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

Annexe A

Données individuelles des patients

Patient	Région explorée	type de réponse	Latence de début (ms)	Latence de fin (ms)	Côté	Nom des contacts	Coordonnées de Talairach		
							X	Y	Z
8	STS postérieur	1	-80	400	G	E'6-8	-42	-53	1
9	GTM postérieur	1	-100	160	G	V'12-14	-46	-63	12
6	GTI/STI postérieur/Gyrus fusiforme	1	-80	140	D	L11	55	-55	0
10	Jonction occipito-temporale	1	-40	600+	G	V'10-12	-35	-65	1
10	Jonction occipito-temporale	1	-40	600+	D	V9-12	34	-64	1
10	Gyrus occipito-temporal ventral/fissure calcarine	1	-40	600+	D	V'7-9	26	-63	2
10	Gyrus occipito-temporal supérieur	1	-20	350	G	W'9-10	-34	-63	11
6	Gyrus occipito-temporal ventral	1	60	350	D	L3-4	32	-55	0
7	Planum temporale/gyrus supra-marginal	2	-20	450	G	G'11-14	-56	-40	21
5	Planum temporale/STS	2	0	600+	G	G'3-5	-57	-39	17
8	STS/GTS	2	-60	500	G	T'8-9	-62	-10	-3
8	Planum temporale	2	0	550	G	H'10-15	-47	-24	6
8	Insula/ planum polare	2	0	600+	G	T'3	-40	-10	-2
3	Fond du STS	2	-40	600+	D	C8	46	-16	-11
3	GTM latéral	2	-40	600+	D	C13	57	-15	-12
3	STS antérieur	2	-120	600+	D	T'7-9	57	-7	1
3	GTS/ gyrus transverse antérieur latéral	2	100	600+	G	T'6-9	-57	-4	4
1	Gyrus transverse postérieur médial/ Planum temporale	2	-20	600+	D	H'7-10	42	-29	7
1	Gyrus transverse antérieur latéral	2	40	600+	D	T'7-10	54	-9	-4
8	Insula/Gyrus transverse médial	2	160	550	D	T2	34	-12	6
2	Gyrus transverse antérieur médial	3	-120	450	D	H6-7	33	-25	11
2	Planum temporale	3	-120	450	D	H8-12	44	-27	10
10	Planum temporale/Gyrus transverse	3	-20	600+	D	H11-14	57	-19	6
10	Gyrus transverse antérieur médial	3	-20	600+	D	H6-9	33	-19	7
3	Planum temporale /STS	3	60	600+	G	H'6-13	-46	-30	5
7	Planum polare	3	140	300	G	T'2-7	43	-10	7
7	STS antérieur	4	60	600+	G	A'10-11	-49	-8	-16
3	GTM latéral	4	60	600+	G	B'11	-62	-20	-14
3	GTS/STS	4	160	600+	D	A11	57	-3	-9
7	STS *	4	160	600+	D	B7-9	49	-23	-9
8	MTG/STI (A'9-10)	4	220	600+	G	A'9-10	-56	0	-19
9	STS	5	40	600+	G	H'11-13	-60	-24	6
6	Bord inférieur STS	5	60	600+	D	B10	49	-16	0
5	GTS	5	220	600+	G	H'8-10	-61	-24	7
4	Fond du STS/ STI *	5	140	300	G	B'7-9	-46	-24	-9
4	GTS/STS antérieur	5	240	600+	G	T'6-9	-54	-15	-2
4	Bord supérieur STS	5	260	600+	G	H'11-15	-58	-28	10
4	GTM/STS	5	300	600+	G	B'10-12	-56	-24	-9

Patient	Région explorée	type de réponse	Latence de début (ms)	Latence de fin (ms)	Côté	Nom des contacts	Coordonnées de Talairach X	Y	Z
3	GTM	6	140	600+	G	L'13-14	-64	-43	-3
1	STS postérieur	6	160	450	D	R2	43	-39	3
10	Gyrus supramarginal/gyrus post-central	7	-20	600+	D	E13	52	-26	33
8	Gyrus supramarginal G	7	80	450	G	G'11-15	50	-33	33
8	Gyrus supramarginal D	7	160	600+	D	G15	57	-31	32
4	STS postérieur (lésion...)	8	100	400	D	X11	47	-53	16
5	Fond du STS postérieur/ gyrus supra marginal	8	100	500	G	Y'8-10	-35	-50	23
10	Gyrus supramarginal/ gyrus post-central/sillon post-central *	8	100	200	D	E8-10	31	-26	33
1	Gyrus angulaire/gyrus supramarginal *	8	220	350	D	V11-12	47	-43	34
10	Gyrus supramarginal/gyrus post-central/sillon post-central	8	240	400	D	E11	46	-25	33
10	Gyrus angulaire/supramarginal *	8	300	600+	D	G11-12	44	-43	26
5	Gyrus supramarginal	8	300	600+	G	Y'13-15	-53	-49	24
7	Gyrus post-central	9	100	260	G	R'1	-47	-18	27
8	Sillon post central D *	9	160	600+	G	G'8	-31	-36	30
5	Opercule post-central	9	220	600+	G	N'8-10	-58	-14	20
2	Opercule post-central	9	160	600+	D	N2	33	-14	12
7	Insula antérieure G/gyrus post-central	9	160	450	G	P'2-3	-37	-3	3
7	Insula/opercule post-centrale	9	240	600+	G	N'2-3	-38	-10	20
5	Insula/opercule post-central*	9	400	600+	G	N'3-4	-38	-14	20
4	Insula*	10	140	400	D	T4-5	38	-15	2
8	Insula/Planum temporale	10	160	550	D	T2	34	-12	6
4	Insula/Planum polare/gyrus transverse antérieur latéral *	10	220	400	G	T'4-5	-42	-15	-2
8	Gyrus cingulaire postérieur/ précunéus*	11	160	600+	G	G'3-4	-14	-37	30
10	Cingulaire postérieur*	11	200	600+	D	G3-4	12	-45	26
10	Gyrus cingulaire postérieur	11	220	600+	G	G'3	-12	-40	28
6	Opercule précentral	12	-60	550	D	N4-7	43	-7	10
1	Opercule précentral	12	140	600+	D	N7-9	52	12	21
7	Opercule précentral	12	240	600+	G	P'8	-58	-1	18
7	Gyrus frontal inférieur postérieur	12	240	600+	G	E'5-8	-51	11	7
9	Opercule précentral	12	240	400	G	P'4-10	-48	6	5

Patient	Région explorée	type de réponse	Latence de début (ms)	Latence de fin (ms)	Côté	Nom des contacts	Coordonnées de Talairach X	Y	Z
7	Hippocampe	13	80	400	G	B'2-5	-35	-23	-5
8	Hippocampe	13	180	600+	D	B'1-2	-30	-20	-10
8	Hippocampe	13	220	600+	D	B2-3	30	-22	-7
10	Gyrus parahippocampique/amygdale	13	220	600+	D	D3	22	-3	-33
10	Hippocampe/insula/gyrus transverse médial	13	260	600+	G	B'2-4	-34	-15	-10
9	Gyrus parahippocampique/gyrus lingual	13	300	600+	G	L'2	-18	-48	5
10	Gyrus lingual *	14	0	120	D	L5	32	-31	-15
10	Fond fissure calcarine *	14	280	400	D	W7	25	-63	9
10	GTM postérieur *	14	450	600+	D	W12	41	-63	7
10	Sillon frontal inférieur	14	160	300	D	K9-11	33	37	24
6	Gyrus frontal médian *	14	400	600+	D	K10-11	37	26	21
2	Opercule post-central inférieur/gyrus transverse	14	80	550	D	N4-5	43	-15	12
5	Sillon intrapariétal *	14	160	600+	G	Q'6	-23	-47	50
6	Gyrus supramarginal/sillon intrapariétal	14	-60	600+	D	R7-11	30	-48	35
8	Sillon colatéral	14	80	600+	G	D'5-7	-44	-5	-30
8	Gyrus temporal ventral antérieur*	14	120	350	G	A'4-7	-39	-1	-19
5	Cunéus/fissure pariéto-occipitale *	14	0	600+	G	W'3-5	-14	-68	36
5	Lobule pariétal inférieur/ fissure pariéto-occipitale *	14	0	600+	G	W'6-7	-26	-66	36
10	Gyrus lingual	14	180	600+	D	L2-5	28	-31	-14
10	Gyrus lingual	14	200	600+	G	L'4-5	-30	-30	-13
9	Sillon colatéral	14	300	600+	D	E'5-6	-42	-30	-9

TAB. A.1 – Coordonnées, localisation et latence des réponses aux syllabes visuelles. Type de réponse : 1. Activité spécifique à la condition visuelle enregistrée autour du GTM postérieur. 2. Activité enregistrée dans le lobe temporal supérieur et dont les sources ressemblent à celles de la réponse auditive entre 50 et 100 ms. 3. Activité enregistrée dans le lobe temporal supérieur et dont les sources ressemblent à celles de la réponse auditive après 100 ms. 4. Réponse autour du STS antérieur, commune aux conditions auditives et visuelles. 5. Activité enregistrée dans le lobe temporal supérieur, spécifique à la condition visuelle. 6. Réponse autour du STS postérieur commune aux conditions auditive et visuelle. 7. Réponse autour du gyrus supramarginal spécifique à la condition visuelle. 8. Réponse autour du gyrus supra marginal commune aux conditions auditives et visuelles. 9. Activité enregistrée autour de l'opercule post-central ou de l'insula, commune aux conditions A et V. 10. Réponse visuelle dans l'insula, sans réponse équivalente en condition auditive (mais peut-être cachée par l'activité provenant du cortex auditif). 11. Activité enregistrée autour du gyrus cingulaire. 12. Réponse enregistrée autour de l'opercule pré-central et du gyrus frontal inférieur. 13. Activité enregistrée autour de l'hippocampe et du gyrus parahippocampique. 14. Activités diverses. Les régions suivies d'une étoile sont celles dans lesquelles la réponse n'était significative qu'en montage bipolaire. 600+ : la réponse continue au-delà de 600 ms post-stimulus. GTM : Gyrus temporal moyen. GTS : Gyrus temporal supérieur. STI : Sillon temporal inférieur. STS : Sillon temporal supérieur.

Patient	Région explorée	type de réponse	Latence de début (ms)	Latence de fin (ms)	Côté	Nom des contacts	Coordonnées de Talairach X	Y	Z
6	Gyrus transverse antérieur médial	1	30	600+	D	H3-5	39	-20	5
2	Planum temporale antérieur	1	40	110	D	T8-10	61	-13	-2
7	GTS / gyrus supramarginal	1	70	600+	G	G'9-14	-54	-39	21
8	Planum temporale	1	70	500	G	H'10-15	-58	-23	6
3	GTS supérieur	1	80	450	G	T'8	-59	-4	4
10	Gyrus transverse médial	1	90	600+	D	H7-10	38	-19	7
3	Matière blanche du GTS	1	100	250	G	H'12-15	-61	-28	5
1	Planum temporal/ gyrus transverse	1	110	250	D	H8-10	44	-28	7
3	Bord supérieur du STS	1	120	600+	D	T7-8	55	-7	2
8	Planum polare/gyrus transverse latéral	1	120	250	D	T6-7	51	-11	8
8	GTS	1	120	180	D	T9-10	62	-11	9
9	GTM/bord inférieur du STS *	1	120	250	G	L'12-14	-56	-47	5
10	Planum Temporale/Gyrus transverse latéral	1	130	600+	D	H11-13	57	-19	6
7	Gyrus précentral	1	130	500	G	N'6-7	-52	-10	20
5	Planum temporale/Gyrus supramarginal	1	130	250	G	G'4	-57	-39	17
8	Bord supérieur du STS/GTS	1	130	200	G	T'8-9	-61	-9	-3
8	Gyrus supramarginal	1	130	300	G	G'13	-50	-35	29
7	Planum polare/bord supérieur du STS	1	140	400	G	T'4-9	-51	-4	2
8	Insula/Planum polare	1	160	450	D	T2-3	37	-11	7
6	Gyrus transverse postérieur latéral **	2	40	120	D	H7-9	53	-20	5
8	Bord supérieur du STS **p<0,005	2	50	90	G	T'7-8	-58	-9	-3
3	GTS ** p<0,01	2	50	100	G	H'12-15	-61	-28	5
8	Planum temporale	2	60	120	G	H'11	-50	-23	7
3	Bord supérieur du STS** p<0,005	2	60	120	D	T7-8	55	-7	2
7	Bord supérieur du STS* p<0,05	2	60	100	G	T'7-8	-57	-4	2
10	Gyrus transverse médial	2	80	160	D	H7-10	38	-19	7
8	Planum temporale	2	80	160	G	H'13-15	-62	-23	7
7	Bord supérieur du STS/planum polare	2	120	200	G	T'5-7	-57	-4	2
1	Gyrus cingulaire postérieur *	3	20	60	D	W4	13	-50	16
10	Fissure calcarine	3	30	80	D	V2	8	-68	5
2	Bord supérieur du STS/GTS	3	40	110	D	C9-13	61	-25	-4
2	GTS/planum temporale	3	50	100	D	H15	63	-26	10
6	Cunéus	3	60	130	D	G5-6	19	-57	16
2	Planum temporale*	3	70	120	D	H12	52	-26	10
9	Bord inférieur STS	3	100	120	G	B'11	-59	-17	-13
7	Gyrus transverse postérieur latéral	3	120	250	G	H'8-9	-57	-22	9
8	Gyrus cingulaire postérieur/précunéus	3	120	300	G	G'4	-15	-36	30
9	STS/GTS *	3	120	250	G	H'13	-62	-22	6
6	Insula	3	120	200	D	T3	38	-10	0
6	GTI ventral postérieur	3	130	200	D	L6-7	39	-55	-8
3	MTG	3	140	550	G	B'12	-66	-19	-14
6	Gyrus précentral	3	140	200	D	N7	50	-7	10
6	STI/GTI	3	140	250	D	L10	51	-54	-8
9	GTM/bord inférieur du STS *	3	160	250	G	V'14-15	-51	-61	12

TAB. A.2 – Coordonnées, localisations et latences des violations du modèles additif commençant entre 0 et 200 ms. Type de violation de l'additivité : 1. le profil spatiotemporel de la violation est identique à celui de la réponse visuelle et de polarité opposée. 2. Le profil spatiotemporel est identique à celui de la réponse auditive et de polarité opposée. 3. autre type de violation. * la violation n'était significative qu'en montage bipolaire. ** la violation n'était significative qu'en montage monopolaire. 600+ : la violation continue au-delà de 600 ms post-stimulus. GTS : Gyrus temporal supérieur. STS : Sillon temporal supérieur.

FIG A.1 (page suivante) - Les représentations tridimensionnelle et bidimensionnelle du ruban cortical sont propres au patient. La représentation tridimensionnelle est celle du lobe temporal et les représentations bidimensionnelle sont faites dans le plan coronal colinéaire à l'axe de de pénétration des électrodes. Sur les cartes de profil spatio-temporel, les zones entourées en jaune sont les échantillons significatifs au seuil corrigé. L'amplitude indiquée sous chaque couple de cartes monopolaire/bipolaire correspond aux couleurs les plus vives aux extrémités de l'échelle (jaune pour une différence de potentiel positive et rouge pour une différence de potentiel négative).

Pour l'électrode T, la réponse visuelle était soutenue entre 40 et 600 ms et le profil spatial ressemble à la première composante auditive entre 50 et 100 ms (foyers négatif sur T6-10 en monopolaire et foyers négatifs sur T6 et T8-10 en bipolaire). Pour l'électrode H, la réponse visuelle est constituée de plusieurs foyers positifs entre -20 et 600 ms que l'on retrouve en condition auditive, en particulier la réponse auditive transitoire vers 100 ms sur H9-10. En montage bipolaire, on retrouve une inversion de polarité autour de H10 dans les deux modalités. De manière générale, la réponse visuelle est plus soutenue que la réponse auditive, comme on peut le voir facilement sur les courbes. Le profil spatio-temporel de la violation présente une ressemblance évidente avec la réponse visuelle sur ces deux électrodes.

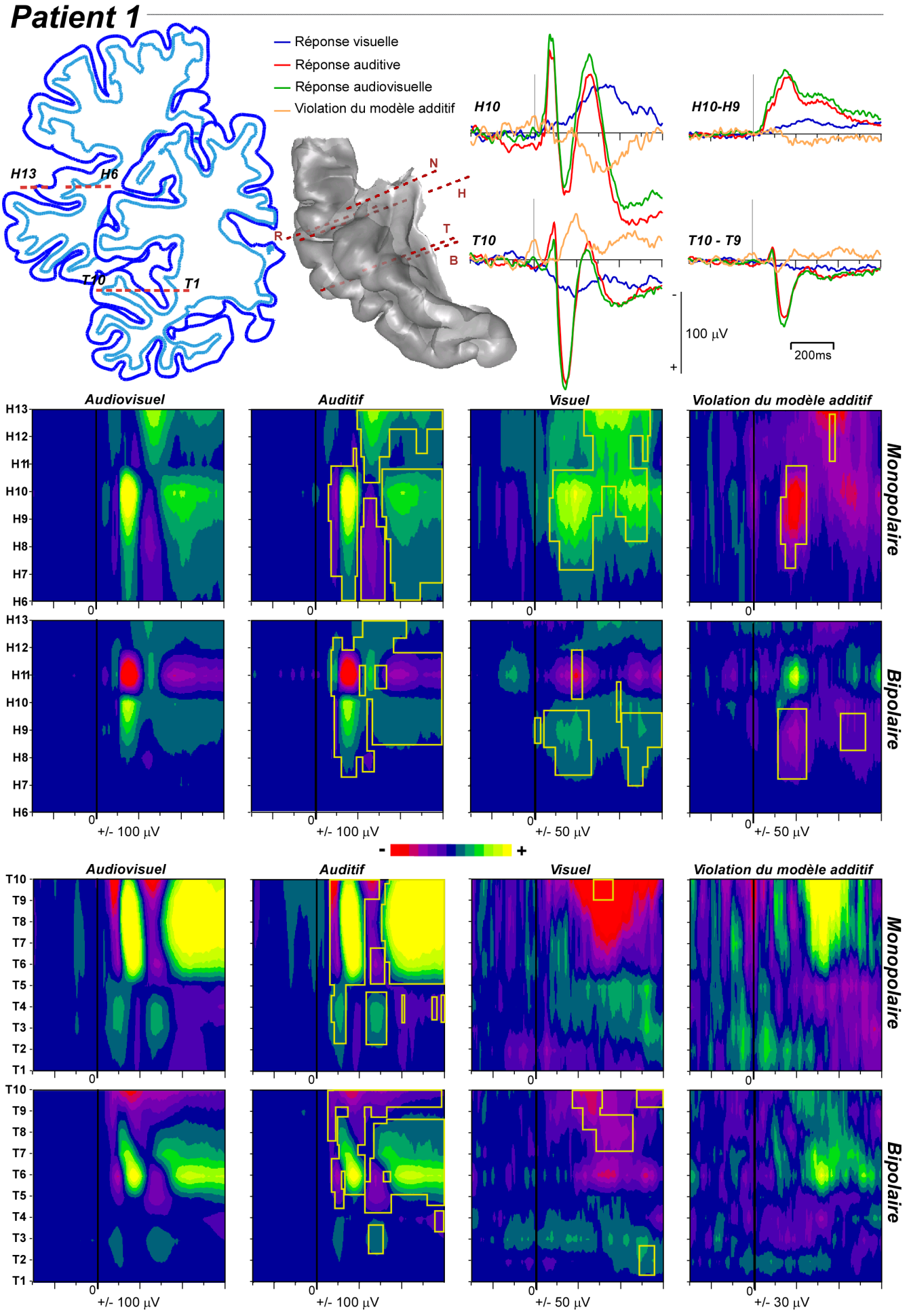

FIG. A.1 – Localisation et activités enregistrées aux électrodes H et T (hémisphère droit) pour le patient 1.

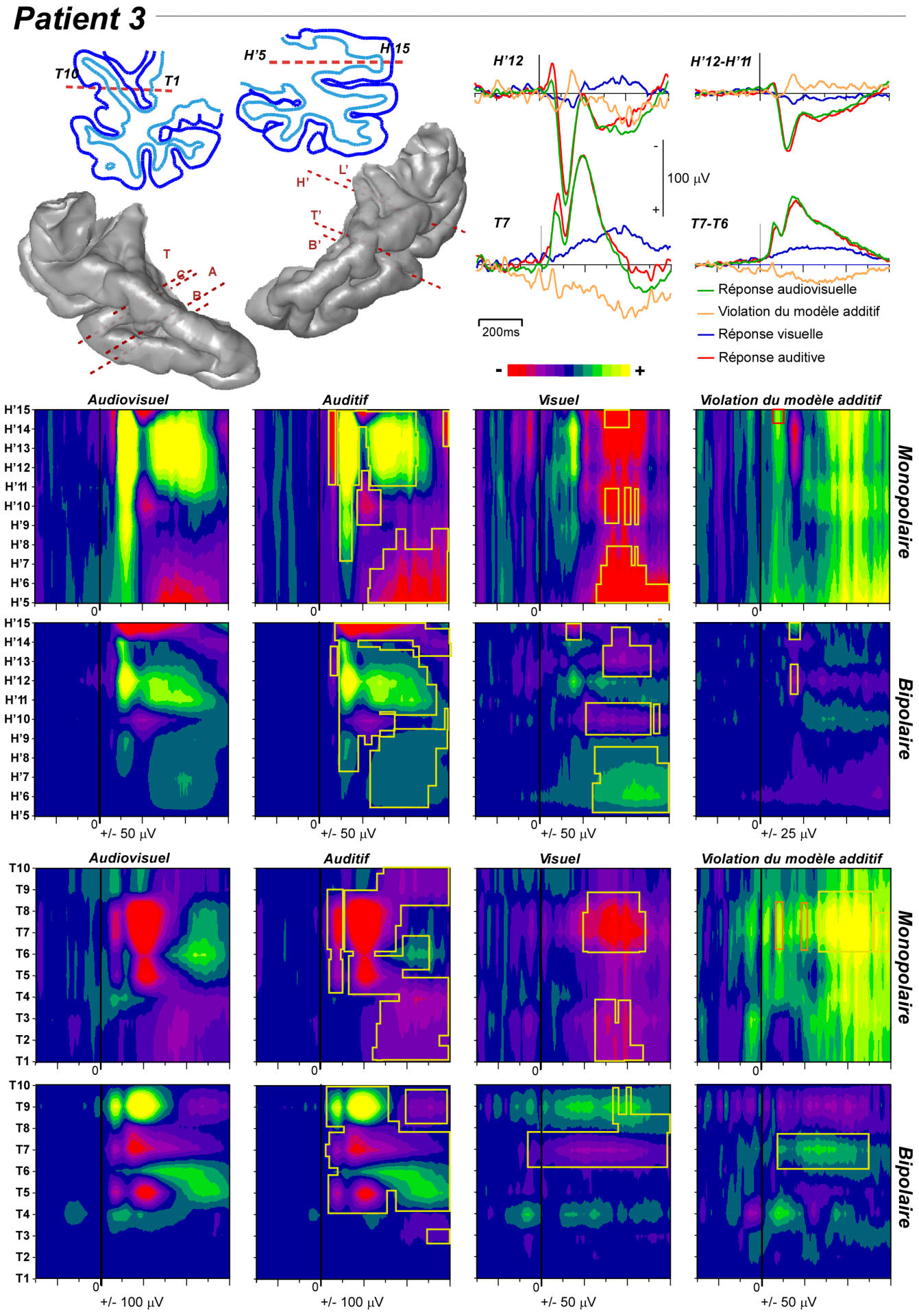

FIG. A.2 – Localisation et activités enregistrées aux électrodes H' (hémisphère gauche) et T (hémisphère droit) pour le patient 3.

FIG A.2 (page ci-contre) - Sur l'électrodes H', la première réponse visuelle significative apparait sur le contact H'15 vers 100 ms comme une composante positive en montage monopolaire ; cette première réponse ressemble à la réponse auditive entre 90 et 160 ms. Cette première réponse visuelle est suivie d'une réponse plus soutenue à partir de 160 ms qui semble ne correspondre à aucune composante auditive. Sur l'électrode T, la réponse visuelle soutenue commençant à -120 ms (elle est significative à partir de -70 sur T7 en bipolaire) sur T7-9 a le même profil spatial que les deux réponses auditives transitoires enregistrées entre 40 et 100 ms puis entre 100 et 300 ms, à la fois en montage monopolaire (foyer négatif sur T7-8) et en montage bipolaire (inversion de polarité entre T7 et T9). Sur les deux électrodes la ressemblance entre le profil spatiotemporel de la violation du modèle additif et la réponse visuelle est évidente (seulement sur les contacts les plus latéraux pour l'électrode T). On observe de plus quelques foyers qui ne peuvent s'expliquer par l'activation visuelle : sur les contacts H'12-15, entre 50 et 100 ms la violation a le même profil spatial que la réponse auditive transitoire à la même latence, mais uniquement en montage monopolaire. Cette modulation est visible sur les courbes du contact H'12. De même le foyer positif sur T7-8 entre 60 et 120 ms (montage monopolaire) correspond à la fois à la réponse auditive transitoire et à la réponse visuelle, mais son amplitude ne peut s'expliquer uniquement par l'activation visuelle. Les zones entourées en orange et rouge correspondent respectivement aux seuils $p < 0,005$ et $p < 0,01$.

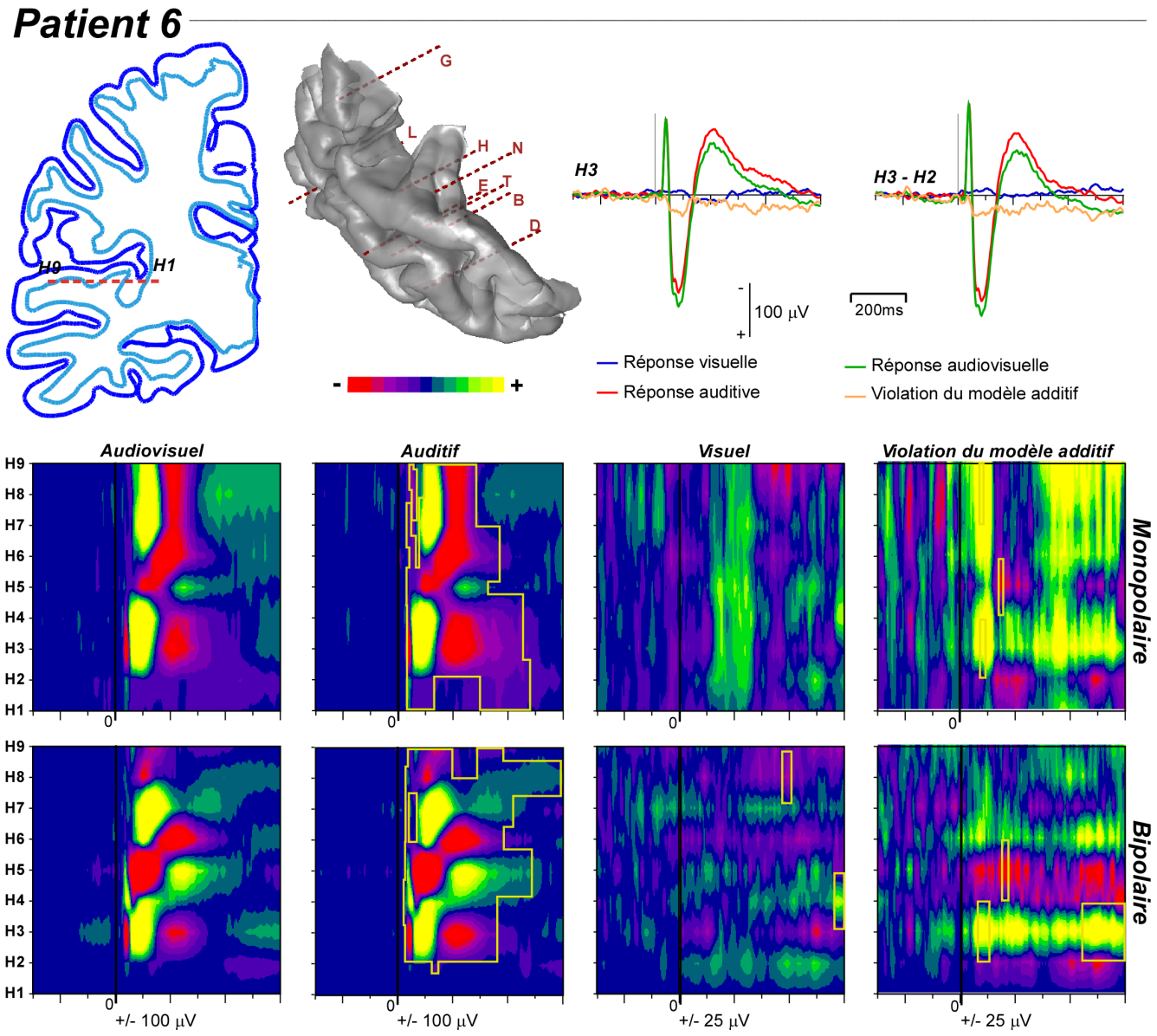

FIG. A.3 – Localisation et activités enregistrées à l'électrode H (hémisphère droit) pour le patient 6. Les premières réponse auditives sur les contacts H3-4 apparaissent dès 23 ms et sont enregistrées en montage monopolaire et bipolaire. La réponse visuelle émerge peu du bruit et ne devient significative que tardivement. On devine cependant l'existence de réponses soutenues dont le profil spatial évoque celui des réponses auditive transitoires, y compris aux niveau des contacts H3-4 où étaient enregistrées des réponses auditives primaires. De même le profil spatiotemporel de la violation ressemble à celui de la réponse visuelle, avec une amplitude plus importante. Le début de la violation de l'additivité sur les contacts H3-4 (entre 40 et 120 ms) pourrait également provenir de la modulation de la réponse transitoire auditive. Mais contrairement aux autres patients, il s'agit ici d'une augmentation de la réponse auditive en condition audiovisuelle.

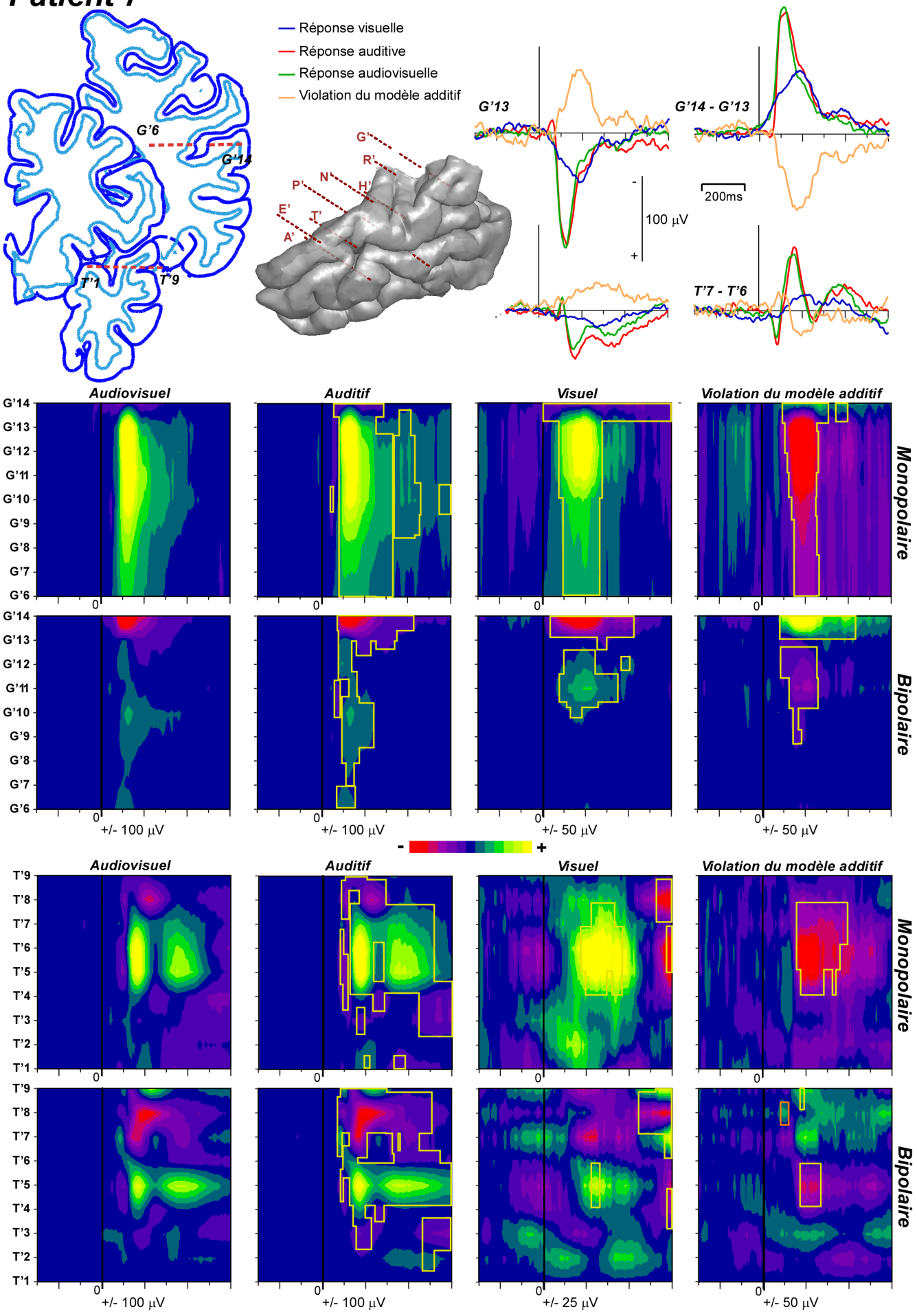

FIG. A.4 – Localisation et activités enregistrées aux électrodes G' et T' (hémisphère gauche) pour le patient 7.

FIG A.4 (page précédente) - Sur l'électrode T, la réponse soutenue commençant vers 100 ms et terminant vers 400 ms sur les contacts T'5-7 a le même profil spatial que la réponse auditive transitoire entre 120 et 200 ms (aussi bien en montage bipolaire que monopolaire). Sur l'électrode G', il existe également une certaine ressemblance entre les réponses visuelles et auditive, notamment en montage bipolaire au niveau du contact G'14. Sur les deux électrodes, la ressemblance entre le profil spatio-temporel de la violation de l'additivité et celui de la réponse visuelle est évidente. De plus sur les contacts T'7-8 entre 60 et 100 ms et T'5-7 entre 120 et 200 ms, la violation a le même profil spatio-temporel que les deux réponses transitoires auditives aux même latences, comme on peut le voir sur la courbe du montage bipolaire T'7-T'6. À ces latences l'amplitude de la réponse visuelle ne suffit pas à expliquer la violation, ce qui suggère l'existence d'une diminution de ces deux réponses auditives en condition audiovisuelle. Les zones entourées en orange correspondent au seuil $p<0,05$.

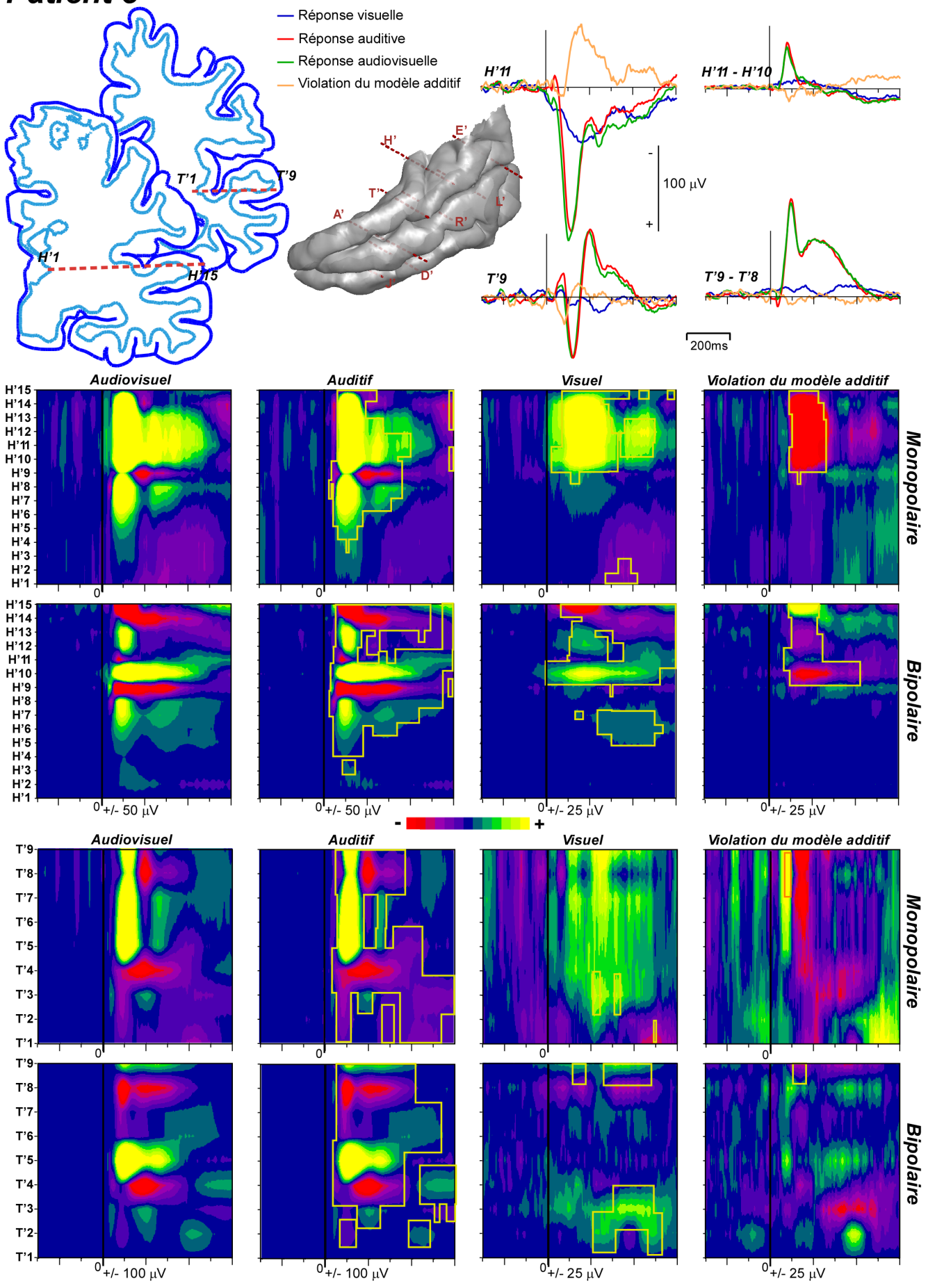

FIG. A.5 – Localisation et activités enregistrées pour les électrodes H' et T' (hémisphère gauche) pour le patient 8.

FIG A.5 (page précédente) - Les premières réponse auditive significatives apparaissent sur les contacts H'8-9 à partir de 25 ms en montage monopolaire et bipolaire. Sur l'électrode H', la réponse visuelle soutenue commençant à 0 et se terminant à 550 ms sur les contacts H'10-15 présente le même profil spatial que la réponse transitoire entre 50 et 150 ms, qui évolue elle-même en réponse soutenu ressemblant beaucoup à la réponse visuelle. Sur l'électrode T', la réponse visuelle soutenue enregistrée en montage bipolaire sur les contacts T'8-9 a également le même profil spatiale que la réponse transitoire/soutenue observée en condition auditive sur le mêmes contacts entre 50 et 400 ms. La réponse visuelle générée dans le cortex auditif est donc pour ce patient enregistrée sur des contacts différents de la réponse auditive primaire. Sur les deux électrodes H' et T', aux mêmes contacts que la réponse visuelle, le profil spatio-temporal de la violation de l'additivité ressemble de manière évidente à celle de la réponse visuelle. En montage bipolaire, un foyer positif au niveau du contact H'11 entre 60 et 120 ms n'est pas présent en condition visuelle mais correspond à la modulation de la réponse transitoire auditive, comme on peut le voir sur la courbe de l'activité bipolaire H'11-H'10. Sur le contact H'13, en montage bipolaire, on note également que la violation semble commencer à une latence inférieure à celle de la réponse visuelle. Cette violation pourrait être due à la diminution de la composante auditive transitoire enregistrée à ce contact entre 80 et 150 ms. De la même façon, la violation positive visible en monopolaire entre 50 et 100 ms sur les contacts T'5-9 suggère l'exitence d'une diminution de la composante négative transitoire auditive entre 50 et 100 ms, comme on peut le voir sur la courbe de l'activité monopolaire au contact T'9. Les zones entourées en orange correspondent au seuil $p<0{,}005$.

Patient 10

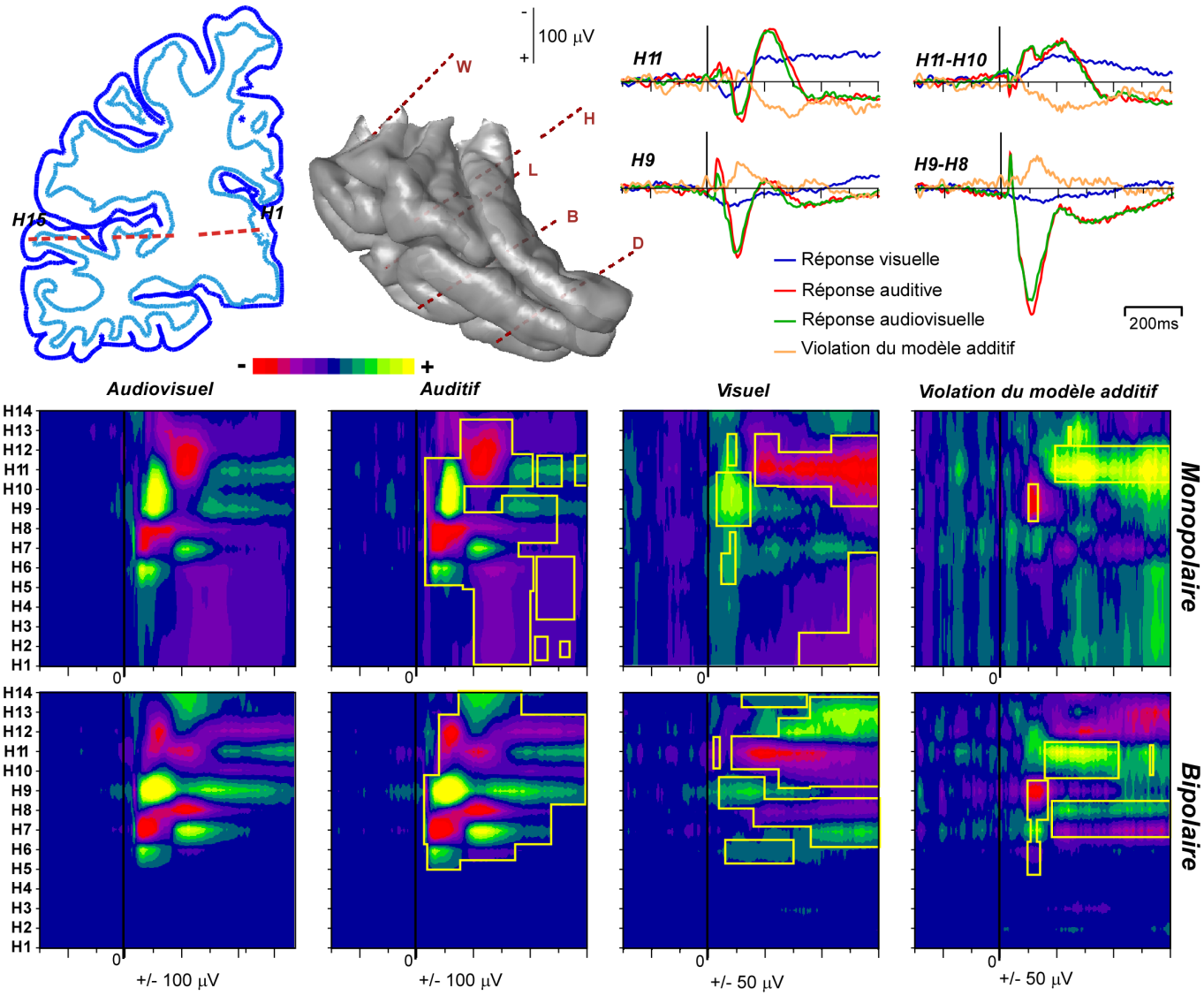

FIG. A.6 – Localisation et activités enregistrées pour l'électrode H (hémisphère droit) pour le patient 10. Les premières réponses auditives apparaissent à partir de 17 ms sur H6, H8 et H10. Les réponses visuelles sont constituées en montage monopolaire d'un réponse transitoire centrée sur les contacts H9-10 dont le profil spatiotemporel correspond à la réponse auditive transitoire entre 80 et 180 ms. En montage bipolaire apparaissent surtout des réponses soutenues dont le profil spatial ressemble à celui de la réponse auditive transitoire entre 50 et 150 ms sur H7-9, mais pas sur les électrodes plus latérales. Le profil spatiotemporel de la violation du modèle additif montrait une ressemblance certaine avec celui de la réponse visuelle aux mouvements articulatoires, excepté sur les contacts H6-8 entre 80 et 160 ms où le profil spatiotemporel était identique à celui de la réponse auditive transitoire et semble refléter une diminution de cette composante en condition audiovisuelle. Notons que pour ce patient la modulation de la réponse auditive transitoire et la réponse visuelle du cortex auditif semblent avoir lieu au niveau du cortex primaire.

Annexe B

Articles

Cogn Process (2004) 5: 189 192
DOI 10.1007/s10339-004-0026-y

LETTER TO THE EDITOR

Julien Besle · Alexandra Fort · Marie-Helene Giard

Interest and validity of the additive model in electrophysiological studies of multisensory interactions

Received: 10 June 2004 / Revised: 21 June 2004 / Accepted: 23 June 2004 / Published online: 22 July 2004
© Marta Olivetti Belardinelli and Springer-Verlag 2004

Over the past decades, the Stein group has provided a fundamental neural model of multisensory integration at the single-neuron level in animals. They have shown in cat and monkey that when inputs from different modalities are presented in close temporal and spatial proximity, multisensory neurons in the superior colliculus (SC) can increase their firing rate to a level exceeding that predicted by summing the responses to each unimodal cue (review in Stein and Meredith 1993).

Although this supra-additive effect applies to the single neuron, it has inspired a wider model that has been used at the integrated level of cortical populations (brain sites) in various functional brain imaging (ERP, MEG, fMRI) studies of multisensory integration. The rationale is that, under certain conditions that will be described below, neural activities induced by a bimodal stimulus (e.g., audiovisual, AV) should be equal to the sum of the responses generated separately by the two unisensory stimuli (e.g., auditory, A, and visual, V), if the two dimensions of the stimulus were to be independently processed. Hence, any neural activity departing from the mere summation of unimodal activities should be attributed to the bimodal nature of the stimulation, that is to interactions between the inputs from the two modalities. Using this model, it is therefore possible to estimate the crossmodal interactions in the differences between the brain responses to bimodal stimuli and the algebraic sum of the unimodal responses.

$$\text{AV Interactions} = \text{Response to (AV)} - [\text{Response to (A)} + \text{Response to (V)}]$$

Note that these interactions may include modulations of unimodal responses as well as new activities in sensory or polysensory areas.

This procedure, first used by Berman (1961) in event-related corticograms of cat, was later more formally expounded by Barth et al. (1995) in a study in which they identified the brain regions that responded (evoked potentials) uniquely to bimodal AV stimuli in rat cortex: "The model assumes that if subpopulations of cells that respond separately to auditory and visual stimulation do not respond uniquely to multisensory stimuli, their contribution to the [AV-ERP] will be the linear sum of their contributions to the [A-ERP] and [V-ERP] respectively. This assumption is valid for extracellular volume conducted potentials in a [sic] purely resistive extracellular media, and is based on the law of superposition of electrical fields. The sum [A-ERP] + [V-ERP] was then subtracted from the actual [AV-ERP] to obtain a difference waveform complex [AV – (A + V)]. The [AV – (A + V)] complex was used to determine cortical regions that were uniquely activated by polysensory stimulation." (Barth et al. 1995, p 179)

Although this model theoretically can be applied to any measure of human brain activity, it has been used mainly in electrophysiological data (scalp ERP and magneto-encephalography, MEG: Miniussi et al. 1998; Giard and Peronnet 1999; Foxe et al. 2000; Raij et al. 2000; Fort et al. 2002a, b; Molholm et al. 2002; Klucharev et al. 2003; Möttönen et al. 2004). On the other hand, its use has been recurrently criticized (Teder-Sälejärvi et al. 2002; Calvert and Thesen 2004) because of the multiple biases it can generate in the estimation of the crossmodal interactions if several important conditions are not fulfilled. We discuss in this note what these biases are and how to avoid or minimize them, with particular emphasis on electromagnetic (EEG/MEG) recordings. Finally, we explain why, in spite of its strict conditions of application, the supra-additive model is particularly interesting in ERP/MEG studies of multisensory integration.

Edited by: Marie-Hélène Giard and Mark Wallace

J. Besle · A. Fort · M.-H. Giard (✉)
Mental Process and Brain Activation Lab,
U280 INSERM, 151 Cours Albert Thomas,
69003 Lyon, France
E-mail: giard@lyon.inserm.fr

Potential biases and artifacts generated by the additive model

1. The additive model is valid only when the brain responses that are analyzed do not include activity common to all conditions. Indeed these activities would be added only once but subtracted twice in the [AV – (A + V)] model, which would confound the derivation of the multisensory interaction. "Common activity" may be of several types. One type is neural responses related to late semantic processes, target processing (e.g., N2b/P3 waves in ERP/MEG recordings), response selection, or motor processes. ERP literature has shown that these activities usually arise about 200 ms post-stimulus, whereas earlier latencies are characterized by sensory-specific responses (review in Hillyard et al. 1998). One way to avoid this problem is to restrict the analysis period to the early time frame (< 200 ms) of stimulus processing. While this procedure is very simple in ERP/MEG recordings since their time resolution is of the order of the millisecond, sorting the response components according to their latency is still virtually impossible in hemodynamic imaging techniques. Second, in paradigms requiring speeded responses with rapidly presented stimuli, "anticipatory" slow responses may arise before each (unimodal and bimodal) stimulus and continue for a time after stimulus onset. These anticipatory responses appear, when present in ERP/MEG recordings, as slow ramp-like deflections in the prestimulus and early poststimulus periods. These deflections can thus give rise to spurious residual effects in the [AV – (A + V)] signals that may be confused with early cross-modal interactions (Teder-Sälejärvi et al. 2002). Note that such anticipatory processes are independent of the technique used and can also be included in fMRI/PET responses. At the level of the experimental design, a procedure that may be applied to avoid or strongly reduce anticipatory processes, whatever the neuroimaging technique, is to present the stimuli at random interstimulus time intervals during data acquisition. In ERP/MEG signal analysis, two further methods have been proposed to control for these effects: modify the latency of the prestimulus period that will be used as the reference baseline, and/or high-pass the data (e.g., 2 Hz cut-off frequency) to remove the slow wave effects.

2. Several functional imaging studies using block-designed paradigms have shown a decrease in activation in sensory-specific cortices (e.g., the auditory cortex) when subjects were presented with continuous stimulation in another (e.g., visual) modality (Haxby et al. 1994; Kawashima et al. 1995; Laurienti et al. 2002). There are two possibilities to explain this. First, these effects may reflect cross-sensory driving and/or inhibition of lower-order sensory areas via direct projections from one sensory cortex to another (Falchier et al. 2002; Rockland and Ojima 2003; review in Schroeder et al. 2004). There is, however, no experimental evidence that such "cross-modal effects" in unimodal conditions may be seen at the integrated level of scalp ERP/MEG or fMRI signals irrespective of the task or stimulus delivery context. In addition, even in this case, the additive model should still apply since any difference in these processes between a unimodal and a bimodal condition should appear—if strong enough—as low-level cross-modal interactions in the model, and further represent one possible neural mechanism for multisensory integration. A second, more likely explanation is that when a particular sensory cortex is continuously and exclusively activated during a whole block, while the other non-matching cortices are not activated, the attentional resources are dedicated to the relevant modality (even in passive tasks or tasks that demand little attention), while the other modalities are more or less voluntarily ignored (deactivated) to optimize the processing in the relevant sensory cortex (see also Ghatan et al. 1998; Kawashima et al. 1999, for similar attentional effects). In studies of multisensory integration, the [AV – (A + V)] model should therefore not be used in experiments based on block-designed paradigms, since these unimodal deactivations would be subtracted from the bimodal activations, resulting in artificial increases of the "crossmodal" effects. One way to eliminate or considerably reduce such attention-related deactivations in unisensory cortices is to consider paradigms in which the stimuli are randomly and equiprobably delivered across all modality conditions (e.g., Giard and Peronnet 1999; Calvert et al. 2000, 2001; Foxe et al. 2000; Raij et al. 2000; Fort et al. 2002a, b; Molholm et al. 2002; Wright et al. 2003).

3. Random mixing of conditions, however, may not be sufficient for a correct control of attention. Although a classical design to avoid attentional biases in the additive model is to require the same task in the three modalities, in some paradigms, the task may be easier and require less effort in one unimodal condition than in the other. This problem can be overcome by equating the levels of difficulty across unimodal conditions (by equating the behavioral performance in both unimodal conditions, e.g., Giard and Peronnet 1999). However, in some particular cases, this may not be possible and using the same task across all the conditions can lead to noticeable spurious effects in the computation of interactions. Consider, for example, speech stimuli (lip movements associated with syllable sounds) randomly presented in the three A, V and AV conditions: if a discrimination task (e.g., respond to target syllables) is required under the three modality conditions, the processing of syllables in the lip-reading condition alone will include an important visual attention effect that will not be eliminated in the AV – (A + V) derivation, since in speech perception (unlike what is likely to occur for bimodal non-speech objects), normal subjects will *naturally* engage much less visual attention to process AV than V stimuli. Alternatively, if the subjects are required to respond only whenever they hear (A and AV conditions) a target syllable, their (selective) auditory attention effect

will be expressed rather similarly for A and AV stimuli and eliminated in the [AV – (A + V)] model; in the same way, a lesser (if any) effect of visual attention (rather similar for V and AV stimuli) should be mostly eliminated in the model. A general principle, therefore, in dealing with attentional problems is, in addition to systematically mixing conditions, to equate the attentional load between each unimodal condition and the bimodal condition (but not necessarily between the two unimodal conditions).

Advantages of the additive model in ERP/MEG studies of cross-modal interactions

All the examples above show that non-biased estimation of multisensory interactions in the human cortex using the additive model requires taking important precautions both in the experimental design and in data analysis. While the constraints relative to the control of attention may be easily respected whatever the neuroimaging technique used, caveats concerning the temporal selection of the response components to be analyzed can be overcome only in EEG/MEG approaches, because of the excellent time information provided by these techniques.

In addition, the additive model has a further fundamental interest in ERP/MEG analysis of crossmodal interactions. Indeed, unlike what is observed at the voxel level in fMRI or PET signals, a significant value at a particular electrode (sensor) in ERP/MEG recordings does not mean that the structure beneath the electrode/sensor is active. Rather what is recorded at the scalp surface results from the diffusion of electrical currents inside the brain originating from *distant* "generators," and the interpretation of the surface signals needs to take into account these volume conduction factors (using topographic analysis, generator modeling, etc.). Interestingly, the additive [AV – (A + V)] model in ERP/MEG has the fundamental property of avoiding the problem of overlaps of volume conduction effects in the different subcomponents of the bimodal response by removing the conduction effects of the corresponding unimodal responses. In this respect, the additive model is not a mere application of the single-cell model used by Stein's group and other authors: it applies not only at the *local* structure level (single cell, voxel), but also at the *distant* electrode/sensor level (volume conduction effects) because it is based on the superposition principle of electrical fields, in which the potentials from separate current sources in a conductive medium sum linearly. If its conditions of application are fulfilled, the additive model will therefore isolate the (volume conduction) effects specifically related to the interactions (which will have to be analyzed in turn in terms of topography and generators).

We therefore believe that the additive model is particularly well suited to ERP/MEG study of multisensory interactions in humans, and that its multiple advantages make it worthwhile dealing with the several constraints it imposes. Provided that its conditions of application are respected, the model can reveal the existence of genuine cross-modal interactions without making a priori assumptions about the congruent/incongruent character of the bimodal inputs, or introducing supra-additive/sub-additive criteria for integration (e.g., Calvert 2001; Calvert et al. 2001). Rather the additive model allows one to access the dynamics of the multisensory interactions and observe both supra-additive and sub-additive modulations of unimodal activities in sensory-specific cortices—which appear to form a highly flexible network of cross-modal operations—as well as to observe new processes specifically activated by the bimodal nature of the stimulus.

References

Barth DS, Goldberg N, Brett B, Di S (1995) The spatiotemporal organization of auditory, visual and auditory visual evoked potentials in rat cortex. Brain Res 678:177 190

Berman AL (1961) Interaction of cortical responses to somatic and auditory stimuli in anterior ectosylvian gyrus of cat. J Neurophysiol 24:608 620

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110 1123

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Par (in press)

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649 657

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio visual integration sites in humans by application of electrophysiological criteria to the bold effect. Neuroimage 14:427 438

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22(13):5749 5759

Fort A, Delpuech C, Pernier J, Giard MH (2002a) Dynamics of cortico-subcortical crossmodal operations involved in audio visual object detection in humans. Cereb Cortex 12(10):1031 1039

Fort A, Delpuech C, Pernier J, Giard MH (2002b) Early auditory visual interactions in human cortex during nonredundant target identification. Cogn Brain Res14:20 30

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Cogn Brain Res 10:77 83

Ghatan PH, Hsieh JC, Petersson KM, Stone-Elander S, Ingvar M (1998) Coexistence of attention-based facilitation and inhibition in the human cortex. Neuroimage 7:23 29

Giard MH, Peronnet F (1999) Auditory visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11(5):473 490

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a pet-rcbf study of selective attention to faces and locations. J Neurosci 14(11):6336 6353

Hillyard SA, Teder-Sälejärvi WA, Munte TF (1998) Temporal dynamics of early perceptual processing. Curr Opin Neurobiol 8:202 210

Kawashima R, O'Sullivan BT, Roland PE (1995) Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the "mind's eye". Proc Natl Acad Sci USA 92:5969 5972

Kawashima R, Imaizumi S, Mori K, Okada K, Goto R, Kiritani S et al (1999) Selective visual and auditory attention toward utterances-a pet study. Neuroimage 10:209 215

Klucharev V, Mottonen R, Sams M (2003) Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Cogn Brain Res 18(1):65 75

Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. J Cogn Neurosci 14(3):420 429

Miniussi C, Girelli M, Marzi CA (1998) Neural site of the redundant target effect: electrophysiological evidence. J Cogn Neurosci 10:216 230

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res 14(1):115 128

Möttönen R, Schurmann M, Sams M (2004) Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. Neurosci Lett 363(2):112 115

Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28(2):617 625

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50(1 2):19 26

Schroeder C, Molholm S, Lakatos P, Ritter W, Foxe JJ (2004) Human-simian correspondence in the early cortical processing of multisensory cues. Cogn Process DOI 10.1007/s10339-004-0020-4

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio visual crossmodal integration by means of event-related potential (erp) recordings. Cogn Brain Res 14(1):106 114

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13(10):1034 1043

European Journal of Neuroscience, Vol. 20, pp. 2225 2234, 2004 © *Federation of European Neuroscience Societies*

Bimodal speech: early suppressive visual effects in human auditory cortex

Julien Besle, Alexandra Fort, Claude Delpuech and Marie-Hélène Giard
INSERM U280, Mental Processes and Brain Activation, 151 Cours Albert Thomas, 69424 Lyon Cedex 03, France

Keywords: audiovisual, electrophysiology, multisensory integration, speech perception,

Abstract

While everyone has experienced that seeing lip movements may improve speech perception, little is known about the neural mechanisms by which audiovisual speech information is combined. Event-related potentials (ERPs) were recorded while subjects performed an auditory recognition task among four different natural syllables randomly presented in the auditory (A), visual (V) or congruent bimodal (AV) condition. We found that: (i) bimodal syllables were identified more rapidly than auditory alone stimuli; (ii) this behavioural facilitation was associated with cross-modal [AV – (A + V)] ERP effects around 120–190 ms latency, expressed mainly as a decrease of unimodal N1 generator activities in the auditory cortex. This finding provides evidence for suppressive, speech-specific audiovisual integration mechanisms, which are likely to be related to the dominance of the auditory modality for speech perception. Furthermore, the latency of the effect indicates that integration operates at pre-representational stages of stimulus analysis, probably via feedback projections from visual and/or polymodal areas.

Introduction

It is commonly known and agreed that vision may improve the comprehension of a talker in a face-to-face conversation or on the television. In behavioural studies, the influence of visual information on auditory speech perception has been particularly explored in the 'McGurk effect' (McGurk & McDonald, 1976), an auditory illusion produced for particular syllables when the lip movements do not match the auditory signal (for example, auditory /ba/ combined with visual /ga/ is perceived as /da/).

Yet, the neural mechanisms by which auditory and visual speech information is combined in normal communication are still poorly understood. Several functional neuroimaging studies have identified possible sites of multisensory convergence and integration for linguistic material with various results. Haemodynamic responses to semantically congruent audiovisual speech stimuli were found to be enhanced in sensory specific auditory and visual cortices, compared to the responses to unimodal or incongruent bimodal inputs (Calvert *et al.*, 1999). However, when only the brain areas presenting supra-additive response enhancement to congruent bimodal inputs and subadditive response to incongruent cues were considered as integration sites, only the left superior temporal sulcus (STS) exhibited significant integration effects. In another functional magnetic resonance imaging experiment, a supra-additive enhancement was found only in the left claustrum/insula whereas activation of the STS occurred for lip-reading alone (Olson *et al.*, 2002; see also Calvert & Campbell, 2003). Whatever the precise sites of multisensory integration, Calvert (2001) hypothesized that increased activity in sensory-specific cortices would be a result of backward projections from polymodal areas such as the STS.

However, this assumption is beyond the reach of haemodynamic imaging techniques because of their poor temporal resolution. In contrast, neuromagnetic (MEG) and event-related potential (ERP) recordings can provide significant insights into the timing of bimodal speech integration.

In three studies using audiovisual oddball paradigms (Sams *et al.*, 1991; Colin *et al.*, 2002; Möttönen *et al.*, 2002), deviant 'McGurk syllables' differing from standard syllables only on the visual dimension were found to elicit a mismatch negativity (MMN) around 150 180 ms post-stimulus, an ERP/MEG component generated for the main part in the auditory cortex. As MMN probably reflects a neuronal mismatch between deviant auditory inputs and a neural representation of the past stimuli in auditory sensory memory (review in Näätänen & Winkler, 1999), it can be concluded from these previous studies that visual speech information has been integrated to the auditory input before the MMN process was triggered, that is before about 150 ms. This McGurk paradigm, however, only put an indirect upper bound on the timing of multisensory integration and the question remains open as to when and where in the sensory processing chain, and by which neural mechanisms auditory-visual speech is combined.

One way to investigate these questions is to compare the electrophysiological responses to bimodal sensory inputs with the sum of the responses to unimodal cues presented separately. This approach was used in humans to analyse the mechanisms of audiovisual integration in bimodal object recognition (Giard & Peronnet, 1999) and revealed the existence of multiple interactions within the first 200 ms post-stimulation, expressed both as modulations (increase and decrease) and as new activations in sensory-specific and polymodal brain areas. Subsequent experiments using this additive model have provided evidence for different integrative operations according to the stimulus type, the modalities involved, or the task required (Foxe *et al.*, 2000; Raij *et al.*, 2000; Fort *et al.*, 2002a,b; Molholm *et al.*, 2002). In the present study we therefore used the same approach to investigate the time-course and neural mechanisms of audiovisual integration in the particular case of speech perception.

Correspondence: Dr M.-H. Giard, as above.
Email: giard@lyon.inserm.fr

Received 14 November 2003, revised 27 July 2004, accepted 3 August 2004

doi:10.1111/j.1460-9568.2004.03670.x

Materials and methods

Subjects

Sixteen right-handed native French speakers (mean age 23.0; eight females) were paid to participate in the study, for which they gave a written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects were free from neurological disease, had normal hearing and normal or corrected-to-normal vision.

Thirteen other subjects (mean age 24.3; nine females) participated in an additional behavioural-only experiment.

Stimuli

ERP study of multisensory integration requires to strictly control the timing of the unimodal input signals, a particularly heavy constraint in the case of natural speech. We therefore proceeded in the following way:

1. A hundred utterances of four different audiovisual syllables (/pa/, /pi/, /po/ and /py/) were produced by a female French speaker and recorded with a DV camera at a video sampling rate of 25 fps and an audio sampling rate of 44.1 kHz.
2. Visual inspection of the video stream showed that for most utterances, six frames (240 ms) separated the first detectable lip movements from the opening of the mouth (corresponding roughly with the beginning of the speech sound). To have stimuli with similar auditory-visual structures, we selected a subset of these syllables. The sound onset was then strictly postsynchronized with the onset of the 7th frame. This point (240 ms after the beginning of lip movements) was taken as time zero for ERP averaging and latency measurements (see Fig. 1B). The voice onset times (the intervals between the consonant burst and the voicing corresponding to the vowel), originally ranging from 15 to 26 ms, were artificially shortened to 15 ms for all the stimuli.
3. Using a unique exemplar of each syllable (/pa/, /pi/, /po/ or /py/) could have led subjects to learn and recognize the stimuli on the basis of low-level sensory features specific to each stimulus but irrelevant for phonetic processing. We therefore selected three exemplars of each syllable, that is 12 different utterances.
4. Eventually, lip movements preceding the sound emission anticipate the shape that will produce the vowel (coarticulation) and can therefore slightly differ between the different syllables. Although the prevowel lip movements were very faint during the first six frames of the video stream, we ensured that they could not allow the subjects to deduce the identity of the syllable before the sound onset (7th frame): we asked seven subjects (who did not participate in the main experiment) to visually identify the syllables on the basis of the first 6, 8 or 13 frames. Results showed that subjects did respond at chance level in the 6-frame condition.

All the images of the video stream were cropped in order to keep only the mouth, the cheeks and the bottom of the nose (see Fig. 1B). In the final frames, the mouth was about 5 cm wide and was presented on a video monitor placed 130 cm in front of the subjects' eyes, subtending a visual angle of 2.2°. The duration of the 12 sounds corresponding to the 12 syllables ranged from 141 to 210 ms; their amplitudes were adjusted to have the same perceived intensity (kept constant for all subjects).

Procedure

Subjects were seated in a dark, sound-attenuating room and were given instructions describing the task along with a practice block of 70 trials (a trial is described in Fig. 1A). Then subjects were presented with 31 repetitions of the 12 syllables in each of the three following conditions: auditory-only (A), visual-only (V) and audiovisual (AV). These 1116 trials were divided into 16 blocks (block duration, about 2 min 35 s; mean ISI, 2210 ms). In all blocks, trials were delivered pseudorandomly with the constraint that two stimuli of the same condition could not occur in a row.

At the beginning of each block, one of the four syllables (/pa/, /pi/, /po/ or /py/) was designated as the target (so that each syllable could be target or nontarget depending on the block). The subjects' task was to press a mouse-button with the right forefinger whenever they heard (A and AV conditions) the target-syllable in the block sequence.

An auditory task alone was chosen because estimation of the crossmodal interactions using the additive [AV − (A + V)] model (see data analysis) requires that the attention level in each modality is similar between the unimodal and the bimodal conditions (but not necessarily between the two unimodal conditions). Indeed, as the subjects are instructed to make an auditory discrimination task, the auditory attention effect will be expressed rather similarly in A and AV

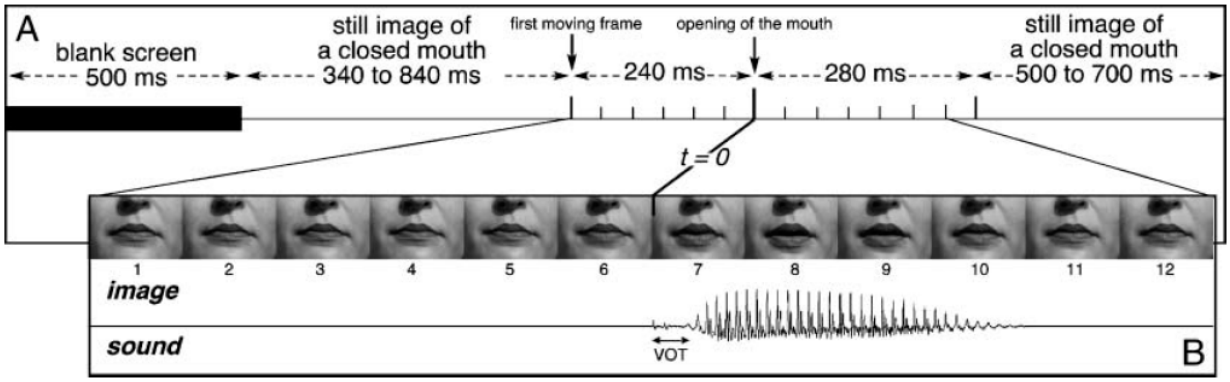

FIG. 1. Time-course of an auditory-visual non-target trial. Each trial began with the presentation of a blank screen for 500 ms; then a still image of a closed mouth was displayed during a random period of 340 840 ms. The mouth began to move for 240 ms (six frames) before opening (time zero). Then, the corresponding sound was played. The lip movement ended 280 ms after time zero with an image of the closed mouth that remained for a random time of 500 700 ms for nontarget trials, and until the key press for target trials (or for 1500 ms if the subject did not respond). In the visual-only condition, the time course was similar except that the sound was not played. In the auditory-only condition, the mouth remained closed all along the trial. VOT, voice onset time.

brain responses and mostly eliminated in [AV − (A + V)]. By contrast, because lip-reading is unnatural and difficult for untrained, normal-hearing subjects, a task in the three A, V and AV conditions would have led the subjects to naturally engage much more visual attention to process visual than bimodal stimuli. As a consequence, a larger visual attention effect in the V than in AV responses would not have been eliminated in the model. On the contrary, the task used here required a rather similar (if any) visual attention effort to process the visual and audio-visual stimuli, then minimizing any attentional bias in the additive model.

Electroencephalogram recording

Electroencephalograms (EEG) were was recorded continuously via a Neuroscan Compumedics system through Synamps DC-coupled amplifiers (0.1 200 Hz analogue band width; sampling rate, 1 kHz) from 36 Ag AgCl scalp electrodes referenced to the nose and placed according to the International 10 20 System: Fz, Cz, Pz, POz, Iz; Fp1, F3, F7, FT3, FC1, T3, C3, TP3, CP1, T7, P3, P7, PO3, O1, and their counterparts on the right hemiscalp; Ma1 and Ma2 (left and right mastoids, respectively); IMa and IMb (midway between Iz-Ma1 and Iz-Ma2, respectively). Electrode impedances were kept below 5 kΩ. Horizontal eye movements were recorded from the outer canthus of the right eye; eye blinks and vertical eye movements were measured in channels Fp1 and Fp2.

Data analysis

EEG analysis was undertaken with the ELAN Pack software developed at the INSERM U280 laboratory (Lyon, France). Trials with signal amplitudes exceeding 100 μV at any electrode from 2000 ms before time zero to 500 ms after were automatically rejected to discard the responses contaminated by eye movements or muscular activities. One subject was excluded from analysis for general noise in EEG at most sites. For seven other subjects, the excessively noisy signals at one or two electrodes were replaced by their values interpolated from the remaining electrodes.

ERPs to nontarget stimuli were averaged offline across the 12 different syllables separately for each modality (A, V, AV), over a time period of 1000 ms including 500 ms prestimulus (the zero time corresponding to the onset of the sound, or the onset of the 7th video frame for visual-only trials). Trials including false alarms were not taken into account when averaging. The mean numbers of averaged trials (by subject) were 155, 157 and 170 in the A, V and AV conditions, respectively (about 40% of the trials were discarded because of important eye movements).

ERPs were finally digitally filtered (bandwidth, 1 30 Hz; slope, 24 dB/octave). The mean amplitude over the [−300 to −200 ms] prestimulus period was taken as the baseline for all amplitude measurements.

Estimation of audiovisual interactions

We assumed that at an early stage of stimulus processing, if auditory (A) and visual (V) dimensions of the stimulus were to be independently processed, the neural activities induced by the audiovisual (AV) stimulus should be equal to the algebraic sum of the responses generated separately by the two unisensory stimuli. Hence, any neural activity departing from the mere summation of unimodal activities should be attributed to the bimodal nature of the stimulation, that is to interactions between the inputs from the two modalities (Barth *et al.*, 1995; Miniussi *et al.*, 1998; Giard & Peronnet, 1999; see Discussion in Besle *et al.*, 2004). This assumption is valid only if the period of analysis does not include nonspecific activities that would be common to all three types of stimuli, and particularly late activities related to semantic processing, response selection or motor processes. ERP litterature shows that these 'nonspecific' components generally arise after about 200 ms, whereas the earlier latencies are characterized by sensory-specific responses (e.g. Hillyard *et al.*, 1998 for a review). We have therefore restricted the analysis period to [0 200] ms and used the following summative model to estimate the AV interactions:

$$\mathrm{ERP\ (AV)} = \mathrm{ERP\ (A)} + \mathrm{ERP\ (V)} + \mathrm{ERP\ (A \times V\ interactions)}$$

This expression is valid whatever the nature, configuration or asynchrony of the underlying neural generators and is based on the law of superposition of electric fields. However, estimation of AV interactions using this procedure further requires that: (i) the levels of modality-specific attention are similar between each unimodal condition and the bimodal condition (see Procedure); and (ii), the effects potentially found in a particular structure cannot be attributed to deactivation processes in that structure under unimodal stimulation of a concurrent modality (see Discussion).

Significant interaction effects were assessed by Student's *t*-tests comparing the amplitudes of the [AV − (A + V)] difference waves to zero for each time sample at each electrode. Student's *t*-maps could then be displayed at each latency. Correction for multiple comparisons was performed using the procedure of Guthrie & Buchwald (1991), which tabulates the minimum number of consecutive time samples that should be significant in ERP differences, in order to have a significant effect over a given time series. As these tables are given for a horizon of 150 time samples, we under-sampled our data at 500 Hz over the [0 200 ms] analysis period, as proposed by the authors. We therefore considered as significant interactions the spatiotemporal patterns having a stable topography with significant amplitude ($P < 0.05$) during at least 12 consecutive time samples (24 ms), which is an upper bound for 15 subjects over 100 time samples (200 ms).

Topographic analysis and dipole modelling

To facilitate the interpretation of the voltage values recorded at multiple electrodes over the scalp surface, we analysed the topographic distributions of the potentials and the associated scalp current densities (SCDs). Scalp potential maps were generated using two-dimensional spherical spline interpolation and radial projection from T3 or T4 (left and right lateral views, respectively), which respects the length of the meridian arcs. SCDs were obtained by computing the second spatial derivative of the spline functions used in interpolation (Perrin *et al.*, 1987, 1989). SCDs do not depend on any assumption about the brain generators or the properties of deeper media, and they are reference free. In addition, SCDs reduce the spatial smearing of the potential fields because of the volume conduction of the different anatomical structures, and thus enhance the contribution of local intracranial sources (Pernier *et al.*, 1988).

Topographic analysis was complemented by spatiotemporal source modelling (Scherg & Von Cramon, 1985, 1986; Giard *et al.*, 1994) based on a three-concentric sphere head model for conductive volumes (brain, skull and scalp) and equivalent current dipoles (ECDs) for generators (local activity of brain regions). Data were modelled using two stationary dipoles with symmetrical positions (one in each hemisphere). The dipole parameters were determined by a

nonlinear iterative procedure (Marquardt minimization method) for the spatial parameters (location and orientation) and with a linear least-mean square algorithm for the time-varying magnitude (Scherg, 1990). The model adequacy was assessed by a goodness-of-fit criterion based on the percentage of experimental variance explained by the model. Note that the modelling procedure was not used here to localize the brain generators involved in the auditory response and/or the cross-modal interactions, but rather to test whether the dipole configuration best explaining the [AV − (A + V)] interactions could also explain most of the auditory response.

Results

Behavioural results

Subjects ($n = 16$) identified the target syllables more rapidly when presented in the audiovisual condition (mean response time 400 ms) than in the auditory-alone condition (423 ms, $F_{1,15} = 18.76$, $P < 0.001$). The error rate was less than 1% in each of the two conditions.

According to the race models (Raab, 1962), a shorter reaction time (RT) in bimodal condition (known as the redundant-stimulus effect) does not necessarily imply the existence of crossmodal interactions before the response, as the first of the two unimodal processes completed could have determined the reaction time. Miller (1982) has shown that under this last hypothesis, particular assumptions can be made on the distribution of RTs:

$$P(\mathrm{RT_{AV}} < T) = P(\mathrm{RT_A} < T) + P(\mathrm{RT_V} < T), \qquad (1)$$

for any reaction time T, where $P(\mathrm{RT} < T)$ is the cumulative probability density function (CDF) of RT.

To test this hypothesis on the speech material used in the ERP study, we performed an additional behavioural-only experiment using the same stimuli and paradigm, except that the subjects ($n = 13$) had to respond to the target syllables in the three (A, V, AV) modalities. The mean RTs to identify the auditory, visual and audiovisual stimuli were 418 ms, 496 ms and 356 ms, respectively (Fig. 2A). Following the procedure proposed by Ratcliff (1979) (see also Miller, 1982), the CDFs of RTs for each subject in each of the three conditions were divided into 19 fractiles (0.05, 0.10, …, 0.90, 0.95) and RTs were group averaged at each fractile, yielding a group distribution (Fig. 2B). Comparison of the AV-CDF with the sum of the A- and V-CDFs using Student t-tests reached statistical significance ($P < 0.001$) for the nine first fractiles (0.05 0.45) showing that inequality (1) was violated for shorter RTs.

Electrophysiological results

Figure 3 presents the ERPs elicited by nontarget unimodal and bimodal stimuli from 150 ms before time zero (onset of the auditory signal) up to 300 ms after, at a subset of electrodes (and corresponding SCDs at Cz). The unimodal A and V waveforms display morphologies typical of activities in sensory-specific areas: the auditory N1 wave was maximum at 136 ms at fronto-central sites (−8.34 μV at Cz) with a small polarity reversal at mastoid electrodes (Ma1, 0.35 μV at 111 ms; Ma2, 0.01 μV at 108 ms). This spatiotemporal configuration is known to reflect neural activity in the supra-temporal auditory cortex (Vaughan & Ritter, 1970; Scherg & Von Cramon, 1986). Auditory N1 was followed by the P2 wave peaking at 221 ms (3.97 μV at Cz) with polarity reversals at mastoids (Ma1, −2.86 μV at 205 ms and Ma2, −2.25 μV at 207 ms).

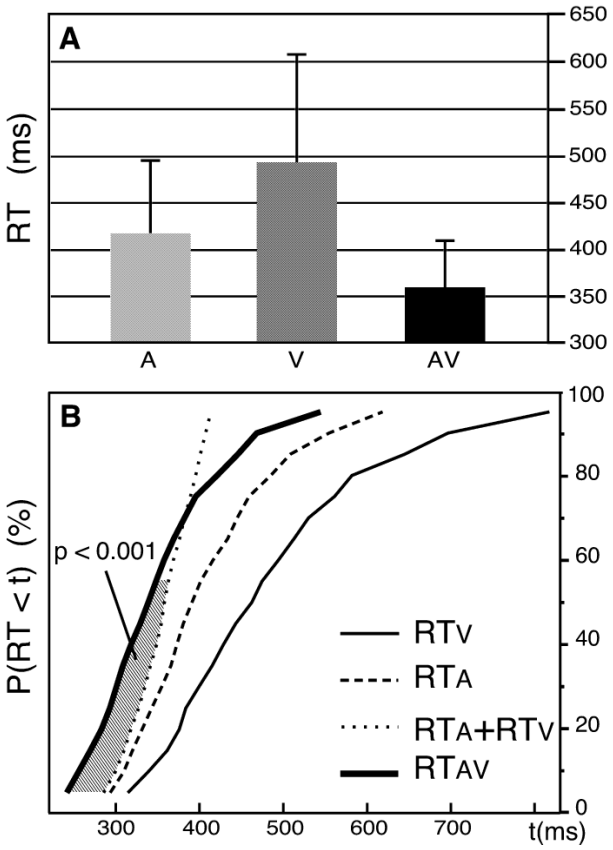

FIG. 2. Violation of the race model inequality in the behavioural-only experiment. (A) Mean reaction times for the auditory, visual and audiovisual trials. (B) Cumulative probability density functions (CDFs) of the reaction times in the three (A, V and AV) conditions of presentation, pooled across subjects. The stimuli and procedure were similar to those used in the main experiment, except that subjects responded to the targets in the three conditions. For shorter reaction times, the CDF for AV responses (thick line) is above the sum of the A and V CDFs (thin dotted line). The hatched area between these two curves illustrates the fractiles for which the violation of the race model inequality [$P(\mathrm{RT_{AV}} < t) \leq P(\mathrm{RT_A} < t) + P(\mathrm{RT_V} < t)$] is statistically significant ($P < 0.001$).

The first deflection in visual ERPs peaked around 40 ms at occipito-parietal electrodes (−3.04 μV at PO3 and −3.60 μV at PO4). Although the onset of the visual stimulus began 240 ms before time zero, this wave is likely to correspond to the visual 'N1' wave, usually peaking around 180 ms post-stimulus. Indeed typical visual N1 responses are usually obtained with stimuli characterized by steep visual energy changes. In our paradigm, the first lip movements were very faint with small progressive changes every 40 ms (see Fig. 1). Therefore the global ERP signal must have also developed progressively, by successive overlaps of small visual responses to each frame delayed by 40, 80, 120 ms, until reaching a 'ceiling' level that appeared about 280 ms after the onset of the first frame. In addition, because we used 12 different visual stimuli (three exemplars of four syllables), the variability of the responses averaged across these stimuli may have reinforced the apparent smoothness of the visual ERP.

A second negative visual component was elicited by lip movements with maximum amplitudes at parieto-central electrodes at about 160 ms (C3, −2.46 μV at 143 ms; C4, −2.16 μV at 179 ms).

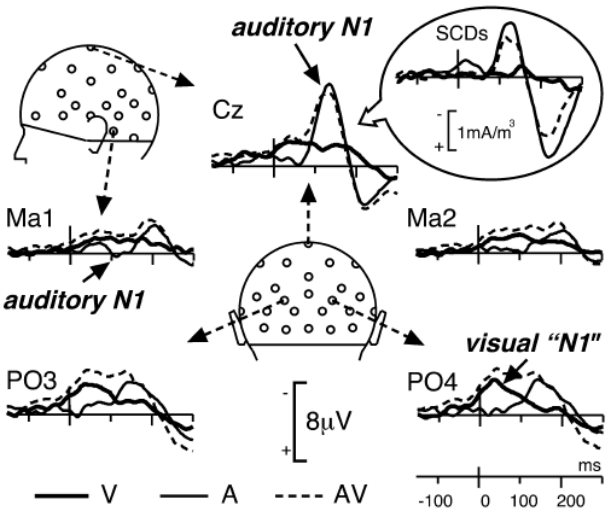

FIG. 3. Unimodal and bimodal responses grand-average ERPs at five illustrative electrodes in each of the three conditions of presentation (A, V and AV) from 150 ms before time zero to 300 ms after. The unimodal auditory N1 wave peaks at 136 ms post-stimulus around Cz with small polarity reversals at mastoid sites (Ma1 and Ma2). The visual 'N1' wave is maximum around occipito-parietal electrodes (PO3 and PO4) at about 40 ms after time zero (this short latency is due to the fact that lip movements began before time zero). Insert: grand-average SCDs at Cz are presented to illustrate the difficulty of interpreting interaction effects locally.

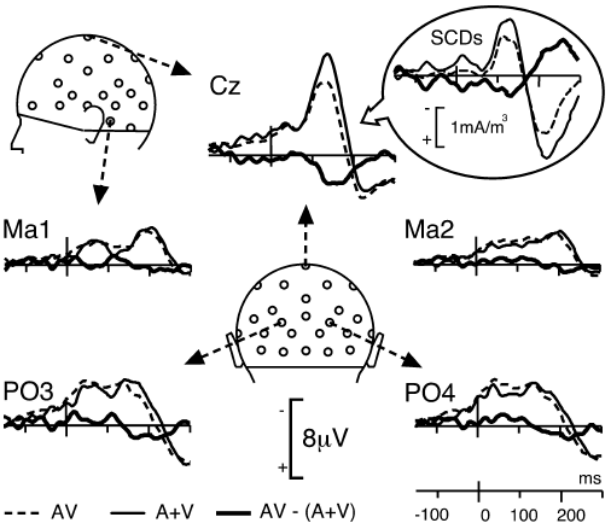

FIG. 4. Bimodal vs. sum of unimodal responses. Comparison of the response to bimodal AV stimuli (dotted lines) with the sum (A + V) of the unimodal responses (thin lines) at five illustrative electrodes, from −150 to +300 ms. The AV response closely follows the A + V trace, except at central sites (illustrated here at Cz) where the two traces significantly differ from about 120 to 190 ms after time zero (see Fig. 5). Insert: for homogeneity with Fig. 3, grand-average SCDs are also presented at this electrode.

Figure 4 displays the superimposition of the ERPs to nontarget bimodal stimuli and the algebraic sum of the responses to unimodal stimuli. Although the morphology of the bimodal response resembles the sum of unimodal ERPs, the differences between the two traces

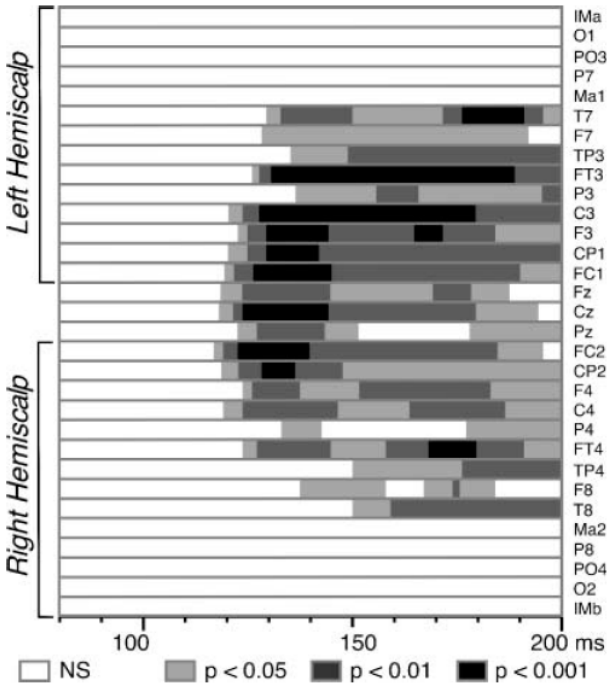

FIG. 5. Statistical significance of the auditory visual interactions. Results of the Student's t-tests (n = 15 subjects) comparing the [AV − (A + V)] amplitudes to zero at each latency from 80 to 200 ms after time zero. Electrodes at the centre of the figure correspond to frontal and central sites and those at the extrema (top and bottom) to more lateral sites. Significant interactions start around 120 ms over fronto-central areas with stronger effects ($P < 0.001$) on the left hemiscalp.

were highly significant over a wide central region within the first 200 ms after time zero. Using the additive criterion [AV − (A + V)] to estimate the interactions, significant patterns were found bilaterally from about 120 to 190 ms at most fronto-central electrodes, that is, in the spatiotemporal range corresponding to the auditory N1 response and the second visual component − {mean amplitude of [AV − (A + V)] over Fz, FC1, FC2, F3, F4, Cz, C3, C4, CP1 and CP2 between 120 and 190 ms: 2.23 (μV)}. The detailed statistical significance of the effect is depicted in Fig. 5. The topography of the cross-modal effect remains roughly stable over the whole 120–190 ms time interval and significance reached the 0.001 threshold at several fronto-central electrodes of the left hemiscalp between 125 and 145 ms latency.

As evidenced by Giard & Peronnet (1999), AV interactions can take multiple forms that are not mutually exclusive: (i) new components that are not present in the unimodal responses; (ii) modulation of the visual response; and (iii), modulation of the auditory response. To assess the nature of the interactions, we therefore compared the topography of [AV − (A + V)] with those of the unimodal responses at the corresponding latencies, with the following reasoning: if the interaction pattern has the same (or inverse) topography as either unimodal response, it is likely to express a modulation (increase or decrease) of that unimodal response. {Note. As ERPs recorded at the scalp surface result from volume conduction activities, it is fundamental to interpret the data from the global topography of the electrical fields and not from a local analysis at one particular electrode. For example, in Fig. 3, the peak amplitudes of the potentials at Cz are similar for A and AV responses, and one could argue that an effect in

[AV − (A + V)] at that electrode could stem only from the (non null) V signal. However, the corresponding SCD traces (Fig. 3) show a different response pattern at Cz in which (i) A and AV traces clearly differ around their peak latency, and (ii) the V signal is close to zero. As the differences between voltage and SCD signals are mainly due to the reduction of the volume conduction effects from distant generators in SCDs, this example illustrates the difficulty of interpreting local measures (potential or SCD) and the need to take the global topography of responses into account.}

Figure 6 displays the topography of the auditory and visual responses, the bimodal responses, the sum of the unimodal responses and the [AV − (A + V)] pattern at the latency of the auditory N1 peak (136 ms). As can be seen, the distributions of the interaction pattern over the left and right hemiscalps (top and bottom panels, respectively) strongly resemble those of the auditory response, but with opposite polarities. This similarity appears not only in the potential maps (Fig. 6, first and third rows), but also in the SCD distributions (Fig. 6, second and fourth rows). Indeed, on the left hemiscalp, the SCD map of [AV − (A + V)] displays sharp, positive current sources at C3, Cz and Pz and a negative current sink around Ma1-T5 (Fig. 6, row 2, col. 5). A similar current sink/source pattern with opposite signs can be observed in the auditory N1 map (Fig. 6, row 2, col. 1). On the right hemiscalp, the polarity reversals at temporal sites are less clear, but the configurations are again very similar in the A and [AV − (A + V)] patterns (Fig. 6, row 4, col. 1 and 5).

By contrast, the latency range of the interaction effects (Fig. 4, Cz) also overlaps that of the second component of the visual response (Fig. 3, PO3/PO4/Cz), suggesting that this component could be modulated by the bimodal inputs. This hypothesis would be supported if the topography of the interactions mirrored that of the unimodal visual response. While the SCD distribution of [AV − (A + V)] could also include part of the topography of the visual reponse around the same latency (particularly over central areas), the overall SCD distributions of the interactions are more complex and differ over occipito-temporal scalp sites.

Finally, we modelled the grand-average [AV − (A + V)] signal in the 110–150 ms period (around the peak of the auditory N1 wave), using two symmetrical (one in each hemisphere) ECDs. The best fitting ECDs were found at an eccentricity of 0.37 and explained the experimental data with a goodness-of-fit of 95.1%. When applied to the same time interval of the auditory response, these ECDs explained 92.3% of the variance of the data whereas they explained only 29.3% of the unimodal visual response within the same latency window. (As a comparison, applied to a 40-ms window in the baseline period, the goodness-of-fit was 39.3%).

Hence, although the pattern of audiovisual interactions observed here may include some contributions from central processes activated by visual stimuli alone, it appears to originate for its main part from the same sources as the auditory N1 response, and may therefore reflect a decrease of activity in the N1 generators in the auditory

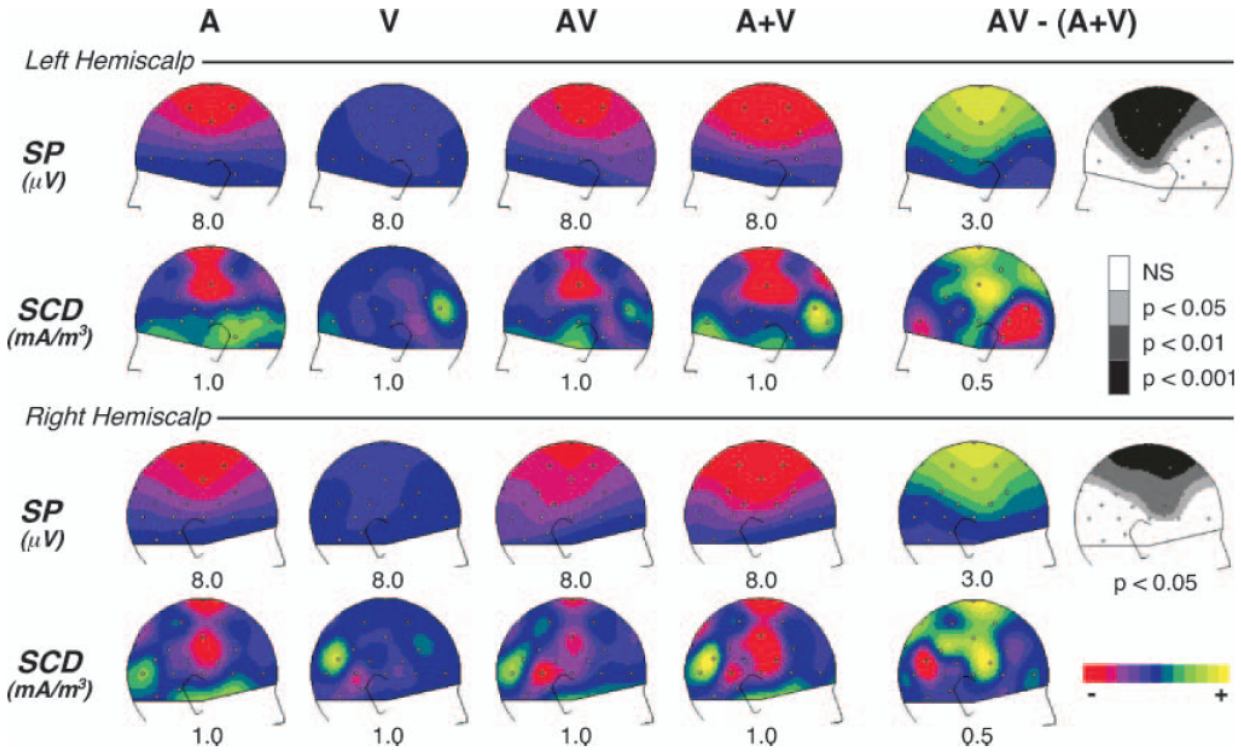

FIG. 6. Comparison of the AV interactions with the auditory N1 wave. Scalp potential (SP) and current density (SCD) topographies over the left and right hemiscalps, at the latency of the unimodal auditory N1 wave (136 ms). In each row the left part displays the distributions of the auditory (A), visual (V), bimodal (AV) responses and the sum of auditory and visual (A + V) responses; the right part shows the distributions of the [AV − (A + V)] interaction pattern with the associated Student's *t*-map estimated on potential values at the same latency (136 ms). The grey colours in *t*-maps indicate the scalp areas where [AV − (A + V)] significantly differs from zero. In potential and SCD maps, half the range of the scale (in μV or mA/m^3) is given below each map. The topography of the crossmodal interaction pattern is similar to that of the unimodal auditory N1 wave, but with opposite polarities. This interaction could therefore reflect a decrease of the unimodal N1 response in auditory cortex.

cortex for the bimodal response compared to the unimodal auditory response.

Furthermore, as we have noted, the SCD maps to auditory alone stimuli display two additional currents sinks on parietal (around Pz) and more anterior (Cz) midline, that both resolve in current sources in the [AV − (A + V)] maps. While a precise interpretation of these current patterns is difficult, the topography of the parietal currents fits with the findings of auditory responses in the posterior intraparietal sulcus (Schroeder *et al.*, 2004). By contrast, the anterior midline currents at Cz could correspond to the 'frontal component' of the auditory N1 response described by Giard's group (Alcaini *et al.*, 1994; Giard *et al.*, 1994). The fact that the same patterns are found with opposite polarities in the auditory-alone condition and in [AV − (A + V)] may indicate that more than the supratemporal component of the auditory response has been modulated by bimodal stimulation.

Discussion

Behavioural facilitation of bimodal speech perception

To date, experimental evidence of behavioural facilitation in bimodal speech perception has been provided almost exclusively on qualitative categorization of phonological continua in situations of sensory conflict (e.g. McGurk illusion: Massaro, 1993) or on detection or intelligibility threshold for degraded unimodal inputs (e.g. speech in noise: Sumby & Pollack, 1954; Grant & Seitz, 2000). Unlike these approaches, our behavioural-only data clearly show that: (i) bimodal information facilitates speech processing also in normal conditions of perception (see also Reisberg *et al.*, 1987; Arnold & Hill, 2001); and (ii), this facilitation can be expressed in chronometric measures, similar to the redundant-stimulus effects widely reported in behavioural studies on cross-modal interactions for nonspeech stimuli (e.g. Hershenson, 1962; Nickerson, 1973; Giard & Peronnet, 1999; Fort *et al.*, 2002a). RT distributions in the three modalities (A, V, AV) falsified the *race* models (Raab, 1962), thereby indicating that unimodal speech inputs interacted somehow during stimulus analysis to speed up the response (coactivation model, Miller, 1982, 1986).

Genuine crossmodal interactions in the ERP paradigm?

The differences in tasks and in the observed RTs for bimodal stimuli preclude a direct application of the previous conclusion to the ERP paradigm on the basis of the sole behavioural measures. It could be argued that the RT effects observed in that paradigm could only result from alertness processes as the visual stimulus started before the auditory stimulus. However, although alerting and spatial orienting of attention are two subcomponents of the attentional system that are probably carried out by separate internal mechanisms (Fernandez-Duque & Posner, 1997), it has been shown that the two processes have similar neural effects on the processing of a subsequent incoming stimulus namely, in the visual modality, an increased activation in extra-striate cortex (Thiel *et al.*, 2004). It is well known in ERP/MEG literature that directing attention to an auditory stimulus results in increased activities in the auditory cortex in a wide latency window including the N1 range (reviews in Näätänen, 1992; Giard *et al.*, 2000). If the auditory processing in the bimodal condition was affected by an alerting process due to the visual signal preceding the acoustic input, the effect would therefore very probably be expressed as a *larger* auditory N1 amplitude for bimodal than for auditory-alone stimuli. Yet we observed a decrease of the auditory N1 amplitude (see next section), strongly suggesting that the alerting hypothesis can be ruled out as a main explanation for the bimodal facilitation in the ERP experiment.

By contrast, several functional imaging studies have reported a decrease of activation in sensory-specific cortices in paradigms where subjects were continuously and exclusively exposed to stimuli in a concurrent modality (Kawashima *et al.*, 1995; Lewis *et al.*, 2000; Bense *et al.*, 2001; Laurienti *et al.*, 2002). Such a cortical deactivation would have led to spurious effects in the [AV − (A + V)] model. However, in the present experiment, all auditory, visual and bimodal stimuli were delivered randomly with equal probability, which should considerably reduce this possibility. Furthermore, because attention was mainly focused on the auditory modality, deactivation processes could have occurred only in the visual cortex (where in fact no [AV − (A + V)] effects were found). Therefore, our significant [AV − (A + V)] effects in temporal areas very probably reflect genuine cross-modal interactions.

Cross-modal depression in the auditory cortex

Both the potential and scalp current density distributions of the crossmodal interactions from about 120 to 150 ms after sound onset mimic those of the unimodal auditory N1 wave in the same latency range. This similarity, also evident in the results of spatio-temporal dipole modelling, strongly suggests that audiovisual integration in speech perception operates at least in part by decreasing the N1 generator activities in supratemporal auditory cortex. This interpretation (that does not preclude the involvement of other additional mechanisms) raises several comments.

First, Miki *et al.* (2004) recently reported no difference in the auditory M100 response (the MEG analogue of the N100 or N1 response) to vowel sounds when they were presented together with the stilled image of a closed mouth or the image of an open mouth pronouncing this vowel. Several reasons may explain these different results: Miki *et al.* used only one stilled image of a mouth pronouncing /a/ in a passive task while we required an auditory discrimination between 12 (three exemplars of four syllables) ecological moving lip movements. First, the use of a passive task prevents from knowing whether their stimuli induced a behavioural facilitation relative to their control condition. In addition, still images and moving speech stimuli may access partly different cortical networks (Calvert & Campbell, 2003). Given that the cross-modal integrative operations are highly sensitive to both the nature of the task and the sensory 'effectiveness' of the unimodal inputs (Fort & Giard, 2004), any of the differences in experimental parameters between the two studies might have explained the differences in results.

Second, the spatial resolution of scalp ERPs does not allow one to rule out the hypothesis that at least part of the interactions are generated in the STS, which has roughly the same orientation as the supratemporal plane. Suppressive effects have indeed been found in the STS in an MEG study comparing the responses to spoken, written and bimodal letters in a recognition task (Raij *et al.*, 2000). However, these effects took place around 380 540 ms and were related to grapheme/phoneme conversion, and do certainly not reflect the same processes as our early interactions occurring at the latency of the auditory sensory N1 response.

Thirdly, congruent bimodal inputs have generally been found to enhance activation in sensory-specific cortices (estimated either by cerebral blood flow measurements: Calvert *et al.*, 1999; Macaluso *et al.*, 2000; or by electric measurements: Giard & Peronnet, 1999; Foxe *et al.*, 2000; Fort *et al.*, 2002b). Yet Giard & Peronnet (1999) reported a decrease of the visual N185 wave (155 200 ms) to

bimodal relative to unimodal visual stimuli in an object discrimination task. This ERP component generated in extra-striate cortex (Mangun, 1995) has been specifically related to visual discrimination processes (Vogel & Luck, 2000). The reduced N185 response was therefore interpreted as reflecting a lesser energetic demand (neural facilitation) from the visual system to discriminate stimuli made more salient by the addition of an auditory cue (see also Fort & Giard, 2004). As the auditory N1 wave is known to be related to stimulus feature analysis in the auditory cortex (Näätänen & Picton, 1987; Näätänen & Winkler, 1999), our results could indicate that lip movements have facilitated feature analysis of the syllables in the auditory cortex by a depression mechanism similar to that found in the visual cortex for object processing. This interpretation makes sense if one considers the general advantage of the cognitive system for visual processing (Posner *et al.*, 1976), and the obvious dominance of the auditory modality in the speech domain: cross-modal facilitation would operate, among other neural mechanisms, as suppressive modulation in the more responsive sensory system.

Finally, a reduced response at the auditory N1 latency appears to be specific to audiovisual speech integration, because this effect differs from those found not only during object recognition (Giard & Peronnet, 1999), but also during the discrimination of verbal material presented in spoken (heard) and written forms (Raij *et al.*, 2000).

Latency of the cross-modal effects

In our paradigm, the first auditory information distinguishing between two different vowels appeared 15 ms after time zero (after the voice onset time; see Material and Methods). The onset latency of the cross-modal effects relative to relevant auditory analysis can therefore be estimated at approximately 105 ms. Several studies on multisensory integration using synchronous nonspeech stimuli have reported very early cross-modal effects (from 40 to 50 ms) in sensory-specific cortices (Giard & Peronnet, 1999; Foxe *et al.*, 2000; Fort *et al.*, 2002a; Molholm *et al.*, 2002). In the speech domain, Lebib *et al.* (2003) recently reported that the processing of congruent and incongruent bimodal inputs generated different ERP effects on the auditory P50 component. Although these and our results might seem hardly compatible with the hypothesis of backprojections from higher-level multisensory areas (Calvert, 2001), one might note that audiovisual speech is special compared with other bimodal objects in that its unimodal inputs are intrinsically asynchronous: coarticulation implies that visual information most often precedes speech sounds, so that visual processing has already started when the sound reaches the auditory system. It is therefore possible that our early effects in the auditory cortex are mediated through visual backprojections from the visual associative system (the visual component peaking bilaterally at occipital sites around 40 ms could well have fed subsequent crossmodal processes in auditory cortex) or from the STS (found to be activated by articulatory lip movements alone and by biological motion in general, review in Calvert & Campbell, 2003). This latter hypothesis fits well with two sets of findings:

At the neural level, although there is growing anatomical and electrophysiological evidence in the primate suggesting that every sensory cortex is likely to receive inputs from each other (from auditory to visual cortex, Falchier *et al.*, 2002; from somatosensory to auditory cortex, Schroeder *et al.*, 2001; Schroeder & Foxe, 2002), no direct pathway from the visual to the auditory cortex has yet been found to our knowledge. However, electrophysiological experiments in monkeys have shown that the associative auditory cortex receives visual inputs with laminar patterns typical of feedback connections, which suggests that visual information is conveyed in the auditory cortex by back-projections from associative areas (Schroeder & Foxe, 2002). The upper bank of the STS (which receives feed-forward auditory and visual information) has been proposed as a candidate for the origin of these visual feedback inputs towards the auditory cortex (see also Pandya *et al.*, 1969; Seltzer & Pandya, 1978).

At a functional level, although we ensured that the subjects could not identify the syllables on the basis of visual information preceding the sound onset, the very first lip movements could have preactivated phonetic units in the auditory cortex via the STS. Several ERP studies have shown that unimodal (e.g. Holcomb & Neville, 1990) and intersensory (e.g. Holcomb & Anderson, 1993) semantic priming effects can decrease the amplitude of the N400 wave, a component associated with late semantic processes. In the same line, the reduced auditory N1 amplitude observed in the present study might reflect an intersensory priming effect on phonetic units at an earlier stage of sensory analysis. Intersensory phonetic priming may therefore be seen as a genuine integrative mechanism by which auditory feed-forward and visual feedback information are combined.

According to Näätänen & Winkler (1999), the auditory N1 component corresponds to a prerepresentational stage of stimulus analysis, during which acoustic features are analysed individually, whereas the first neural correlate of an integrated auditory trace is the MMN, the latency onset of which closely follows that of the N1 wave. Several studies have shown that MMN for speech stimuli is sensitive to phonological (categorical) information (review in Näätänen, 2001). If the MMN is an index of the first phonological trace in the auditory processing chain, then our early cross-modal interactions may reflect online binding of audiovisual information at a prerepresentational stage of stimulus analysis, before the phonological (categorical) trace is built. This chronology of events is in agreement both with the observation of MMN to McGurk syllables (Sams *et al.*, 1991; Colin *et al.*, 2002; Möttönen *et al.*, 2002) and with psycholinguistic models of speech perception by ear and eye (Summerfield, 1987; Massaro & Cohen, 2000).

Abbreviations

A, auditory-only; AV, audiovisual; ECD, equivalent current dipoles; EEG, electroencephalogram; ERP, event-related potential; MEG, neuromagnetic; MMN, mismatch negativity; RT, reaction time; SCD, scalp current density; STS, superior temporal sulcus; V, visual only.

References

Alcaini, M., Giard, M.H., Thevenet, M. & Pernier, J. (1994) Two separate frontal components in the N1 wave of the human auditory evoked response. *Psychophysiology*, **31**, 611–615.

Arnold, P. & Hill, F. (2001) Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *Br. J. Psychol.*, **92**, 339–355.

Barth, D.S.N., Goldberg, B., Brett, B. & Di, S. (1995) The spatiotemporal organization of auditory, visual and auditory-visual evoked potentials in rat cortex. *Brain. Res.*, **678**, 177–190.

Bense, S., Stephan, T., Yousry, T.A., Brandt, T. & Dieterich, M. (2001) Multisensory cortical signal increases and decreases during vestibular galvanic stimulation (fMRI). *J. Neurophysiol.*, **85**, 886–899.

Besle, J., Fort, A. & Giard, M.-H. (2004) Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, **5**, 189–192.

Calvert, G.A. (2001) Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb. Cortex*, **11**, 1110–1123.

Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iversen, S.D. & David, A.S. (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, **10**, 2619–2623.
Calvert, G.A. & Campbell, R. (2003) Reading speech from still and moving faces: The neural substrates of visible speech. *J. Cogn. Neurosci.*, **15**, 57–70.
Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F. & Deltenre, P. (2002) Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clin. Neurophysiol.*, **113**, 495–506.
Falchier, A., Clavagnier, S., Barone, P. & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.*, **22**, 5749–5759.
Fernandez-Duque, D & Posner, M.I. (1997) Relating the mechanisms of orienting and alerting. *Neuropsychologia*, **35**, 477–486.
Fort, A., Delpuech, C., Pernier, J. & Giard, M.H. (2002a) Dynamics of cortico-subcorical crossmodal operations involved in audio-visual object detection in humans. *Cereb. Cortex*, **12**, 1031–1039.
Fort, A., Delpuech, C., Pernier, J. & Giard, M.H. (2002b) Early auditory–visual interactions in human cortex during nonredundant target identification. *Brain Res. Cogn. Brain Res.*, **14**, 20–30.
Fort, A. & Giard, M.-H. (2004) Multiple electrophysiological mechanisms of audio-visual integration in human perception. In Calvert, G., Spence, C. & Stein, B. (eds), *The Handbook of Multisensory Processes*, MIT Press, Cambridge, pp. 503–514.
Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C. & Schroeder, C.E. (2000) Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.*, **10**, 77–83.
Giard, M.-H., Fort, A., Mouchetant-Rostaing, Y. & Pernier, J. (2000) Neurophysiological mechanisms of auditory selective attention in humans. *Front. Biosci.*, **5**, 84–94.
Giard, M.H. & Peronnet, F. (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.*, **11**, 473–490.
Giard, M.H., Perrin, F., Echallier, J.F., Thevenet, M., Froment, J.C. & Pernier, J. (1994) Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalogr. Clin. Neuro.*, **92**, 238–252.
Grant, K.W. & Seitz., P.F. (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.*, **108**, 1197–1208.
Guthrie, D. & Buchwald, J.S. (1991) Significance testing of difference potentials. *Psychophysiology*, **28**, 240–244.
Hershenson, M. (1962) Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.*, **63**, 289–293.
Hillyard, S.A., Teder-Salejarvi, W.A. & Munte, T.F. (1998) Temporal dynamics of early perceptual processing. *Curr. Opin. Neurobiol.*, **8**, 202–210.
Holcomb, P.J. & Anderson, J.E. (1993) Cross-modal semantic priming – a time-course analysis using event-related brain potentials. *Lang. Cogn. Proc.*, **8**, 379–411.
Holcomb, P.J. & Neville, H.J. (1990) Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Proc.*, **5**, 281–312.
Kawashima, R., O'Sullivan, B.T. & Roland, P.E. (1995) Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: Closing the 'mind's eye'. *Proc. Natl. Acad. Sci. USA.*, **92**, 5969–5972.
Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S. & Stein, B.E. (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cogn. Neurosci.*, **14**, 420–429.
Lebib, R., Papo, D., de Bode, S. & Baudonnière, P.M. (2003) Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci. Lett.*, **341**, 185–188.
Lewis, J.W., Beauchamp, M.S. & DeYoe, E.A. (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex*, **10**, 873–888.
Macaluso, E., Frith, C. & Driver, J. (2000) Modulation of human visual cortex by crossmodal spatial attention. *Science*, **289**, 1206–1208.
Mangun, G.R., (1995) Neural mechanisms of visual selective attention. *Psychophysiology*, **32**, 4–18.
Massaro, D.W. (1993) Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Comm.*, **13**, 127–134.
Massaro, D.W. & Cohen, M.M. (2000) Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *J. Acoust. Soc. Am.*, **108**, 784–789.
McGurk, H. & McDonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, 746–748.
Miki, K., Watanabe, S. & Kakigi, R. (2004) Interaction between auditory and visual stimulus relating to the vowel sounds in the auditory cortex in humans: a magnetoencephalographic study. *Neurosci. Lett.*, **357**, 199–202.
Miller, J.O. (1982) Divided attention: Evidence for coactivation with redundant signals. *Cognit. Psychol.*, **14**, 247–279.
Miller, J.O. (1986) Time course of coactivation in bimodal divided attention. *Percept. Psychophys.*, **40**, 331–343.
Miniussi, C., Girelli, M. & Marzi, C.A. (1998) Neural site of the redundant target effect: Electrophysiological evidence. *J. Cogn. Neurosci.*, **10**, 216–230.
Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2002) Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.*, **14**, 115–128.
Möttönen, R., Krause, C.M., Tiippana, K. & Sams, M. (2002) Processing of changes in visual speech in the human auditory cortex. *Brain Res. Cogn. Brain Res.*, **13**, 417–425.
Näätänen, R. (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, **38**, 1–21.
Näätänen, R. (1992) *Attention and Brain Function*, Lawrence Erlbaum Associates, Hillsdale, NJ.
Näätänen, R. & Picton, T.W. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, **24**, 375–425.
Näätänen, R. & Winkler, I. (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.*, **125**, 826–859.
Nickerson, R.S. (1973) Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychol. Rev.*, **80**, 489–509.
Olson, I.R., Gatenby, J.C. & Gore, J.C. (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Brain Res. Cogn. Brain Res.*, **14**, 129–138.
Pandya, D.N., Hallett, M. & Kmukherjee, S.K. (1969) Intra- and interhemispheric connections of the neocortical auditory system in the rhesus monkey. *Brain. Res.*, **14**, 49–65.
Pernier, J., Perrin, F. & Bertrand, O. (1988) Scalp current density fields: Concept and properties. *Electroencephalogr. Clin. Neuro.*, **69**, 385–389.
Perrin, F., Pernier, J., Bertrand, O. & Echallier, J.F. (1989) Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neuro.*, **72**, 184–187.
Perrin, F., Pernier, J., Bertrand, O. & Giard, M.H. (1987) Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr. Clin. Neuro.*, **66**, 75–81.
Posner, M.I., Nissen, M.J. & Klein, R.M. (1976) Visual dominance: An information-processing account of its origins and significance. *Psychol. Rev.*, **83**, 157–171.
Raab, D.H. (1962) Statistical facilitation of simple reaction times. *Trans. NY Acad. Sci.*, **24**, 574–590.
Raij, T., Uutela, K. & Hari, R. (2000) Audiovisual integration of letters in the human brain. *Neuron*, **28**, 617–625.
Ratcliff, R. (1979) Group reaction time distributions and an analysis of distribution statistics. *Psychol. Bull.*, **86**, 446–461.
Reisberg, D.J., McLean, J & Goldfield, A. (1987) Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In Dodd, B. & Campbell, R. (eds), *Hearing by Eye: the Psychology of Lipreading*. Lawrence Erlbaum Associates, London, pp. 93–113.
Sams, M., Aulanko, R., Hamalainen, H., Hari, R., Lounasmaa, O.V., Lu, S.T. & Simola, J. (1991) Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.*, **127**, 141–145.
Scherg, M. (1990) Fundamentals of dipole source potential analysis. In Grandori, F., Hoke, M. & Romani, G.L. (eds), *Auditory Evoked Magnetic Fields and Electric Potentials. Advances in Audiology*, Vol. 5. Karger, Basel, pp. 40–69.
Scherg, M. & Von Cramon, D. (1985) A new interpretation of the generators of BAEP waves I–V: Results of a spatio-temporal dipole model. *Electroencephalogr. Clin. Neuro.*, **62**, 290–299.
Scherg, M. & Von Cramon, D. (1986) Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr. Clin. Neuro.*, **65**, 344–360.
Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F. & Javitt, D.C. (2001) Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.*, **85**, 1322–1327.
Schroeder, C., Molholm, S., Lakatos, P., Ritter, W. & Foxe, J.J. (2004) Human-simian correspondence in the early cortical processing of multisensory cues. *Cognitive Processing*, **5**, 140–151.

Schroeder, CE & Foxe, J.J. (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res. Cogn. Brain Res.*, **14**, 187–198.
Seltzer, B. & Pandya, D.N. (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain. Res.*, **149**, 1–24.
Sumby, W.H. & Pollack, I. (1954) Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.*, **26**, 212–215.
Summerfield, Q. (1987) Some preliminaries to a comprehensive account of audio-visual speech perception. In Dodd, B. & Campbell, R., eds. *Hearing by Eye: the Psychology of Lipreading*. Lawrence Erlbaum Associates, Location, pp. 3–52.
Thiel, C.M., Zilles, K. & Fink, G.R. (2004) Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: An event-related fMRI study. *Neuroimage*, **21**, 318–328.
Vaughan, H.G. & Ritter, W. (1970) The sources of auditory evoked responses recorded from the human scalp. *Electroencephalogr. Clin. Neuro.*, **28**, 360–367.
Vogel, E.K. & Luck, S.J. (2000) The visual N1 component as an index of a discrimination process. *Psychophysiology*, **37**, 190–203.

Exp Brain Res (2005) 166: 337 344
DOI 10.1007/s00221-005-2375-x

RESEARCH ARTICLE

Julien Besle · Alexandra Fort · Marie-Hélène Giard

Is the auditory sensory memory sensitive to visual information?

Received: 4 August 2004 / Accepted: 9 November 2004 / Published online: 23 July 2005
© Springer-Verlag 2005

Abstract The mismatch negativity (MMN) component of auditory event-related brain potentials can be used as a probe to study the representation of sounds in auditory sensory memory (ASM). Yet it has been shown that an auditory MMN can also be elicited by an illusory auditory deviance induced by visual changes. This suggests that some visual information may be encoded in ASM and is accessible to the auditory MMN process. It is not known, however, whether visual information affects ASM representation for any audiovisual event or whether this phenomenon is limited to specific domains in which strong audiovisual illusions occur. To highlight this issue, we have compared the topographies of MMNs elicited by non-speech audiovisual stimuli deviating from audiovisual standards on the visual, the auditory, or both dimensions. Contrary to what occurs with audiovisual illusions, each unimodal deviant elicited sensory-specific MMNs, and the MMN to audiovisual deviants included both sensory components. The visual MMN was, however, different from a genuine visual MMN obtained in a visual-only control oddball paradigm, suggesting that auditory and visual information interacts before the MMN process occurs. Furthermore, the MMN to audiovisual deviants was significantly different from the sum of the two sensory-specific MMNs, showing that the processes of visual and auditory change detection are not completely independent.

J. Besle
Univ. Lyon 2, Lyon, France

J. Besle · A. Fort · M.-H. Giard
Univ. Lyon 1, Lyon, France

J. Besle · A. Fort · M.-H. Giard
IFNL IFR19, Lyon, France

J. Besle (✉) · A. Fort · M.-H. Giard
INSERM U280, Mental Processes and Brain Activation,
69675 Bron Cedex, France
E-mail: besle@lyon.inserm.fr
Tel.: +33-472-138907
Fax: +33-472-138901

Keywords Electrophysiology · Audiovisual · MMN · Multisensory integration · Memory

Introduction

The most counter-intuitive effect of audiovisual interactions in the brain is, perhaps, the fact that sensory-specific cortices (e.g. the auditory cortex) seem to be sensitive to information from other modalities, even in primary cortices (Bental et al. 1968) and at very early stages of sensory processing (Fort and Giard 2004).

The mismatch negativity (MMN) component of event-related potentials (ERPs) is elicited in the auditory cortex when incoming sounds are detected as deviating from a neural representation of acoustic regularities and is computed by subtracting the responses to frequent standard sounds from those to infrequent deviant sounds. MMN implies the existence of an auditory sensory memory (ASM) that stores a neural representation of the standard against which any incoming auditory input is compared (Ritter et al. 1995). It is mainly generated in the auditory cortex (Kropotov et al. 1995; Alain et al. 1998) and has long been regarded as specific to the auditory modality (Nyman et al. 1990; Näätänen 1992).

It has, however, recently been discovered that the MMN is not completely impervious to crossmodal influences. For example, in bimodal speech processing, an MMN has been shown to be elicited by deviant syllables differing from the standards only on their visual dimension. In this so-called McGurk illusion (McGurk and McDonald 1976), the same physical sound is therefore differently perceived and processed in ASM, depending on the lip movements that are simultaneously seen (Sams et al. 1991; Möttönen et al. 2002; Colin et al. 2002b, 2004). To keep in line with the auditory-specificity assumption, several non-exclusive explanations have been proposed, that are related to the special status of speech. Either there would exist a phonetic MMN

process that is sensitive to the phonetic nature of articulatory movements (Colin et al. 2002b) or visual speech cues could have specific access to the MMN generators in auditory cortex because, like auditory speech, they carry time-varying information (Möttönen et al. 2002).

Nonetheless, generation of an MMN by visual-only deviants is not restricted to the speech domain, because it can also be observed with the ventriloquist illusion, in which the perceived location of a sound is shifted by a spatially disparate visual stimulus (Stekelenburg et al. 2004; see also Colin et al. 2002a). Rather, what these two phenomena have in common is that they give rise to irrepressible audiovisual illusions that seem to occur at a sensory level of representation (McGurk effect: Soto-Faraco et al. 2004; ventriloquist effect: Bertelson and Aschersleben 1998; Vroomen et al. 2001).

The question therefore arises whether any visual change of an audiovisual event, even in the absence of perceived audiovisual illusion, is likely to access the ASM indexed by the MMN. In other words, does the ASM encode more than the auditory part of an audiovisual event?

When replacing articulatory lip-movements by non-speech visual stimuli in a McGurk MMN paradigm, Sams et al. (1991) found no evidence of an auditory MMN elicited by visual variations alone of the audiovisual event. However, it is very possible that in the absence of strong illusion, the effect is of much less amplitude. Moreover, the effect of visual deviance on the MMN process could occur only in a suitable auditory-deviance context: thus the MMN elicited by both auditory and visual deviances of an audiovisual event should be different from the MMN elicited by auditory deviances alone, while a visual deviance alone would not be detected by the auditory system.

We therefore conducted an audiovisual oddball paradigm in which audiovisual deviants differed from audiovisual standards (AV) either on the visual dimension (AV′), on the auditory dimension (A′V) or on both dimensions (A′V′), with the following hypothesis: If visual information is represented in ASM, AV′ deviants should elicit an auditory MMN, or A′V′ deviants should at least elicit an MMN different from those elicited by A′V deviants.

This would be the whole story if there were not a spoilsport: visual mismatch negativity (vMMN). Several studies have recently shown that visual stimuli deviating from repetitive visual standards can elicit a visual analogue of the MMN in the same latency range (review in Pazo-Alvarez et al. 2003). This vMMN seems to be mainly generated in occipital areas (Berti and Schroger 2004) with possibly a more anterior component (Heslenfeld 2003; Czigler et al. 2002), to be independent of attention (Heslenfeld 2003), and to rely also on memory processes (Stagg et al. 2004; Czigler et al. 2002; see however Kenemans et al. 2003). However, it seems that a greater amount of deviance is necessary to evoke a vMMN than an auditory MMN (Pazo-Alvarez et al. 2003).

If visual-specific components are evoked by visual deviances, then it is necessary in our audiovisual paradigm to separate them from the effect of visual information on the auditory-specific MMN process. To disentangle the contributions of each unisensory process (vMMN and auditory MMN) and isolate the effect of visual information on the auditory MMN, we have therefore:

1. conducted an additional visual oddball paradigm[1]. using the same visual inputs as in our main experiment, so as to elicit a genuine visual MMN (V′MMN); and
2. analyzed the voltage and scalp current density (SCD) distributions of that V′MMN relative to the AV′ MMN elicited in the audiovisual paradigm.

If, on the other hand, the two unisensory MMN processes do not somehow interact, then the two unisensory MMNs should be strictly additive.

Methods

Participants

Fifteen right-handed adults (eight female, ages 20 25 years, mean age 23.1 years) were paid to participate in the study, for which they gave a written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects were free of neurological disease, had normal hearing, and normal or corrected-to-normal vision.

Stimuli

The stimuli were inspired from those previously used by our group in various experiments which revealed a variety of crossmodal interactions in the first 200 ms of processing (Giard and Peronnet 1999; Fort et al. 2002a, b; Fort and Giard 2004).

Visual stimuli consisted in the deformation of a circle into an ellipse either in the horizontal or in the vertical direction (Giard and Peronnet 1999). The basic circle had a diameter of 4.55 cm and was displayed on a video screen placed 130 cm in front of the subjects' eyes, subtending a visual angle of 2°. The amount of deformation in either direction relative to the diameter of the circle was 33% and lasted 140 ms. Between each deformation, the circle remained present on the screen; a cross at its centre served as the fixation point.

Auditory stimuli were rich tones (the fundamental and the second and the fourth harmonics) shifting

[1]As, on the one hand, the topography of the auditory MMN is well known and on the other hand, it would have needlessly lengthened the recording session, we chose not to conduct an auditory oddball paradigm

linearly in frequency (fundamental) either from 500 to 540 Hz or from 500 to 600 Hz. Their duration was also 140 ms, including 14 ms rise/fall time.

All stimuli consisted in the synchronous presentation of a visual and an auditory feature. One association (e.g. an elongation in the horizontal direction and a frequency shift from 500 to 540 Hz) was delivered in 76% of the trials (AV standard). Each remaining association was presented in 8% of the trials: the A′V deviant had the same visual feature as the standard but a different auditory feature, the AV′ deviant had the same auditory but a different visual feature, and the A′V′ deviant differed from the AV standard on both dimensions.

To ensure that the MMNs obtained could not be attributed to physical differences between the standard and deviants, the features of the standard and the A′V′ deviant were exchanged in half of the experimental blocks (and so were the features of A′V and AV′ deviants)

Distractive task

An important characteristic of the auditory (and visual) MMN is that it is automatic and pre-attentive. A "pure" MMN (that is not contaminated by attentional processes) is elicited by stimuli that are irrelevant to the subject. This is a more difficult constraint for visual or audiovisual oddball paradigms, because visual stimuli have to be presented in the visual field of the subjects, but outside their attentional focus. We therefore required a task on the fixation cross. From time to time (13% of the trials) the fixation cross disappeared for 120 ms. This disappearance occurred unpredictably within a standard trial but it was desynchronised relative to the trial's onset and could not occur in a standard preceding a deviant trial. Subjects had to stare at the fixation cross and click a button as quickly and accurately as possible when the cross disappeared.

Visual control

To control for the existence of a vMMN and to study its topography, a visual oddball paradigm was conducted. The experimental parameters were those used in the audiovisual paradigm with the sound off: standard visual stimuli occurred in 84% of the trials (V standards) and the other visual feature occurred in 16% of the trials (V′ deviants). Standards and deviants were exchanged in half of the experimental blocks.

Procedure

After setting the ERP recording apparatus, subjects were seated in a dark, sound-attenuating room and were given instructions describing the distractive task along with an audiovisual practice block of 267 trials. They were told to stare at the fixation cross at the centre of the screen, to respond as accurately and as quickly as possible to the cross disappearance and not to pay attention to the circle or the tones. In the audiovisual paradigm, 256 deviant trials of each type (AV′, A′V and A′V′) were randomly delivered among 2432 AV-standard trials, over 12 blocks including 267 trials each (except the last block that included 263 trials) at a fixed ISI of 560 ms, with the constraint that two deviants could not occur in a row. In the visual oddball paradigm, 256 deviant trials were randomly delivered among 1344 V-standard trials, over 6 blocks including 267 trials each, except the last block that included 265 trials (same ISI). The 12 audiovisual and the six visual blocks were randomly presented to the subjects.

EEG recording

EEG was continuously recorded via a Neuroscan Compumedics system through Synamps AC coupled amplifiers (0.1 200 Hz analogue bandwidth; sampling rate: 1 kHz) from 36 Ag to AgCl scalp electrodes referred to the nose and placed according to the International 10 20 System: Fz, Cz, Pz, POz, Iz; Fp1, F3, F7, FT3, FC1, T3, C3, TP3, CP1, T7, P3, P7, PO3, O1, and their counterparts on the right hemi scalp; Ma1 and Ma2 (left and right mastoids, respectively); IMa and IMb (midway between Iz-Ma1 and Iz-Ma2, respectively). Electrode impedances were kept below 5 kΩ. Horizontal eye movements were recorded from the outer canthus of the right eye; eye blinks and vertical eye movements were measured in channels Fp1 and Fp2.

Data analysis

The EEG analysis was undertaken with the Elan Pack software developed at the INSERM U280 laboratory (Lyon, France). Trials with signal amplitudes exceeding 100 μV at any electrode from 300 ms before time 0 to 500 ms after were automatically rejected to discard the responses contaminated by eye movements or muscular activities.

The ERPs to audiovisual stimuli were averaged offline separately for the six different stimulus types (AV and V standards, A′V, AV′, A′V′ and V′ deviants), over a time period of 800 ms including 300 ms pre-stimulus. Trials including disappearance of the fixation cross were not taken into account when averaging. The mean numbers of averaged trials (by subject) were 1299, 649, and 204 for AV-standard, V-standard, and deviants of each type, respectively (about 20.4% of the trials were discarded because of eye movements).

The ERPs were finally digitally filtered (bandwidth: 1.5 30 Hz, slope: 24 dB/octave). The mean amplitude over the [−100 to 0 ms] pre-stimulus period was taken as the baseline for all amplitude measurements.

340

Topographic analysis

To facilitate interpretation of the voltage values recorded at multiple electrodes over the scalp surface, we analyzed the topographic distributions of the potentials and the associated scalp current densities. Scalp potential maps were generated using two-dimensional spherical spline interpolation and radial projection from T3, T4 or Oz (left, right and back views, respectively), which respects the length of the meridian arcs. The SCDs were obtained by computing the second spatial derivative of the spline functions used in interpolation (Perrin et al. 1987, 1989). The SCDs do not depend on any assumption about the brain generators or the properties of deeper media, and they are reference-free. In addition, SCDs reduce the spatial smearing of the potential fields due to the volume conduction of the different anatomical structures, and thus enhance the contribution of local intracranial sources (Pernier et al. 1988).

Statistical analysis

The MMNs were statistically assessed by *t*-tests comparing the averaged amplitude of the deviant minus standard difference waveform to zero in the 40 ms time-window around the latency of the peak in the grand-average responses. Results are displayed as statistical probability maps associated with the *t*-tests at each electrode.

Results

Behavioral measures

Mean reaction time to respond to the disappearance of the fixation cross was 404 ms (SD = 51 ms) in the audiovisual oddball paradigm and 409 ms (SD = 52 ms) in the visual paradigm. The mean error ratios were respectively 3.51% (SD = 3.13%) and 3.24 % (SD = 3.11%). Neither the reaction times nor the error rates significantly differed between the two paradigms.

A′V MMN

The response to A′V deviants in the audiovisual paradigm began to differ from AV standards at about 120 ms of processing, being more negative at Fz, and more positive at mastoid sites (Ma1 and Ma2) until about 250 ms (Fig. 1). As can be seen from the difference waveforms (Fig. 2), the MMN elicited by A′V deviants was maximum at 199 ms at Fz (−3.75 μV) and at 190 ms around the mastoids sites (2.77 μV at Ma1 and 2.14 μV at Imb).

Scalp potential and current density topographies (Fig. 3. upper left panel) display a clear-cut polarity reversal around the supra-temporal plane, as expected from an auditory MMN generated in the auditory cortex (Giard et al. 1990). Student *t*-tests on the MMN amplitude around its peak latency (199 ms) are significant at most electrodes around the reversal plane.

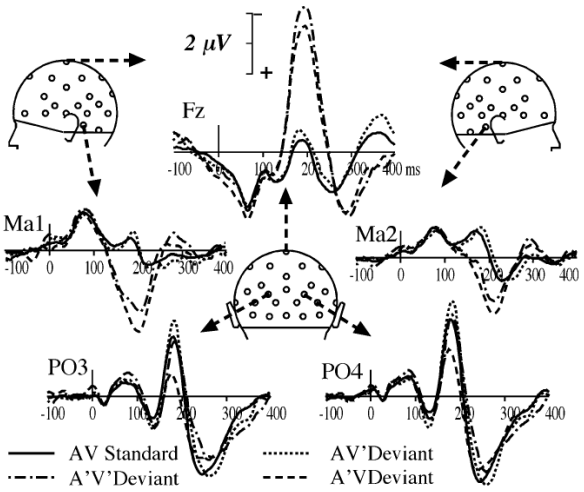

Fig. 1 Potential waveforms elicited by AV standards, AV′, A′V, and A′V′ deviants at a subset of five electrodes (Fz, Ma1, Ma2, PO3, and PO4) from 100 ms pre-stimulus to 400 ms post-stimulus. Negative values are plotted upwards

AV′ MMN

Responses to AV′ deviants and to AV standards are hardly different (Fig. 1). However, the deviant minus standard difference curves (Fig. 2) reveal an occipital deflection that peaks bilaterally at a latency of 192 ms (−0.87 μV at PO4 and −0.63 μV at PO3), with a second peak around 215 ms (−0.72 μV at PO3 and PO4). Figure 3 (upper right panel) illustrates the bilateral occipital

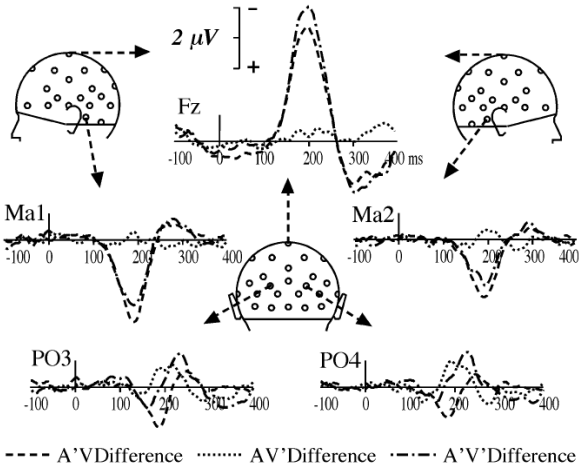

Fig. 2 Deviant minus AV standard difference waveforms for each AV′, A′V, and A′V′ deviant type in the audiovisual oddball paradigm at the same subset of electrodes as in Fig. 1

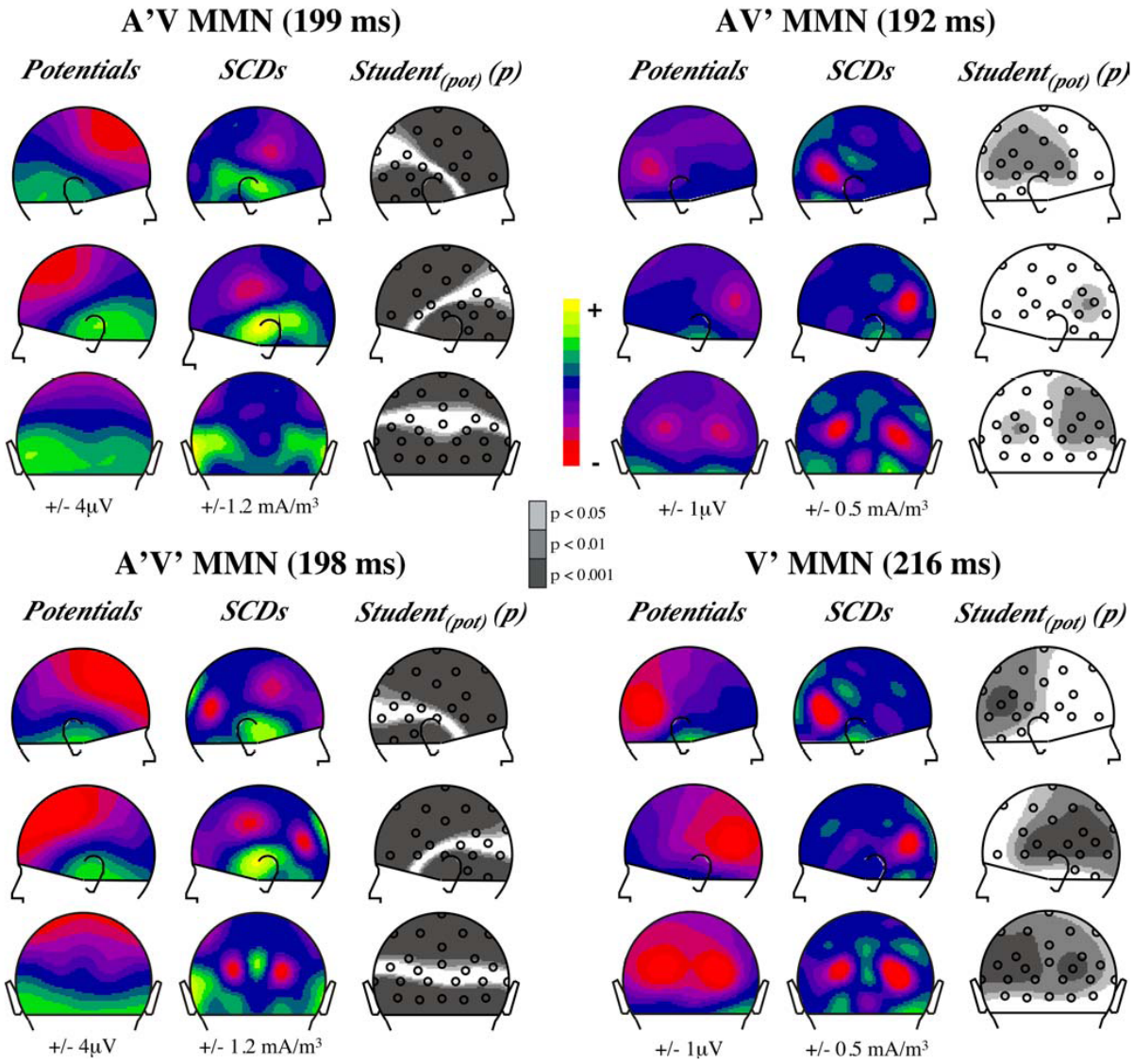

Fig. 3 Topographies of the MMNs elicited by each deviant type in the audiovisual oddball paradigm and by the visual deviant in the visual oddball paradigm (*lower right panel*). Scalp potentials (1st column of each panel), scalp current densities (2nd column) and probability maps associated to Student *t*-tests (3rd column) are presented in right, left, and back views at the latency of the MMN peak indicated above each panel. In potential and SCD maps half the range of the colour scale is given below each column. In Student *t*-maps *grey* areas include electrodes where the averaged potential in a 40 ms time-window around the indicated latency significantly differs from zero

topography of the AV′ MMN at the latency of its largest peak and the statistical significance of its amplitude on the scalp around its peak latency.

A′V′ MMN

Although the responses elicited by A′V′ deviants most resemble those elicited by A′V deviants at fronto-central and mastoid sites (Figs. 1, 2), they tend to come near the curves elicited by AV′ deviants at occipital sites (Fig. 2). Figure 3 (lower left panel) displays the SCD distribution of A′V′ MMN at the latency of its peak in ERPs (198 ms). It clearly shows that it consists of the SCD patterns observed in both the auditory MMN component and the component elicited by AV′ deviants at occipital sites.

Additivity of the MMNs

The additivity of the MMNs elicited by each deviant type was tested by Student *t*-tests comparing the

A′V′ MMN to the sum of the two AV′ MMN and A′V MMN, averaged in the 178 to 218 ms latency window (around the peak latency of both the A′V and the A′V′ MMNs at Fz). Figure 4A shows that additivity is violated at several left parieto-temporal electrodes.

Visual oddball paradigm

Figure 5 displays the ERPs elicited by V standards and V′ deviants in the visual oddball paradigm, and the deviant minus standard difference curve. As in the audiovisual paradigm, the V′ deviants elicited a bilateral occipital component that peaked at a latency of 215 ms (that is at the latency of the second peak of the AV′ MMN), with a larger amplitude (−1.19 μV at PO3 and −1.21 μV at PO4). Figure 3 (lower right panel) displays the topography of the vMMN and the brain areas where its amplitude is statistically significant. There was no hint of an anterior component as was found in other studies investigating the visual MMN.

We further compared the vMMN elicited in the visual paradigm (V′ deviants) and the audiovisual paradigm (AV′ deviants) with Student *t*-tests in the 40 ms time-window around the peak latency that was common to both vMMNs (216 ms). As shown in Fig. 4B, tests were significant at several electrodes over the left hemi scalp.

Discussion

Auditory deviance of an audiovisual event elicited a classical MMN with topography typical of activities in the auditory cortex. Visual deviance of an audiovisual object elicited a bilateral occipital component in the same latency range as the auditory MMN. This component was very similar to that found in the visual oddball paradigm. The spatio-temporal characteristics of these two components are consistent with previous reports of an analogue of the MMN in the visual modality (vMMN) and, especially, with the study by Berti and Schroger (2004) who reported a vMMN with a bilateral occipital topography at a latency of 240 ms.

Thus in our data, the visual variation of an audiovisual event does not seem to elicit an auditory MMN, unlike what has been observed in McGurk and ventriloquist illusions (e.g. Möttönen et al. 2002; Stekelenburg et al. 2004), suggesting that real or illusory perception of an auditory change is necessary to elicit a clear MMN response in the auditory cortex.

In addition, we found that the MMN to deviance on both the auditory and visual dimensions of a bimodal event includes both supratemporal and occipital components, suggesting that the deviance detection processes operate separately in each modality.

However, the vMMNs elicited in the visual (V′ MMN) and the audiovisual (AV′ MMN) oddball paradigms were found to significantly differ, while the only difference between these paradigms was the presence or absence of the same sound that was constantly associated with the visual standards and deviants. Two mutually non-exclusive explanations could account for this finding:

1. either an auditory MMN of small amplitude induced by visual change of the audiovisual event by the same phenomenon as in audiovisual illusions is superimposed to the vMMN and alters its topography;

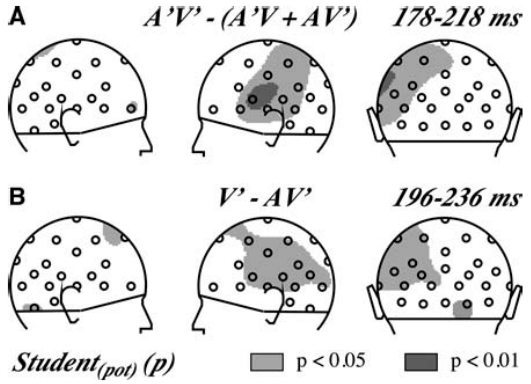

Fig. 4 **A** Test of the additivity of the three MMNs elicited in the audiovisual oddball paradigm presented as probability maps associated to Student *t*-tests in right, left, and back views. *Grey* areas include electrodes where the averaged potential in the indicated time-window significantly differs between the MMN elicited by A′V′ deviants and the sum of the MMNs elicited by A′V and AV′ deviants. **B** Comparison of the MMNs elicited in the audiovisual and the visual oddball paradigm presented as probability maps associated to Student *t*-tests in right, left, and back views. *Grey* areas include electrodes where the averaged potential in the indicated time-window significantly differs between the AV′ MMN and V′ MMN

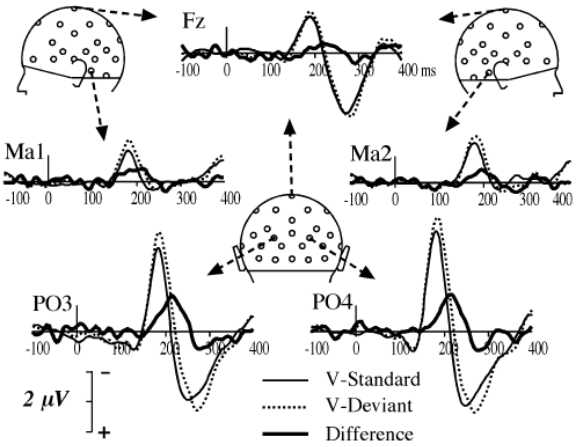

Fig. 5 ERPs elicited by V standards and V′ deviants and deviant minus standard difference waveforms at a subset of five electrodes (Fz, Ma1, Ma2, PO3 and PO4) from 100 ms pre-stimulus to 400 ms post-stimulus

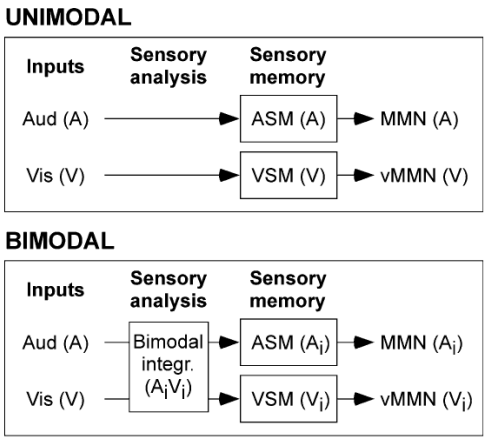

Fig. 6 A schematic model of the MMN processes for unimodal and bimodal inputs. In unimodal parameters, the model refers to that proposed by Näätänen (1992) for auditory MMN. When auditory (*A*) and visual (*V*) inputs are synchronously presented, crossmodal interactions underlying the construction of a multimodal percept can start at early stages of analysis in the afferent sensory systems and modify the input signals before they are encoded in the respective sensory memories (ASM and VSM). The auditory and visual MMN processes would operate on these new sensory signals (A_i and V_i). Note that, although the stage of crossmodal interactions is outlined exclusively before the MMN processes in this figure for simplification, these interactions may persist for several hundred of milliseconds

2. or, if the vMMN reflects a memory-based process (Czigler et al. 2002), an audiovisual event would be encoded differently from a visual-only event in that memory.

The first possibility is hardly supported by the topographies displayed in Fig. 3, and the scalp distribution of the significant differences between the V′ MMN and the AV′ MMN is difficult to interpret regarding either hypothesis.

Whichever hypothesis is correct, the difference between V′ MMN and AV′ MMN implies that the auditory and visual features of the bimodal input have been already partly combined in the afferent sensory systems before the MMN process occurs. This assumption fits with several recent observations that the construction of an integrated percept from bimodal inputs begins at very early stages of sensory analysis, well before the latency of the MMN processes (e.g. Giard and Peronnet 1999; Fort et al. 2002a; Molholm et al. 2002; Lebib et al. 2003; Besle et al. 2004). In addition, the fact that MMN is sensitive to the perceptual dimension of the stimulus (here its multimodal status) rather than to its physical dimension has been well documented in the auditory modality (review in Näätänen and Winkler 1999) and can also explain the existence of an auditory MMN in the McGurk illusion.

However, neither the auditory nor the visual sensory memory seems to encode an integrated trace of the audiovisual deviants. Indeed, in this case, assuming that the source of a MMN is likely to be close to or at the location of the memory upon which that MMN is based, the three kinds of deviant (A′V, AV′ and A′V′) should have elicited MMNs with similar topographies, because they would be all based on the same integrated memory representation. Rather, the three MMNs were found to have clearly different topographies with components typical of activities in the respective sensory-specific cortices, indicating that the MMN processes would operate mostly separately in each modality on the different sensory components of the multimodal representation under construction (Fig. 6).

Nonetheless, the hypothesis of complete independence of auditory and visual MMN processes is unlikely because the MMN to the double deviants in the audiovisual paradigm departs hardly from the mere addition of its unisensory components. (For the sake of simplification, this interpretation has not been shown in Fig. 6.) Future experimentation should be conducted to further assess the relationships between the auditory and visual sensory memories and MMN processes.

References

Alain C, Woods DL, Knight RT (1998) A distributed cortical network for auditory sensory memory in humans. Brain Res 812:23 37

Bental E, Dafny N, Feldman S (1968) Convergence of auditory and visual stimuli on single cells in the primary visual cortex of unanesthetized unrestrained cats. Exp Neurol 20:341 351

Bertelson P, Aschersleben G (1998) Automatic visual bias of perceived auditory location. Psychon Bull Rev 5:482 489

Berti S, Schroger E (2004) Distraction effects in vision: behavioral and event-related potential indices. Neuroreport 15:665 669

Besle J, Fort A, Delpuech C, Giard M-H (2004) Bimodal speech: early suppressive visual effects in the human auditory cortex. Eur J Neurosci 20:2225 2234

Bruneau N, Roux S, Garreau B, Martineau J, Lelord G (1990) Cortical evoked potentials as indicators of Auditory-Visual Cross-Modal Association in young adults. Pavlov J Biol Sci 25:189 204

Cahill L, Ohl F, Scheich H (1996) Alteration of auditory cortex activity with a visual stimulus through conditioning: a 2-deoxyglucose analysis. Neurobiol Learn Mem 65:213 222

Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P (2002b) Mismatch negativity evoked by the McGurk MacDonald effect: a phonetic representation within short-term memory. Clin Neurophysiol 113:495 506

Colin C, Radeau M, Soquet A, Dachy B, Deltenre P (2002a) Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. Clin Neurophysiol 113:507 518

Colin C, Radeau M, Soquet A, Deltenre P (2004) Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. Clin Neurophysiol 115:1989 2000

Czigler I, Balazs L, Winkler I (2002) Memory-based detection of task-irrelevant visual changes. Psychophysiology 39:869 873

Fort A, Giard M-H (2004) Multiple electrophysiological mechanisms of audio-visual integration in human perception. In: Calvert G, Spence C, Stein B (eds) The handbook of multisensory processes. MIT Press, Cambridge

Fort A, Delpuech C, Pernier J, Giard MH (2002a) Dynamics of cortico-subcortical crossmodal operations involved in audiovisual object detection in humans. Cereb Cortex 12:1031 1039

Fort A, Delpuech C, Pernier J, Giard MH (2002b) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14:20 30

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473 490

Giard MH, Perrin F, Pernier J (1990) Brain generators implicated in processing of auditory stimulus deviance A topographic ERP study. Psychophysiology 27:627 640

Heslenfeld DJ (2003) Visual mismatch negativity. In: Polich J (ed) Detection of change: event-related potential and fMRI findings. Kluwer Academic Publishers, Dordrecht, pp 41 60

Kenemans JL, Jong TG, Verbaten MN (2003) Detection of visual change: mismatch or rareness?. Neuroreport 14:1239 1242

Kropotov JD, Näätänen R, Sevostianov AV, Alho K, Reinikainen K, Kropotova OV (1995) Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex. Psychophysiology 32:418 422

Lebib R, Papo D, de Bode S, Baudonniere PM (2003) Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. Neurosci Lett 341:185 188

McGurk H, McDonald J (1976) Hearing lips and seeing voices. Nature 264:746 748

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115 128

Möttönen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. Brain Res Cogn Brain Res 13:417 425

Näätänen R (1992) Attention and Brain Function. Hillsdale, NJ, USA

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826 859

Nyman G, Alho K, laurinen P, Paavilainen P, Radil T, Rainikainen K, Sams M, Näätänen R (1990) Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. Electroencephalogr Clin Neurophysiol 77:436 444

Pazo-Alvarez P, Cadaveira F, Amenedo E (2003) MMN in the visual modality: a review. Biol Psychol 63:199 236

Pernier J, Perrin F, Bertrand O (1988) Scalp current density fields: concept and properties. Electroencephalogr Clin Neurophysiol 69:385 389

Perrin F, Pernier J, Bertrand O, Giard M-H (1987) Mapping of scalp potentials by surface spline interpolation. Electroencephalogr Clin Neurophysiol 66:75 81

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalogr Clin Neurophysiol 72:184 187

Ritter W, Deacon D, Gomes H, Javitt DC, Vaughan HG Jr (1995) The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review. Ear Hear 16:52 67

Rosenblum LD, Fowler CA (1991) Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. J Exp Psychol Hum Percept Perform 17:976 985

Saldana HM, Rosenblum LD (1993) Visual influences on auditory pluck and bow judgments. Percept Psychophys 54:406 416

Sams M, Aulanko R, Hamalainen H, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 127:141 145

Soto-Faraco S, Navarra J, Alsius A (2004) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. Cognition 92:B13 B23

Stagg C, Hindley P, Tales A, Butler S (2004) Visual mismatch negativity: the detection of stimulus change. Neuroreport 15:659 663

Stekelenburg JJ, Vroomen J, de Gelder B (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci Lett 357:163 166

Vroomen J, Bertelson P, de Gelder B (2001) The ventriloquist effect does not depend on the direction of automatic visual attention. Percept Psychophys 63:651 659

Audiovisual events in sensory memory

Julien Besle[1], Anne Caclin[1], Romaine Mayet[1], Françoise Bauchet[2], Claude Delpuech[2], Marie-Hélène Giard[1], Dominique Morlet[1].

[1]INSERM, U821, Brain dynamics and cognition, Lyon, F-69500, France;
[1]Institut Fédératif des Neurosciences, Lyon, F-69000, France;
[1]Université Lyon 1, Lyon, F-69000, France
[2]CERMEP – Centre MEG, Lyon, F69000, France

Article in press in *Journal of Psychophysiology*

Corresponding author :

Dominique Morlet
INSERM U281, Brain dynamics and cognition
Centre Hospitalier Le Vinatier
Bâtiment 452
69675 BRON Cedex
Tel: +33 (0)4 72 13 89 03
Fax: +33 (0)4 72 13 89 01

e-mail : morlet@lyon.inserm.fr

ABSTRACT

The functional properties of the auditory sensory memory have been extensively studied using the Mismatch Negativity (MMN) component of the auditory Event-Related Potential (ERP) and its magnetic counterpart recorded using Magneto-encephalography (MEG). It has been found that distinct auditory features (such as frequency or intensity) are encoded separately in sensory memory. Nevertheless, the conjunction of these features (auditory "gestalts") can also be encoded in auditory sensory memory.

Here we investigated how auditory and visual features of bimodal events are represented in sensory memory by recording audiovisual MMNs in two different audiovisual oddball paradigms. The results of a first ERP experiment showed that the sensory memory representations of auditory and visual features of audiovisual events lie within the temporal and occipital cortex respectively, yet with possible interactions between the processing of the unimodal features. In a subsequent MEG experiment, we found some evidence that audiovisual feature conjunctions could also be represented in sensory memory. These results thus extend to the audiovisual domain a number of properties of sensory memory already established within the auditory system.

INTRODUCTION

The Mismatch Negativity is elicited in the auditory cortex when incoming sounds are detected as deviating from a neuronal representation of acoustic regularities. This neuronal representation is likely to form the neurophysiological basis of the Auditory Sensory Memory (ASM) (e.g. Näätänen, 1992; Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). If one assumes that the mismatch process between the deviant input and the neural trace of the regular (« standard ») stimuli occurs where the deviating feature is stored, then the MMN can be used to study the functional organization of ASM and the representation of sounds in that ASM. For example, it has been shown that the MMNs to sounds deviating in frequency, intensity, or duration, or along different dimensions of timbre, originate from different locations in the auditory cortex, indicating that these different acoustic features are processed in separate registers in ASM (Caclin, Brattico, Tervaniemi, Näätänen, Morlet, Giard, & McAdams, 2006; Giard, Lavikainen, Reinikainen, Perrin, Bertrand, Pernier, & Näätänen, 1995; Rosburg, 2003). On the other hand, it has been found that ASM can also store, besides the separate acoustic features of a sound, the conjunction of those features, suggesting the existence of a « gestalt » representation of sounds in ASM (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997; Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998; Takegata, Huotilainen, Rinne, Näätänen, & Winkler, 2001; Takegata, Paavilainen, Näätänen, & Winkler, 1999; Winkler, Czigler, Sussman, Horvath, & Balazs, 2005). The principle of these studies was to use several standard sounds created by combining different values of individual features (e.g., location and frequency), and one or several deviants having the very same individual features as the standards, but using different pairings (conjunctions) of these features. Hence the only difference between the standards and the deviants was the particular combination of otherwise identical individual features.

An important question regarding the functional organization of ASM is whether this memory encodes only acoustic features –separately and in conjunction– or if the memory traces can be affected by visual information. A number of studies have indeed established that there exist early interactions between the processing of simultaneous auditory and visual information (Besle, Fort, Delpuech, & Giard, 2004; Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999; Lebib, Papo, de Bode, & Baudonniere, 2003; Molholm, Ritter, Murray, Javitt, Schroeder, & Foxe, 2002; review in Fort & Giard, 2004), which opens the possibility that the content of ASM could be modified when a visual event accompanies the auditory event.

This last hypothesis is supported by several studies showing an influence of visual cues on the auditory MMN process in particular situations where the presence of visual information

gives rise to auditory perceptual illusions like the McGurk effect or the ventriloquist illusion. In the McGurk illusion (McGurk & McDonald, 1976), the very same physical sound of a syllable can be perceived differently depending on the lip movements that are simultaneously seen (e.g. auditory /ba/ associated with visual /ga/ is perceived as /da/). An auditory MMN can be elicited by audiovisual McGurk syllables deviating from standards only on their visual dimension (Colin, Radeau, Soquet, & Deltenre, in press; Colin, Radeau, Soquet, Demolin, Colin, & Deltenre, 2002b; Möttönen, Krause, Tiippana, & Sams, 2002; Sams, Aulanko, Hamalainen, Hari, Lounasmaa, Lu, & Simola, 1991). Several explanations have been proposed, that are related to the functional specificity of speech: either there would exist a phonetic MMN process that is sensitive to the phonetic nature of articulatory movements (Colin et al., 2002b), or visual speech cues could have a specific access to the MMN generators in auditory cortex because, like auditory speech, they carry time-varying information (Möttönen et al., 2002). Generation of an auditory MMN by visual-only deviants can also be observed with the ventriloquist illusion in which the perceived location of a sound is shifted by a spatially disparate visual stimulus (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002a; Stekelenburg, Vroomen, & de Gelder, 2004). As underlined above however, these two phenomena are highly peculiar in that they give rise to irrepressible audiovisual illusions that seem to occur at a sensory level of representation (ventriloquist effect: Bertelson & Aschersleben, 1998; McGurk effect: Soto-Faraco, Navarra, & Alsius, 2004; Vroomen, Bertelson, & de Gelder, 2001).

Yet in everyday life, perceptual events often occur in multiple sensory systems at once, and the brain coordinates and integrates redundant information from different sensory – particularly auditory and visual – modalities to produce coherent and unified representations of the external world (Calvert, Spence, & Stein, 2004). The question thus arises of how, in the general case, are audiovisual events processed in sensory memory?

Recently, a visual homologue of the auditory MMN, the vMMN has been observed on posterior scalp sites (Berti & Schröger, 2004) around the same latency range as the auditory MMN (review in Pazo-Alvarez, Cadaveira, & Amenedo, 2003). More specifically, a vMMN has been found in response to stimuli deviating from a regular visual sequence either in color, spatial frequency, stimulus contrast, motion direction, shape, line orientation, or stimulus location (review in Czigler, this issue). Although less extensively studied than in the auditory modality, the vMMN might also rely on memory-based processes (Czigler, Balazs, & Winkler, 2002; Stagg, Hindley, Tales, & Butler, 2004) (see however Kenemans, Jong, & Verbaten, 2003) ; Czigler, this issue) The questions therefore are: Is ASM sensitive to general visual information? Are the auditory and visual features of a bimodal event encoded

separately in the memory system underlying the MMN process? Or is a bimodal event processed in a Gestalt manner in this memory?

We have addressed these questions in two experiments, one using event-related potentials (ERPs), the other using Magnetoencephalography (MEG). The first (ERP) study, already published in detail elsewhere (Besle, Fort, & Giard, 2005), will only be briefly recalled here.

EXPERIMENT 1

In the first experiment, we used a bimodal oddball paradigm in which audiovisual deviant stimuli differed from audiovisual standards (AV) either on the visual dimension only (AV'), or on the auditory dimension only (A'V), or on both dimensions simultaneously (A'V'), in order to test the following non-mutually exclusive hypotheses: (i) if the visual dimension of a bimodal event is represented in ASM, then AV' deviants should elicit an auditory MMN (similarly to visual deviants in the McGurk or ventriloquist effects); (ii) if the visual and auditory features of a bimodal event are encoded in separate memory systems, then the MMN to A'V' deviants (A'V'-MMN) should present separate components over temporal and posterior scalp areas; and (iii) if the auditory and visual MMN processes are independent, then the A'V'-MMN should be equal to the sum of the MMNs to unimodal deviants (A'V-MMN + AV'-MMN).

The bimodal stimuli consisted in the deformation of a circle into an horizontal (standard V) or vertical (deviant V') ellipse associated with a synchronously presented rich tone (standard A, deviant A'). Standards (AV) were presented with a probability of 76%, and deviants (A'V, AV', A'V') with a probability of 8% each. In half of the experimental blocks, the visual and auditory features of the standards and deviants were exchanged to insure that the resulting MMNs could not be attributed to physical differences between the standard and deviants. The subjects' (N=15) task was to respond to short and unpredictable disappearance of the fixation cross at the centre of the circle. ERPs were recorded from 36 scalp electrodes with a nose reference. (See Besle et al., 2005, for a detailed description of the stimuli, paradigm, and ERP analysis).

In addition, we conducted as a control a visual oddball experiment in which the visual stimuli were identical to those used in the bimodal paradigm, in order to compare a genuine vMMN (elicited by deviants in a visual sequence) with the MMN elicited by AV' deviants (i.e., visual-only deviants in a bimodal sequence).

The main results were the following: (i) A'V deviants elicited an auditory MMN with a typical temporo-frontal topography; (ii) AV' deviants elicited a "visual MMN" with a bilateral occipital topography; (iii) A'V' deviants elicited an MMN with both temporal and occipital components;

(iv) A'V'-MMN significantly differed from the sum A'V-MMN + AV'-MMN at several temporo-parietal electrodes; and (v) the occipital topography of the unimodal vMMN (recorded in the visual-only paradigm) partly differed from that of AV'-MMN on the left hemiscalp. Figure 1 illustrates some of these results by showing the scalp current density distributions of A'V-MMN, AV'-MMN, A'V'-MMN, and vMMN.

These results already have several important consequences. First, the fact that the AV'-MMN had an occipital (and not temporo-frontal) topography in our protocol could indicate that, in the general case, the visual deviant of a bimodal stimulus elicits a visual MMN-like response; an auditory MMN would be elicited only if that visual deviance gives rise to the (illusory) perception of an auditory change.

However we also found that the visual MMNs elicited in the visual (vMMN) and the audiovisual (AV'-MMN) paradigms significantly differed while the only difference between these paradigms was the presence or absence of repetitive identical sounds associated with both the visual standards and deviants. If the visual MMN reflects a memory-based process (Czigler, this issue), this result may indicate that an audiovisual event is encoded differently than a visual-only event in that visual memory, and thus that the processing of the unisensory features of the bimodal input have already interacted in the afferent sensory systems before the vMMN process occurs. This interpretation fits with the repeated findings that the crossmodal operations underlying the construction of an integrated multimodal percept can begin at very early stages of sensory processing, well before the latency of the MMN processes (Besle et al., 2004; Fort et al., 2002; Giard & Peronnet, 1999; Lebib et al., 2003; Molholm et al., 2002). This is also in agreement with the fact that MMN is sensitive to the perceptual dimensions of the stimulus rather than to its physical dimensions (review in Näätänen & Winkler, 1999), a result which could also explain the elicitation of an auditory MMN in the McGurk illusion.

Nevertheless, the fact that the MMN to simultaneous deviances on both the auditory and visual dimensions of a bimodal event includes supratemporal and occipital components indicates that the MMN processes operate mostly separately in each modality; this would mean that the different sensory components of the multimodal representation under construction are encoded separately in the transient memory systems of their respective modality. These results fit with the repeated findings of separation of elementary feature encoding in sensory memory that have been established in the auditory modality (Caclin et al., 2006; Giard et al., 1995; Rosburg, 2003), and extend these findings to the case of the constituent elements of bimodal events.

However, the MMN elicited by the double auditory and visual deviants partly differed from the sum of the MMNs to single deviants (A'V-MMN + AV'-MMN). This non-additivity of the MMNs could be accounted for by two non-mutually exclusive hypotheses: either the two deviance-detection processes are not entirely independent, or the MMN generating processes can also access, besides the separate auditory and visual stimulus features, the conjunction of these features. Indeed, in the auditory modality, it has been shown that both single features and feature conjunctions may be processed by the MMN system (e.g. Takegata et al., 2001). If this principle holds for audiovisual regularities, one should be able to observe an MMN to the violation of the conjunction of the auditory and visual features of repetitive bimodal events. Note that such an interpretation could also account for the difference between the vMMN recorded in the visual-only paradigm and the AV'-MMN recorded in the bimodal paradigm.

EXPERIMENT 2

This second experiment thus aimed at testing whether, in bimodal events, the conjunction of auditory and visual features may also be encoded in the transient memory system used by the MMN processes. In addition, to further compare the MMN to audiovisual feature conjunction – if existing – to that of the « classical » MMN originating in the auditory cortex, we ran a control auditory-only experiment using the same sounds.

Method

Subjects

Ten right-handed adults (5 female, mean age 29 years) were paid to participate. All were free of any neurological disease, and had normal hearing and normal or corrected-to-normal vision. All participants gave a written informed consent prior to their inclusion in the study in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Audiovisual experiment: Stimuli and protocol

We ran an audiovisual oddball paradigm inspired from those used to study the memory representation of auditory feature conjunctions (e.g. Takegata et al., 1999). Four stimuli (two «standards»: A1V1, A2V2, and two «deviants»: A1V2, A2V1) were used with the following hypothesis: since the deviants and standards had the same auditory features (A1, A2) and the same visual features (V1, V2), the deviants should elicit an MMN only if the frequently occuring conjunctions of auditory and visual features have been detected and encoded in the memory representations involved in the MMN process.

The four stimuli were randomly delivered with a probability of 0.44 for each standard type and 0.06 for each deviant type. The visual features consisted in the deformation of a circle into an horizontal (V1) or a vertical (V2) ellipse formed by a 23% reduction of the horizontal (vertical) diameter of the circle. The basic circle had a diameter of 3 cm and was presented permanently on a dark screen placed 85 cm in front of the subjects' eyes (visual angle: 2°). A cross at its centre served as the fixation point. The auditory features consisted in rich harmonic tones (fundamental, 2^{nd} and 4^{th} harmonics) with the fundamental frequency rising linearly from 500 to 540 Hz (A1) or from 500 to 600 Hz (A2). The sounds were delivered binaurally through plastic tubes and earpieces with an intensity adjusted for each subject at 35 dB SL. The auditory and visual features were synchronously presented with a duration of 167 ms (including 7ms of rise/fall times for sounds).

The experiment included 10 blocks of 260 stimuli delivered with an interstimulus interval (onset to onset) of 583 ms. Each block began with the presentation of three standards, and each deviant was preceded by at least 3 standards. The subject's task was to ignore the stimuli and press a key to the disappearance of the fixation cross (pseudo-random disappearance of 120 ms occuring in about 10% of the trials, always during standard trials).

Control auditory experiment

The stimuli were the A1 and A2 sounds used in the audiovisual paradigm. The experiment included 4 blocks of 425 stimuli each delivered with an ISI of 590 ms. In two blocks, the standards (p= 0.88) were A1 and the deviants (p=0.12) were A2; in the other two blocks, the standards and deviants were exchanged. The subjects' task was to read a book of their choice. The auditory-only experiment was always run after the bimodal experiment.

Recordings and data analysis

Recordings were carried out in a magnetically shielded room with a whole-scalp 275 channel CTF system at the MEG-EEG CERMEP Department in Lyon. The magnetic signals were continuously acquired with a 150-Hz low-pass filter and a sampling rate of 600 Hz. EOG activity was recorded from a bipolar montage of two electrodes placed at the outer canthi of both eyes.

Data analysis was performed using the ELAN pack software developed at the INSERM U821 (former U280) laboratory (Lyon). The MEG signals were digitally filtered offline (1-40 Hz bidirectional Butterworth filter, slopes 12 dB/octave). The signals (event-related fields, ERFs) were then averaged separately for each stimulus type over a time period of 500 ms including a 100-ms prestimulus baseline. Responses to the first three standards of each block and to the standards immediately following a deviant were excluded from averaging. For each

subject, a signal rejection threshold was chosen so as to keep about 85% of the remaining trials for averaging.

The auditory MMN (A-MMN) and the MMN to audiovisual feature conjunctions (AV_{conj}-MMN) were measured in the differences between the brain responses to the deviant and the standard stimuli in the auditory-alone and audiovisual experiments, respectively. At the group level, the existence of an MMN (i.e., non-null amplitudes in the difference wave) was assessed at each channel in consecutive 10-ms periods in the time window usually found for the MMN (140-320 ms). We used permutation tests for paired data (the distribution of the deviant-minus-standard response amplitudes under the null hypothesis of an equal amplitude for standards and deviants is estimated by randomly permuting the standard and deviant ERPs within each subject). To further investigate how individual responses relate to the grand average, we also analyzed the significance of the MMNs at the individual level using comparable randomization (permutation) procedures.

Results

Subjects responded to the disappearance of the fixation cross with a mean reaction time of 418 ms (± 50 ms) and less than 1% of errors, showing that they performed the distractive task adequately.

Group analysis

Figure 2.A displays the superimposition of the grand average ERFs across the 10 subjects at all channels for standard and deviant stimuli in the auditory-only and audiovisual paradigms.

In the auditory-only paradigm, the ERF traces for standard and deviant stimuli present a first peak around 60 ms after stimulus onset and begin to differ from each other from about 165 ms with a maximum difference around 200 ms. Statistical analysis showed that the amplitudes of the deviant-minus-standard difference waves at temporal sensors were highly significant between about 155 and 250 ms of latency. These activities showed a stable topography within this time range with polarity reversals over the temporal sites of each hemiscalp, highlighting the presence of an auditory MMN with a main origin in the auditory cortex. Figure 2.B (left) illustrates the mean topography over a 10 ms-period around the MMN peak latency, together with the associated probability map showing the scalp sites of significant MMN amplitudes.

In the audiovisual paradigm, the differences between the responses to standard and deviant stimuli were globally much smaller (Fig 2.A, right) but detailed statistical analysis revealed several sensors presenting significant ERF amplitudes in the difference waveforms between about 210 and 300 ms. Although the scalp areas of significant amplitude spread out less and

the peak latency of the effect over temporal areas was later (280 ms) in the audiovisual paradigm than in the auditory-only condition, the topography of the deviant-minus-standard responses in the audiovisual paradigm resembled that of the auditory MMN on both the left and right temporal sites (Fig. 2.B right). Furthermore, compared to the auditory MMN, it presented an additional component over the occipital sites between about 235 and 265 ms latency with a peak around 250 ms (Fig. 2.B right, bottom line), suggesting the presence of an additional source in the audiovisual condition.

Individual subject analysis

For the auditory-only paradigm, all subjects presented a significant MMN with a topography typical of activities in the auditory cortex. Figure 2.C (left) illustrates the data for one subject (S10) and Table 1 gives the latency window of significant amplitudes for each subject.

The results were much more variable in the audiovisual paradigm: 3 subjects out of 10 presented significant amplitudes with a corresponding MMN topography on both temporal and occipital sites, 5 subjects had an instable and/or unilateral MMN topography with marginally significant amplitudes, and 2 subjects did not present any significant amplitude nor MMN topography (Table 1).

Discussion

The MEG experiment clearly evidenced the presence of an auditory MMN in all subjects when using a classical auditory oddball paradigm with frequency glides as standard and deviant stimuli. Although the relative position of the head within the MEG system varied from one subject to the other (from 5 to 40 mm, because of the variability in head sizes), the presence of a significant MMN signal in the grand-average data underlines the robustness of the auditory MMN process.

While the differences between the responses elicited by the standards and the conjunction deviants in the audiovisual paradigm were much less pronounced, the significant amplitudes and the topography of the grand-average deviant-minus-standard signals strongly suggest that an MMN-like response has been elicited by a change in the conjunction of auditory and visual features of bimodal events. Cross-modal feature conjunctions may thus be represented somehow in the transient memory system used by the MMN processes. Furthermore, this representation would include both temporal and occipital components as suggested by the topography of the AV_{conj}-MMN.

The weak amplitude, associated with a limited statistical significance, of the brain responses to the conjunction of audiovisual features may be explained by several factors. First, in the group analysis, the variability in the subjects' head size and position in the MEG system

might have had a greater effect on the grand-average AV_{conj}-MMN because of its smaller amplitude and the poorer spatial and temporal spreading out of significant field patterns in individual subjects (compared to the auditory-alone condition).

In addition, two other possible explanations can be found in previous MMN studies in the auditory modality. The amplitude and latency of the MMN strongly depend on the strength of the memory trace encoding an auditory regularity (Ritter et al., 1995), as well as on the subject's ability to discriminate sounds deviating from this auditory regularity (e.g. Pakarinen, Takegata, Rinne, Huotilainen, & Näätänen, 2007; Tiitinen, May, Reinikainen, & Näätänen, 1994): the more difficult the discrimination, the later the latency and the smaller the amplitude of the associated MMN. When subjects cannot detect auditory changes, no MMN is elicited. Concerning our experiment, these findings lead to two predictions.

First, the MMN elicited in the auditory-alone paradigm should present a larger amplitude and a shorter latency than the MMN to audiovisual feature conjunction since there is only one type of standard in the first paradigm vs. two in the audiovisual paradigm. Therefore a stronger memory trace is expected in the former than in the latter paradigm, this is indeed what we observed.

Second, a change in audiovisual feature conjunction should not elicit an MMN if the subjects cannot explicitly discriminate the deviants A1V2 and A2V1 from the standards A1V1 and A2V2. To assess whether the absence of an AV_{conj}-MMN in some subjects could be due to the difficulty in discriminating the deviants from the standard stimuli in our paradigm, we have performed an additional behavioural experiment in 6 subjects (mean age: 30 years), including 3 subjects that had participated in the MEG experiment. The experimental set-up was similar to that used in the MEG experiment except that the sounds were delivered through headphones. There were 6 sequences of 100 stimuli delivered in each paradigm (auditory-alone and audiovisual). The subject's task was to press a key as quickly as possible upon the detection of a deviant stimulus. Table 2 presents the results for each subject. In the auditory paradigm, the percentage of correct detections was on average of 98% with a mean response time of 411 ms. In the audiovisual paradigm, the mean percentage of correct responses was much lower (67%) and the response times longer (mean: 723 ms); 3 subjects out of 6 detected less than 53% of the deviants. These behavioural results confirm that the detection of the audiovisual-conjunction deviants was very difficult in our protocol and may thus explain the small amplitude or the absence of MMN elicited by these deviants. This hypothesis is further supported by the results from the 3 subjects who participated in both the MEG and behavioural experiments (S1, S10 and S9 in Table 1, corresponding to S'2, S'3 and S'6, respectively, in Table 2). The first two subjects could correctly discriminate the audiovisual deviants and presented a significant AV_{conj}-MMN

with a temporal and occipital topography, while the third subject did not present any MMN pattern at temporal sites.

To sum up, our experiments support the view that bimodal events are encoded in the transient memory system used by the MMN processes with anatomically separate representations in modality-specific cortices. In addition, the transient memory system could encode not only the single sensory features of bimodal events, but also their conjunction. These data would generalize some of the « rules » established in the auditory modality, namely that (i) both the single features of bimodal events and their conjunction are encoded in the transient memory system used by the MMN processes (Takegata et al, 1999, 2001); and (ii) an MMN would be elicited only if the deviants, whatever their nature can be detected by the subject. Further experiments using audiovisual feature conjunction deviants easier to discriminate from audiovisual standards should be conducted to confirm the existence and the topography of the MMN to audiovisual conjunctions.

Acknowledgements

We thank Claude Delpuech and Françoise Lecaignard at the MEG-EEG CERMEP Department in Lyon for their much appreciated help during the realization of the MEG experiment.

FIGURE LEGENDS

Figure 1

Summary of Experiment 1. Scalp current densities of the deviant-minus-standard ERPs at the latency of their respective maximum amplitude, for the different deviances in the audiovisual paradigm (A'V, AV', and A'V') and in the visual-only paradigm (V'). The range of the colour scale used is indicated below each figure.

Figure 2

Results of the auditory-only (left column) and audiovisual (right column) paradigms in Experiment 2.

A. Superimposition of the MEG responses at all the 275 channels for standard (green lines) and deviant (blue lines) stimuli in each paradigm.

B. Mean topographies of the deviant-minus-standard grand-average responses over a 10-ms window around the peak latency in each paradigm, with the corresponding statistical maps (randomization tests). Although the responses in the audiovisual paradigm are much less significant than in the auditory-only paradigm, the topographies in both paradigms present a polarity reversal over the temporal sites typical of activities in the auditory cortex. In addition, the responses in the audio-visual paradigm present an additional occipital component peaking about 30 ms earlier than the temporal component.

C. Same as in B for one subject (S10).

The range of the colour scale used is indicated below each figure. Significant areas ($p < 0.05$) are depicted in white in the statistical maps.

TABLE 1

	Auditory		Audiovisual Conjunction		
	Left temporal	Right temporal	Left temporal	Right temporal	Occipital
S1	150-240	150-250	190-265	195-265	215-235
S2	190-260	200-275	-	285-330	-
S3	160-230	160-230	-	-	-
S4	170-230	160-230	245-260	245-295	245-265
S5	170-250	170-240	265-275	270-300	280-305
S6	180-260	160-270	-	-	-
S7	180-250	180-270	220-270	175-205	245-255
S8	170-240	180-230	265-295	-	-
S9	205-215	200-250	-	-	230-270
S10	150-230	170-230	275-295	275-295	245-255

Latency windows (ms) of significant MMN amplitudes for each subject in Experiment 2, in each of the two paradigms.

TABLE 2

	Correct deviance detection (%)		Mean response times (ms)	
	Auditory	Audiovisual	Auditory	Audiovisual
S'1	98	53	406	750
S'2 (=S1)	10	98	407	673
S'3 (=S10)	98	84	425	628
S'4	98	50	378	778
S'5	95	77	390	771
S'6 (=S9)	100	44	461	739
Mean ± sd	98 ± 2	67 ± 22	411 ± 29	723 ± 60

Results of the behavioural experiment complementing Experiment 2. Six subjects performed a speeded detection of the deviants in the auditory-only and audiovisual paradigms. The correspondence with the subjects' numbers in Experiment 2 (see Table 1) is indicated for the three subjects concerned.

REFERENCES

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5*, 482-489.

Berti, S., & Schröger, E. (2004). Distraction effects in vision: behavioral and event-related potential indices. *NeuroReport, 15*(4), 665-669.

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in the human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225-2234.

Besle, J., Fort, A., & Giard, M.-H. (2005). Is the auditory sensory memory sensitive to visual information ? *Experimental Brain Research, 166*(3-4), 337-334.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M. H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience, 18*(12), 1959-1972.

Calvert, G. A., Spence, C., & Stein, B. (Eds.). (2004). *The Handbook of Multisensory Processes*. Cambridge: The MIT Press.

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002a). Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology, 113*(4), 507-518.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (in press). Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. *Clin Neurophysiol*.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002b). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology, 113*(4), 495-506.

Czigler, I. Visual Mismatch Negativity: violation of non-attended environmental regulations. *(This issue).*

Czigler, I., Balazs, L., & Winkler, I. (2002). Memory-based detection of task-irrelevant visual changes. *Psychophysiology, 39*(6), 869-873.

Czigler, I., Balazs, L., & Winkler, I. (2002). Memory-based detection of task-irrelevant visual changes. *Psychophysiology, 39*(6), 869-873.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcorical crossmodal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12*(10), 1031-1039.

Fort, A., & Giard, M.-H. (2004). Multiple electrophysiological mechanisms of audio-visual integration in human perception. In G. Calvert, C. Spence & B. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 503-514). Cambridge: MIT Press.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representations of stimulus frequency, intensity, and duration in auditory sensory memory: An Event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience, 7,2*, 133-143.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans : a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G., & Miller, J. (1997). Storage of feature conjonctions in transient auditory memory. *Psychophysiology, 34*, 712-716.

Kenemans, J. L., Jong, T. G., & Verbaten, M. N. (2003). Detection of visual change: mismatch or rareness? *NeuroReport, 14*(9), 1239-1242.

Lebib, R., Papo, D., de Bode, S., & Baudonniere, P. M. (2003). Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neuroscience letters, 341*(3), 185-188.

McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115-128.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13*(3), 417-425.

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125*(6), 826-859.

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clin Neurophysiol, 118*(1), 177-185.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological Psychology, 63*(3), 199-236.

Ritter, W., Deacon, D., Gomes, H., Javitt, D. C., & Vaughan, H. G., Jr. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review. *Ear and Hearing, 16*, 52-67.

Rosburg, T. (2003). Left hemispheric dipole locations of the neuromagnetic mismatch negativity to frequency, intensity and duration deviants. *Cognitive Brain Research, 16*(1), 83-90.

Sams, M., Aulanko, R., Hamalainen, H., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141-145.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition, 92*(3), B13-23.

Stagg, C., Hindley, P., Tales, A., & Butler, S. (2004). Visual mismatch negativity: the detection of stimulus change. *NeuroReport, 15*(4), 659-663.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters, 357*(3), 163-166.

Sussman, E., Gomes, H., Nousak, J. M., Ritter, W., & Vaughan, H. G., Jr. (1998). Feature conjunctions and auditory sensory memory. *Brain Research, 793*(1-2), 95-102.

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *NeuroReport, 12*(3), 525-529.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters, 266*(2), 109-112.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*(6501), 90-92.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics, 63*(4), 651-659.

Winkler, I., Czigler, I., Sussman, E., Horvath, J., & Balazs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience, 17*(2), 320-339.