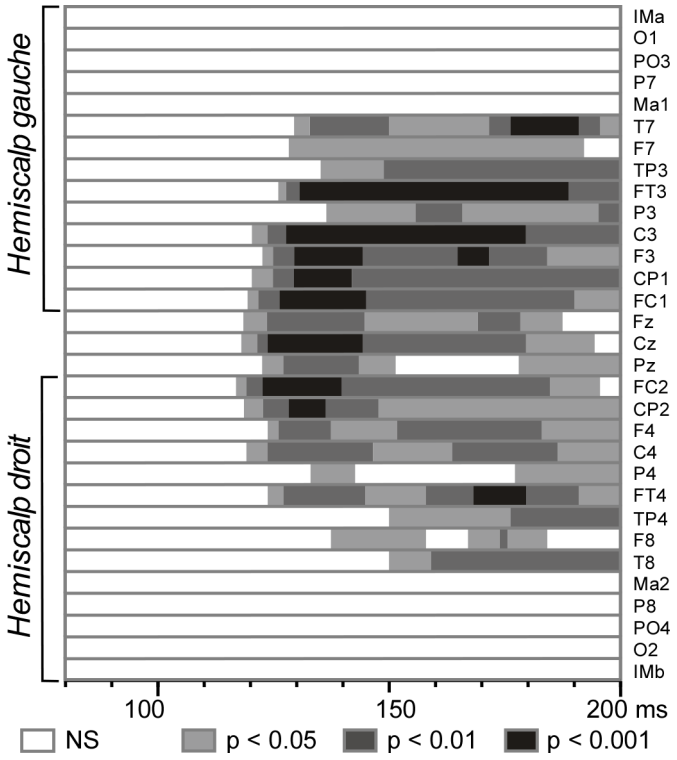

FIG. 9.4 – Résultat des tests statistiques de la violation du modèle additif sur l'ensemble des électrodes entre 80 et 200 ms. Le niveau de gris indique la significativité

au pic de l'onde N1 auditive. La topographie des interactions ressemble clairement plus à celle de l'activité auditive unimodale qu'à celle de l'activité visuelle. En particulier, la configuration des puits et des sources de courant reproduit assez fidèlement celle de l'onde N1 auditive, avec des polarités inversées. Cela suggère que les interactions audiovisuelles observées autour de 135 ms reflètent une diminution d'activité des générateurs de l'onde N1 auditive dans la condition audiovisuelle par rapport à la condition auditive seule.

9.4 Discussion

9.4.1 Comportement

Les résultats comportementaux de l'expérience d'EEG, ainsi que ceux de l'expérience comportementale, montrent que le traitement de la parole peut être accéléré par des indices visuels, même lorsque la performance des sujets a atteint un plafond en termes de pourcentage de réponses correctes. À notre connaissance, c'est la première fois qu'un tel résultat est montré. Peu d'études se sont en fait intéressé aux temps de réactions à des stimuli de parole audiovisuelle. Deux études ont mesuré les TR auditifs et audiovisuels dans un tâche de catégorisation de syllabes commençant par des consonnes différentes, mais elles ont soit rapporté des différences faibles et non reproductibles (Massaro & Cohen, 1983), soit des TR audiovisuels supérieurs à l'un des TR unimodaux (K. P. Green & Gerdeman, 1995).

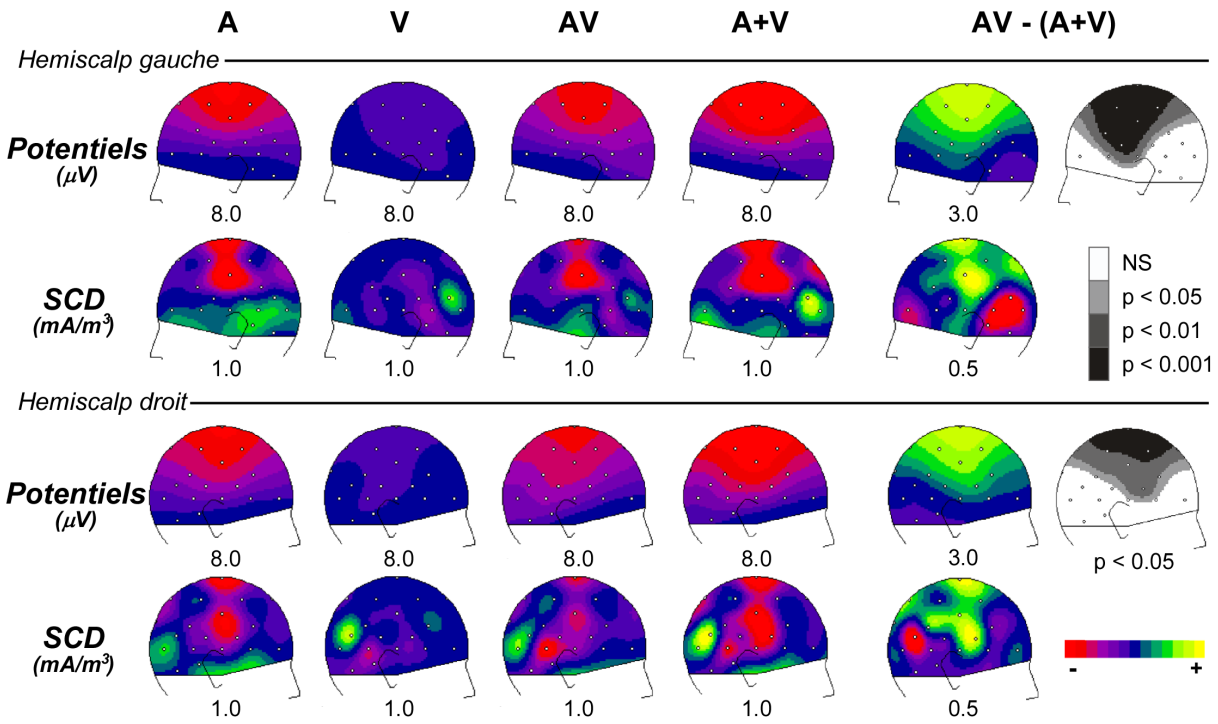

FIG. 9.5 – Topographies des réponse auditives (A), visuelles (V), audiovisuelles (AV), ainsi que de la somme des réponses unimodales (A+V) et de la violation du modèle additif (AV - (A+V)) sur les hémiscalps droit et gauche. La valeur maximale de l'échelle de couleur est indiquée sous chaque carte. La couleur jaune correspond aux potentiels ou aux courants positifs, tandis que la couleur rouge correspond aux potentiels ou aux courants négatifs. Les deux cartes en niveaux de gris à droite donnent la significativité des PE d'interaction (AV-(A+V)). SCD : Densité radiale de courant.

Une étude précédente, rapportée par Massaro (1987), avait cependant testé un modèle d'activations séparées chez deux sujets (sous l'hypothèse d'indépendance des distributions des TR unimodaux, voir la partie 7.1.1 page 101). Les auteurs trouvaient des TR audiovisuels inférieurs en moyenne aux TR unimodaux mais qui restaient prédictibles par le modèle. Dans notre expérience complémentaire, nous montrons au contraire que le gain en temps de réaction observé avec les même stimuli que ceux utilisés dans notre expérience d'EEG ne peut s'expliquer par un tel modèle. Ce résultat implique (dans les limites exposées dans la partie 7.1 page 99) que les canaux auditifs et visuels ont échangé des informations.

On notera cependant que le gain de temps de réaction dans la condition audiovisuelle par rapport à la condition auditive seule est beaucoup plus important dans l'expérience comportementale que dans l'expérience électrophysiologique. Le fait d'avoir utilisé des groupes de sujets différents limite les conclusions que l'on peut tirer de ce résultat. Il est cependant probable que le fait d'attirer l'attention des sujets vers les indices visuels en leur demandant d'effectuer une tâche de lecture labiale ait augmenté la contribution des indices visuels au traitement de l'identité de la syllabe.

Il faut aussi souligner que la tâche de l'expérience d'EEG s'apparente plus à paradigme du stimulus accessoire dans lequel le sujet n'a pas à analyser les informations visuelles pour discriminer les syllabes auditives. On ne peut donc exclure le fait que le gain de TR en condition audiovisuelle représente un effet d'alerte dû à la présence d'un stimulus visuel, d'autant plus que le mouvement des lèvres précédait la syllabe auditive.

9.4.2 Résultats électrophysiologiques

Ce gain comportemental pour la discrimination de syllabes était associé à ce que l'on a interprété comme une diminution d'activité des générateurs de l'onde N1 auditive. Avant de tenter d'interpréter cette diminution et son rôle dans l'intégration des indices auditifs et visuels de parole, soulignons que des résultats analogues ont été rapportés dans la littérature soit en même temps, soit à la suite de notre étude.

Ainsi, Klucharev, Möttönen et Sams (2003) ont testé le modèle additif pour des syllabes auditives, visuelles et audiovisuelles et ont trouvé des interactions audiovisuelles vers 125 ms de traitement, dont la topographie radiale suggère la diminution de certaines composantes seulement, de l'onde N1 auditive (notons que les sujets réalisaient des tâches différentes dans des blocs de stimulations auditifs, audiovisuels et visuels séparés, ce qui pose quelques problèmes pour l'application du modèle additif, voir la partie 7.2.1 page 108).

van Wassenhove, Grant et Poeppel (2005) ont mis en évidence une diminution d'amplitude importante de l'onde N1 auditive en condition audiovisuelle par rapport à un condition auditive seule, dans une tâche de discrimination phonologique, dans des blocs séparés pour les différentes conditions auditives, visuelles et audiovisuelles. Cette diminution d'amplitude était doublée d'une diminution de latence, difficile à interpréter, cependant, en l'absence d'utilisation du modèle additif.

En MEG, Möttönen, Schurmann et Sams (2004) ont montré la même diminution. Si l'utilisation de la MEG limite en pratique le besoin de recourir au modèle additif en raison de la moindre diffusion des champs magnétiques sur le scalp, ces auteurs ne présentent toutefois pas de réponses visuelles seules permettant de s'assurer que la diminution observée était bien due à une modulation de l'activité auditive.

Une étude de Miki, Watanabe et Kakigi (2004) en MEG n'a en revanche pas rapporté une telle diminution. Plusieurs raisons peuvent expliquer cette absence, comme par exemple le fait que les sujets étaient totalement passifs ou que les mouvements de lèvres consistaient simplement en la présentation d'une image de bouche ouverte et non en de véritables mouvements filmés. Malgré les nombreux problèmes méthodologiques que présentent ces études, elle convergent presque toutes vers le même résultat, ce qui suggère que l'effet trouvé est assez robuste.

Une telle diminution de l'onde N1 auditive ne semble pas exister dans des expériences de discrimination ou de détection de stimuli non langagiers dans lesquels une diminution du TR audiovisuel était observé (Fort et coll., 2002a, 2002b ; Giard & Peronnet, 1999 ; Molholm et coll., 2002 ; Teder-Sälejärvi et coll., 2002). Cette effet pourrait donc bien être spécifique de l'intégration audiovisuelle des indices de parole, ou plus généralement

d'évènements bimodaux dans lesquels le stimulus visuel précède le stimulus auditif (notons cependant que Möttönen et coll., 2004 trouvent une diminution de l'onde N1 auditive avec des stimuli de paroles auditifs et visuels dont les débuts sont synchrones).

En revanche, une diminution de l'onde N1 visuelle (vers 180 ms de latence) a été trouvée pour la discrimination de stimuli audiovisuels par rapport à des stimuli visuels seuls (Giard & Peronnet, 1999). Cette onde, générée dans le cortex visuel extrastrié (Mangun, 1995) serait liée à des processus de discrimination visuelle (Vogel & Luck, 2000). Cette réduction avait été interprétée comme le reflet d'une demande énergétique moindre pour discriminer les stimuli visuels, rendu plus saillants par la présence et l'utilisation d'informations auditives. De la même manière, l'onde N1 auditive serait liée à l'analyse séparée des traits acoustiques du stimulus dans le cortex auditif (Näätänen & Picton, 1987 ; Näätänen & Winkler, 1999). La diminution observée pourrait donc refléter la facilitation de traitement des syllabes auditives due à la présence d'informations phonétiques visuelles, à une latence où les différents traits acoustiques n'ont pas encore abouti à une représentation intégrée du stimulus sonore (Näätänen & Winkler, 1999).

Bien que traditionnellement, on situe les générateurs de l'onde N1 auditive dans le cortex auditif, c'est-à-dire sur la partie supérieure du cortex temporal, il est possible que l'onde N1 en réponse à des sons de parole, beaucoup moins étudiée, inclue d'autres générateurs. Plusieurs études ont montré l'implication du STS dans le traitement de sons complexes, avec une préférence pour les sons de parole, qu'ils soient intelligible ou non (revue dans Hickok & Poeppel, 2004). Étant donné que le STS a été impliqué dans plusieurs études de neuroimagerie sur l'intégration audiovisuelles des indices de parole (Beauchamp, Argall et coll., 2004 ; Calvert et coll., 2000 ; Wright et coll., 2003) et que son orientation est parallèle au plan supratemporal, c'est un candidat possible pour la localisation de l'effet observé. Toutefois, le fait que tous les générateurs, visibles sur la carte des densités radiales de courant de l'onde N1 auditive, apparaissent également sur la topographie des interactions suggère qu'il s'agit d'une diminution globale de l'activité auditive seule à cette latence et non d'un seul générateur spécifique au traitement de la parole.

Plusieurs interprétations alternatives de l'effet observé peuvent être proposées. Tout d'abord, cette diminution pourrait refléter une facilitation du traitement due à une meilleure préparation du sujet pour traiter les indices auditifs lorsque ceux-ci ont été précédés de mouvements lui indiquant qu'un son va peut-être lui être présenté dans les 240 ms. En effet, bien que la violation de l'inégalité de Miller suggère l'existence d'échanges d'informations auditives et visuelles, elle a été appliquée à des TR enregistrés dans des conditions différentes qui font qu'on ne peut exclure un pur effet d'alerte dans l'expérience d' EEG. Si tel était le cas, cependant, on s'attendrait à observer plutôt une augmentation de l'onde N1 auditive, analogue aux effets d'attention auditive qui se manifestent sur plusieurs ondes sensorielles auditives, dont l'onde N1 (revue dans Näätänen, 1992 ; Giard, Fort, Mouchetant-Rostaing & Pernier, 2000). De façon intéressante, si des effets d'un indice visuel spatial sur les potentiels évoqués auditifs ont été mis en évidence (McDonald, Teder-Sälejärvi, Heraldez & Hillyard, 2001), ils prennent la forme d'une négativité accrue à la latence de nos effets. De tels effets auraient donc résulté en une augmentation de l'onde N1 auditive. Il n'est toutefois pas dit que les effets d'alerte se manifestent de la même

manière que les effets d'attention spatiale sur les PE auditifs, même si certaines études indiquent des effets analogues pour les deux phénomènes sur la réponse visuelle dans le cortex extrastrié (Thiel, Zilles & Fink, 2004).

Ensuite, plusieurs études ont montré que la lecture labiale pouvait, en elle-même, activer le cortex auditif (par exemple Calvert et coll., 1997). Même si les sujets n'avaient pas pour tâche de lire les syllabes sur les lèvres, certains sujets ont rapporté avoir tenté de le faire. Et, quoiqu'il en soit, la vision des mouvements articulatoires, même sans tentative d'en comprendre le contenu pourrait également activer le cortex auditif en condition visuelle seule. Cette activation du cortex auditif par les mouvements labiaux, si elle avait lieu à la latence de l'effet observé, pourrait apparaitre comme une violation du modèle additif et expliquer la topographie auditive des interactions prenant place entre 120 et 190 ms. Cette explication est cependant très peu probable dans la mesure où l'on n'observe pas de réponse ayant une topographie auditive dans cette fenêtre de latence dans la condition visuelle seule.

Une autre explication enfin serait que les informations visuelles sur l'identité de la syllabe sont disponibles avant les informations auditives, par un phénomène de coarticulation. Ces informations pourraient alors pré-activer des unités phonologiques dans le cortex auditif (ou le STS). Plusieurs expériences ont montré que l'amorçage sémantique, aussi bien unimodal qu'intermodal, pouvait se manifester au niveau neuronal par des diminutions d'activité (Badgaiyan, Schacter & Alpert, 1999 ; Holcomb & Anderson, 1993 ; Holcomb & Neville, 1990). De façon analogue, la diminution d'activité dans le cortex auditif pourrait refléter un effet d'amorçage des informations phonétiques visuelles sur le traitement phonétique ou phonologique auditif (voir aussi Jaaskelainen et coll., 2004). Bien que nous ayons montré dans une pré-expérience que les informations visuelles au moment de l'arrivée du son étaient insuffisantes pour identifier les syllabes (voir la partie 9.2.2 page 120), le traitement intégral de l'amorce n'est pas nécessaire pour observer des effets d'amorçage. Il est toutefois probable que les informations visuelles présentes avant l'ouverture complète de la bouche soient trop subtiles pour participer à l'amélioration audiovisuelle. Munhall, Kroos, Jozan et Vatikiotis-Bateson (2004) ont en effet montré que les fréquences spatiales des informations visuelles participant à l'amélioration de l'intelligibilité de la parole dans le bruit sont assez grossières (inférieures à 7 cycles/visage).

Si des informations phonétiques visuelles ont permis de moduler l'activité auditive de traitement des syllabes, ce sont sans doute celles portées par la forme de l'ouverture de la bouche, qui sont disponibles au même moment que les informations auditives. Dans ce cas, les informations visuelles mettent environ 100 ms à venir moduler l'activité dans les structures traitant la parole auditive.

Chapitre 10

Étude en sEEG

10.1 Introduction

Notre expérience en EEG de scalp a montré l'existence d'importants effets d'interactions audiovisuelles dans la perception de la parole bimodale entre 120 et 190 ms de traitement de la syllabe, reflétant vraisemblablement une diminution d'activité auditive. Contrairement aux études précédentes utilisant le modèle additif dans l'effet du stimulus redondant avec des stimuli non-langagiers et qui avaient mis en évidence des interactions complexes, de topographies différentes à différentes latences, nous n'avons trouvé que cet effet de modulation de l'activité auditive. Il était cependant possible que l'amplitude importante de l'effet de diminution de l'onde N1 auditive ait caché d'autres effets d'interaction dans d'autres structures. Par ailleurs la résolution spatiale limitée de l'EEG de scalp ne permettait pas de s'assurer de la localisation exacte de la diminution d'activité. Celle-ci aurait pu avoir lieu aussi bien dans le planum temporale que sur l'une des aires bordant le STS.

Afin d'étudier plus en détail les interactions audiovisuelles ayant lieu lors de la perception de syllabes bimodales, nous avons fait passer cette expérience à des patients épileptiques portant des électrodes intracérébrales, en collaboration avec O. Bertrand (U821) et le Docteur C. Fischer (Hôpital Neurologique de Lyon). La plupart de ces patients étaient suivis pour des épilepsies d'origine temporale et avaient donc un certain nombre d'électrodes traversant le planum temporale, le gyrus de Heschl, le gyrus temporal supérieur (GTS), le STS et le gyrus temporal moyen (GTM) (ces structures sont indiquées sur la figure 10.4.B, page 138). À l'occasion, d'autres structures ont pu être explorées (insula, gyrus supra-marginal, opercules pré-central et post-central, gyrus temporal moyen postérieur, etc...). Les emplacements de toutes les électrodes de tous les patients ont été reportées sur un cerveau commun dans les figures 10.1 page suivante et 10.2 page 133.

Bien que nous n'ayons pas observé d'activation de type auditif en réponse aux mouvements labiaux présentés isolément dans l'expérience d'EEG, les enregistrements sEEG constituaient aussi une occasion de vérifier l'existence de traitements des indices visuels de parole dans le cortex auditif. La plupart des études d'IRMf ayant étudié la lecture labiale ont montré l'implication, entre autres structures corticales, d'une partie importante du cortex auditif Il existe cependant un débat concernant l'implication du cortex auditif

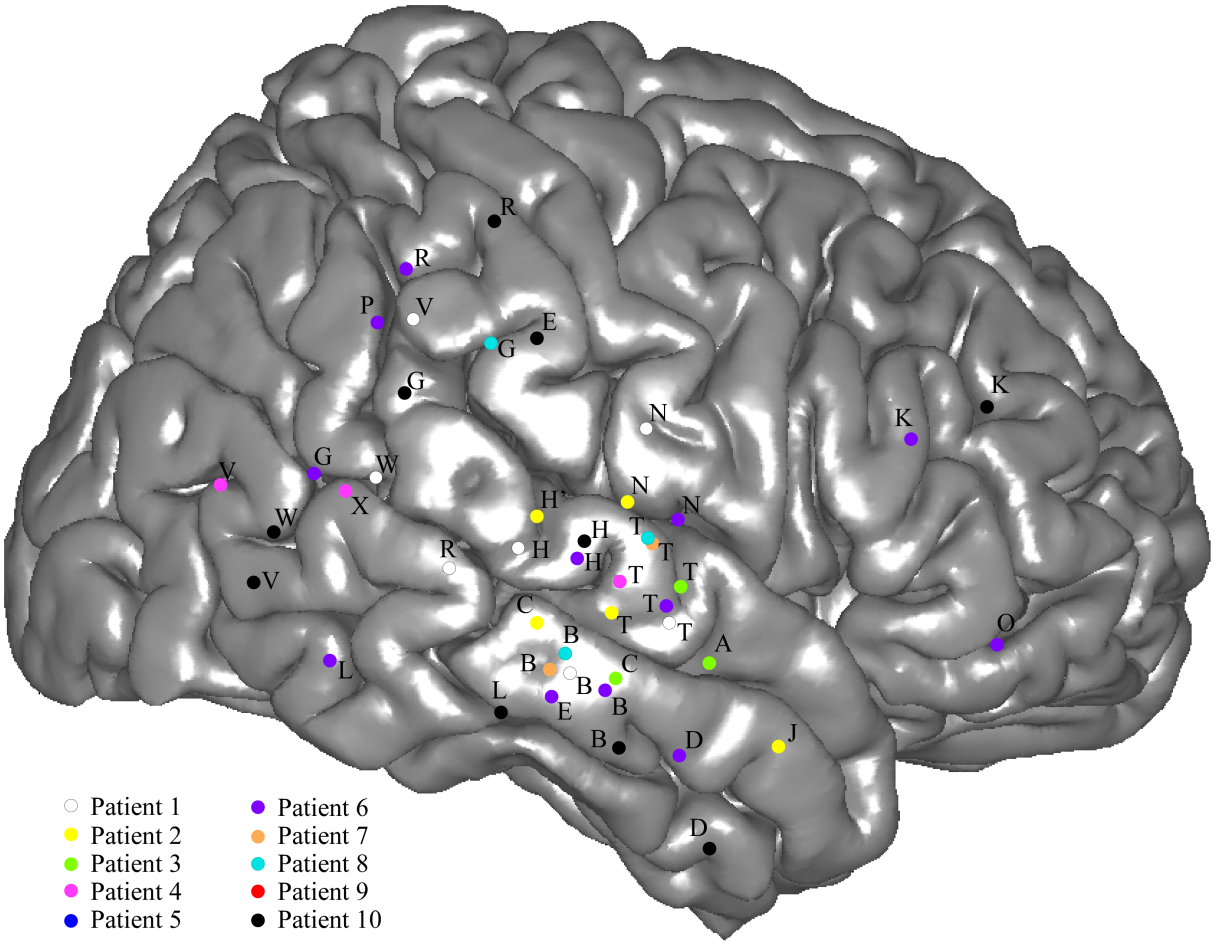

FIG. 10.1 – Emplacements des électrodes de l'hémisphère droit reportés à la surface d'un cerveau standard (single-subject du MNI). Le recalage des électrodes des différents patients a été réalisé par la méthode de Talairach (transformation linéaire par cadrans). Chaque électrode comprend entre 5 et 15 contacts explorant les structures situées à la perpendiculaire du plan de la figure

primaire (aire 41 de Brodmann) dans cette activation. Certaines études ont montré une activation de la partie médiale du gyrus transverse (ou gyrus de Heschl), où se situe le cortex auditif primaire (Calvert et coll., 1997 ; Ludman et coll., 2000 ; MacSweeney et coll., 2001). D'autres ont trouvé une activation de sa partie latérale (Calvert & Campbell, 2003), qui ne correspond déjà plus au cortex primaire ou seulement des cortex secondaires (L. E. Bernstein et coll., 2002 ; Campbell et coll., 2001 ; MacSweeney et coll., 2000, 2002 ; Olson et coll., 2002 ; Paulesu et coll., 2003), dont le planum temporale (aire 42) et le GTS latéral (aire 22). Le cortex auditif primaire étant une structure de petite taille, la variabilité anatomique inter-individuelle est cependant susceptible de cacher des activations dans une étude de groupe et c'est seulement récemment qu'une étude a défini, chez chaque sujet, les zones activées par la lecture labiale d'une part et la position anatomique du gyrus transverse d'autre part : chez 7 sujets sur 10, une activation du cortex auditif primaire a été trouvée (Pekkola et coll., 2005).

Un autre débat concerne la signification fonctionnelle de cette activation. Ainsi l'ac-

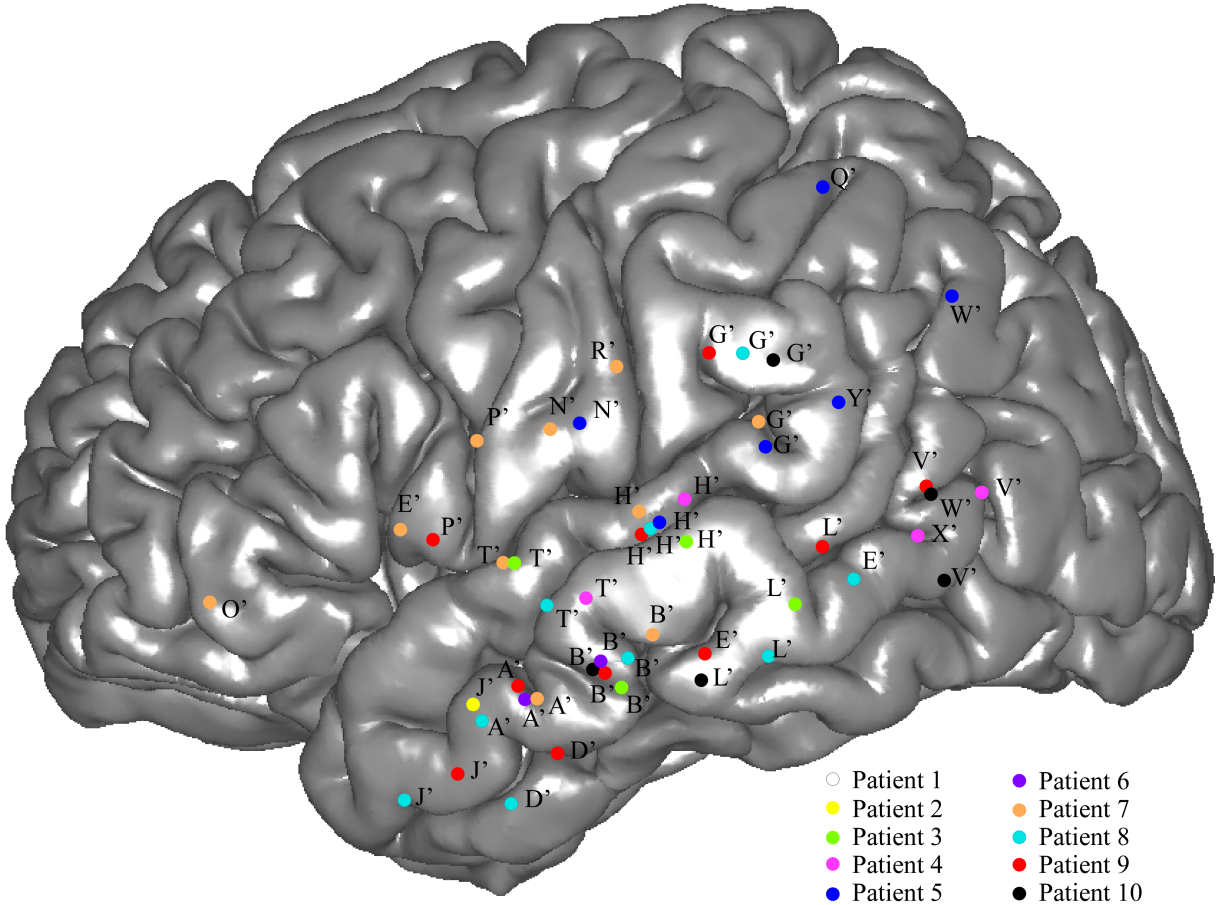

FIG. 10.2 – Emplacements des électrodes de l'hémisphère gauche reportés à la surface d'un cerveau standard (single-subject du MNI). Le recalage des électrodes des différents patients a été réalisé par la méthode de Talairach (transformation linéaire par cadrans). Chaque électrode comprend entre 5 et 15 contacts explorant les structures situées à la perpendiculaire du plan de la figure

tivation du cortex auditif (primaire ou secondaire) pourrait correspondre à de l'imagerie auditive et avoir lieu à une latence tardive : la vision des articulateurs pourrait activer des représentations phonologiques et l'accès à ces représentations permettraient aux sujets d'imaginer les sons de parole correspondant, ce qui pourrait activer le cortex auditif. Certains éléments suggèrent cependant que ce scénario est peu plausible : en effet deux études ont montré une activation du cortex auditif par des mouvements labiaux ressemblant à de la parole mais ne correspondant à aucun mot ou son connu (phonèmes étrangers : Calvert et coll., 1997 ou vidéos passées à l'envers : Paulesu et coll., 2003). Il se pourrait donc que le cortex auditif (primaire ou secondaire) participe au décodage phonologique de la parole visuelle comme il participe à celui de la parole auditive.

La résolution temporelle de la sEEG, ainsi que sa résolution spatiale devraient permettre d'apporter des éléments concernant la signification fonctionnelle des activations du cortex auditif, en donnant la latence d'activation de ses différentes parties, ainsi qu'une preuve directe de l'implication ou non du cortex auditif primaire dans la lecture labiale.

Patient	1	2	3	4	5	6	7	8	9	10	moyenne
V	91	130	125	120	135	81	103	79	92	82	104
A	87	128	130	130	144	84	103	90	81	81	106
AV	92	129	130	135	142	81	106	87	94	77	107

TAB. 10.1 – Nombre d'essais pris en compte pour le calcul des potentiels évoqués et des tests statistiques. V : condition visuelle. A : condition auditive. AV : condition audiovisuelle.

10.2 Méthodes

10.2.1 Patients

10 patients ont participé à cette étude. Aucun de ces patients ne souffrait de troubles auditifs (excepté le patient 1 qui était capable de lire sur les lèvres) ou visuels.

10.2.2 Stimuli et procédure

Les stimuli, la procédure et la tâche des patients étaient identiques à ceux employés dans l'étude d'EEG de scalp, excepté que seuls 8 blocs de 66 stimuli (d'une durée de 2 minutes 15 chacun) étaient présentés. Le nombre total de stimuli non-cibles présentés était de 150 dans chacune des conditions de présentation.

Pour 6 des patients (patients 5 à 10), nous avons ajouté des essais audiovisuels incongruents. Les résultats pour cette condition expérimentale ne seront pas rapportés ici. Afin de ne pas rallonger la durée de l'expérience, le nombre total de stimuli était identique avec et sans syllabe incongruentes, si bien que le nombre d'essais moyen par condition pour les 6 derniers patients était diminué d'un quart (108 essais par condition expérimentale).

10.2.3 Calcul des potentiels évoqués

Les méthodes de calcul des PE intracérébraux ayant été exposées dans la partie 6.4 page 92, nous nous contenterons de rappeler que les essais comprenant des valeurs d'amplitude, supérieures en valeur absolue à 5 écart-types de la distribution des amplitudes sur l'ensemble des essais dans une condition donnée, étaient rejetés avant le moyennage, afin d'éviter la contamination des données par les pointes inter-critiques. Le nombre d'essais retenus après rejet des artéfacts pour l'analyse par conditions et par patients est donné dans la table 10.1.

Les contacts qui participaient à plus de 6% de rejet étaient considérés comme mauvais et exclus de l'analyse. Le nombre de contacts retenus par patients après rejet des artéfacts est donné dans la table 10.2 page suivante. Pour les tests d'émergences des activitis unisensorielles, nous n'avons pas appliqué cette contrainte (voir la partie 10.2.4 page ci-contre).

Rappelons également que, comme pour l'étude en EEG de surfaces, le temps 0 pour le calcul des PE correspondait au début de la syllabe auditive, et que la ligne de base était prise entre -300 et -150 ms.

Patient	1	2	3	4	5	6	7	8	9	10	moyenne
Modèle additif	63	45	63	63	63	65	40	62	51	42	56
A ou V vs 0	63	63	63	63	63	127	124	127	112	127	93

TAB. 10.2 – Nombre de contacts considérés pour les tests statistiques. A ou V vs 0 : test d'émergence

10.2.4 Analyses statistiques

Pour les données comportementales, nous avons comparé le TR pour les syllabes auditives et les syllabes audiovisuelles cibles, pour chaque patient et au niveau du groupe. Les TR moyens de chaque patient étaient comparés par un test de Student pour groupes indépendants et les TR moyens du groupe étaient comparés par un test de Student pour mesures appariées.

Pour tous les tests statistiques portant sur les données électrophysiologiques, le signal a été sous-échantillonné à 50 Hz, l'amplitude à un échantillon temporel donné étant égal à la moyenne du signal dans une fenêtre de 40 ms autour de cet échantillon.

Pour le calcul des interactions audiovisuelles, nous avons testé la violation du modèle additif à chaque échantillon temporel de chaque contact retenu pour l'analyse entre 0 et 200 ms (20 échantillons temporels ; les tests ont en fait été réalisés entre -300 et 600 ms après le stimulus auditif, mais nous ne considèrerons que les violations du modèle additif qui commençaient avant 200 ms post stimulus, voir la partie 7.2.1 page 108). Le nombre moyen de contacts retenus par patient était de 56 (voir la table 10.2), ce qui donne un total de 1120 tests en moyenne par patient.

Les tests multiples étaient pris en compte indépendamment pour chaque patient, dans les dimensions spatiales et temporelles. Dans la dimension temporelle, nous avons utilisé la méthode du minimum d'échantillons consécutifs significatifs (voir la partie 8.1 page 112). Pour tenir compte des tests multiples dans la dimension des capteurs, nous avons appliqué la correction de Bonferroni (voir la partie 8.1 page 111) et exigé que les violations du modèle additif soient significatives à $p<0{,}001$, ce qui correspond à un seuil classique de 0,05 divisé par 50 (le nombre approximatif de contacts par patient). En réalité ce seuil est sans doute trop conservateur car les signaux enregistrés sur des contacts voisins sont souvent corrélés (mais pas toujours, en particulier dans le cas de gradients locaux importants).

Pour l'analyse des réponses visuelles seules et des réponse auditives seules, nous n'avons considéré que les réponses qui différaient significativement de la ligne de base. La significativité était testée par un test non paramétrique apparié (test de Wilcoxon) à chacun des échantillons entre -150 et 600 ms (38 échantillons temporels). À la différence du test du modèle additif, l'émergence des réponses sensorielles a été testée sur l'ensemble des capteurs enregistrés et pas seulement sur ceux conservés lors du rejet des artéfacts, de façon à augmenter l'échantillonnage spatial et parce que l'on s'attend à observer des effets moins sensibles au bruit dans ce cas. Le nombre de tests réalisés par patient était donc en moyenne de 38 échantillons $\times$ 93 capteurs, c'est-à-dire environ 3500 tests. Pour ces tests, je n'ai pas eu le temps d'implémenter la méthode du minimum d'échantillons significatifs consécutifs, nous avons donc corrigé le seuil de significativité par la méthode

de Bonferroni dans les dimensions temporelles et spatiales, c'est-à-dire utilisé une seuil égal à $0,05/3500 = 1,4 \times 10^{-5}$. La conservativité de cette approche est cependant moins problématique ici que dans le cas du modèle additif car les effets sont de manière générale plus robustes.

Afin de localiser le cortex auditif primaire, nous avons en particulier recherché les premières réponses sensorielles auditives corticales qui apparaissent à partir de 10-15 ms. Ces réponses étant des réponses transitoires rapides, les tests de significativité (test de Wilcoxon par rapport à la ligne de base) ont été menés sur les données échantillonnées à 1000 Hz (512 Hz pour la seconde moitié des sujets) entre 10 et 40 ms (respectivement 30 et 15 échantillons), sur les électrodes traversant les gyrus temporal supérieur. Pour ce test, un seuil de $p < 10^{-5}$ était suffisant.

Tous les tests ont été menés à la fois sur les données monopolaires et les données bipolaires. Mais seules les données bipolaires ont été prises en considération pour l'application des critères statistiques, de manière à pouvoir attribuer l'effet à la région traversée par le contact concerné (en particulier pour l'analyse de groupe). Les données monopolaires n'étaient donc utilisées que pour la description et l'interprétation des résultats (excepté dans un cas, qui sera signalé).

Dans tous les cas, lorsqu'un effet (violation du modèle additif ou émergence de la réponse unisensorielle) remplissait les critères statistiques requis, c'est l'ensemble de l'effet présentant une unité spatiale et temporelle qui était pris en compte dans l'interprétation, même s'il ne remplissait pas les critères à tous les contacts et à tous les échantillons temporels concernés. En d'autres termes, lorsqu'un effet était significatif sur un certain nombre d'échantillons consécutifs et sur un certain nombre de contacts voisins, il suffisait qu'au moins un échantillon remplisse les critères, pour que cet effet soit retenu et/ou décrit dans son intégralité.

10.3 Résultats

10.3.1 Données comportementales

La figure 10.3 page suivante montre les temps de réactions des 10 patients pour les syllabes auditives et audiovisuelles. En moyenne, les TR étaient plus rapides en condition audiovisuelle, mais cette différence n'était pas significative (p=0,13). Au niveau individuel, seul le patient 4 montrait une facilitation significative pour détecter les syllabes en condition audiovisuelle. Aucune des autres différences n'était significative.

10.3.2 Réponses évoquées auditives

Les réponses auditives évoquées par les syllabes se manifestaient comme une succession d'ondes transitoires enregistrées principalement dans le gyrus temporal supérieur et d'amplitude beaucoup plus importante que celles enregistrées dans les mêmes régions pour les activités visuelles. Ces activités n'étant pas l'objet principal de cette étude, on se contentera de les décrire de façon globale en négligeant les aspects propres à chaque patient.

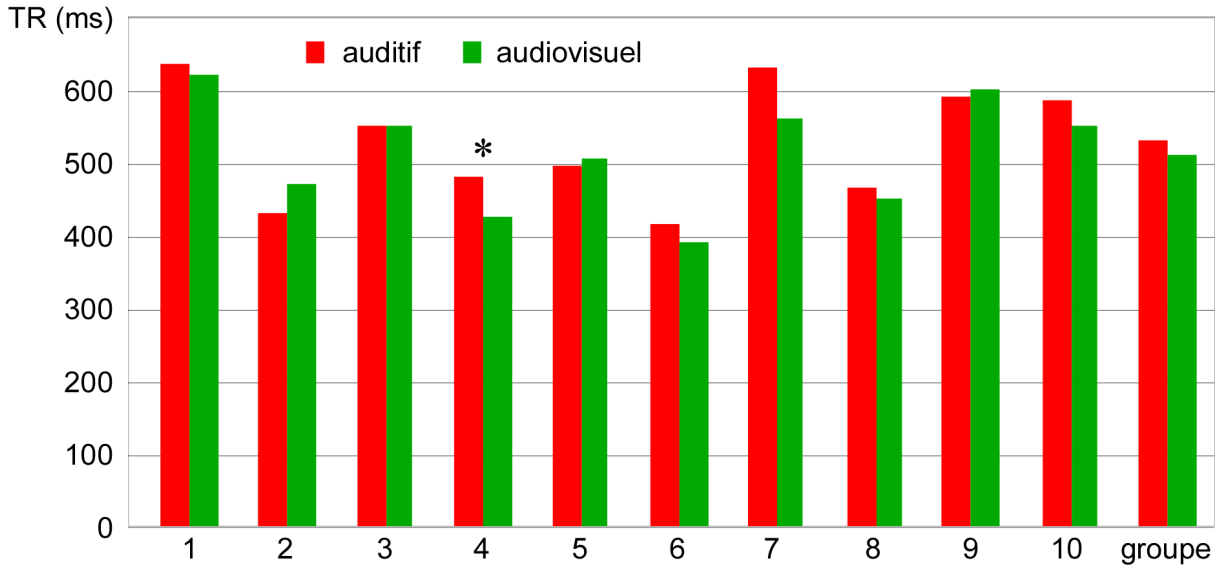

FIG. 10.3 – TR moyens auditifs et audiovisuels par patient et pour le groupe de patients. L'étoile indique une différences significative au seuil $p<0,05$.

Soulignons simplement que la variabilité des réponses peut être attribuée tout autant à des différences d'implantation, qu'à une variabilité anatomique et fonctionnelle. Malgré cette variabilité, on peut aisément distinguer plusieurs composantes communes à la plupart des patients (figure 10.4 page suivante).

Les premières réponses étaient enregistrées dans la partie médiane du gyrus transverse antérieur (ou gyrus de Heschl) à partir de 15 ms. Le détail des réponses enregistrées dans les 30 premières millisecondes est donné dans la table 10.3 page 139.

Les réponses s'étendaient ensuite dans les parties plus latérales du gyrus transverse ainsi que vers l'arrière sur le planum temporale, à partir de 40 ms post-stimulus. Toutes ces réponses étaient de polarité aussi bien positive que négative (en montage monopolaire). À partir de 70 ms commençait une réponse enregistrée majoritairement comme positive et dont l'amplitude culminait vers 100-130 ms. Cette réponse était enregistrée au niveau du gyrus transverse, du planum temporale, ainsi que sur la partie latérale du gyrus temporal supérieur (GTS) jusqu'à des zones assez postérieures jouxtant le gyrus supramarginal (et correspondant à l'aire Wernicke). Cette composante était suivie par une autre composante d'origine similaire, de polarité majoritairement négative dont le maximum d'amplitude avait lieu autour de 200 ms. Des exemples de ces différentes réponses sont visible chez le patient 6 sur les contacts H3-5 (figure A.3 page 232) ou chez le patient 8, électrode T9 (figure A.5 page 235). Des réponses d'amplitude beaucoup plus faible étaient également visibles dans plusieurs autres régions corticales à partir de 70 ms.

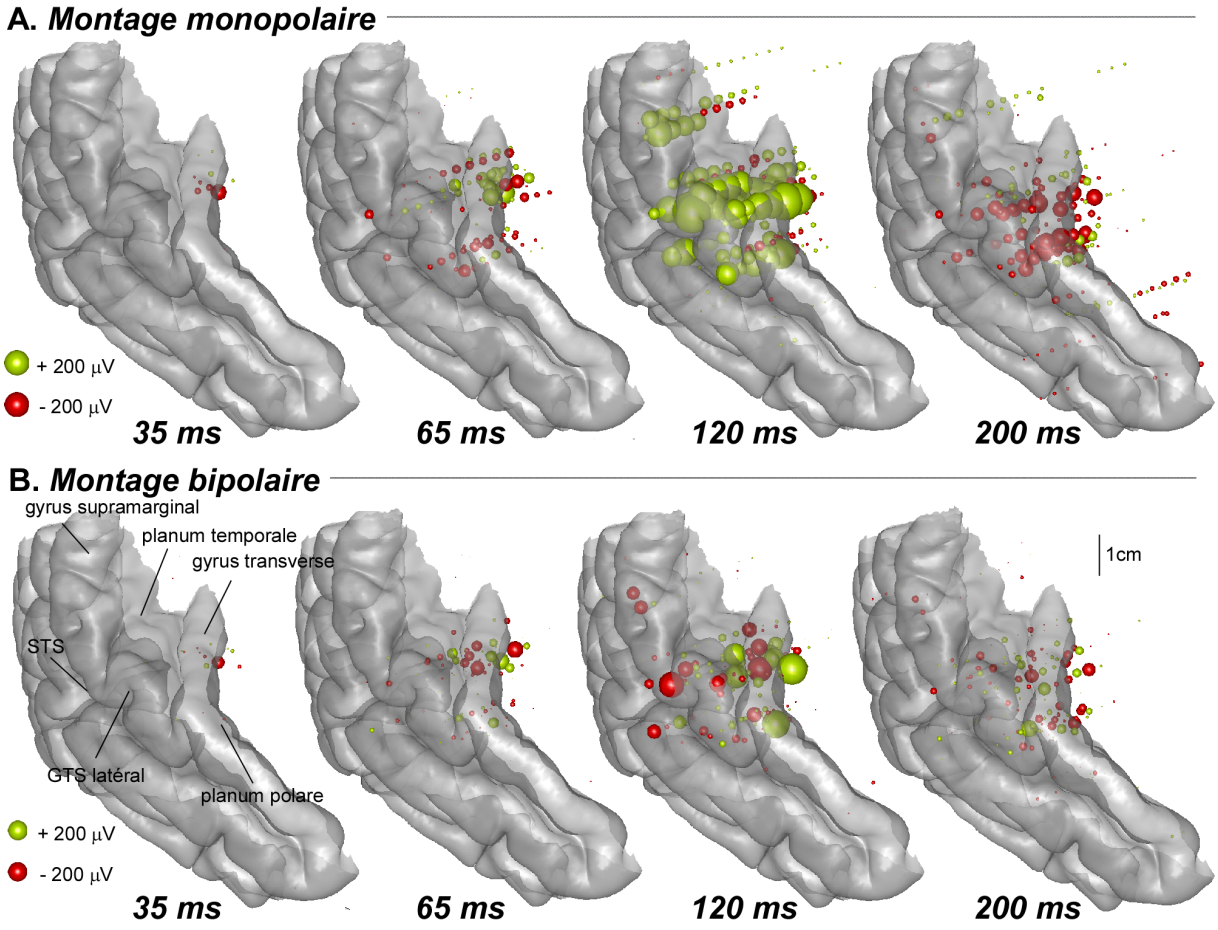

FIG. 10.4 – Réponses auditives de l'ensemble des patients enregistrées dans le cortex temporal, présentés sur une représentation 3D du lobe temporal droit du cerveau du MNI. Les activités enregistrées dans l'hémisphère gauche et droit ont été reportées sur un même hémiphère. Chaque sphère représente la différence de potentiel enregistrée à un contact en montage monopolaire (A) ou bipolaire (B). Le diamètre de la sphère est proportionnel à l'amplitude du potentiel évoqué et la couleur code la polarité. Les coordonnées des contacts des différents patients ont été normalisées et converties dans le repère du cerveau du MNI.

10.3.3 Réponses évoquées visuelles

Les réponses visuelles au mouvement des articulateurs étaient d'amplitude plus faible et avait un caractère moins transitoire que la réponse auditive aux syllabes. Cette différence peut s'expliquer par plusieurs facteurs : d'une part aucune électrode n'explorait les zones visuelles sensorielles (en tous cas pas primaires), d'autre part les stimuli employés dans la modalité visuelle présentaient un départ beaucoup moins abrupt que ceux utilisées dans la modalité auditive, ce qui ne facilite pas l'obtention de réponses élémentaires synchronisées permettant l'observation d'un potentiel évoqué net. Cette différence rappelle celle obtenue en PE de scalp dans l'expérience précédente.

La table A.1 (pages 224–226) rapporte l'ensemble des activités significatives enregistrées en réponse aux mouvements labiaux présentés seuls, que nous avons regroupé par type d'activité présentant des caractéristiques temporelles, spatiales et fonctionnelles communes

Patient	Région explorée	Côté	Latence de début (ms)	Nom des contacts	Coordonnées de Talairach X	Y	Z
5	Gyrus transverse antérieur médial	G	14	H'2	-35	-24	7
10	Gyrus transverse antérieur médial	D	17	H10	43	-19	7
4	Gyrus transverse antérieur médial	G	19	H'6-7	-33	-28	10
7	Gyrus transverse antérieur/planum temporale	G	23	H'6	-48	-22	9
9	Gyrus transverse antérieur médial	G	23	H'6-8	-35	-22	6
6	Gyrus transverse antérieur médial	D	23	H4	39	-20	5
10	Gyrus transverse antérieur médial	D	23	H6	30	-19	7
10	Gyrus transverse antérieur médial	D	23	H8	36	-19	7
4	Gyrus transverse antérieur/planum temporale	G	25	H'8-9	-41	-28	10
8	Gyrus transverse antérieur latéral	D	25	T3-5	42	-11	7
6	Gyrus transverse antérieur médial	D	27	H3	36	-20	5
8	Gyrus transverse antérieur médial	G	27	H'8	-39	-23	7
7	Gyrus transverse antérieur	G	29	H'2-3	-36	-22	9

TAB. 10.3 – Coordonnées, localisation et latence des réponses auditives commençant avant 30 ms chez les différents patients. Les réponses sont classées par latence. Les structures traversées ont été déterminées visuellement sur l'IRM anatomique de chaque patient. Le nom des contacts est constitué de la lettre désignant l'électrode (localisation sur les figures 10.1 page 132 et 10.2 page 133) et du numéro du contact, les nombres les plus petits indiquant les contacts les plus profonds.

et ayant été trouvé chez au moins 3 patients.

On se contentera ici de décrire les 3 premiers types de réponses, les plus précoces, enregistrées dans le MTG postérieur et le STG. D'autres réponses visuelles, généralement plus tardives (à partir de 80 ms après le début du son) ont été enregistrées dans de nombreuses régions. Les régions trouvées chez au moins trois patients étaient : le gyrus supramarginal, le STS antérieur et postérieur, l'opercule post-central, l'insula, le gyrus cingulaire postérieur, l'opercule pré-central/gyrus frontal inférieur (pouvant correspondre à l'aire de Broca), l'hippocampe/ gyrus parahippocampique.

Rappelons que le mouvement des lèvres commençait à partir de 240 ms préstimulus (le temps zéro correspondant au début de la syllabe auditive). Il ne faut donc pas s'étonner que les réponses les plus précoces apparaissaient dès 120 ms pré-stimulus. Ces réponses ont été enregistrées d'une part au niveau de la jonction occipito-temporale et du GTM postérieur et d'autre part au niveau du GTS sur des électrodes explorant aussi bien le gyrus transverse, le planum temporale, le planum polare, le GTS latéral et le bord supérieur du STS.

Concernant la zone occipito-temporale, une réponse y a été enregistrée chez tous les patients dont l'implantation était aussi postérieure. Cette réponse était spécifique à la condition visuelle, et lorsqu'une réponse auditive était enregistrée plus tard dans la même zone, son profil spatio-temporel était clairement différent.

Concernant la partie supérieure du lobe temporal, des réponses visuelles ont été enregistrées sur les mêmes contacts que ceux sur lesquels ont été observés les potentiels évoqués auditifs sensoriels entre 50 et 200 ms. L'un des buts de cette étude étant de vérifier si on peut enregistrer une réponse aux mouvements articulatoires dans le cortex auditif, il nous faut nous assurer que ces réponses proviennent bien du plan supérieur du GTS et non

du STS. En effet, la localisation des contacts ne suffit pas puisque l'activité enregistrée, même en montage bipolaire peut correspondre à la diffusion des potentiels dans le milieu extracellulaire. Cette ambigüité est clairement illustrée dans l'implantation du patient 3 (figure A.2 page 230) : l'électrode H' passe entre le bord supérieur du STS et le planum temporale : il est impossible de dire si une activité enregistrée sur un des contacts de l'électrode H' provient du cortex situé en-dessous ou au-dessus des contacts. Pour répondre à cette question, nous avons comparé le profil spatiale des réponses visuelles à celui des premières réponses auditives transitoires. Il est en effet bien établi que ces réponses auditives précoces sont générées dans le cortex auditif (Liégeois-Chauvel, Musolino, Badier, Marquis & Chauvel, 1994 ; Yvert, Fischer, Bertrand & Pernier, 2005), comme le montre également la figure 10.4 page 138. Si nous pouvons montrer que les réponses visuelles enregistrées dans le lobe supérieur temporal possèdent le même gradient spatial que cette réponse auditive, on pourra en conclure qu'elle est bien générée dans le cortex auditif.

Nous avons classé les différents types de réponse visuelle enregistrées dans le cortex temporal supérieur en fonction de leur ressemblance spatiale avec la réponse auditive transitoire. Sur 12 sites répartis parmi 5 patients, le gradient spatial de la réponse visuelle ressemblait à celui d'une réponse auditive générée à partir de 50 ms, donc à une réponse dont l'origine dans le cortex auditif ne fait guère de doute (type 2 dans la table A.1 page 226 (pages 224–226). On peut voir des exemples d'une telle réponse chez le patient 3 (figure A.2 page 230) au niveau des contacts T4-5 et T7-9 (correspondant respectivement au bord supérieur du STS et au gyrus transverse latéral), chez le patient 8 (figure A.5 page 235), au niveau des contacts H'11-15 et T'8-9 (Planum temporale et STS/GTS latéral). Dans d'autres cas, la ressemblance est plus vague (patient 1, contacts T7-9, gyrus transverse latéral, figure A.1 page 229). Notons que dans le cas du patient 3, la réponse auditive entre 50 et 100 ms était enregistrée avec un gradient plus fort sur le bord supérieur du STS que sur le gyrus transverse, ce qui suggère que le cortex auditif s'étend dans les aires corticales bordant le STS dans le cas de ce patient.

Sur 6 sites répartis sur 4 patients, la réponse visuelle montrait une ressemblance frappante avec une réponse auditive transitoire aux syllabes commençant après 100 ms (type 3 dans la table A.1 pages 224–226). On peut en voir des exemples chez le patient 1 (figure A.1 page 229) sur les contacts H8-10 (gyrus transverse médial/planum temporale) et chez le patient 7 (figure A.4 page 233) au niveau des contacts T'5-7 (planum polare). D'autres sites ne montrent pas le même profil spatial dans les deux modalités, mais l'on observe de forts gradients spatiaux au niveau des mêmes électrodes dans les deux conditions : c'est le cas pour le patient 10 (figure A.6 page 237) au niveau des contacts H7-15 (gyrus transverse médial et planum temporale) et pour le patient 6 (figure A.1 page 229) au niveau des contacts H3-9 (gyrus transverse antérieur médial et postérieur latéral, mais dans ce dernier cas, la réponse visuelle n'était pas significative avec le critère requis).

Au total une telle activation visuelle du cortex auditif a été trouvée chez 7 patients, sur 18 sites. Une telle affirmation n'est pas basée sur une délimitation anatomique du cortex auditif, mais plutôt sur une définition fonctionnelle assez large : le cortex auditif est défini comme la zone du cortex temporal dans laquelle on enregistre une réponse évoquée transitoire à un son ; en effet ces 18 sites comprennent aussi bien le planum polare, le gyrus

transverse, le STG latéral, le bord supérieur du STS que le planum polare jusqu'au gyrus supramarginal.

Un autre argument permettant d'affirmer que cette réponse visuelle venait de la partie supérieure du GTS et non du STS est qu'elle n'était pas enregistrée dans le GTM, ou avec une amplitude beaucoup plus faible, alors que les implantations dans cette région étaient assez nombreuses, comme on peut le voir sur les figures 10.1 à 10.2 pages 132–133 (données non illustrées).

Un autre but de cette étude était de savoir si la réponse visuelle dans le cortex auditif pouvait être générée dans le cortex auditif primaire. Une façon de répondre à cette question est de comparer l'emplacement des sites d'enregistrement de cette réponse visuelle avec la position des sites d'enregistrement des réponses auditives transitoires générées avant 30 ms, probablement dans le cortex auditif primaire. Une telle réponse auditive a été enregistrée sur 13 sites chez 7 patients, exclusivement dans la partie médiale du gyrus transverse, comme c'est illustré dans la figure 10.5 page suivante (aux erreurs de normalisation près). Considérées au niveau individuel, toutes ces réponses étaient enregistrées dans le gyrus transverse médial (voir la table 10.3 page 139). Une comparaison de ces activations auditives primaires avec les réponses visuelles enregistrées dans le cortex auditif (figure 10.5) suggère que les réponses visuelles étaient toujours enregistrées en dehors de la zone définie par les réponses auditives précoces. Cependant, les erreurs de localisation dues à la normalisation des coordonnées et à l'utilisation d'un cerveau standard ne permettent pas d'être catégorique sur ce point.

Si l'on regarde individuellement chaque patient, seuls deux d'entre eux montraient les deux types de réponse sur des contacts voisins : pour le patient 8 (figure A.5 page 235), les foyers étaient clairement différents puisque la réponse visuelle était enregistrée uniquement sur les contacts H'11-15 (planum temporale) alors que la réponse auditive précoce était enregistrée sur H'8-10 (gyrus transverse médial). Quant au patient 10, si le profil spatial de la réponse auditive précoce sur H'7-9 était bel et bien différent de celui de la réponse visuelle, ces deux réponses étaient enregistrées sur les même contacts (c'est également vrai pour le patient 6, contacts H3-4, mais la réponse visuelle émergeait à peine du bruit dans ce cas et n'était pas significative avec le critère requis). L'analyse qualitative de groupe suggère donc que les réponses visuelles dans le cortex auditif sont en général générées hors du cortex auditif primaire. Toutefois, les données d'un (ou peut-être deux) patients suggèrent une activation visuelle du cortex auditif primaire.

10.3.4 Violations du modèle additif

La table A.2 page 227 rapporte les violations significatives de l'additivité des réponses auditives et visuelles, qui signent l'existence d'interactions. Ces violations peuvent être classées en deux catégories selon leur profil spatio-temporel. La figure 10.6 page 143 montre la localisation de ces deux types de violation, qui avaient toutes lieux dans le cortex temporal supérieur. D'autres violations du modèle additif ont été trouvées dans des régions diverses, en dehors du cortex temporal, sans qu'il soit possible d'en dégager une unité fonctionnelle, temporelle ou anatomique (voir la table A.2 pour des détails).

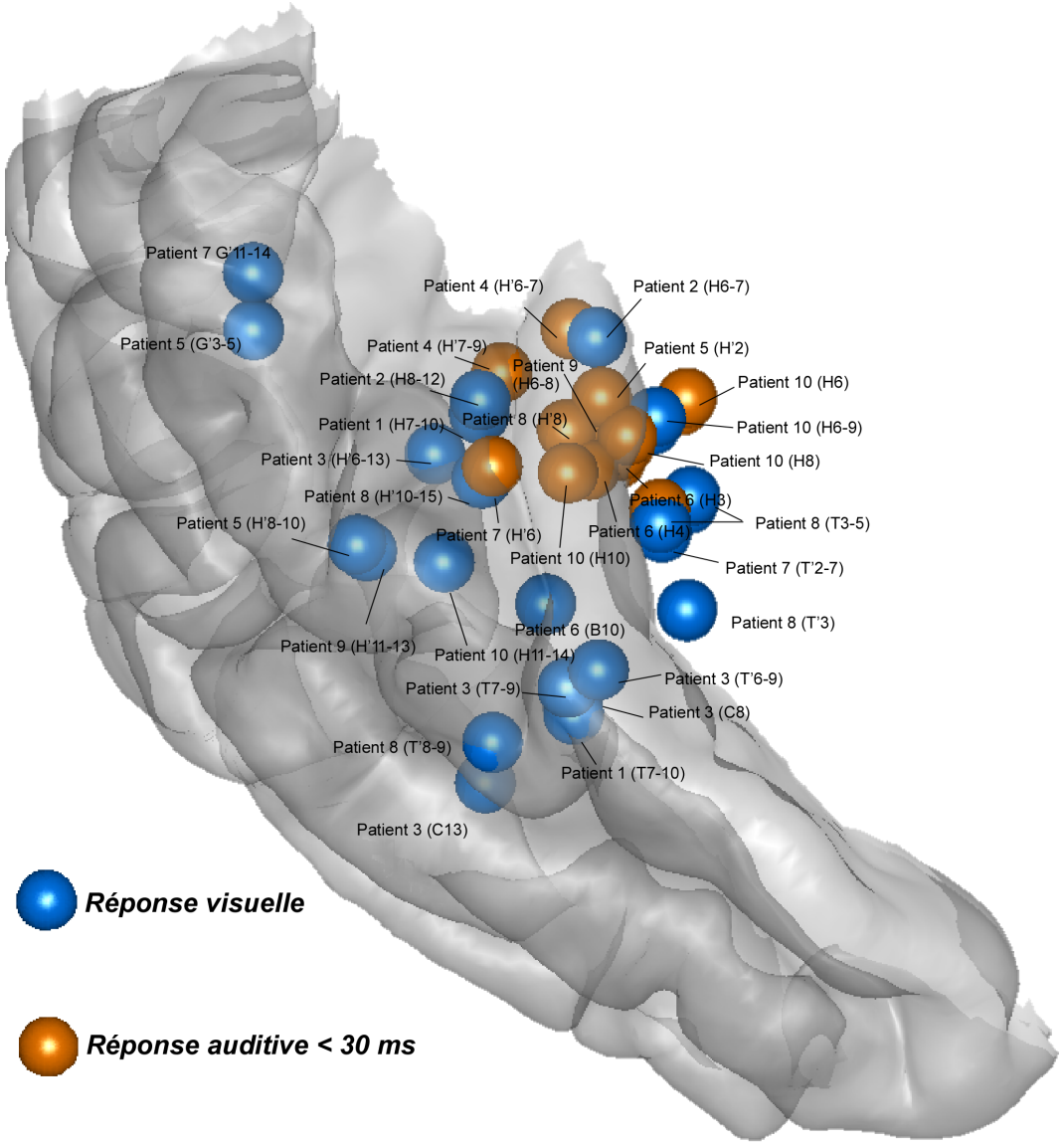

FIG. 10.5 – Sites d'enregistrement des réponses visuelles générées dans le cortex auditif et des réponses auditives précoces générées dans le cortex auditif primaire, présentés sur une représentation 3D du lobe temporal droit du cerveau du MNI. Les activités enregistrées dans l'hémisphère gauche et droit ont été reportées sur un même hémiphère. On considère que la réponse visuelle était générée dans le cortex auditif lorsque le profil spatial de la réponse le long des contacts d'une même électrode était identique à celui d'une réponse auditive transitoire générée entre 50 et 200 ms. On considère qu'une réponse auditive était primaire lorsqu'elle apparaissait avant 30 ms de traitement.

Le premier type de violation du modèle additif a été observé sur 19 sites chez 9 patients. Ces sites étaient tous situés dans la partie supérieure du GTS, dans la région que nous avons définie plus haut comme le cortex auditif au sens large. Ce type de violation de

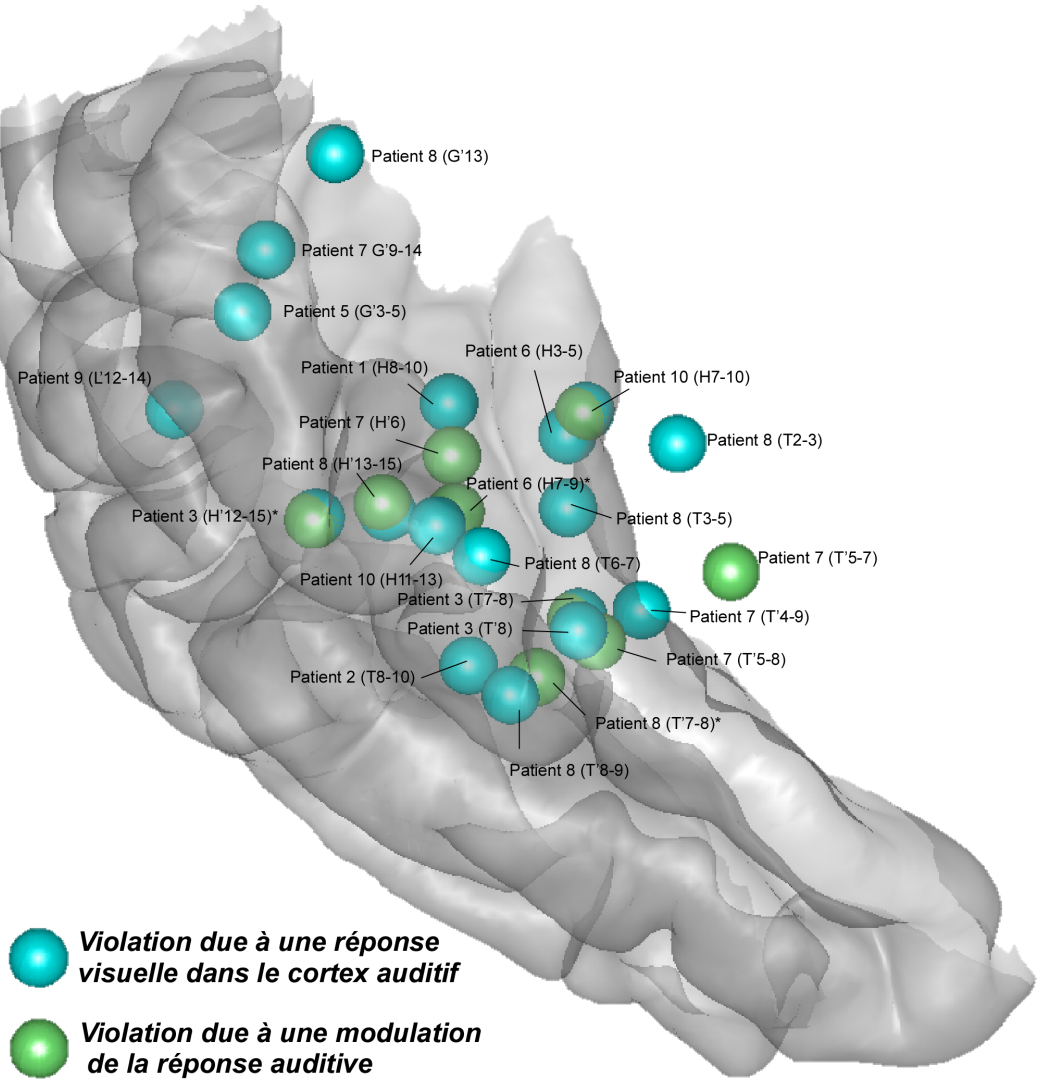

FIG. 10.6 – Deux types principaux de violations du modèle additif commençant avant 200 ms de traitement, présentés sur une représentation 3D du lobe temporal droit du cerveau du MNI. Les activités enregistrées dans l'hémisphère gauche et droit ont été reportées sur un même hémiphère. Les contacts sur lesquels étaient observées ces violations sont indiqués entre parenthèse.

l'additivité est visible chez tous les patients dont les résultats sont illustrés (figures A.1 à A.6 pages 229–237). Ces interactions se présentaient sous la forme suivante : la violation de l'additivité commençait entre 30 et 160 ms après la présentation de la syllabe auditive pour continuer au-delà de la fenêtre d'analyse (200 ms) et souvent au-delà de 600 ms. Le profil spatio-temporel de la violation est exactement celui de la réponse visuelle, mais de polarité opposée. Cela est probablement dû au fait que la réponse en condition audiovisuelle diffère peu de la réponse en condition auditive, autrement dit que la réponse visuelle dans le cortex auditif semble ne pas exister lorsque la stimulation est audiovisuelle, mais seulement lorsque les mouvements articulatoires sont présentés seuls.

Le second type de violation avait lieu entre 40 et 200 ms après la syllabe auditive, au

niveau du gyrus transverse et du planum temporale. Ici, le profil spatio-temporel correspond à celui de la réponse auditive transitoire avec une polarité opposée. Ce type de violation correspond apparemment à une diminution de la réponse auditive transitoire en condition audiovisuelle. On voit clairement cette modulation chez 2 patients.

Chez le patient 8 (figure A.5 page 235), sur le contact H' 11 (planum temporale), en montage bipolaire, on voit clairement un foyer identique à l'activité auditive et audiovisuelle entre 60 et 120 ms, qui n'est pas présent en visuel. La diminution est visible sur les courbes et la violation du modèle additif montre un rebond qui est absent de la réponse visuelle. Chez le patient 10 (figure A.6 page 237), l'activité bipolaire montre une triple inversion de polarité entre 80 et 160 ms aux contacts H6, 7 et 9 (gyrus transverse médial), identique aux inversions observées en conditions auditives et audiovisuelles. À cette latence, on n'observe pas de réponse visuelle dans cette zone. Chez d'autres patients, l'interprétation est plus ambigüe puisque cette forme de violation se superpose au premier type : la violation semble être due à la fois à l'absence de réponse visuelle en condition audiovisuelle et à une diminution de la réponse auditive, à la même latence (patient 7, contacts T'5-7 entre 120 et 200, patient 8, contact H'13 entre 80 et 160 ms).

Enfin, chez certains patients il a fallu augmenter le seuil pour observer cette diminution, tout en conservant l'exigence d'un nombre minimal d'échantillons consécutifs significatifs (patient 7, contacts T'7-8, bord supérieur du STS, entre 60 et 100ms ; patient 8 contacts T'7-8 bord supérieur du STS ; patient 3, contacts H'12-15, GTS latéral entre 50 et 100 ms ; patient 3 contacts T7-8 bord supérieur du STS entre 60 et 120 ms). Notons que pour ces 3 dernières violations, la diminution n'était observée que sur les données monopolaires. L'augmentation du seuil statistique reste raisonnable si l'on considère que ces effets ne pouvaient se produire que sur les contacts sur lesquels étaient enregistrées des réponses transitoires, ce qui réduit en principe le nombre de tests à effectuer (nous reconnaissons le caractère a posteriori de cette affirmation).

La localisation de ce deuxième type d'effet ne diffère guère de celle du premier type, comme on peut le voir sur la figure 10.6 page précédente. Les modulations étaient en fait souvent superposées aux violations dues à l'activation visuelle sur les même contacts décrites plus haut, ce qui rend difficile leur description. Pour la plupart des patients (patients 3, 6, 7, 8 et 10), lorsque l'on compare les courbes de la violation aux courbes de l'activité visuelle, on constate que l'amplitude de la violation est supérieure celle de l'activité visuelle, ce qui suggère que les deux types d'interaction co-existent.

10.3.5 Relations entre réponses auditives, visuelles et interactions audiovisuelles

On peut tenter de décrire les relations existant entre l'activation auditive et visuelle du lobe temporal (supérieur) et les interactions audiovisuelles mises en évidence par l'application du modèle additif, au moins pour les activités communes à plusieurs patients. La table 10.4 page ci-contre donne, pour chaque patient, les latences de début et de fin des 4 principaux effets mis en évidence : l'activation visuelle de la jonction occipito-temporale, l'activation visuelle du cortex auditif, la modulation des ondes audiovisuelles transitoires en condition audiovisuelle et la violation du modèle additif due à la disparition de la réponse

Patient	Réponse V GTM post. JOT		Réponse V Cortex Auditif		Modulation réponse auditive		Disparition réponse V cortex auditif	
	début	fin	**début**	fin	**début**	fin	**début**	fin
1	**-**	-	**-20**	600+	**-**	-	**110**	250
2	**-**	-	**-120**	450	**-**	-	**40**	110
3	**-**	-	**-120**	600+	**50**	120	**80**	600+
4	**-**	-	**-**	-	**-**	-	**-**	-
5	**-**	-	**0**	600+	**-**	-	**130**	250
6	**-80**	350	**-**	-	**40**	120	**30**	600+
7	**-**	-	**-20**	450	**60**	200	**70**	600+
8	**-80**	400	**-70**	600+	**50**	120	**70**	500
9	**-100**	160	**-**	-	**-**	-	**120**	250
10	**-40**	600+	**-30**	600+	**80**	160	**80**	600+

TAB. 10.4 – Latence de début (en gras) et de fin des 4 types d'effets mis en évidence, chez chaque sujet. Réponse V : réponse visuelle significativement différente de la ligne de base. GTM post. : gyrus temporal moyen postérieur. JOT : jonction occipito-temporale. Modulation réponse auditive : violation significative du modèle additif due à une diminution d'une onde auditive transitoire en condition auditive. Disparition réponse V cortex auditif : violation significative du modèle additif due à la disparition de la réponse visuelle du cortex auditif en condition audiovisuel. 600+ : l'effet se prolonge au-delà de 600 ms.

visuelle du cortex auditif en condition audiovisuelle.

Malgré la variabilité des latences, l'enchainement des différentes activations se vérifient chez chacun des patients : lors d'une stimulation audiovisuelle, les indices visuels, qui sont disponibles plus tôt, activent tout d'abord les régions autour de la jonction occipito-temporale (patients 6, 8, 9 et 10), puis immédiatement après le cortex auditif (patients 8 et 10). Cette activation du cortex auditif peut commencer jusqu'à 100 ms avant la présentation de la syllabe auditive (patients 2 et 3). Lorsque les indices auditifs sont présentés, ils activent tout d'abord le cortex auditif primaire puis à partir de 50 ms post-stimulus des zones du cortex auditif qui ont déjà été activées par les indices visuels (voir la partie 10.3.2 page 136). C'est à ce moment que prennent place les deux types d'interaction audiovisuelle : l'amplitude de la réponse auditive est diminuée par rapport à la condition auditive seule alors que le cortex a déjà été activé par les indices visuels (patients 3, 7, 8 et 10). Immédiatement après, ou à la même latence, l'activation soutenue et faible du cortex auditif observée en modalité visuelle seule prend fin pour être dominée par le traitement des indices auditifs (patients 3, 6, 7, 8 et 10). Cette chronologie relative se vérifie en particulier chez les 2 patients chez lesquels nous avons observé les 4 effets (patients 8 et 10).

10.4 Discussion

Les données intracérébrales chez les patients épileptiques donnent des informations précieuses sur le fonctionnement du cerveau, mais proviennent de sujets dont on ne sait pas s'ils représentent un bon modèle du fonctionnement cognitif normal étant donné leur

pathologie. Nous avons donc privilégié, dans notre description des résultats, ceux qui pouvaient être caractérisés de manière fonctionnelle, anatomique et/ou temporelle de la même manière chez plusieurs patients.

10.4.1 Activité du cortex auditif en réponse aux indices visuels de parole

La vision des mouvements articulatoires active de nombreuses aires cérébrales dont la jonction occipito-temporale, le GTS (gyrus transverse, planum temporale, planum polare, GTS latéral), le STS antérieur et postérieur, le gyrus supra-marginal, le STS postérieur, l'opercule post-central, l'opercule pré-central, le gyrus frontal inférieur postérieur, l'insula, l'hippocampe ou le gyrus para-hippocampique. La liste n'est bien évidemment pas exhaustive, d'autant plus que nombre d'aires cérébrales n'étaient pas explorées. Parmi ces aires, on peut en particulier distinguer la jonction occipito-temporale et le GTS dont l'activation, bien que la plupart du temps assez soutenue, commençait avant celle des autres aires cérébrales mentionnées (à partir de 100 ms avant le stimulus auditif, c'est-à-dire 140 ms post-stimulus visuel).

La jonction occipito-temporale faisant partie du cortex visuel, il n'est pas étonnant qu'elle soit la première aire que nous voyions activée par un stimulus visuel. En revanche, il est frappant de voir que le GTS est activé presque à la même latence. La comparaison des profils spatiaux de cette activation avec les réponses auditives transitoires montre qu'il s'agit d'une activation visuelle du cortex auditif. Cette activation avait déjà été rapportée par la plupart des études IRMf sur la lecture labiale, mais c'est la première fois à ma connaissance que l'on a accès à sa dimension temporelle. Il semble qu'elle soit donc relativement précoce puisqu'elle suit de très peu les traitements dans le cortex visuel (ce qu'on en voit en tous cas) et il est donc peu probable qu'elle représente un phénomène d'imagerie auditive. L'analyse de groupe suggère cependant que cette activation a en général lieu hors du cortex auditif primaire, contrairement à ce qui a été montré en IRMf par un certain nombre d'auteurs (Calvert et coll., 1997 ; Ludman et coll., 2000 ; MacSweeney et coll., 2001 ; Pekkola et coll., 2005). Une telle activation est cependant observée chez au moins un patient (le patient 10). Ce résultat peut être attribué soit à un défaut de couverture spatiale chez les autres patients, soit à une réponse atypique chez ce patient.

Les autres aires étaient activées en condition visuelle plus tardivement (en général après 50 ms post-stimulus auditif — 300 ms post-stimulus visuel — pour le STS antérieur, après 100 ms post-stimulus auditif , pour le STS postérieur et le gyrus supra-marginal et après 200 ms post-stimulus auditif dans les autres structures ; voir la figure 10.7 page ci-contre). Il est cependant hasardeux d'établir une chronologie étant donné la variabilité importante des latences entre les patients, sans doute due à la variabilité des implantations.

Notre protocole ne nous permet pas de distinguer parmi les activations trouvées celles qui sont propres à la perception visuelle de la parole et celles qui pourraient être évoquée par tout type de mouvements labiaux, contrairement aux expériences en IRMf ayant utilisé comme contrôle des mouvements labiaux non langagiers (Calvert et coll., 1997 ; Campbell et coll., 2001 ; Paulesu et coll., 2003). Les figures 10.7 page suivante et 10.8 page 148 comparent les activations visuelles trouvées dans notre étude aux résultats des études

IRMf sur la lecture labiale.

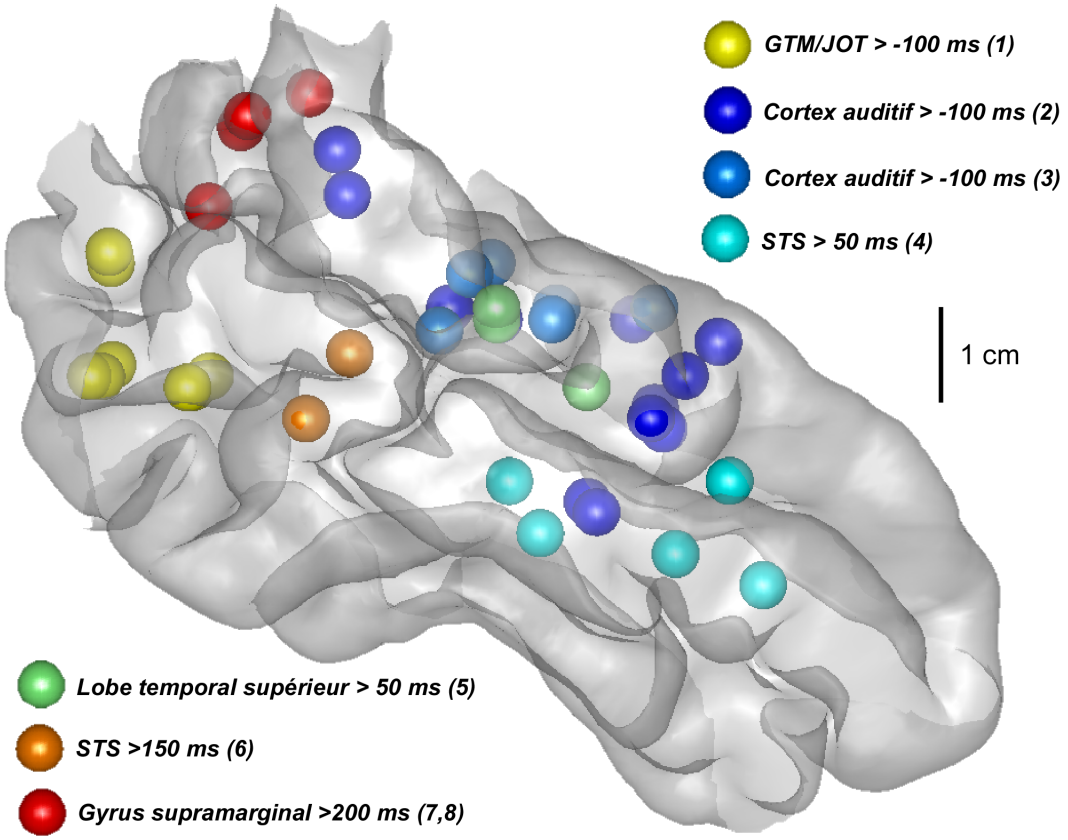

FIG. 10.7 – Activités enregistrées dans le lobe temporal en réponse aux mouvements articulatoires dans la présente étude. Les catégories de réponse correspondent à celles données dans la table A.1 page 226. Les latences sont données par rapport au début de la syllabe auditive. Il existe une discordance entre la localisation indiquée dans la légende et la situation effective sur le cerveau du MNI, due aux erreurs de normalisation. Les activations dans l'hémisphère gauche et droit ont été reportées sur un même hémiphère. JOT : jonction occipito-temporale. STS : sillon temporal supérieur. GTM : gyrus temporal moyen.

Le contraste le plus couramment utilisé dans ces études a pour but d'identifier les zones du cerveau présentant une réponse hémodynamique plus grande pour des mouvements articulatoires langagiers que pour la vision d'une bouche au repos. Il est analogue à la comparaison que nous avons effectuée entre la ligne de base et la réponse au mouvement. La localisation de ces activations en IRMf (sphères de couleur bleu foncé dans la figure 10.8) correspondent grossièrement à celles des activations que nous avons rapporté (figure 10.7), si l'on prend en compte la diffusion de potentiels en sEEG.

Certaines études IRMf ont comparé les activations induites par des mouvements labiaux non langagiers et une bouche au repos. Ces activations sont toutes regroupées au niveau de la jonction occipito-temporale et du GTM postérieur (sphères de couleur turquoise dans la figure 10.8). Il est donc vraisemblable que les premières activations que nous observons au niveau occipito-temporal ne sont pas spécifique de la parole.

Logiquement, les études IRMf qui ont testé un contraste entre mouvements langagiers et non langagiers (sphères de couleur jaune dans la figure 10.8) ont trouvé des activa-

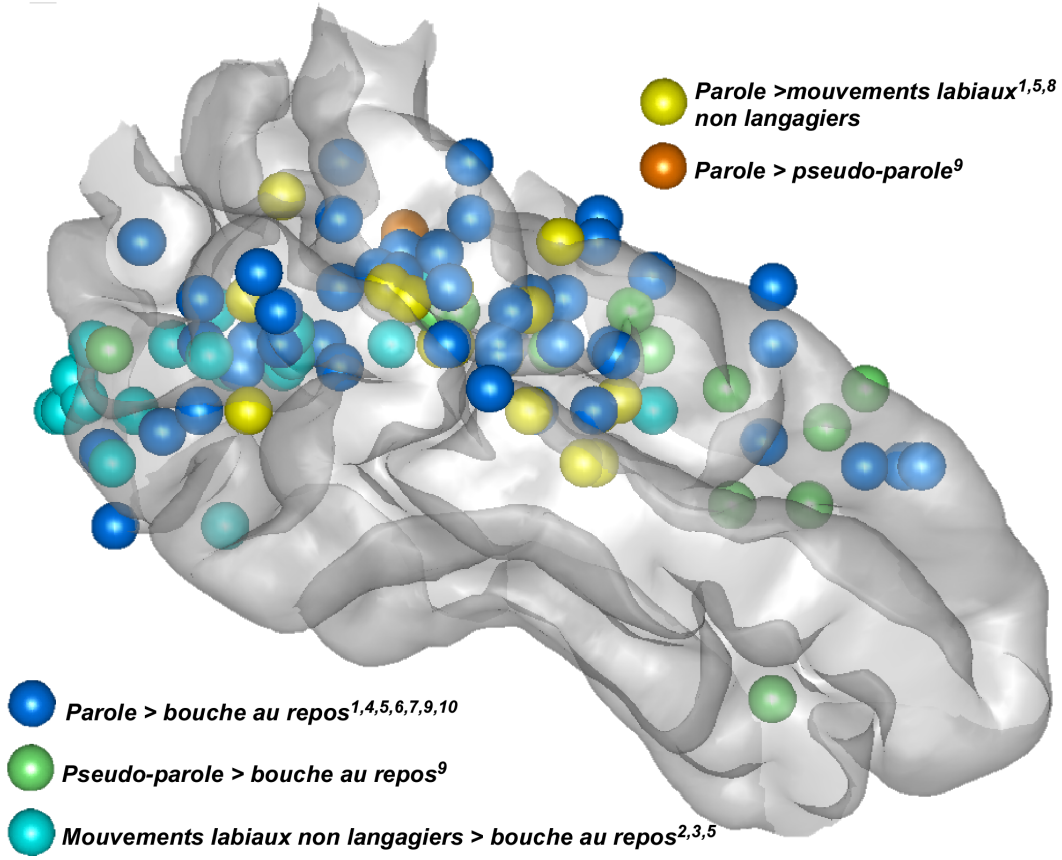

FIG. 10.8 – Activation du lobe temporal en lecture labiale. Les activités reportées proviennent de différentes études en IRMf, dont les résultats étaient reportés en coordonnées de Talairach ou directement en coordonnées du MNI. Les coordonnées de Talairach ont été converties dans le repère du cerveau du MNI. Les chiffres en exposant à côté de chaque contraste indiquent de quelle(s) étude(s) proviennent les activations : 1. Calvert et coll. (1997) 2. Puce et coll. (1998) 3. Puce et Allison (1999) 4. MacSweeney et coll. (2000) 5. Campbell et coll. (2001) 6. MacSweeney et coll. (2001) 7. Olson et coll. (2002) 8. MacSweeney et coll. (2002) 9. Paulesu et coll. (2003) 10. Calvert et Campbell (2003). Les activations dans l'hémisphère gauche et droit ont été reportées sur un même hémiphère.

tions autour du STS, du STG latéral et du planum temporale. Nous pensons donc que nos activations du cortex auditif sont spécifiques au traitement langagier des mouvements articulatoires de la bouche.

En revanche, excepté une activation dans le planum temporale rapporté par Paulesu et coll. (2003, sphère orange dans la figure 10.8), la comparaison entre des mouvements de parole ayant un sens pour le locuteur et des mouvements présentés à rebours (et n'étant donc pas interprétables phonétiquement par le sujet, pseudo-parole) n'active que des zones en dehors du lobe temporal (Calvert et coll., 1997 ; Paulesu et coll., 2003). La pseudo-parole active d'ailleurs largement le cortex auditif (sphères vertes dans la figure 10.8).

N'oublions toutefois pas que les études IRMf n'ont pas accès à la dimension temporelle et que les activations reportées dans la figure 10.8 sont susceptibles de correspondre à des activations plus tardives que celles décrites dans notre étude.

Soulignons enfin une différence fondamentale entre notre expérience et les études IRMf :

dans notre étude, les patients n'avaient pas explicitement à lire sur les lèvres. Cela n'empêche pas que les indices visuels aient eu une certaine pertinence dans la mesure où ils étaient susceptibles d'aider à réaliser la tâche demandée.

Les activations ayant lieu hors des lobes occipitaux et temporaux (opercules pré-central, post-central, le gyrus frontal inférieur, l'insula, l'hippocampe) à des latences plus tardives ont été rapportées de façon récurrente dans les études en IRMf. L'activation de l'opercule pré-central et gyrus frontal inférieur en particulier est intéressante puisqu'elle pourrait correspondre à l'aire de Broca ou à l'aire motrice correspondant aux articulateurs faciaux, dont il a été proposé qu'elle participe au décodage phonologique des sons de parole (Ojanen et coll., 2005 ; K. E. Watkins, Strafella & Paus, 2003 ; Wilson, Saygin, Sereno & Iacoboni, 2004) et/ou lors de la lecture labiale (Blasi et coll., 1999 ; MacSweeney et coll., 2001 ; Paulesu et coll., 2003 ; Sundara, Namasivayam & Chen, 2001). Nous avons cependant peu d'éléments permettant de dire que cette activation était suffisamment précoce pour remplir cette fonction. Certes, chez un patient (patient 6), l'activation de cette région commençait dès 60 ms avant l'arrivée du son. Mais, d'une part, il existe une certaine ambigüité due au fait que l'électrode sur laquelle a été enregistrée cette activité se trouvait juste au dessus du cortex auditif et, d'autre part, chez les autres patients, elle n'avait lieu qu'à partir de 200 ms après le début du son. De plus elle avait lieu dans l'hémisphère droit, alors que l'aire de Broca est censée être fortement latéralisée à gauche.

10.4.2 Interactions audiovisuelles

L'application du modèle additif a révélé de nombreuses violations du modèle additif avant 200 ms de traitement des syllabes, et ce en dépit du fait que les patient n'ont en général pas tiré parti des indices visuels pour améliorer leurs performances. Les violations observées au niveau individuel sans être reproduites chez plusieurs patients ne seront pas discutées plus avant. Ces résultats individuels peuvent être attribués à la fois à la spécificité des implantations des électrodes chez chaque patient et peut-être au caractère idiosyncratique de certaines formes d'interactions audiovisuelles.

Là où l'implantation était la plus fournie, à savoir au niveau du lobe temporal supérieur, nous avons pu mettre en évidence deux formes de violation de l'additivité, qui semblent refléter la non additivité des réponses du cortex auditif aux indices auditifs et visuels de parole. La forme de violation la plus indiscutable semble être due au fait que les indices de chaque modalité active le cortex auditif d'une manière qui lui est propre : logiquement, les activités dues aux indices auditifs sont beaucoup plus nettes, amples et transitoire que celles dues aux indices visuels. Lorsque les indices des deux modalités sont présentés (essais audiovisuels), la réponse visuelle semble complètement s'effacer au profit de la réponse auditive, ce qui résulte en des interactions dont le profil spatio-temporel imite exactement celui de l'activation visuelle avec des polarités opposées. Cette violation de l'additivité représente indubitablement une forme d'intégration des informations auditives et visuelles dans la mesure où, en condition audiovisuelle, l'activation visuelle du cortex auditif semble ne pas continuer dès lors que les même zones sont activées par les indices auditifs. Le traitement visuel semble donc influencé par la présence des indices auditifs dès 30 ms de

traitement auditif. Notons que ce type de violation de l'additivité pourrait être dû à un effet plafond de l'activation du cortex auditif.

Mais ce qui nous intéresse plus encore est de savoir si les indices visuels ont réciproquement une influence sur le traitement des syllabes auditives dans le cortex auditif. Il semble bien que ce soit le cas (bien que l'effet soit moins robuste dans ce cas) : chez 5 patients la violation de l'additivité présente un profil spatio-temporel ressemblant à celui d'une réponse auditive transitoire et ne peut être expliquée par la réponse visuelle sur ces contacts et à cette latence. Chez tous les patients, cette modulation prend place à une latence à laquelle une réponse aux indices visuels a déjà pris place, sur les même contacts. Il parait vraisemblable que la préactivation visuelle est responsable de la diminution de la réponse auditive. On peut imaginer que le traitement des indices auditifs est ici facilité par le traitement déjà réalisé sur les indices visuels. Mais, pas plus qu'en EEG, ces données ne nous permettent de dire si les informations auditives et visuelles intégrées à ce niveau sont de nature phonétique ou non ou si cette facilitation représente un amorçage phonologique ou un effet d'indiçage attentionnel.

De même que les activations visuelles décrites plus haut, ces deux types d'interaction semblent avoir lieu majoritairement dans le cortex auditif secondaire (GTS, Planum temporale, Gyrus transverse latéral, Planum polare). Quant au cortex auditif primaire, on y retrouve logiquement la première forme de violation chez deux patients (6 et 10) qui montraient également une réponse visuelle au niveau du cortex auditif primaire. On observe également une diminution de la réponse auditive transitoire au niveau du cortex auditif primaire chez le patient 10, mais il s'agit d'une réponse transitoire générée entre 80 et 160 ms et non d'une composante auditive précoce. Nous n'avons donc pas d'éléments permettant de dire que le traitement auditif des syllabes peut être modulé par les indices visuels avant 50 ms de traitement auditif.

Le cortex auditif primaire a été impliqué dans plusieurs études IRMf de l'intégration des indices auditifs et visuels de parole. Une expérience de L. M. Miller et D'Esposito (2005) a montré par exemple qu'il était plus activé lorsque la syllabe audiovisuelle était perçue comme un évènement audiovisuel unitaire que lorsque les indices auditifs et visuels n'étaient pas subjectivement fusionnés. Son activité serait également liée à l'amélioration de l'intelligibilité de la parole dans le bruit sous l'influence des indices visuels (Callan et coll., 2003). Cependant nos résultats sont contradictoires avec des données IRMf ayant utilisé un critère de super-additivité (voir la partie 4.5 page 72) pour mettre en évidence une implication du cortex auditif primaire(Calvert et coll., 2000) ou du GTS (Wright et coll., 2003). En effet, les effets observés chez nos deux patients suggèrent plutôt un effet de type sous-additif puisque l'activité visuelle semble disparaitre et que l'activité auditive semble diminuer en condition audiovisuelle. Il se peut que l'activité observée dans les études IRMf correspondent à une activité plus tardive du cortex auditif.

10.4.3 Comparaison avec l'expérience EEG de surface

Comparons maintenant les données obtenues dans cette expérience sEEG à celle obtenues en EEG de scalp. Rappelons que les stimuli étaient identiques dans les deux expériences, à ceci près que les syllabes étaient présentées dans un casque aux patient et en champ ouvert aux sujets de l'expérience EEG.

On peut faire deux constats : la réponse générée dans le cortex auditif par les indices visuels de parole n'a pas été observée en scalp, et les latences des violations de l'additivité dans les deux expériences ne correspondent pas. La réponse visuelle, tout comme les violations du modèle additif provenant du cortex auditif (types 1 et 2), devrait en principe apparaitre sur le scalp comme des inversions de polarité entre les mastoïdes et le vertex. Or, on n'observe pas une telle topographie en EEG dans la condition visuelle seule. Par ailleurs, la violation ne prend la forme d'une inversion de polarité qu'à partir de 120 ms en EEG de scalp alors qu'en sEEG le premier type de violation du modèle apparait dès 30 ms et les modulations de l'activité auditive sont visibles principalement sur des composantes générées entre 50 et 120 ms.

On peut avancer plusieurs explications pour cette divergence de résultats : Tout d'abord, il est possible que les patients épileptiques ne constituent pas un bon modèle du fonctionnement cognitif normal. Cette explication parait cependant insuffisante étant donné d'une part la reproductibilité chez plusieurs patients des résultats rapportés et d'autre part le fait qu'aucun d'entre eux ne présentait de difficulté de compréhension ou de production de la parole.

Une possibilité plus convaincante est que l'EEG de scalp n'accède qu'à une partie des composantes générées dans le cortex auditif, notamment du fait que les activités avant 100 ms sont de polarités variées en montage monopolaire. On peut donc s'attendre à ce que la résultante de ces activations, et donc de leurs modulations par les informations visuelles, aient une amplitude assez faible sur le scalp et n'émergent pas du bruit. De la même façon les réponses visuelles dans le cortex auditif, qui présentaient souvent le même profil spatial que les réponses auditives transitoires avant 100 ms présentaient des polarités variées qui pourraient expliquer qu'elles soient invisibles en EEG de scalp. Étant donné que cette réponse visuelle n'est pas visible en EEG de scalp, cela permet d'exclure que la violation de l'additivité observée dans l'expérience précédente corresponde au premier type de violation observée en sEEG.

En revanche, la composante qui apparait à partir de 70 ms sur une large part du planum temporale et du gyrus transverse médian et qui présente un pic d'activation entre 100 et 130 ms selon les patients pourrait correspondre à l'onde N1, bien que le pic de cette dernière avait lieu vers 135 ms en EEG de scalp. En sEEG, la polarité de cette composante en montage monopolaire, positive sur des contacts situés sous le cortex correspond bien à la polarité de l'onde N1, qui est positive au niveau des mastoïdes en EEG (voir aussi Godey, Schwartz, Graaf, Chauvel & Liégeois-Chauvel, 2001 ; Yvert et coll., 2005). Une modulation de cette composante, visible entre 80 et 200 ms chez trois patients, dont au moins deux sur une composante positive en montage monopolaire, pourrait donc fort bien correspondre à l'effet trouvé en EEG de scalp.

Chapitre 11

Étude comportementale de l'effet d'indiçage temporel des stimuli visuels sur le traitement de la parole

11.1 Introduction

Nous avons montré que voir les mouvements de lèvres accompagnant une syllabe auditive permet de la traiter plus rapidement dans une tâche de discrimination et que cet avantage temporel était associé dans les potentiels évoqués à la diminution de l'onde N1 auditive évoquée par la syllabe plosive. Les données sEEG ont montré, d'une part, que les informations visuelles de parole pouvaient activer le cortex auditif avant la présentation de la syllabe auditive et, d'autre part, que cette activation modifiait l'activation du cortex auditif par la syllabe auditive. Ces résultats ont d'abord été interprétés comme un effet de l'intégration des informations phonétiques visuelles données par la configuration des articulateurs faciaux (ouverture de la bouche notamment) aux informations auditives, permettant de faciliter le traitement phonétique de la syllabe auditive. Il existe cependant d'autres explications plausibles. Elles tiennent principalement au fait que, dans les syllabes plosives utilisées, le mouvement des lèvres précède toujours le son. En effet les lèvres doivent préparer l'explosion du /p/. Bien que ce mouvement soit de faible amplitude par rapport à l'ouverture de la bouche qui accompagne le son et qui donne une véritable information phonétique, il est néanmoins clairement perceptible et commence entre 200 et 100 ms avant l'explosion. Ce mouvement précoce peut donner deux types d'informations :

- il informe le locuteur du moment précis auquel se produira le son.
- par le phénomène de co-articulation, il peut informer le locuteur sur la nature phonétique de la voyelle qui suit.

C'est le premier phénomène qui peut mettre en défaut notre interprétation : en effet si le mouvement des lèvres indique au locuteur que la syllabe arrive, il réduit l'incertitude sur le début de ce son et peut permettre de le traiter plus efficacement. L'effet observé au niveau de l'onde N1 auditive pourrait alors refléter cet effet d'indiçage temporel. Ce phénomène pourrait alors être l'équivalent intermodal et temporel de l'indiçage périphérique dans le

domaine spatial.

De nombreuses études ont montré l'existence d'effets attentionnels exogènes intermodaux en dehors du champ de la parole. Ainsi, il a été montré qu'un indice visuel spatial facilite le traitement d'un stimulus auditif présenté subséquemment au même emplacement (Ward, 1994 ; Ward, McDonald & Lin, 2000). L'existence d'un tel effet attentionnel intermodal a cependant longtemps été controversé (Buchtel & Butter, 1988 ; Spence & Driver, 1997) et semble plus difficile à démontrer expérimentalement que celui d'un indice auditif spatial sur le traitement visuel.

Au niveau des potentiels évoqués de scalp, les bénéfices attentionnels d'un indice visuel sur le traitement auditif se manifestent par une négativité accrue (McDonald et coll., 2001), contrairement à ce que nous avions observé dans l'étude EEG. Cependant dans notre cas, il s'agit non pas d'attention spatiale exogène, mais d'un effet d'alerte du stimulus visuel sur le traitement auditif, et les manifestations de ce type d'attention sur les potentiels évoqués pourraient être différents ce ceux de l'attention spatiale exogène intermodale. Contrairement à l'effet d'alerte d'un stimulus auditif accessoire sur la vitesse de traitement visuel, ce phénomène intersensoriel a été très peu étudié. On dispose de quelques données comportementales sur l'amélioration du seuil de perception auditive (Child & Wendt, 1938 ; Howarth & Treisman, 1958) et sur une diminution de temps de détection de stimuli auditifs par un stimulus accessoire visuel, qui suggèrent qu'un effet d'alerte d'un stimulus auditif sur la vitesse de traitement visuel pourrait exister (L. K. Morrell, 1968a ; Posner et coll., 1976 ; I. H. Bernstein et coll., 1973, expérience 2). Mais il n'existe pas à ma connaissance de données en électrophysiologie.

En ce qui concerne la perception de la parole, il semble bien que l'information temporelle (non phonétique) apportée par la vision des articulateurs puisse être utilisée pour faciliter le traitement de l'information auditive. Ainsi Grant et Seitz (2000) ont montré que les informations visuelles permettaient de diminuer le seuil de perception d'une phrase dans le bruit. Ils avancent que cela est dû à la corrélation temporelle existant entre la variation de la surface d'ouverture de la bouche et l'enveloppe du signal auditif. Cependant il se pourrait que les zones de fortes corrélations correspondent aux zones temporelles où le visage donne le plus d'informations, auquel cas l'effet ne serait pas dû à un effet d'indiçage temporel mais à une intégration des informations phonétiques auditives et visuelles.

Schwartz et coll. (2004) ont tenté d'isoler la contribution des indices visuels temporels d'un possible effet des informations visuelles phonétiques sur l'amélioration de l'intelligibilité de la parole. Dans leur expérience, ils utilisaient 10 syllabes différant soit sur leur lieu d'articulation, soit sur leur mode, soit sur leur voyelle (/gy/, /gu/, /dy/, /du/, /ty/, /tu/, /ky/, /ku/, /y/, /u/). Ces 10 syllabes présentent toutes un mouvement articulatoire identique si bien qu'elles sont impossible à distinguer visuellement. La tâche des sujets consistait, à chaque essai, à identifier la syllabe présentée dans le bruit (un bruit de foule), accompagnée ou non des indices visuels. Les résultats montrent que les indices visuels, bien que non discriminants, améliorent l'intelligibilité du voisement dans le bruit, mais pas des autres traits phonétiques (dans leur expérience 3, la même vidéo était artificiellement montée sur les 10 syllabes auditives pour s'assurer que les indices visuels ne fournissent aucune information phonétique pour la réalisation de la tâche). C'est donc que l'information tem-

porelle portée par le mouvement a facilité la détection du pré-voisement, dont la présence ou l'absence détermine la nature voisée ou non voisée de la syllabe. Il s'agit donc d'un pur effet d'indiçage temporel par le mouvement de lèvres. Cette facilitation semble toutefois être spécifique aux indices visuels de parole, puisque lorsque la bouche est remplacée par un rectangle de surface variant proportionnellement à la surface d'ouverture de la bouche, cet effet disparait.

La question se pose alors de savoir quels sont les corrélats neurophysiologiques de cet effet d'indiçage temporel. Se manifestent-ils de la même manière que les interactions audiovisuelles que nous avons mises en évidence en EEG et en sEEG ? Si tel était le cas, les effets observés dans ces expériences pourraient refléter cet effet d'indiçage intermodal et ne pourraient plus être considérés comme un corrélat de l'intégration audiovisuels d'informations phonétiques auditives et visuelles. Une question intéressante est alors de savoir si cet effet d'indiçage est spécifique à la parole ou peut s'observer avec n'importe quel indice temporel visuel.

Nous avons donc voulu explorer par une méthode électrophysiologique les mécanismes à l'œuvre dans cet effet d'indiçage temporel. Notre projet était à l'origine de réaliser une expérience en MEG en utilisant les stimuli de Schwartz et coll. (2004), présentés dans les modalités auditive, visuelle et audiovisuelle et d'utiliser le modèle additif pour mettre en évidence d'éventuels effets d'interaction audiovisuelle associés à cet effet d'indiçage. Les expériences comportementales présentées dans cette thèse étaient destinées à voir comment on peut adapter l'expérience de Schwartz et coll. (2004) à une étude MEG, afin de mettre en évidence à la fois l'effet comportemental de facilitation et des interactions audiovisuelles. L'expérience en MEG n'a pu être réalisée, faute de temps.

11.2 Expérience comportementale 1

L'application du modèle additif en électrophysiologie nécessite un nombre d'essais important avec des stimuli identiques présentés dans trois conditions (auditive, visuelle et audiovisuelle). Or, dans le protocole de Schwartz et coll. (2004), les sujets devaient identifier 12 syllabes assez différentes d'un point de vue acoustique. Il fallait donc limiter le nombre de syllabes différentes présentées aux sujets. Le résultat principal de leur étude étant que l'indiçage visuel temporel facilite la discrimination du voisement, nous avons décidé de n'utiliser qu'une paire de syllabes différant sur leur voisement (par exemple /du/-/tu/), la tâche étant de simplement discriminer ces deux syllabes. Ainsi, le processus de discrimination sur lequel influe la modalité visuelle reste présent et devrait engager à peu près les mêmes processus sensoriels dans un protocole plus simple et adapté à la MEG.

Pour optimiser le temps d'expérience et réduire les problèmes liés à la réponse motrice, l'idéal est d'utiliser une des deux syllabes comme stimulus non-cible fréquent et l'autre comme stimulus cible rare. Nous avions donc besoin de savoir si l'influence des informations visuelles sur la discrimination s'exerce sur l'un, l'autre ou les deux types des syllabes afin de choisir quelles seraient la syllabe cible et la syllabe non-cible.

Un autre problème de la MEG/EEG est que le rapport signal/bruit des réponses cérébrales doit être le plus grand possible. On doit donc éviter de présenter les stimuli dans le bruit car celui-ci risque de rajouter un bruit neuronal à l'activité MEG de fond dont on tente de se débarrasser en moyennant les essais individuels. Or, à supposer que l'on observe des résultats analogues à ceux de Schwartz et coll. (2004) au même niveau de bruit, il n'est pas garanti qu'ils seraient toujours observés sans bruit car la performance dans ce cas atteint un plafond, d'autant qu'une tâche de discrimination entre deux syllabes est plus facile que la tâche d'identification parmi 12 syllabes. Nous avons donc testé 3 conditions de bruit (pas de bruit, un niveau de bruit de équivalent à celui utilisé dans le protocole original et un niveau intermédiaire) et nous avons mesuré à la fois les performances dans la tâche de discrimination et les TR de discrimination, car l'effet de facilitation était plus susceptible de s'exprimer sur les TR dans les conditions où le bruit était plus faible.

De plus nous voulions savoir si les effets éventuellement mis en évidence étaient spécifiques aux mouvements des lèvres ou s'ils pouvaient exister si les lèvres étaient remplacées par le mouvement d'un rectangle donnant les même informations temporelles.

On a donc une expérience manipulant 4 facteurs : le voisement, le niveau de bruit, la modalité et la nature de l'information visuelle (lèvres ou rectangle). Nous avons émis l'hypothèse que l'on devrait observer un taux d'erreurs moins important dans la condition audiovisuelle que dans la condition auditive seule, mais seulement lorsque les informations temporelles étaient données par les lèvres, et non par les rectangles. Cette configuration d'effets devrait être observé au moins dans la condition la plus bruité. En ce qui concerne les temps de discrimination, ils devraient être plus courts dans la condition audiovisuelle que dans la condition auditive, et cet effet devrait être plus important pour la bouche que pour le rectangle.

Ces deux effets devraient interagir avec le niveau de bruit puisqu'il est connu que l'influence des informations visuelles est d'autant plus important que le rapport signal sur bruit est faible. On espère cependant qu'ils seront toujours présents dans la modalité sans bruit, contrairement au taux d'erreurs.

11.2.1 Méthodes

Sujets

Onze sujets droitiers (dont 8 de sexe féminin), d'une moyenne d'âge de 27,7 ans (écart-type : 4 ans) ont passé cette expérience. Aucun ne souffrait de troubles auditifs ou visuels.

Stimuli

Les vidéos utilisées dans cette expérience ont été adaptées de celles utilisées par Schwartz et coll. (2004). Les syllabes étaient prononcées par un homme de langue maternelle française aux lèvres peintes en bleu (pour une raison indépendante de notre volonté), dont seule la partie inférieure du visage était visible. La taille de la bouche correspondait à 2,2° d'angle visuel. Une séquence visuelle commençait par une l'image fixe d'une bouche au repos et se terminait par la même image fixe. Les mouvements labiaux présentés étaient

identiques, quelle que soit l'identité de la syllabe auditive et consistaient en une suite de 20 images d'une durée de 33 millisecondes chacune. Dans la condition "rectangle", le visage était remplacé par un rectangle rouge dont la surface variait de façon inversement proportionnelle à l'aire d'ouverture de la bouche. La largeur de ce rectangle était identique à celle de la bouche, sa hauteur minimale était de 0,12° et sa hauteur maximale de 0,52° d'angle visuel.

Les stimuli visuels était présentés dans les mêmes conditions que notre première étude en EEG. En prévision de l'étude MEG, dans laquelle on utilise un vidéo projecteur ayant une fréquence de rafraichissement, non modifiable, de 60 Hz, nous avons dû présenter chaque image à une cadence de 30 images par seconde, alors qu'elles avaient été enregistrées à 25 images par seconde. La vitesse était donc accélérée d'un facteur 6/5 par rapport aux mouvements naturels présentés dans l'étude originale. En conséquence, les syllabes auditives ont dû être compressées d'un facteur équivalent afin de conserver la synchronisation des indices auditifs et visuels, tout en conservant le spectre fréquentiel du signal acoustique original. Cette compression temporelle a été réalisée grâce au logiciel Soundforge. Les syllabes résultant de cette transformation semblaient tout aussi naturelles que les syllabes originales, aussi bien sur le plan visuel qu'auditif.

Nous avons utilisé 4 couples de syllabes (/gu/-/ku/, /gy/-/ky/, /du/-/tu/, /dy/-/ty/) qui étaient toujours présentés dans des blocs expérimentaux différents, dans le but de conserver pour l'expérience MEG uniquement le couple de syllabes montrant l'effet comportemental le plus net. Chacune des 8 syllabes présentait une structure audiovisuelle différente, mais pour chaque paire de syllabe, le son de la syllabe voisée commençait toujours systématiquement plus tôt par rapport au début du mouvement des lèvres que celui de la syllabe non voisée, en raison du pré-voisement. Le schéma temporel des stimulations est illustré pour les syllabes /ku/ et /gu/ dans la figure 11.1. L'intensité de chacune des syllabes était ajusté de façon à ce que la puissance acoustique moyenne de la partie stationnaire du signal, correspondant à la voyelle, soit la même.

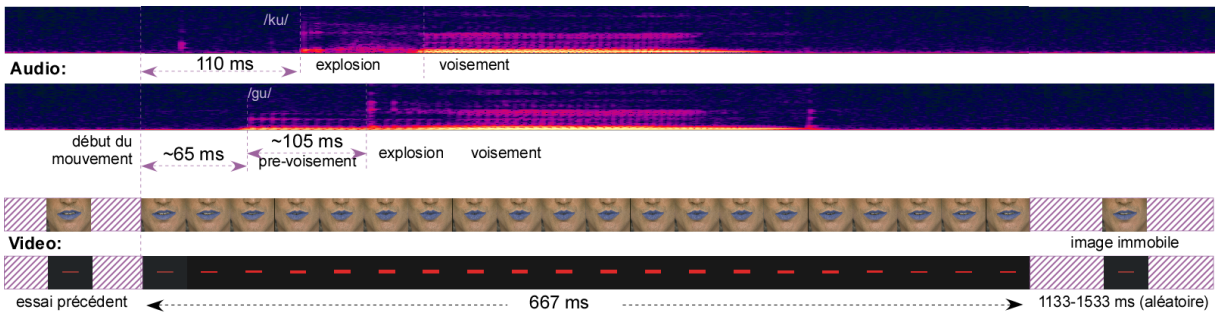

FIG. 11.1 – Structure temporelle des syllabes /ku/ et /gu/. Audio : variation temporelle du spectre fréquentiel entre 0 et 10000 Hz. Pour tous les couples de syllabes, le pré-voisement de la syllabe voisée commençait toujours avant l'explosion de la consonne de la syllabe non voisée. Les délais temporels indiqués sont les valeurs moyennes sur l'ensemble de 4 syllabes voisées et des 4 syllabes non voisées.

Dans les deux conditions bruitées, un bruit de foule continu était présenté pendant tout le bloc de stimulation. Le rapport signal (syllabe) sur bruit (foule) était calculé comme le rapport de la puissance moyenne pendant la partie stationnaire, correspondant à la voyelle,

sur la puissance moyenne du bruit. Dans la condition la plus bruitée, le rapport signal sur bruit était de -9 dB, dans la condition intermédiaire de 0 dB et dans la conditions sans bruit aucun bruit n'était présenté. Contrairement à nos premières études comportementales, les sons ont été présentés dans un casque à écouteurs afin d'imiter les conditions de stimulation dans la MEG.

Procédure

Dans tous les blocs expérimentaux, un essai commence avec la présentation d'un visage (ou un rectangle) au repos. Avec un intervalle interstimulus variant aléatoirement entre 1800 et 2200 ms, il entend une syllabe parmi deux syllabes possibles (variable "Voisement" : voisée ou non voisée). Cette syllabe est accompagnée ou non de l'articulation visuelle (variable "Modalité" : auditif et audiovisuel). Donc, en condition auditive seule, le sujet voit un visage (ou un rectangle) immobile.

La tâche du sujet consiste à cliquer le plus rapidement possible sur l'un des 2 boutons de la souris, chacun des boutons correspondant à une des 2 syllabes, ce qui revient à discriminer le voisement, sans que cela soit explicitement dit au sujet. Les sujets n'étaient pas informés que les mouvements labiaux ne donnaient aucune information sur l'identité de la syllabe et il leur était seulement demandé de fixer la bouche pendant toute l'expérience, sans préciser s'il fallait ou non se servir des indices visuels. Les associations bouton/voisement étaient constantes pour tous les couples de syllabes pour un sujet donné, mais contrebalancées entre les sujets.

Chaque bloc expérimental contenait 40 stimuli (10 syllabes de chacune des conditions suivantes : voisée auditive, voisée audiovisuelle, non voisée auditive et non voisée audiovisuelle)

En plus de ces 2 variables intrabloc, on manipulait 3 variables interbloc :

- le niveau de Bruit (-9dB, 0dB, sans bruit).
- la Nature de l'information visuelle (visage ou rectangle).
- les couples de syllabes voisée/non voisée (/gu/-/ku/, /gy/-/ky/, /du/-/tu/ ou /dy/-/ty/). Cette dernière variable n'entrait pas dans l'analyse statistique et les performances et TR étaient moyennés à travers les 4 couples.

Chaque sujet était donc soumis à 24 blocs de stimuli, dont l'ordre était aléatoire et différent pour chaque sujet.

Analyses

Deux ANOVA avec, pour facteurs, le niveau de bruit, le voisement, la modalité et la nature des informations visuelles on été réalisées, l'une sur le pourcentage d'erreurs moyen sur l'ensemble des 4 couples de syllabes et l'autres sur le TR moyen dans les essais justes. Les degrés de libertés ont été corrigés selon la méthode de Greenhouse-Geisser pour prendre en compte la non homogénéité éventuelle des variances. Lorsqu'une interaction était significative, des ANOVA était réalisées sur chacune des modalités de l'un des facteurs impliqués dans l'interaction, pour tester l'effet des autres facteurs impliqués, et ceci jusqu'à aboutir à des ANOVA à un seul facteur, où jusqu'à ce qu'aucune interaction ne soit significative.

11.2.2 Résultats

Performances

La figure 11.2 montre les performances de sujets en fonction des 4 facteurs expérimentaux. Comme on aurait pu le prédire, le pourcentage d'erreurs augmente significativement avec le niveau de bruit ($p < 0,001$). On observe un effet significatif du voisement sur le pourcentage d'erreur ($p < 0,04$), les sujets se trompant plus souvent sur les non-voisées que sur les voisées. Enfin, on observe une interaction significative entre les facteurs voisement et modalité ($p < 0,04$) indiquant que si les informations visuelles améliorent les performances pour les syllabes non-voisées, elles les dégradent pour les voisées. Mais si on teste maintenant l'effet de la modalité séparément pour les syllabes voisées et non voisées, il n'est significatif pour aucun des 2 types de syllabe. Aucun autre effet ou interaction n'est significatif.

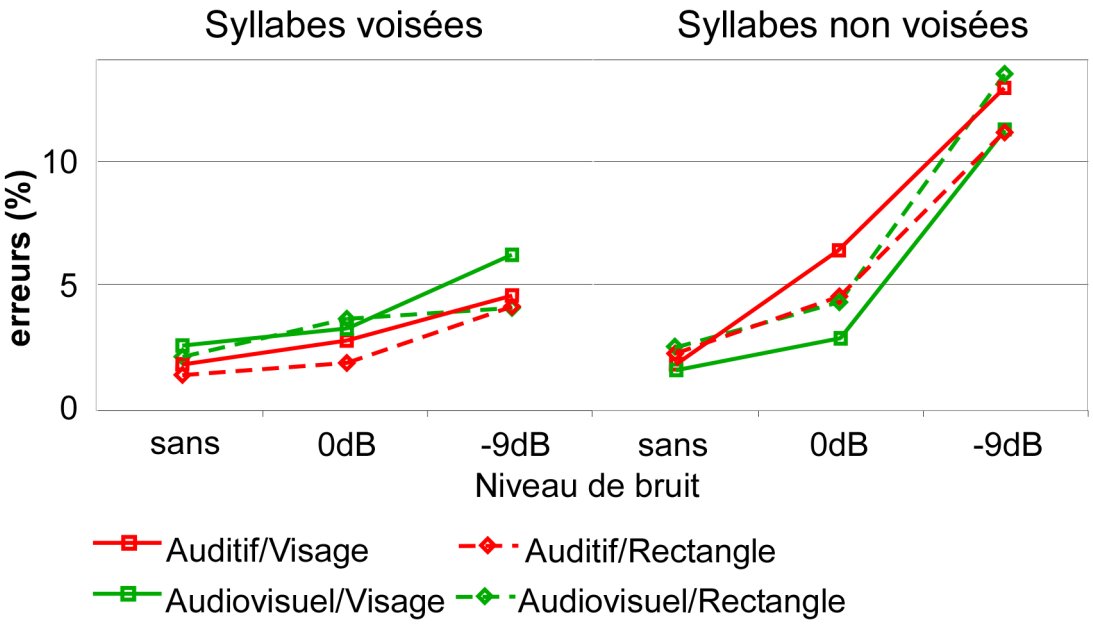

FIG. 11.2 – Pourcentage d'erreur dans la tâche de discrimination du voisement, en fonction du voisement, du niveau bruit, de la modalité de présentation et de la nature des informations visuelles.

Temps de réaction

La figure 11.3 page suivante présente les TR moyens pour les 24 conditions expérimentales testées. Comme nous l'avions prédit, on trouve un effet très significatif du bruit ($p < 0,0001$) sur le temps de traitement des syllabes qui augmente avec le niveau de bruit. L'effet du voisement est également présent ($p = 0,008$), les voisées donnant lieu à des temps de réaction plus courts, comme c'était prédictible étant donné que le début du son commençait plus tôt par rapport à l'instant où est mesuré le TR (le début du mouvement des lèvres), dans ces syllabes.

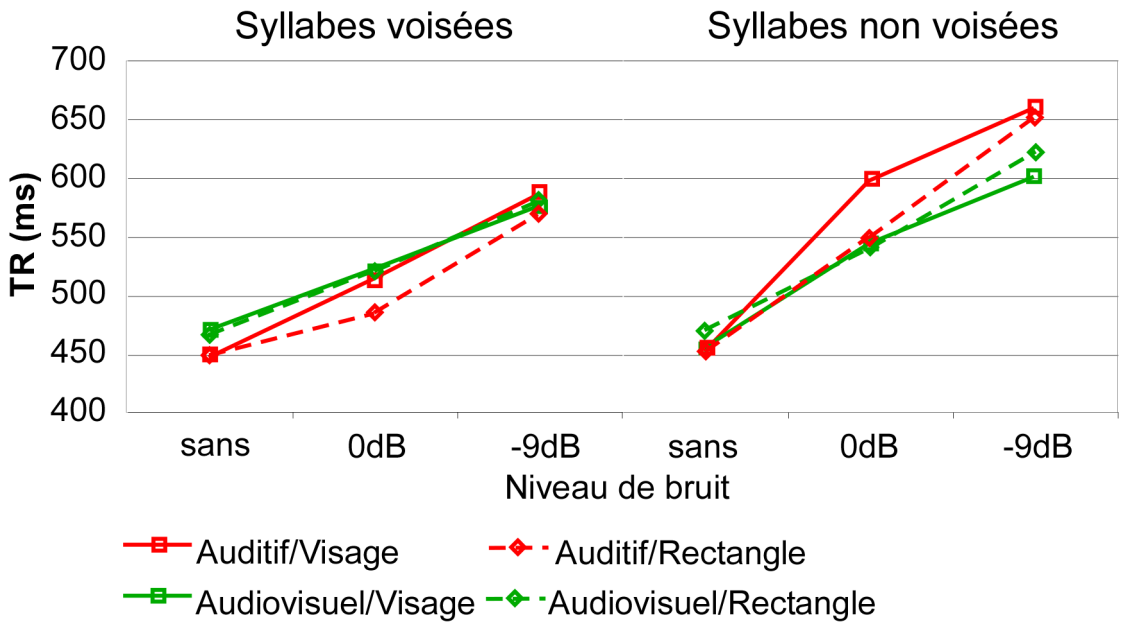

FIG. 11.3 – Temps de réaction dans la tâche de discrimination du voisement, en fonction du voisement, du niveau bruit, de la modalité de présentation et de la nature des informations visuelles.

Contrairement à nos hypothèse, l'effet principal de la modalité de présentation n'est pas significatif. Mais il semble, si l'on examine la figure 11.3, que cela soit dû au fait que l'effet de la modalité était différent selon le type de syllabe et le niveau de bruit. De fait, la triple interaction Voisement × Bruit × Modalité était marginalement significative ($p < 0,06$)

La figure 11.4 page ci-contre décrit cette interaction. Dans les 2 conditions bruitées, l'interaction entre les variables Modalité et Voisement est significative (-9dB : $p < 0,02$; 0dB : $p < 0,0004$). Dans la condition la plus bruitée, cette interaction indique un effet bénéfique des informations visuelles temporelles sur le temps de traitement, présent pour les syllabes non voisées ($p < 0,005$), mais pas pour les voisées. Dans la condition de bruit intermédiaire, l'interaction peut se décrire comme un effet opposé de la modalité sur les syllabes voisées et non voisées : on observe une diminution du TR avec les informations visuelles pour les syllabes non voisées ($p < 0,005$) et une augmentation du TR pour les syllabes voisées ($p < 0,06$).

Dans la condition sans bruit, l'interaction entre les facteurs Voisement et Modalité n'est pas significative, mais on observe un effet principal de la modalité se traduisant par une augmentation du TR dans la condition audiovisuelle par rapport à la condition auditive ($p < 0,04$). On n'observe en revanche pas d'effet significatif du voisement dans cette condition sans bruit.

Concernant l'interaction entre la présence d'informations visuelles temporelles et la nature de ces informations, nous avions prédit, sur la base des résultats antérieurs de Schwartz et coll., que la diminution de TR devrait être plus forte pour le visage que pour le rectangle, et que cette relation pouvait évoluer en fonction du niveau de bruit. La triple interaction Bruit × Modalité × Nature de informations était marginalement significative ($p < 0,07$).

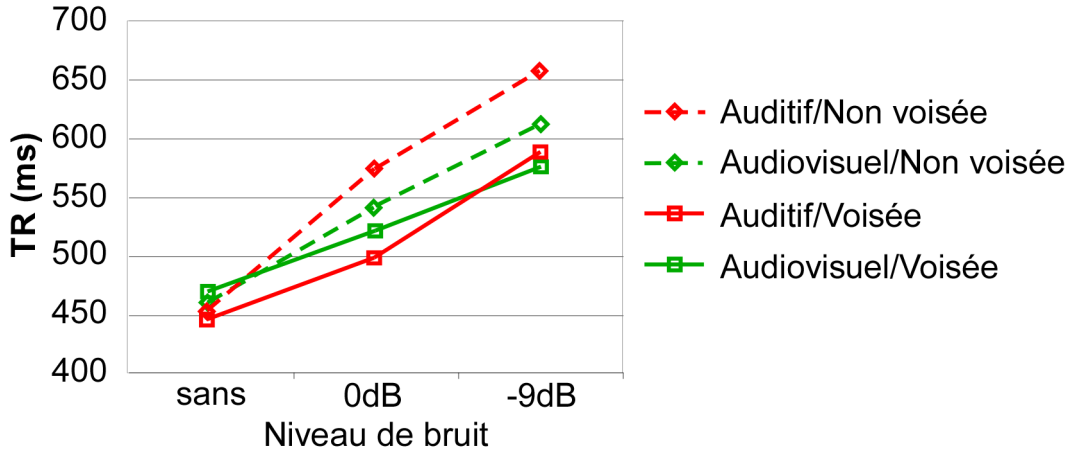

FIG. 11.4 – Temps de réaction dans la tâche de discrimination du voisement, en fonction du voisement, du niveau bruit et de la modalité de présentation.

La représentation graphique de cette interaction (figure 11.5) suggère en effet que le schéma d'interaction entre la présence et la nature de informations visuelles variait en fonction du niveau de bruit, mais d'un manière différente de celle à laquelle on aurait pu s'attendre.

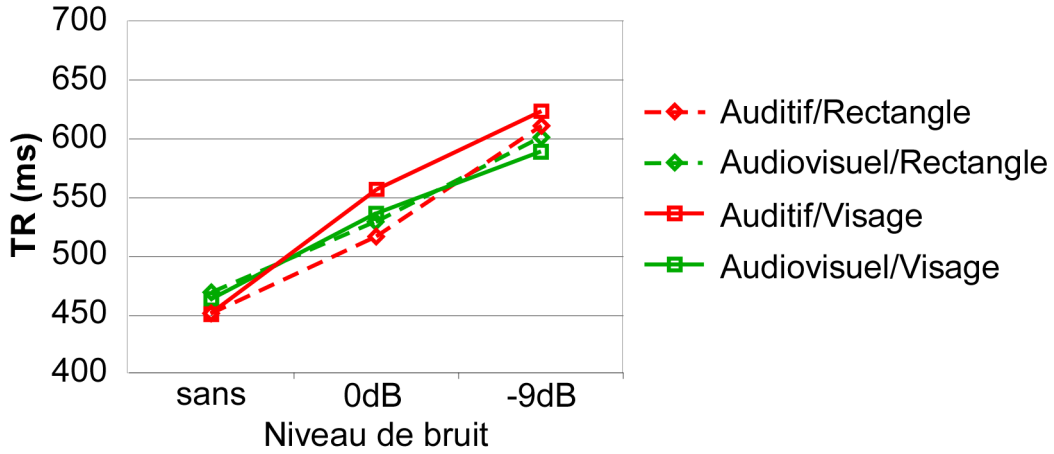

FIG. 11.5 – Temps de réaction dans la tâche de discrimination du voisement, en fonction du niveau bruit, de la modalité et de la nature des informations visuelles.

Dans les deux conditions de bruit, on trouve une interaction significative entre les facteurs Modalité et Nature des informations (-9dB : $p < 0,05$; 0dB : $p < 0,002$). Dans les deux cas, cette interaction va dans le sens prédit puisque le TR en audiovisuel est significativement inférieur au TR auditif dans le cas du visage (-9dB : $p < 0,02$; 0dB : $p < 0,03$) mais pas dans le cas du rectangle. Cependant, un aspect troublant de l'interaction est que dans la condition 0dB, l'effet semble être dû plus à une différence entre les temps de traitement des syllabes auditives (c'est-à-dire accompagnées par un rectangle ou un visage immobile $p < 0,02$) que par une facilitation plus forte du visage que du rectangle en condition audiovisuelle. Il est probable que ces différences entre conditions auditives seules

aient contribué en grande partie à la présence d'une interaction entre les facteurs Modalité et Nature des informations pour ces deux conditions de bruit.

Dans la condition sans bruit, on n'observait pas d'interaction entre les facteurs Modalité et Nature des informations, ni d'effet principal de la nature des informations visuelles, mais un effet principal de la modalité, sous la forme d'un cout de la condition audiovisuel ($p < 0,04$), déjà décrit plus haut.

Aucune autre interaction ou effet principal que ceux décrits n'étaient significatif. Jamais nous n'avons observé d'interaction entre les facteurs Voisement et Nature de l'information visuelle.

11.2.3 Discussion

L'analyse des performances n'indique que des effets faibles et peu significatifs de la modalité de présentation. Cette quasi-absence d'effet de la modalité pourrait s'expliquer par la variabilité intersujet importante du taux d'erreur.

En tout état de cause, le pourcentage d'erreur moyen observé était plus faible que celui trouvé par Schwartz et coll. (2004) avec les même stimuli et pourrait refléter la différence de tâche demandée au sujet. Discriminer entre deux syllabes est en effet plus facile qu'identifier une syllabe parmi 12, à niveau de bruit équivalent, et la simplicité de notre tâche pourrait être une seconde raison pour laquelle on n'a pas observé pas de facilitation de la performance avec l'apport d'information visuelle temporelle.

L'aide apportée par les informations visuelles temporelles a en revanche été répliquée sur les TR, mais uniquement pour les syllabes non voisées dans les deux conditions de bruit. De plus, dans ces deux conditions, on trouvait une interaction entre la présence d'informations visuelles temporelles et la nature de ces informations, mais cet effet semblait autant venir d'une diminution du TR pour le visage en mouvement par rapport au visage immobile que d'une augmentation du TR pour le rectangle en mouvement par rapport au rectangle immobile.

Par ailleurs, l'effet des informations visuelles temporelles change selon le niveau de bruit. De manière générale, il semble rester vrai que plus le niveau de bruit est important, plus les informations visuelles sont utiles, mais ces effets s'expriment différemment pour les syllabes voisées et non voisées. Pour les syllabes non voisées, en augmentant le niveau de bruit, on passe d'une situation où les indices visuels n'aident pas à une situation où ils diminuent le TR. À l'inverse, pour les syllabes voisées, en augmentant le niveau de bruit, on passe d'une situation où les indices visuels augmentent le TR à une situation où le TR pour les syllabes auditives et audiovisuelles est équivalent. Peut-être en augmentant encore le niveau de bruit, observerait-on une amélioration du TR pour les syllabes voisées également.

Cette triple interaction peut avoir plusieurs explications : d'une part les syllabes voisées ont une puissance spectrale totale plus importante que celle des syllabes non-voisées (la zone stationnaire du signal dure plus longtemps), ce qui peut expliquer pourquoi elles sont plus facilement détectables dans le bruit, comme on peut le constater au niveau des

performances. De ce fait il est possible que leur traitement bénéficie moins de la présence des indices visuels. D'autre part, le délai séparant le début des indices visuels et auditifs est différent pour les voisées et les non voisées. Or plusieurs études ont montré des effets d'intégration multisensorielle différents selon le délai séparant les informations des deux modalités (Ghazanfar, Maier, Hoffman & Logothetis, 2005 ; Lakatos, Chen, O'Connell, Mills & Schroeder, 2007).

Enfin, la triple interaction entre voisement, bruit et modalité se traduit également par une convergence des TR des différentes combinaisons voisement/modalité dans la condition sans bruit : cet effet pourrait s'expliquer soit par un effet plancher, soit une différence de stratégie. Selon la seconde explication, les mécanismes de discrimination du voisement seraient des plus efficaces dans la condition sans bruit et, par conséquent, le traitement des voisées et non voisées prendrait des temps équivalents tout en laissant peu l'occasion aux mécanismes d'intégration de se manifester. Dans les conditions bruitées, au contraire, la discrimination du voisement reposerait beaucoup plus sur la détection de la présence ou de l'absence d'un prévoisement, qui pourrait être plus sensible à la présence d'informations visuelles temporelles.

Deux aspects des données jettent toutefois le doute sur l'interprétation des résultats obtenus. Il s'agit d'une part du fait que dans certaines conditions (dans la condition sans bruit et, pour les syllabes voisées, dans la condition de bruit intermédiaire), on observait une augmentation des TR dans la condition audiovisuelle par rapport à la condition auditive, et d'autre part, de la différence de TR observée entre les deux conditions auditives seules.

Ces effets suggèrent que les conditions auditives choisies n'étaient pas de bons contrôles, dans la mesure où le type d'informations visuelles présentes à l'écran semble influer sur le temps de traitement de la syllabe bien qu'il ne donne aucune information sur le voisement, pas même une information temporelle.

Cette différence entre les conditions rectangle et visage pourrait être due à des différences de stratégie : en effet la variable Nature des informations est une variable interbloc et il est tout à fait envisageable que les sujets aient traité différemment les stimuli (A et AV) selon que le contexte était celui d'un visage ou celui d'un rectangle. Un visage qui prononce une syllabe une fois en remuant les lèvres, une fois sans les bouger n'a pas le même sens que des syllabes accompagnées ou non du mouvement d'un rectangle. L'interaction entre la présence d'informations visuelles et leur nature est donc difficile à interpréter du fait de la présence possible d'un effet de bloc. Afin de mieux étudier l'effet de la nature des informations visuelles sur l'effet de la modalité et de confirmer la présence d'un cout de l'ajout d'informations temporelles visuelles, nous avons mené une nouvelle expérience comportementale.

11.3 Expérience comportementale 2

Dans l'expérience précédente, la présence d'un cout audiovisuel et d'un effet de la nature des informations statiques sur le temps de traitement des syllabes "auditives" nous a incité à la prudence quant à nos conclusions.

En effet, dans la mesure où les conditions auditives montraient des différences significatives entre les conditions visage et rectangle, on peut mettre en doute l'interprétation des effets en termes de bénéfices ou de cout des informations temporelles visuelles. Il se pourrait en effet que la simple présence d'un visage au repos, même immobile, accélère la discrimination du voisement. Il nous fallait donc trouver un meilleur contrôle auditif seul. Nous avons ajouté une condition auditive seule dans laquelle l'écran était totalement vide pendant la présentation de la syllabe. Et, pour éviter les effets de blocs, nous avons présenté les 5 conditions dans un même bloc expérimental : la condition auditive seule, les deux conditions audiovisuelles statiques dans laquelle seule une bouche ou un rectangle au repos était présenté pendant la stimulation auditive (conditions auditives de l'expérience précédente) et les deux conditions audiovisuelles dynamiques dans lesquelles le mouvement du visage ou du rectangle donnaient une information temporelle sur la syllabe.

11.3.1 Méthodes

Sujets

Neuf sujets droitiers (dont 5 de sexe féminin) âgés en moyenne de 27,5 ans (écart-type : 4 ans) ont passé cette expérience. Huit de ces sujets avait passé l'expérience 1 deux mois auparavant.

Stimuli

Les stimuli utilisés étaient identiques à ceux de l'expérience 1, excepté que nous n'avons employé qu'un seul couple couple de syllabe (les syllabes /ku/ et /gu/), afin d'éliminer une source de variabilité des TR. Ce couple a été choisi pour la ressemblance des effets sur les TR présentés par ce seul couple avec les effets estimés sur la moyenne des 4 couples de syllabes dans l'expérience précédente. Pour des raisons qui seront exposées ci-dessous, le rectangle rouge était présenté sur un fond gris au lieu d'un fond noir. Dans la condition auditive seule, la syllabe auditive était présentée avec un fond visuel gris uni.

Procédure

Le sujet devait donc réaliser la tâche de discrimination des syllabes voisées et non voisées dans 5 conditions visuelles mélangées aléatoirement : écran gris (auditif seul), visage statique, rectangle statique, visage dynamique, rectangle dynamique.

Contrairement à la première expérience, le changement de condition au sein d'un bloc nécessitait l'apparition et la disparition brusque des stimuli visuels (passage d'un essai visage à un essai rectangle ou auditif seul, par exemple). Pour éviter que le début d'un essai donne plus d'informations temporelles dans une condition que dans une autre, les essais au sein d'un bloc étaient séparés par un écran noir pendant 150 ms. Ainsi, la prédictibilité du stimulus auditif était identique pour les cinq conditions : au moment ou l'écran noir disparait, apparait soit un visage, soit un rectangle sur fond gris, soit un fond gris seul. Cependant, on ne voulait pas que cette information temporelle à elle seule aide le sujet à détecter le voisement. Quel intérêt y aurait-il alors à exploiter les informations visuelles dynamiques ? Afin de limiter la prédictibilité temporelle de la syllabe et de favoriser la

capacité du mouvement visuel (que ce soit celui du rectangle ou du visage) à fournir de l'information temporelle, nous avons introduit une période aléatoire (variant entre 300 et 750 ms) entre l'apparition de l'image immobile et le début de la syllabe auditive et/ou du mouvement articulatoire. Un essai se terminait par un période aléatoire et l'intervalle interstimulus moyen était de 2000 ms. La structure temporelle d'un essai est illustrée dans la difugre 11.6.

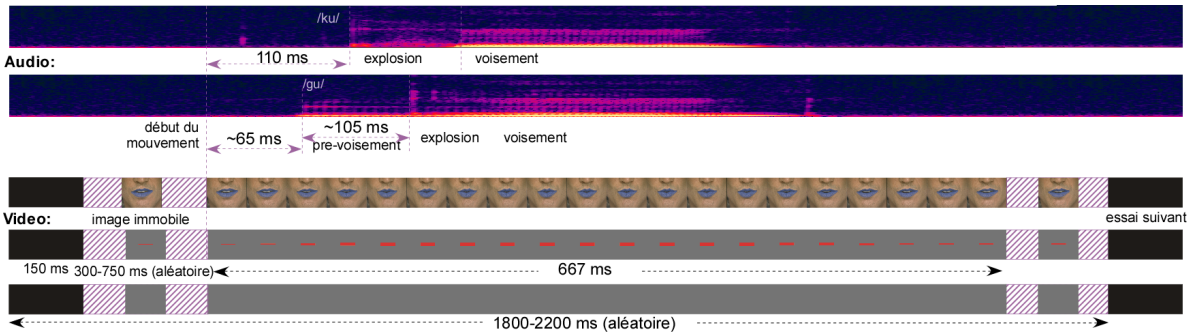

FIG. 11.6 – Structure temporelle des syllabes /ku/ et /gu/. Audio : variation temporelle du spectre fréquentiel entre 0 et 10000 Hz. Pour tous les couples de syllabes, le pré-voisement de la syllabe voisée commençait toujours avant l'explosion de la consonne de la syllabe non voisée. Vidéo : stimuli visuels des conditions audiovisuelles dynamiques (visage ou rectangle) et de la condition auditive seule. dans tous les cas, un écran noir précédait la présentation du visage, du rectangle ou de l'écran gris. Les délais temporels indiqués sont les valeurs moyennes sur l'ensemble de 4 syllabes voisées et des 4 syllabes non voisées.

Afin d'étudier plus finement la variation des effets avec le niveau de bruit, nous avons utilisé 5 niveaux de bruit : sans bruit, 0dB, -4,5dB, -9dB et -13,5dB. Le niveau de bruit le plus fort devrait permettre d'observer une facilitation de TR pour les syllabes voisées. Les différents niveaux de bruit étaient présentés dans des blocs différents.

Chaque sujet passait 20 blocs de stimulation, soit 4 blocs de chaque niveau de bruit. Un bloc comprenait 5 syllabes voisées (/gu/) et 5 syllabes non voisées (/ku/) dans chacune des 5 conditions de présentation, pour un total de 50 syllabes.

Analyses

Pour cette expérience nous n'avons analysé que les TR, dans les essais où les sujets n'avaient pas commis d'erreur. On a effectué deux types d'analyse sur les temps de réaction.

- on a analysé les données des 4 conditions déjà présentes dans l'expérience 1 avec la même ANOVA à 4 facteurs : Bruit × Voisement × Modalité × Nature des informations visuelles, sans prendre en compte les essais auditifs seuls. Cela permet d'évaluer l'effet de la présentation aléatoire par rapport à la présentation par bloc des rectangles et des visages. Notons tout de même quelques différences supplémentaires entre les 2 protocoles : utilisation d'un seul couple de syllabes, présence de 5 niveaux de bruit et sujets plus familiers avec les stimuli et la tâche (les mêmes sujets ont en effet en majorité participé aux deux expérience).
- Afin d'évaluer l'existence de bénéfices et éventuellement de couts dans les conditions audiovisuelles dynamiques et statiques, on a testé la significativité de la différence

entre chacune des 4 combinaisons Modalité × Nature des informations et la condition auditive seule, ainsi que l'interaction de cet effet avec les variables bruit et voisement. On a donc réalisé, pour chacune des conditions visage dynamique, visage statique, rectangle dynamique et rectangle statique, une ANOVA Présence d'informations visuelles (statique ou dynamique) × Bruit × Voisement.

Tous les tests ont été corrigés pour la non sphéricité des données par la méthode de Greenhouse-Geisser.

11.3.2 Résultats

ANOVA Bruit × Voisement × Modalité × Nature

On retrouve l'effet attendu du bruit sur les temps de réaction ($p < 0,0001$), ainsi que l'effet du voisement, les voisées donnant lieu à des TR plus rapides que les non voisées (p=0,0007). Ces deux effets et leur interaction sont décrits dans la figure 11.7.

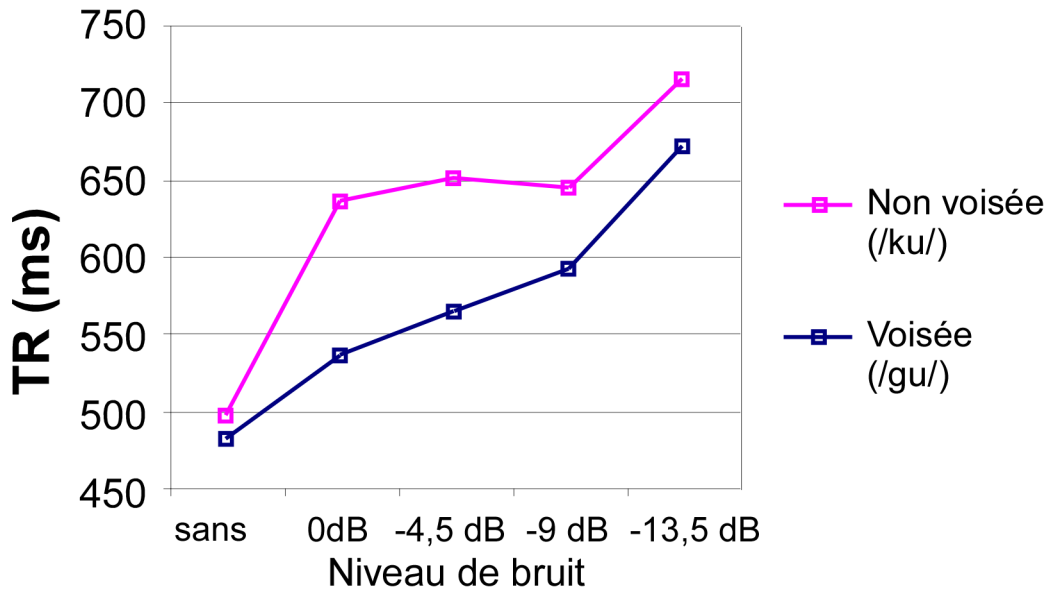

FIG. 11.7 – Temps de réaction dans la tâche de discrimination du voisement, en fonction du voisement et du niveau bruit.

L'interaction entre ces deux facteurs est significative ($p < 0,0001$) et semble s'expliquer par le fait que la différence entre voisées n'existe que pour les conditions bruitées (0dB : $p < 0,0001$; 4,5dB : $p = 0,0004$; 9dB : $p < 0,02$; 13,5 dB : $p < 0,02$).

Contrairement à l'expérience 1, le facteur voisement n'interagissait avec aucun autre facteur de l'analyse.

Par contre, comme dans l'expérience 1, la triple interaction Modalité × Nature × Bruit était marginalement significative ($p < 0,08$). Cette interaction est décrite dans la figure 11.8 page ci-contre.

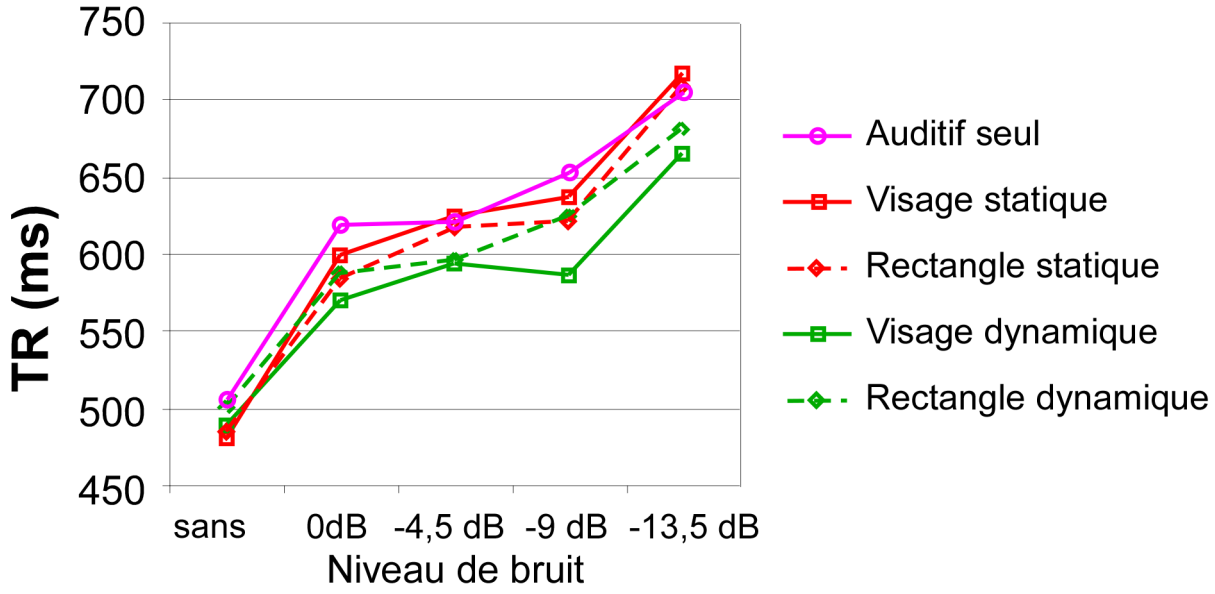

FIG. 11.8 – Temps de réaction dans la tâche de discrimination du voisement, en fonction du niveau bruit, de la modalité et de la nature des informations visuelles.

Nous avons testé l'interaction Modalité × Nature des informations dans chacune des conditions de bruit. Dans la condition sans bruit, l'interaction n'est pas significative. L'effet de la modalité est marginalement significatif et s'exprime par une augmentation du TR pour les conditions audiovisuelles dynamiques par rapport aux conditions audiovisuelles statiques ($p < 0,08$).

Dans la condition 0dB, l'interaction significative (p=0,01) se manifeste autant par un cout significatif du visage immobile par rapport au rectangle immobile ($p < 0,02$) que par un gain du visage en mouvement par rapport au rectangle en mouvement ($p = 0,05$).

Dans la condition 4,5dB, l'interaction n'est pas significative et on trouve un effet principal de la modalité de présentation qui s'exprime par une diminution des TR avec les informations visuelles dynamiques ($p < 0,02$).

Dans la condition 9dB, l'interaction est significative ($p < 0,01$) et se traduit par un cout marginalement significatif du visage immobile ($p < 0,07$) et un gain très significatif du visage mobile (p=0,004) par rapport au rectangle.

Enfin dans la condition 13,5 dB, l'interaction était marginalement significative ($p < 0,07$) et s'expliquait par un avantage marginalement significatif du visage dynamique par rapport au rectangle dynamique ($p < 0,04$), le cout pour le visage immobile par rapport au rectangle immobile n'étant pas significatif.

Test des couts et bénéfices

Dans chacune des 4 conditions visage dynamique, visage statique, rectangle dynamique et rectangle statique, on a retrouvé les effets significatifs du bruit, du voisement ainsi que leur interaction, déjà décrits. De même que dans l'analyse précédente, le voisement n'interagissait jamais avec le facteur Présence d'information visuelle, dans aucune des 4

conditions. La figure 11.9 présente donc la différence de TR entre les 4 conditions audiovisuelles et la condition auditive seule en fonction du niveau de bruit, moyennée sur le type de voisement.

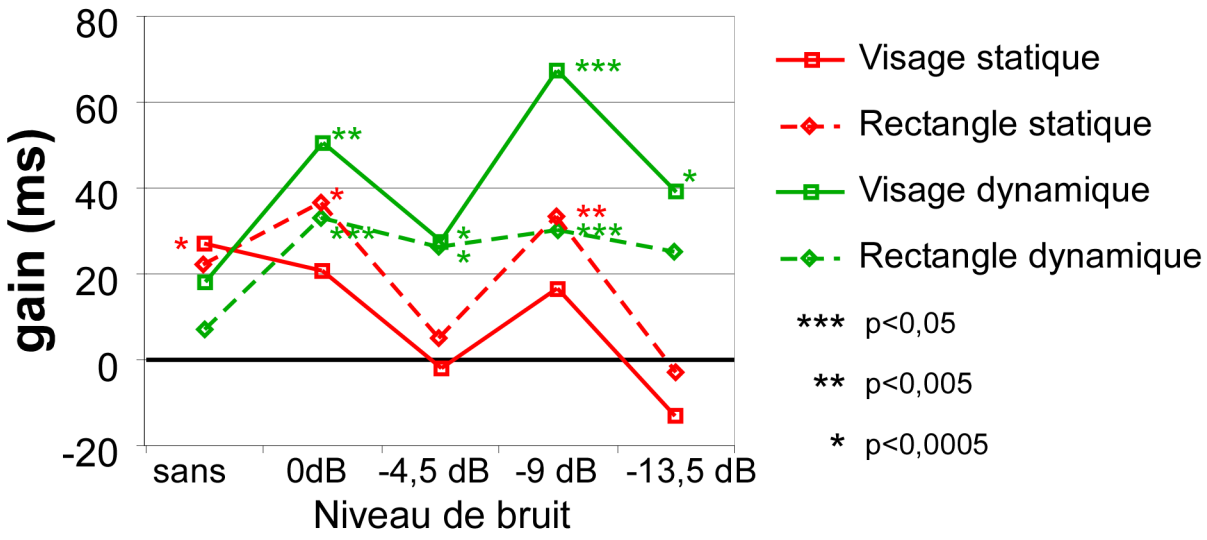

FIG. 11.9 – Bénéfices et cout du TR par rapport à la condition auditive seule dans la tâche de discrimination du voisement, en fonction du niveau bruit, de la modalité et de la nature des informations visuelles. Les étoiles indiquent les conditions dans lesquelles l'effet de la présence d'un stimulus visuel (statique ou dynamique) est significatif.

On peut constater que, dans presque toutes les conditions, cette différence prenait l'aspect d'un bénéfice. Pour la condition visage dynamique, même si l'effet de la présence d'informations visuelles était très significatif ($p < 0,0001$), il interagissait significativement avec le niveau de bruit ($p < 0,04$) : le bénéfice apporté par les informations visuelles était significatif pour toutes les conditions de bruit mais pas dans la conditions sans bruit (0dB : $p < 0,003$; 4,5dB : $p < 0,04$; 9dB : $p < 0,0001$; 13,5dB : $p < 0,008$).

Pour la condition rectangle dynamique, l'interaction entre la présence d'information visuelle et le bruit n'était pas significative et l'effet global de l'information visuelle était très significatif ($p < 0,0001$). Donc excepté dans la condition sans bruit pour le visage, la présence d'informations visuelles temporelles dynamiques diminuait le TR par rapport à une condition auditive seule.

Dans les deux conditions où l'information visuelle consistait en la simple présence d'une image immobile, l'interaction entre la présence d'information visuelle et le niveau de bruit était significative (visage $p < 0,02$; rectangle : $p < 0,008$).
Dans la condition visage statique, le bénéfice était plus ou moins significatif selon le niveau de bruit (sans bruit : $p < 0,04$; 0dB : $p < 0,10$; 9dB : $p < 0,10$).
Dans la condition rectangle statique, on observe exactement le même schéma d'interaction, en plus significatif (sans bruit : $p < 0,06$; 0dB : $p < 0,03$; 9dB : $p < 0,003$).

Ajoutons enfin qu'aucune de ces 20 conditions Modalité × Nature × Bruit ne montre de cout de la condition audiovisuelle (statique ou dynamique) par rapport à la condition auditive seule.

11.3.3 Discussion

En dépit du fait que toutes les conditions audiovisuelles (visage, rectangle, statique, dynamique) étaient mélangées dans cette expérience, on continue à observer d'une part des TR plus rapides pour les conditions audiovisuelles statiques que pour les conditions audiovisuelles dynamiques dans la condition sans bruit et d'autre part, des TR globalement plus rapides lorsque les syllabes sont présentées associées à un rectangle statique que lorsqu'elles sont présentées avec un visage statique.

Ces effets sont toutefois moins significatifs que dans l'expérience 1, peut-être parce que le nombre de sujets est moins élevé et/ou parce qu'ils sont atténués par le mélange des conditions visage et rectangle.

Cependant la condition auditive seule nous a permis de montrer que l'ajout d'informations visuelles, qu'elles soient temporellement informatives (dynamiques) ou (statiques), ne se traduit jamais par un cout en termes de temps de traitement. Si l'on ne considère que les conditions audiovisuelles dynamiques et la condition auditive seule, nous avons donc montré une diminution du temps de réaction dans la discrimination du voisement lorsque les sujets disposent d'informations temporelles pouvant les aider à détecter le prévoisement par rapport à une condition où aucun stimulus visuel n'est présenté. Cette effet existe aussi bien pour des informations temporelles dynamiques fournies par un rectangle que par un visage, mais uniquement lorsque la discrimination est rendue plus difficile par la présence de bruit. Cependant lorsque ces informations temporelles sont fournies par un visage, la diminution du temps de réaction est plus importante que lorsqu'elles sont fournies par un rectangle, au moins dans deux conditions de bruit[1]. Cet effet semble donc être en partie spécifique aux indices visuels de parole et représente un réplication de l'effet mis en évidence par Schwartz et coll. (2004) sur les performances.

Toutefois on observe également une diminution du TR par rapport à la condition auditive seule lorsque l'on ajoute un visage ou un rectangle statique (les anciennes conditions "auditives" de l'expérience 1), au moins dans les conditions les moins bruitées. On peut en conclure, d'une part, que ces conditions ne constituaient vraisemblablement pas de bon contrôles pour étudier l'effet d'indiçage temporel et, d'autre part, que cette diminution du TR n'est pas de aux informations temporelles. En effet les deux conditions audiovisuelles statiques donnaient exactement les mêmes informations temporelles que la condition auditive seule. On peut donc exclure qu'il s'agisse de quelque effet d'indiçage temporel.

De plus ce bénéfice inattendu des stimuli visuels statiques semble être plus important pour les rectangles que pour les visages. Il ne s'agit donc pas simplement d'un effet attentionnel non spécifique, ou alors il faudrait expliquer pourquoi cet effet est plus fort pour un rectangle qu'une visage. Il se pourrait que cet effet représente la conjonction d'un effet attentionnel non spécifique qui aurait tendance à diminuer le TR et d'un effet d'incongruité des stimuli auditifs et visuels qui aurait tendance à augmenter le TR, l'incongruité d'un

[1] Curieusement, cet effet d'interaction n'est observé que pour les conditions de bruit 0db et -9dB, alors qu'il n'existe pas dans la condition -4,5dB et n'est pas significatif dans la condition -13,5dB. Dans ces deux dernières conditions, le bénéfice associé à la présences d'informations dynamiques est d'ailleurs plus faibles que dans les deux autres. La seule différence entre ces conditions était que les sujet avaient déjà été confrontés au niveaux de bruit 0dB et -9dB dans la première expérience.

visage immobile et d'un son de parole étant plus forte que celle d'un rectangle et d'un son de parole.

En tout état de cause, ce bénéfice semble diminuer avec le niveau de bruit, au contraire du bénéfice dû aux informations visuelles dynamiques, ce qui suggère que les indices visuels dynamiques améliorent spécifiquement la détection du prévoisement dans le bruit, alors que l'effet de la présence d'un stimulus visuel statique influencerait plutôt des processus plus généraux et non liés à la perception de la parole.

Une autre différence entre l'expérience 1 et l'expérience 2 est la disparition de l'effet d'interaction entre le voisement et la présence d'informations visuelles : cette disparation peut être due à une perte de puissance statistique due au nombre moins important de sujets, mais également au fait que les TR ont été mesurés pour un seul couple de syllabe.

11.4 Discussion générale

L'objectif initial des ces expériences comportementales étaient d'adapter le protocole de Schwartz et coll. (2004) à une expérience électrophysiologique. Nous avons montré que l'effet d'indiçage temporel des mouvements pré-phonatoires sur la perception du voisement pouvait être mis en évidence sur les temps de réaction dans une tâche de discrimination entre une syllabe voisée et une syllabe non voisée. Ce paradigme, plus simple, pourrait permettre d'étudier les corrélats électrophysiologiques à l'origine de cet effet, en enregistrant les potentiels évoqués par les mouvements articulatoires, une syllabe voisée et une syllabe voisée accompagnée des mouvements articulatoires.

On pourrait, à l'aide du modèle additif, étudier l'influence des informations visuelles temporelles non phonétiques sur le potentiel évoqué par le prévoisement dans le cortex auditif ou d'autres structures temporales. Si cet effet se traduit par une diminution de l'onde N1 auditive, on aurait un argument pour dire que l'effet observé dans notre première expérience électrophysiologique représenterait plutôt un effet d'indiçage temporel qu'une véritable intégration audiovisuelle phonétique. Dans le cas contraire, il serait plus difficile de conclure, étant donné la différence de structure audiovisuelle et acoustique des stimuli utilisés dans les deux paradigmes. L'expérience MEG que nous avions prévue au départ n'a malheureusement pas pu être réalisée, faute de temps.

Néanmoins, nos résultats comportementaux suggèrent que l'effet de pur indiçage temporel ne s'observe que lorsque la tâche des sujets consiste à détecter le pré-voisement dans le bruit et non lorsqu'il s'agit de discriminer le voisement dans de bonnes conditions acoustiques. À ce stade de nos investigations, c'est un argument supplémentaire pour dire que la diminution du TR observée dans nos expériences d'EEG était bien due à une intégration audiovisuelle phonétique et non à cet effet d'indiçage temporel, car notre expérience électrophysiologique était réalisée sans bruit acoustique et montrait néanmoins une diminution robuste du TR pour la discrimination des syllabes audiovisuelles par rapport aux syllabes auditives. À l'appui de cette affirmation, Callan et coll. (2004) ont montré en IRMf des effets d'interaction audiovisuelle dans la perception de la parole dans le STG/STS spécifiques aux informations visuelles de haute fréquence spatiale et qui ne sont pas trouvées pour des informations visuelles basse-fréquence, qui pourtant donnent une information tem-

porelle. À l'inverse, certains effets d'interaction audiovisuels très précoces sur les potentiels évoqués auditifs du tronc cérébral (Musacchia, Sams, Nicol & Kraus, 2006), similaires à des effets attentionnels, peuvent difficilement s'expliquer par une intégration phonétique et sont probablement dus à l'avance temporelle des informations visuelles sur les informations auditives.

Un résultat frappant et inattendu de nos deux expériences comportementales est que la simple présentation d'un stimulus visuel, ne fournissant aucune information pertinente, même temporelle, pour la tâche auditive à réaliser, semble diminuer le TR pour effectuer cette tâche. Cet effet n'est pas sans rappeler l'effet d'un stimulus accessoire sur le temps de traitement d'un stimulus dans une autre modalité (voir partie 2.3.2 page 34). Il était néanmoins assez faible et nécessiterait d'être répliqué et étudié plus en détail. Tout ce qu'on peut en dire pour l'instant c'est qu'il constitue une nouvelle preuve de l'interdépendance des traitements auditifs et visuels.