

· THÈSE ·

présentée pour obtenir le grade de  
Docteur de l'Université Louis Pasteur - Strasbourg I

École doctorale : Sciences pour l'ingénieur  
Discipline : Électronique, électrotechnique, automatique  
Spécialité : Traitement d'images et vision par ordinateur

**Analyse et fusion markovienne de séquences en  
imagerie 3D+t. Application à l'analyse de séquences  
d'images IRM fonctionnelles cérébrales**

Soutenue publiquement  
le 9 décembre 2004  
par

Sylvain FAISAN

Membres du jury :

Mme	Régine	ANDRÉ-OBRECHT	Rapporteur externe
M.	Jérôme	IDIER	Rapporteur externe
M.	Michel	DE MATHELIN	Rapporteur interne
M.	Laurent	THORAVAL	Examineur
M.	Fabrice	HEITZ	Directeur de thèse
M.	Jean-Paul	ARMSPACH	Directeur de thèse



# Remerciements

Je tiens tout d'abord à adresser tous mes remerciements à :

- Madame Régine André-Obrecht, Professeur des Universités,
- Monsieur Jérôme Idier, Chargé de Recherches au CNRS,
- Monsieur Michel de Mathelin, Professeur des Universités,

d'avoir accepté la tâche fastidieuse d'être les rapporteurs de cette thèse. Je leur suis très reconnaissant d'avoir pris le temps d'examiner mon manuscrit.

Ce travail a été réalisé au LSIIT (Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection), dans l'équipe MIV (Modèles, Images et Vision) en collaboration avec l'IPB (Institut de Physique Biologique) dans le cadre de l'équipe-projet multi-laboratoires IRMC (Imagerie et Robotique Médicale et Chirurgicale). J'exprime ma gratitude à Monsieur Jean-François Dufourd, directeur du LSIIT, et à Monsieur Daniel Grucker, directeur de l'IPB, de leur accueil dans leur laboratoire respectif.

Je remercie mes directeurs de thèse :

- Monsieur Fabrice Heitz, Professeur des Universités. Ses conseils ont toujours été clairs, constructifs et riches d'enseignement.
- Monsieur Jean-Paul Armspach, Ingénieur de Recherches à l'IPB. Je lui suis reconnaissant de m'avoir fait partager son expérience du monde de l'imagerie médicale. Le travail est loin d'être fini mais je le remercie beaucoup de son aide.

Je remercie Laurent Thoraval d'avoir initié et guidé mon travail. J'ai beaucoup apprécié son aide, ses conseils judicieux et avisés, et la foi qu'il a en son travail.

Il convient maintenant de remercier les neurologues avec lesquels j'ai pu travailler, à savoir Marie-Noëlle Metz-Lutz et Jack Foucher. J'espère que la collaboration entreprise durant la thèse se renforcera avec le temps.

Merci enfin à tous les membres de l'équipe MIV, en particulier à son responsable, Monsieur Ernest Hirsch, de leur gentillesse et de leur soutien.

Je remercie mes collègues de bureau : leur présence, leur bonne humeur ont été pour moi un cadeau de tous les jours.

Merci à Nadège sans qui je n'aurais pas été présentable le jour de la soutenance. Par pudeur, je la remercie uniquement de ... et surtout de ... et enfin de ...

L'ingratitude des enfants veut que je finisse par remercier ma famille en dernier. Merci à eux de me laisser poursuivre mes rêves.

*Merci à toutes les étoiles qui brillent dans le ciel*



# Table des matières

Notations	xiii
Introduction générale	1
<b>I Cadre de travail</b>	<b>5</b>
Introduction	7
<b>1 L'imagerie par résonance magnétique fonctionnelle cérébrale</b>	<b>9</b>
1.1 Le système nerveux . . . . .	9
1.1.1 Rôle . . . . .	9
1.1.2 Structures anatomo-fonctionnelles du système nerveux . . . . .	10
1.1.3 L'unité fonctionnelle du système nerveux : le neurone . . . . .	11
1.1.4 Les tissus cérébraux du SNC . . . . .	12
1.1.5 Les régions fonctionnelles du cortex . . . . .	14
1.2 Modalités d'observation de l'activité cérébrale . . . . .	15
1.2.1 Mesure de l'activité neuronale . . . . .	16
1.2.2 Mesure de l'activité métabolique . . . . .	17
1.2.3 Mesures basées sur le système cérébrovasculaire . . . . .	17
1.2.4 Comparaison . . . . .	18
1.3 L'IRMf par contraste BOLD . . . . .	19
1.3.1 Principe physique de l'IRM . . . . .	19
1.3.1.1 Principe physique de la résonance magnétique nucléaire . . . . .	19
1.3.1.2 Formation d'une image . . . . .	21
1.3.2 L'effet BOLD . . . . .	21
1.3.3 Mise en œuvre : du paradigme d'activation à la carte d'activité cérébrale	22
1.3.4 Types de paradigmes d'activation cérébrale . . . . .	23
1.3.4.1 Le déroulement temporel du paradigme d'activation . . . . .	24
1.3.4.2 Choix des tâches . . . . .	25
1.3.5 Caractéristiques des données recueillies . . . . .	27
1.3.5.1 Les différentes sources d'artéfacts . . . . .	27
1.3.5.2 La variabilité des données acquises . . . . .	28
<b>2 Méthodes de cartographie en IRMf cérébrale : état de l'art</b>	<b>29</b>
2.1 Prétraitements . . . . .	29
2.2 Les méthodes inférentielles . . . . .	30
2.2.1 La méthodologie SPM (Statistical Parametric Mapping) . . . . .	30

2.2.1.1	Le modèle linéaire général . . . . .	30
2.2.1.2	Les cartes paramétriques $\{t\}$ . . . . .	31
2.2.1.3	Les cartes paramétriques $\{f\}$ . . . . .	32
2.2.1.4	Seuillage des cartes de p-valeur ou de z-score . . . . .	33
2.2.2	De l'applicabilité des méthodes inférentielles en IRMf . . . . .	34
2.2.2.1	Limitations . . . . .	34
2.2.2.2	Solutions . . . . .	35
2.2.3	De nouvelles méthodologies . . . . .	37
2.2.3.1	Cadre d'inférence bayésien . . . . .	37
2.2.3.2	Autres méthodes inférentielles . . . . .	37
2.3	Les méthodes exploratoires . . . . .	38
2.3.1	La décomposition en valeurs singulières . . . . .	39
2.3.2	L'analyse en composantes indépendantes . . . . .	39
2.3.3	L'analyse de corrélation canonique . . . . .	40
2.3.4	Les méthodes de clustering . . . . .	40
2.4	Bilan . . . . .	40
<b>3</b>	<b>Approches markoviennes cachées : une synthèse</b>	<b>43</b>
3.1	Le modèle markovien caché standard . . . . .	44
3.1.1	Définition . . . . .	44
3.1.2	Graphe d'états . . . . .	44
3.1.3	Aspects algorithmiques . . . . .	44
3.1.3.1	Evaluation . . . . .	44
3.1.3.2	Décodage . . . . .	45
3.1.3.3	Apprentissage . . . . .	46
3.1.3.3.1	L'algorithme de Baum-Welch . . . . .	46
3.1.3.3.2	La procédure de Viterbi . . . . .	47
3.1.3.3.3	Autres procédures d'apprentissage . . . . .	48
3.2	Le modèle semi-markovien caché . . . . .	49
3.2.1	Origine . . . . .	49
3.2.2	Définition . . . . .	49
3.2.3	Graphe d'états . . . . .	49
3.2.4	Aspects algorithmiques . . . . .	50
3.3	Graphes de dépendance . . . . .	50
3.3.1	Définitions . . . . .	50
3.3.1.1	Graphe de dépendance non orienté . . . . .	51
3.3.1.2	Graphe de dépendance orienté . . . . .	51
3.3.2	Les réseaux bayésiens dynamiques . . . . .	52
3.3.3	Problèmes d'inférence : évaluation et décodage . . . . .	54
3.3.3.1	Calcul exact . . . . .	54
3.3.3.1.1	Algorithme du passage de message . . . . .	54
3.3.3.1.2	Algorithme de l'arbre de jonction . . . . .	56
3.3.3.2	Méthodes approximatives . . . . .	58
3.3.3.2.1	Les méthodes de Monte-Carlo . . . . .	58
3.3.3.2.2	Les méthodes variationnelles . . . . .	59
3.3.4	Problèmes d'inférence : apprentissage . . . . .	60
3.3.4.1	L'algorithme EM . . . . .	60
3.3.4.2	Autres méthodes d'apprentissages . . . . .	60

3.3.4.2.1	Variantes de l'EM . . . . .	60
3.3.4.2.2	L'algorithme ICE . . . . .	61
3.3.4.2.3	Segmentation courante . . . . .	61
3.4	Fusion markovienne de processus en interaction . . . . .	61
3.4.1	Introduction . . . . .	61
3.4.2	Cadre formel . . . . .	62
3.4.3	Approches de modélisation . . . . .	62
3.4.3.1	Ensemble de processus en interaction avec un unique canal d'observation . . . . .	63
3.4.3.2	Ensemble de processus en interaction avec un canal d'observa- tion par processus . . . . .	63
	<b>Conclusion partielle</b>	<b>67</b>
	 <b>II Le modèle semi-markovien caché de séquence d'événements : application en cartographie IRMf cérébrale</b>	 <b>69</b>
	<b>Introduction</b>	<b>71</b>
	 <b>4 Principes et méthodes</b>	 <b>73</b>
4.1	Modélisation markovienne et signal IRMf . . . . .	73
4.1.1	Modèle markovien caché auto-régressif . . . . .	74
4.1.2	Modèle markovien caché segmental . . . . .	74
4.1.3	Modèle proposé dans le cadre de l'IRMf . . . . .	75
4.2	Solution proposée . . . . .	75
4.2.1	Un nouveau modèle markovien . . . . .	76
4.2.2	Principe d'alignement . . . . .	76
4.3	Problématiques . . . . .	77
4.3.1	Que faut-il modéliser ? . . . . .	77
4.3.1.1	Le modèle biologique . . . . .	77
4.3.1.2	Le modèle du détecteur . . . . .	77
4.3.1.3	Un modèle multicouche ? . . . . .	78
4.3.2	Intérêt de la modélisation markovienne cachée . . . . .	79
	 <b>5 Le modèle semi-markovien caché de séquence d'événements</b>	 <b>81</b>
5.1	Séquence d'événements, séquence d'observations . . . . .	81
5.2	Éléments d'un MSMCSE . . . . .	82
5.2.1	Espace d'états . . . . .	82
5.2.2	Matrice de transition . . . . .	82
5.2.3	Lois de durée . . . . .	83
5.2.4	Lois d'observation . . . . .	83
5.3	Graphe d'états d'un MSMCSE . . . . .	83
5.4	Résolution des problèmes d'inférence . . . . .	85
5.4.1	Evaluation . . . . .	85
5.4.2	Décodage . . . . .	86
5.4.3	Apprentissage . . . . .	86

<b>6</b>	<b>Application en cartographie IRMf cérébrale</b>	<b>89</b>
6.1	Présentation générale . . . . .	89
6.2	Séquence d'événements IRMf, séquence d'observations . . . . .	90
6.2.1	Détection des événements . . . . .	90
6.2.2	Caractérisation des événements . . . . .	91
6.3	Modélisation par MSMCSE de l'activité cérébrale . . . . .	92
6.3.1	Principe . . . . .	92
6.3.2	Paramètres du modèle et liaison de paramètres . . . . .	94
6.3.2.1	La matrice de transition . . . . .	94
6.3.2.2	Les lois de durée . . . . .	95
6.3.2.3	Les lois d'observation . . . . .	95
6.4	Apprentissage et cartographie non supervisée de signaux d'activation cérébrale	96
6.4.1	Principe général . . . . .	96
6.4.2	Initialisation . . . . .	97
6.4.3	Cartographie . . . . .	97
6.4.4	Algorithme d'apprentissage . . . . .	98
6.5	Décodage des signaux d'activation : vers de nouvelles fonctionnalités d'analyse	99
6.5.1	Analyse des retards d'activation . . . . .	99
6.5.2	Analyse des modes d'activation . . . . .	99
6.5.3	Analyse de la réponse hémodynamique . . . . .	99
<b>7</b>	<b>Résultats expérimentaux</b>	<b>101</b>
7.1	Evaluation sur données synthétiques . . . . .	101
7.1.1	Données . . . . .	101
7.1.2	Méthodologie d'évaluation . . . . .	103
7.1.2.1	Courbes COR, $\alpha$ et $\beta$ . . . . .	103
7.1.2.2	Méthode de comparaison utilisée : SPM . . . . .	105
7.1.3	Résultats . . . . .	106
7.1.3.1	Procédure d'apprentissage-cartographie . . . . .	106
7.1.3.2	Cartes d'activité . . . . .	107
7.1.3.2.1	Influence du bruit . . . . .	107
7.1.3.2.2	Variabilité du signal actif . . . . .	109
7.2	Evaluation sur données réelles . . . . .	110
7.2.1	Données . . . . .	110
7.2.2	Résultats . . . . .	112
7.2.2.1	Procédure d'apprentissage-cartographie . . . . .	112
7.2.2.2	Cartes d'activité et retard d'activation . . . . .	113
7.2.2.3	Modes d'activation . . . . .	115
7.2.2.4	Réponse hémodynamique cérébrale . . . . .	115
<b>8</b>	<b>Discussion</b>	<b>117</b>
8.1	Influence des paramètres sur données synthétiques . . . . .	117
8.1.1	Etape de détection . . . . .	117
8.1.2	Influence des paramètres sur les résultats de cartographie . . . . .	119
8.2	Influence des paramètres sur données réelles . . . . .	121
8.3	Bilan de l'approche de cartographie par MSMCSE . . . . .	123
8.3.1	Les points positifs . . . . .	123
8.3.2	Les améliorations envisageables . . . . .	124



8.3.2.1	Un unique modèle d'activité? . . . . .	124
8.3.2.2	Modélisation des durées inter-événements . . . . .	124
8.3.2.3	Caractérisation ou discrimination? . . . . .	125
	<b>Conclusion partielle</b>	<b>127</b>
	<b>III Le modèle markovien caché de multiples séquences d'événements : application en cartographie IRMf cérébrale</b>	<b>129</b>
	<b>Introduction</b>	<b>131</b>
<b>9</b>	<b>Principes et méthodes</b>	<b>133</b>
9.1	Principe d'alignement multi-séquences . . . . .	133
9.2	Problématique . . . . .	133
9.3	Stratégie de résolution . . . . .	134
9.3.1	Le modèle biologique . . . . .	135
9.3.2	Le modèle des détecteurs . . . . .	135
9.3.3	Le modèle développé . . . . .	135
<b>10</b>	<b>Le modèle markovien caché de multiples séquences d'événements</b>	<b>137</b>
10.1	Séquences d'événements, signatures, scénario et séquence d'observations . . . . .	137
10.1.1	Séquences d'événements . . . . .	137
10.1.2	Signatures . . . . .	138
10.1.3	Scénario . . . . .	138
10.1.4	Séquence d'observations . . . . .	139
10.2	Définition du MMCMSE . . . . .	140
10.2.1	Éléments d'un MMCMSE . . . . .	140
10.2.2	Lois d'observation et MMCMSE . . . . .	141
10.3	Problèmes d'inférence . . . . .	141
10.3.1	Présentation . . . . .	141
10.3.1.1	Evaluation . . . . .	142
10.3.1.2	Décodage . . . . .	142
10.3.1.3	Apprentissage . . . . .	143
10.3.2	Stratégie de résolution . . . . .	143
10.3.2.1	Couple observation-état . . . . .	143
10.3.2.2	Association des événements détectés . . . . .	144
10.3.2.3	Construction de l'ensemble des couples observation-état . . . . .	144
10.3.2.4	Chemin caché $\mathbf{c}_\psi$ . . . . .	145
10.3.2.5	Ensemble $\gamma_k$ et $\gamma_k^{-1}$ . . . . .	145
10.4	Résolution du problème d'évaluation . . . . .	145
10.4.1	Evaluation directe . . . . .	146
10.4.2	Evaluation récursive . . . . .	146
10.4.3	Exemple . . . . .	147
10.5	Résolution du problème de décodage . . . . .	148
10.6	Résolution du problème d'apprentissage . . . . .	149
10.6.1	Définition et décomposition de la fonction auxiliaire . . . . .	149
10.6.2	Optimisation de la fonction auxiliaire . . . . .	150

10.7	Stratégie de diminution de la combinatoire . . . . .	150
10.7.1	Contrainte de nombre . . . . .	151
10.7.2	Contrainte de forme . . . . .	151
10.7.3	Contrainte de maximalité . . . . .	152
10.7.4	Conclusion . . . . .	153
<b>11</b>	<b>Application en cartographie IRMf cérébrale</b>	<b>155</b>
11.1	Présentation générale . . . . .	155
11.2	Prétraitement du signal IRMf . . . . .	156
11.3	Modélisation . . . . .	156
11.3.1	Processus caché . . . . .	156
11.3.2	Processus observable . . . . .	159
11.3.2.1	Contrainte de sommation des lois d'observation . . . . .	159
11.3.2.2	Prise en compte du retard d'activité . . . . .	160
11.3.2.3	Expression finale des lois d'observation . . . . .	160
11.3.3	Liaison de paramètres . . . . .	162
11.3.3.1	Matrice de transition . . . . .	163
11.3.3.2	Lois d'observation . . . . .	164
11.4	Algorithme de cartographie . . . . .	164
11.4.1	Estimation des paramètres . . . . .	164
11.4.1.1	Critère utilisé . . . . .	164
11.4.1.2	Optimisation du critère . . . . .	165
11.4.1.3	Stratégie de diminution de la complexité algorithmique . . . . .	166
11.4.2	Cartographie . . . . .	167
11.4.2.1	Problématique . . . . .	167
11.4.2.2	Mise à jour de la carte d'activité . . . . .	168
11.4.2.3	Carte de p-valeur . . . . .	168
<b>12</b>	<b>Résultats expérimentaux : paradigmes en blocs</b>	<b>171</b>
12.1	Résultats . . . . .	171
12.1.1	Résultats sur données synthétiques . . . . .	171
12.1.1.1	Rappel . . . . .	171
12.1.1.2	Influence du bruit . . . . .	172
12.1.1.3	Variabilité du signal actif . . . . .	174
12.1.2	Résultats sur données réelles . . . . .	176
12.2	Influence des paramètres . . . . .	178
12.2.1	Etude sur données synthétiques . . . . .	178
12.2.1.1	Etape de détection . . . . .	178
12.2.1.2	Cartographie . . . . .	180
12.2.1.2.1	Les paramètres utilisés . . . . .	180
12.2.1.2.2	Résultats . . . . .	180
12.2.2	Etude sur données réelles . . . . .	183
12.2.2.1	Protocole . . . . .	183
12.2.2.2	Analyse . . . . .	186

---

<b>13 Résultats expérimentaux : paradigmes événementiels</b>	<b>187</b>
13.1 Données . . . . .	187
13.1.1 Séquence IRMf synthétiques . . . . .	187
13.1.2 Séquences IRMf réelles . . . . .	188
13.2 Résultats . . . . .	190
13.2.1 Résultats sur données synthétiques . . . . .	190
13.2.2 Résultats sur données réelles . . . . .	192
13.3 Influence des paramètres . . . . .	193
13.3.1 Influence des paramètres sur données synthétiques . . . . .	193
13.3.1.1 Etape de détection . . . . .	193
13.3.1.2 Cartographie . . . . .	195
13.3.1.2.1 Les paramètres utilisés . . . . .	195
13.3.1.2.2 Résultats . . . . .	195
13.3.2 Influence des paramètres sur données réelles . . . . .	196
<b>Conclusion partielle</b>	<b>201</b>
<b>Conclusion générale et perspectives</b>	<b>203</b>
<b>A Liste des publications</b>	<b>209</b>
<b>Bibliographie</b>	<b>211</b>



# Notations

Les notations de la partie II sont résumées dans les tableaux 1 et 2. Les notations du tableau 1 concernent plus spécifiquement le modèle semi-markovien caché de séquence d'événements alors que celles du tableau 2 concernent son application en cartographie fonctionnelle cérébrale.

Les notations de la partie III sont résumées dans trois tableaux. Les tableaux 3, 4, 5 concernent respectivement, les notations relatives au modèle markovien caché de multiples séquences d'événements, celles relatives à la stratégie de résolution des problèmes d'inférence et celles de l'application en cartographie fonctionnelle cérébrale.

Tout au long du manuscrit, il est à noter que les variables en gras représentent des matrices ou des vecteurs. Ainsi,  $\mathbf{X}$  peut dénoter un vecteur,  $X_t$  le t-ème élément du vecteur  $\mathbf{X}$ ,  $\mathbf{X}_{t_1}^{t_2}$  les éléments  $X_{t_1}, X_{t_1+1}, \dots, X_{t_2}$  du vecteur  $\mathbf{X}$ .

Enfin, il est à noter que l'on fait rarement la distinction entre la variable aléatoire et sa réalisation.

Symbole	Signification	Renvoi
MSMCSE	modèle semi-markovien caché de séquence d'événements.	sec. 5, p. 81
$\mathbf{E} =$ $e_1 e_2 \dots e_L$	Séquence d'événements	sec. 5.1, p. 81
$\tau_E =$ $t_1 = 1, t_2, \dots, t_L = T$	Ensemble des instants de détection associé à $\mathbf{E}$ .	sec. 5.1, p. 81
$\tau_O =$ $1, 2, \dots, T$	Ensemble des instants d'observation du processus à analyser.	sec. 5.1, p. 81
$\mathbf{O} =$ $O_1 O_2 \dots O_T$	Séquence d'observation du MSMCSE.	sec. 5.1, p. 81
$\emptyset$	Trou d'observation résultat d'une non-détection ( <i>null</i> ).	sec. 5.1, p. 81
<i>evp</i>	Événement vrai positif.	sec. 5.2.1, p. 82
<i>efp</i>	Événement faux positif.	sec. 5.2.1, p. 82
$C$	En raison des deux <i>evps</i> fictifs, $C - 2$ représente le nombre de classes d' <i>evp</i> caractérisant le processus modélisé.	sec. 5.2.1, p. 82
$\mathcal{Q} =$ $\mathcal{Q}_{evp} \cup \mathcal{Q}_{efp/null}$	Espace d'état du MSMCSE. L'espace d'état résulte de l'union de deux sous-espaces : l'espace d'état $\mathcal{Q}_{evp} = \{q_i\}_{1 \leq i \leq C}$ et de l'espace d'état $\mathcal{Q}_{efp/null} = \{q_{ij}\}_{1 \leq i < C, 1 < j \leq C}$ où $q_{ij} = \{q_{ij}^o, q_{ij}^\emptyset\}$ , $q_{ij}^o$ modélise les <i>efps</i> et $q_{ij}^\emptyset$ les trous d'observation.	sec. 5.2.1, p. 82
$\mathbf{A} =$ $\{a_{ij}\}_{1 \leq i < C, 1 < j \leq C}$	Matrice de transition du MSMCSE. Elle est définie sous le sous-espace $\mathcal{Q}_{evp}$ . En raison de la spécificité des états $q_1$ et $q_C$ , elle est de taille $C - 1 \times C - 1$ .	sec. 5.2.2, p. 82
$\mathbf{D} =$ $\{p_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$	Ensemble des lois de durée inter-événements.	sec. 5.2.3, p. 83
$\mathbf{B} =$ $\{b_j(\cdot)\}_{1 \leq j \leq C}$	Ensemble des lois d'observation associés aux états <i>evp</i> .	sec. 5.2.4, p. 83
$\mathbf{B}' =$ $\{b_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$	Ensemble des lois d'observation associés aux états <i>efp</i> .	sec. 5.2.4, p. 83
$\mathbf{E} =$ $\{e_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$	Matrice de probabilité d'émission d' <i>efps</i> .	sec. 5.2.4, p. 83
$\lambda =$ $\{\mathcal{Q}, \mathbf{A}, \mathbf{B}, \mathbf{B}', \mathbf{D}, \mathbf{E}\}$	Paramètres du MSMCSE.	sec. 5.2.4, p. 83

TAB. 1 – Liste de notations pour la partie II relative au MSMCSE.

Symbole	Signification	Renvoi
RD	Rupture de dynamique.	sec. 4.2.1, p. 76
DRH	Début de réponse hémodynamique.	sec. 4.2.1, p. 76
$t'_i$	Instant d'occurrence de la (i-1)-ème transition du paradigme.	sec. 6.1, p. 89
$P$	Nombre de transitions « off-on » dans le paradigme, transition fictive incluse.	sec. 6.1, p. 89
$\mathcal{F} = \{f^v(\cdot)\}$	Ensemble des données IRMf. $f^v(\cdot)$ est le signal IRMf au voxel $v$ .	sec. 6.1, p. 89
$\mathcal{O} = \{\mathcal{O}^v\}$	Ensemble des séquences d'observation déterminé à partir de $\mathcal{F}$ . $\mathcal{O}^v$ est la séquence d'observation déterminée à partir de $f^v(\cdot)$ .	sec. 6.1, p. 89
$k$	Taux de détection. $k \times (P-2)$ RDs sont détectées en moyenne par signal IRMf sur l'ensemble $\mathcal{F}$ .	sec. 6.2.1, p. 90
$\mathcal{F}_i, \mathcal{A}_i, \lambda_i$	$\mathcal{F}_i$ est l'ensemble de signaux déclarés actifs de $\mathcal{F}$ par le modèle d'activation $\lambda_i$ (i-ème passe de l'algorithme d'apprentissage et de cartographie). $\mathcal{A}_i$ est la carte d'activation associée à $\mathcal{F}_i$ .	sec. 6.4.1, p. 96
$\mathcal{F}_{opt}, \mathcal{A}_{opt}, \lambda_{opt}$	Résultats de l'algorithme d'apprentissage et de cartographie.	sec. 6.4.1, p. 96
$m$	Rapport signal sur bruit.	sec. 6.1, p. 89
$\mathcal{F}_{1a}, \mathcal{F}_{1b}$	Séquences d'images IRMf synthétiques (paradigme en blocs) avec différents rapports signal sur bruit ( $m=2$ pour $\mathcal{F}_{1a}$ , 1 pour $\mathcal{F}_{1b}$ ).	sec. 7.1.1, p. 101
$\mathcal{F}_{2a}, \mathcal{F}_{3a}, \mathcal{F}_{4a}, \mathcal{F}_{5a}$	Séquences d'images IRMf synthétiques (paradigmes en blocs) avec une grande variabilité de signaux actifs et un rapport signal sur bruit $m=2$ .	sec. 7.1.1, p. 101

TAB. 2 – Liste de notations pour la partie II relative à l'utilisation du MSMCSE en cartographie fonctionnelle cérébrale.

Symbole	Signification	Renvoi
MMCMSE	Modèle markovien caché de multiples séquences d'événements.	sec. 10, p. 137
$N$	Nombre de canaux d'observation.	sec. 10.1.1, p. 137
$\mathbf{E} = \{\mathbf{E}^{(i)}\}_{1 \leq i \leq N}$	Ensemble des séquences d'événements. $\mathbf{E}^{(i)}$ est la séquence d'événements sur le $i$ -ème canal d'observation.	sec. 10.1.1, p. 137
$\boldsymbol{\tau}^{(i)}$	Instants de détection associés à $\mathbf{E}^{(i)}$ .	sec. 10.1.1, p. 137
$evpo$	Événement vrai positif observé.	sec. 10.1.2, p. 138
$evp\emptyset$	Événement vrai positif manquant.	sec. 10.1.2, p. 138
$efp$	Événement faux positif.	sec. 10.1.2, p. 138
$u$	Temps associé au processus profond.	sec. 10.1.3, p. 138
$\mathcal{S}$	Scénario explicatif de l'ensemble des séquences d'événements $\mathbf{E}$ .	sec. 10.1.3, p. 138
$\tau_u$	Instant d'occurrence du $u$ -ème événement profond.	sec. 10.1.3, p. 138
$\boldsymbol{\tau}_u^* = \{\tau_u^{(i)}\}_{1 \leq i \leq N}$	Ensemble des instants d'occurrence des événements de surface associé au $u$ -ème événement profond. $\tau_u^{(i)}$ est l'instant d'occurrence sur le $i$ -ème canal d'observation de l'événement de surface associé au $u$ -ème événement profond.	sec. 10.1.3, p. 138
$\mathbf{m}_u^* = \{m_u^{(i)}\}_{1 \leq i \leq N}$	Ensemble des mesures des événements de surface associé au $u$ -ème événement profond. $m_u^{(i)}$ est la mesure sur le $i$ -ème canal d'observation de l'événement de surface associé au $u$ -ème événement profond.	sec. 10.1.3, p. 138
$\mathbf{s}_u = \{\boldsymbol{\tau}_u^*, \mathbf{m}_u^*\}$	Signature associé au $u$ -ème événement profond.	sec. 10.1.3, p. 138
$\mathbf{O} = \{\mathbf{O}_u\}_{1 \leq u \leq U}$	Séquence d'observation d'un MMCMSE. On a par définition : $\mathbf{O}_u = \{\tau_u, \mathbf{s}_u, \mathbf{o}_{u-1,u}^*\}$ .	sec. 10.1.4, p. 139
$\mathbf{o}_{u-1,u}^*$	Fausse alarmes détectées sur les $N$ canaux d'observation entre $\mathbf{s}_{u-1}$ et $\mathbf{s}_u$ .	sec. 10.1.4, p. 139
$C$	$C - 2$ représente le nombre de classes d'événements profonds.	sec. 10.2.1, p. 140
$\mathcal{Q} = \{q_i\}_{1 \leq i \leq C}$	Espace d'état du MMCMSE.	sec. 10.2.1, p. 140
$\mathbf{A} = \{a_{ij}\}_{1 \leq i < C, 1 < j \leq C}$	Matrice de transition du MMCMSE.	sec. 10.2.1, p. 140
$\mathbf{w} = (w_1 \dots w_N)^T$	Poids des canaux.	sec. 10.2.1, p. 140
$\mathbf{B} = \{b_{ij\mathbf{w}}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$	Lois d'observation.	sec. 10.2.1, p. 140
$\boldsymbol{\lambda} = \{\mathcal{Q}, \mathbf{A}, \mathbf{B}\}$	Paramètres du MMCMSE. Le vecteur de poids $\mathbf{w}$ n'est pas inséré dans le vecteur de paramètres. Il est supposé connu.	sec. 10.2.1, p. 140

TAB. 3 – Liste de notations pour la partie III relative au MMCMSE.



Symbole	Signification	Renvoi
$\omega_k = \{\tau_k, \tau_k^*, m_k^*, n_k\}$	couple observation-état.	sec. 10.3.2.1, p. 143
$\mathcal{W} = \{\omega_1, \dots, \omega_K\}$	Ensemble des couples observation-état.	sec. 10.3.2.1, p. 143
$K$	Cardinal de l'ensemble $\mathcal{W}$ .	sec. 10.3.2.1, p. 143
$\tau_k$	Temps en profondeur associé au couple $\omega_k$ . En toute rigueur, la notation $\tau_k$ n'est pas satisfaisante car cette dernière est déjà utilisée (cf. $\tau_u$ ). Cependant, quand les vecteurs sont indicés par $u$ , on suppose que l'on fait obligatoirement référence au $u$ -ème événement du processus profond.	sec. 10.3.2.1, p. 143
$\tau_k^*$	Ensemble des instants d'occurrence des événements de surface associé au couple $\omega_k$ . En toute rigueur, la notation $\tau_k^*$ n'est pas satisfaisante (voir remarque pour $\tau_k$ ).	sec. 10.3.2.1, p. 143
$m_k^*$	Ensemble des mesures des événements de surface associé au couple $\omega_k$ . En toute rigueur, la notation $m_k^*$ n'est pas satisfaisante (voir remarque pour $\tau_k$ ).	sec. 10.3.2.1, p. 143
$n_k$	Etat caché associé au couple $\omega_k$ .	sec. 10.3.2.1, p. 143
$\mathcal{V}$	Ensemble des associations entre événements de surface. Cet ensemble permet de déterminer l'ensemble $\mathcal{W}$ .	sec. 10.3.2.2, p. 144
$\psi_c(\cdot)$	fonction injective de $[1, l_c]$ dans $[1, K]$ permettant de définir un chemin caché $c_\psi$ .	sec. 10.3.2.4, p. 145
$\gamma_i, \gamma_i^{-1}$	Reformulation locale de la notion de chemin caché ( $i \in [1, K]$ ).	sec. 10.3.2.5, p. 145

TAB. 4 – Liste de notations pour la partie III relative à la stratégie de résolution des problèmes d'inférence.

Symbole	Signification	Renvoi
RD	Rupture de dynamique.	sec. 4.2.1, p. 76
DRH	Début de réponse hémodynamique.	sec. 4.2.1, p. 76
DRHI	Début de réponse hémodynamique d'intérêt.	sec. 9.3, p. 134
$t'_i$	Instant d'occurrence de la (i-1)-ème transition du paradigme.	sec. 6.1, p. 89
$P$	Nombre de transitions « off-on » dans le paradigme, transition fictive incluse.	sec. 6.1, p. 89
$\mathcal{F} = \{f^v(\cdot)\}$	Ensemble des données IRMf. $f^v(\cdot)$ est le signal IRMf au voxel $v$ .	sec. 6.1, p. 89
$k$	Taux de détection. $k \times (P-2)$ RDs sont détectées en moyenne par signal IRMf sur l'ensemble $\mathcal{F}$ .	sec. 6.2.1, p. 90
$\mathbf{E}^v$	Ensemble des séquence d'événements associées au voxel $v$ .	sec. 11.1, p. 155
$l$	Retard d'activité.	sec. 11.3.2.2, p. 160
$m$	Rapport signal sur bruit.	sec. 6.1, p. 89
$\mathcal{F}_{1a}, \mathcal{F}_{1b}, \mathcal{F}_{1c}, \mathcal{F}_{1d}$	Séquences d'images IRMf synthétiques (paradigme en blocs) avec différents rapports signal sur bruit ( $m=2$ pour $\mathcal{F}_{1a}$ , 1 pour $\mathcal{F}_{1b}$ , 0.5 pour $\mathcal{F}_{1c}$ et 0.25 pour $\mathcal{F}_{1d}$ ).	sec. 12.1.1.1, p. 171
$\mathcal{F}_{2b}, \mathcal{F}_{3b}, \mathcal{F}_{4b}, \mathcal{F}_{5b}$	Séquences d'images IRMf synthétiques (paradigme en blocs) avec une grande variabilité de signaux actifs et un rapport signal sur bruit $m=1$ .	sec. 12.1.1.1, p. 171
$\mathcal{F}_{Eb}, \mathcal{F}_{Ec}, \mathcal{F}_{Ed}$	Séquences d'images IRMf synthétiques (paradigme événementiel) avec différents rapports signal sur bruit ( $m=1$ pour $\mathcal{F}_{Eb}$ , 0.5 pour $\mathcal{F}_{Ec}$ et 0.25 pour $\mathcal{F}_{Ed}$ ).	sec. 13.1.1, p. 187
$\mathbf{a}_i$	Détecteur déterminant une séquence de RDs par signal IRMf. Ce détecteur est donc caractérisé par un unique taux de détection $k$ .	sec. 12.2.1.2.1, p. 180
$\mathbf{b}_i$	Détecteur déterminant deux séquences de RDs par signal IRMf. Ce détecteur est donc caractérisé par deux taux de détection $k$ .	sec. 12.2.1.2.1, p. 180
$\mathbf{c}_i$	Détecteur déterminant trois séquences de RDs par signal IRMf. Ce détecteur est donc caractérisé par trois taux de détection $k$ .	sec. 12.2.1.2.1, p. 180
$C_0, C_6$ et $C_{26}$	Connexité 0 (aucune prise en compte du voisinage), 6-connexité et 26-connexité.	sec. 12.2.1.2.1, p. 180
$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$	Trois algorithmes différents de création de l'ensemble $\mathcal{W}$ classé du moins contraint au plus contraint.	sec. 12.2.1.2.1, p. 180

TAB. 5 – Liste de notations pour la partie III relative à l'utilisation du MSMCSE en cartographie fonctionnelle cérébrale.

# Introduction générale

## Cadre de travail, Objectifs

L'imagerie par résonance magnétique fonctionnelle (IRMf) tient une place prépondérante parmi les différentes techniques permettant d'imager l'activité cérébrale. Elle possède de nombreux avantages dont le fait d'être non invasive et de présenter une bonne résolution spatiale et temporelle. L'IRMf cérébrale permet en particulier de cartographier les différentes fonctions cérébrales, en d'autres termes de localiser les aires impliquées dans la réalisation d'un processus sensoriel, moteur ou cognitif. Elle permet également de caractériser l'activité cérébrale en termes d'amplitude et de décours temporel du signal IRMf observé en réponse à une stimulation externe.

L'analyse de séquences d'images IRMf se confronte à différents problèmes. En premier lieu, appréhender le signal IRMf reste difficile tant d'un point de vue physique que d'un point de vue biologique. En particulier, la nature de l'information portée par le signal IRMf fait toujours débat. Ensuite, le signal IRMf cérébral présente un faible rapport signal à bruit auquel s'ajoute une forte variabilité anatomo-fonctionnelle intra- et inter-individus. Enfin, les nombreux prétraitements appliqués aux données brutes, tel que le recalage spatial et temporel des images, sont susceptibles de modifier le contenu informatif du signal IRMf originel.

Les méthodes actuelles de cartographie IRMf se répartissent en deux grandes classes : les méthodes inférentielles et les méthodes exploratoires. Les méthodes inférentielles s'appuient le plus souvent sur une modélisation paramétrique du signal IRMf. La détection d'activation cérébrale repose alors sur un test statistique des paramètres estimés du modèle. Ces méthodes sont généralement univariées, les signaux IRMf cérébraux étant traités indépendamment les uns des autres. Leur principal désavantage tient à l'utilisation d'un modèle non adapté aux données, risque d'autant plus élevé compte tenu de la forte variabilité du signal actif en termes de forme de la réponse hémodynamique et de chronologie d'activation cérébrale. Les méthodes exploratoires cherchent, quant à elles, à extraire des composantes d'intérêt à partir des données IRMf. Leur principal avantage réside dans le fait de ne pas faire d'hypothèses, ni sur la forme, ni sur la chronologie du signal IRMf actif. En revanche, l'interprétation des résultats reste délicate.

Le travail présenté dans cette thèse vise à développer et à valider de nouvelles méthodes en cartographie IRMf cérébrale prenant en compte la grande variabilité du signal IRMf actif. Dans ce cadre, les approches privilégiées sont de nature inférentielle, et plus précisément de type modèles de Markov cachés. Rarement utilisée dans le contexte de la détection d'activation cérébrale, la modélisation markovienne cachée se révèle bien adaptée à l'analyse du signal IRMf. Il s'agit en effet d'analyser un processus *caché*, le processus d'activation neuronale, au travers d'un processus aléatoire *observable* qui lui est corrélé : le signal IRMf cérébral.

Ce travail de recherche a été mené en étroite collaboration entre le Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (ULP / UMR CNRS 7005) et l'Institut

de Physique Biologique (ULP - Hôpitaux Universitaires / UMR CNRS 7004) dans le cadre du projet multi-laboratoires « Imagerie et Robotique Médicale et Chirurgicale » (EPML IRMC).

## Contributions

Deux méthodes de cartographie ont été proposées, toutes deux fondées sur une modélisation markovienne cachée de la distribution temporelle des ruptures de dynamique (RDs) détectées dans le signal IRMf. Le nombre d'hypothèses *a priori* est ainsi fortement restreint permettant aux modèles statistiques de représenter une grande variété de signaux actifs. De plus, les capacités d'apprentissage et de généralisation des modèles de Markov cachés permettent aux deux méthodes de cartographier les aires fonctionnelles cérébrales de façon non supervisée.

Notre première méthode de cartographie repose sur le principe suivant : en présence d'activation au voxel  $v$ , la séquence de RDs détectée dans le signal IRMf associé à  $v$  s'aligne, dans une certaine mesure sur la séquence de transitions « off-on » du paradigme d'activation dans le cas d'un paradigme en blocs, ou sur la séquence de stimuli du paradigme dans le cas d'un paradigme événementiel. Les deux séquences d'événements, ainsi que leur couplage temporel, sont modélisés par un modèle de Markov caché spécialement adapté : le modèle semi-markovien caché de séquence d'événements (MSMCSE).

La seconde méthode de cartographie repose sur le même principe d'alignement temporel sur le paradigme d'activation. Cependant, afin de gagner en robustesse, ce principe est étendu aux séquences de RDs détectées dans le voisinage spatial du voxel  $v$  et à une approche de détection de ruptures multidétecteurs. En définitif, à un voxel est désormais associé un ensemble de séquences de RDs. La problématique d'alignement de ces multiples séquences d'événements sur le paradigme d'activation devient complexe. Elle relève, par les asynchronismes entre événements détectés, par l'absence possible de détections, ainsi que par la multiplicité des canaux d'observation à prendre en compte, de la fusion de données multicapteurs. Pour la résoudre, un nouveau type de modèles de Markov cachés a été développé : le modèle de Markov caché de multiples séquences d'événements (MMCMSE).

Les deux méthodes de cartographie ont été appliquées sur données synthétiques et données réelles, ainsi que pour des paradigmes en bloc et des paradigmes événementiels. Les résultats obtenus sont comparés avec ceux obtenus par la méthode faisant référence dans le domaine, en l'occurrence SPM2 (Statistical Parametric Mapping). Ils valident la stratégie d'ensemble consistant à analyser le signal IRMf au travers de ses non-stationnarités. Ils soulignent par ailleurs l'aptitude des modèles markoviens cachés développés (MSMCSEs et MMCMSEs) à prendre en compte la forte variabilité du signal IRMf.

En dehors des développements théoriques sur les modèles statistiques, les deux méthodes de cartographie ont été implantées dans la plate-forme logicielle d'analyse d'images médicales *Medimax* de l'IPB (Institut de Physique Biologique) et du LSIIT (Laboratoire des Sciences de l'Image de l'Informatique et de la Télédétection). Cette plate-forme est accessible aux médecins et chercheurs de ces deux équipes de recherche.

## Organisation du document

Le mémoire de thèse est divisé en trois parties. La première partie décrit le contexte et l'objectif principal des travaux réalisés. Tout d'abord, après une brève introduction sur le système nerveux, nous présentons l'IRMf du point de vue de l'expérimentation, des sciences

physiques et des neurosciences. Nous mettons ainsi en valeur la pluridisciplinarité de ce domaine, pluridisciplinarité que le traicteur d'images se doit d'appréhender pour définir des algorithmes pertinents. Nous présentons ensuite un état de l'art des méthodes de cartographie en IRMf. Nous insistons en particulier sur les principales difficultés rencontrées lors de l'analyse des séquences d'images fonctionnelles et sur les limitations des méthodes proposées dans la littérature. L'introduction des modèles de Markov cachés dans ce contexte est innovant. Nous nous concentrons ensuite sur un état de l'art général de ces représentations statistiques. Nous commençons par présenter le modèle markovien caché standard (la chaîne de Markov cachée), les problèmes d'inférence associés et leur résolution. En raison de ses limitations, le modèle standard est inadapté à la modélisation du signal IRMf cérébral d'activation et à la modélisation de processus en interaction (cas de signaux IRMf voisins). Par conséquent, des modèles plus sophistiqués sont ensuite présentés.

La seconde partie de la thèse est consacrée à la présentation de la première méthode de cartographie proposée. Dans un premier temps, nous rappelons les objectifs et les principes de modélisation. Nous présentons ensuite de façon détaillée le modèle markovien utilisé : le modèle semi-markovien caché de séquence d'événements (MSMCSE). Ce modèle markovien, développé par L. Thoraval [208], a été utilisé pour l'analyse du signal électrocardiographique. Il est dédié à l'analyse de processus aléatoires de nature événementielle. Le processus observable n'est plus de nature segmentale mais issu d'une étape de détection et de caractérisation des événements d'intérêt. Nous voyons ensuite comment ce modèle peut s'appliquer à la cartographie en IRMf cérébrale. Nous présentons et nous discutons enfin les résultats obtenus. Bien que donnant des résultats prometteurs, nous notons que la méthode proposée est très sensible au bruit. Nous tentons alors d'en cerner les causes.

La troisième partie constitue la principale contribution de la thèse, c'est à dire la présentation du modèle de Markov caché de multiples séquences d'événements (MMCMSE) et sa première application dans le cadre de la cartographie en IRMf. Le MMCMSE est un modèle probabiliste à deux niveaux d'information, dédié à la modélisation et à l'analyse d'un processus aléatoire composé d'événements non directement observables, le processus profond. Ce dernier est observé au travers de  $N$  processus stochastiques de nature événementielle, les processus de surface. Les processus de surface sont corrélés car ils correspondent à des observations différentes du processus profond. Ils apportent ainsi des informations complémentaires ou redondantes sur une même réalité. Nous présentons une première application de ce modèle dans le cadre de l'IRMf. Le processus profond du MMCMSE est le processus neuronal prenant place au voxel  $v$  et les processus de surface sont les signaux IRMf observés au voxel  $v$  et dans son voisinage. Les résultats sur données synthétiques et réelles montrent clairement que la méthode proposée est plus robuste au bruit que la méthode qui fait référence dans le domaine (Statistical Parametric Mapping) tout en étant représentative d'un nombre de signaux actifs beaucoup plus important.



Première partie

Cadre de travail





# Introduction

Cette partie présente les concepts fondamentaux de notre domaine d'étude, l'imagerie par résonance magnétique fonctionnelle (IRMf), et de la modélisation statistique utilisée, la modélisation markovienne cachée. A ce titre, cette partie permet de mieux appréhender les travaux de thèse présentés dans la seconde et troisième partie du manuscrit.

Dans un premier chapitre, nous nous intéressons à l'IRMf. Nous présentons brièvement le rôle, la composition et le fonctionnement du système nerveux. Nous voyons notamment que le cortex se divise en différentes régions fonctionnelles. Nous citons ensuite les grandes méthodes permettant d'observer l'activité cérébrale. Parmi celles-ci, nous nous intéressons en particulier à l'IRMf par contraste BOLD (Blood Oxygenation Level Dependent). Dans le but de comprendre pourquoi l'activité fonctionnelle d'une zone corticale engendre une augmentation locale du signal IRMf, nous étudions les principes physiques de l'IRM. Nous regardons enfin comment un protocole en IRMf se déroule en pratique.

Le second chapitre aborde un état de l'art des méthodes de cartographie en IRMf. Nous y distinguons notamment les approches inférentielles des approches exploratoires. Les méthodes inférentielles nécessitent de faire des hypothèses restrictives sur la forme et sur la chronologie du signal IRMf actif. Les méthodes exploratoires fournissent, quant à elles, des résultats dont l'interprétation reste délicate. Dans le but de développer de nouvelles méthodes de cartographie en IRMf cérébrale qui prennent en compte la grande variabilité du signal IRMf actif, nous avons privilégié une approche de type inférentiel basée sur les modèles de Markov cachés.

Le troisième et dernier chapitre est une synthèse des modèles markoviens cachés. Les principaux concepts nécessaires à la compréhension des parties suivantes y sont donc présentés. Nous commençons par présenter la chaîne de Markov cachée pour finir par les réseaux bayésiens dynamiques et leur application en fusion de données.



# Chapitre 1

## L'imagerie par résonance magnétique fonctionnelle cérébrale

Ce premier chapitre traite de l'imagerie par résonance magnétique fonctionnelle (IRMf) du cerveau. Le principe de l'IRMf consiste à acquérir une séquence d'images volumiques (en trois dimensions) du cerveau pendant que le patient effectue différentes tâches selon un protocole très précis : le paradigme d'activation. Cette technique permet ainsi d'observer le fonctionnement du cerveau *in vivo* dans l'espace et dans le temps. Elle dévoile ainsi l'architecture spatio-temporelle des fonctions supérieures du cerveau. On sait en effet depuis le milieu du XIX<sup>e</sup> siècle que les différentes fonctions du cerveau sont localisées anatomiquement. L'IRMf permet de cartographier précisément ces différentes zones et d'en caractériser l'activité en termes d'amplitude et de décours temporel de la réponse.

Dans une première section, nous présentons succinctement le système nerveux, son rôle, sa composition et son fonctionnement. La seconde section aborde les principales techniques permettant d'imager l'activité cérébrale. Dans une dernière section, nous nous intéressons plus spécifiquement à l'IRMf. Nous exposons brièvement son principe tant du point de vue de l'expérimentation que du point de vue des sciences physiques et des neurosciences.

### 1.1 Le système nerveux

#### 1.1.1 Rôle

Le système nerveux permet à un organisme de percevoir le monde qui l'environne et d'agir sur lui. Ses récepteurs sensoriels lui fournissent tout d'abord des informations sur les changements (*stimuli*) qui se produisent à l'intérieur et à l'extérieur de l'organisme. Certains récepteurs sont en effet sensibles aux stimuli provenant de l'environnement extérieur à l'organisme (récepteurs des organes des sens) alors que d'autres sont sensibles à ceux produits à l'intérieur de l'organisme (changement chimique dans les viscères ou les vaisseaux, étirement des tendons et des muscles). L'information délivrée par les récepteurs est appelée *information sensorielle*. Le système nerveux interprète cette dernière, élabore l'action à entreprendre et fournit finalement une réponse motrice permettant d'activer les muscles ou les glandes. Ainsi, lorsqu'un automobiliste voit un feu rouge devant lui (information sensorielle), son système nerveux analyse cette information (le feu rouge implique l'arrêt) et fournit finalement la réponse motrice adéquate (enfoncer la pédale de frein).

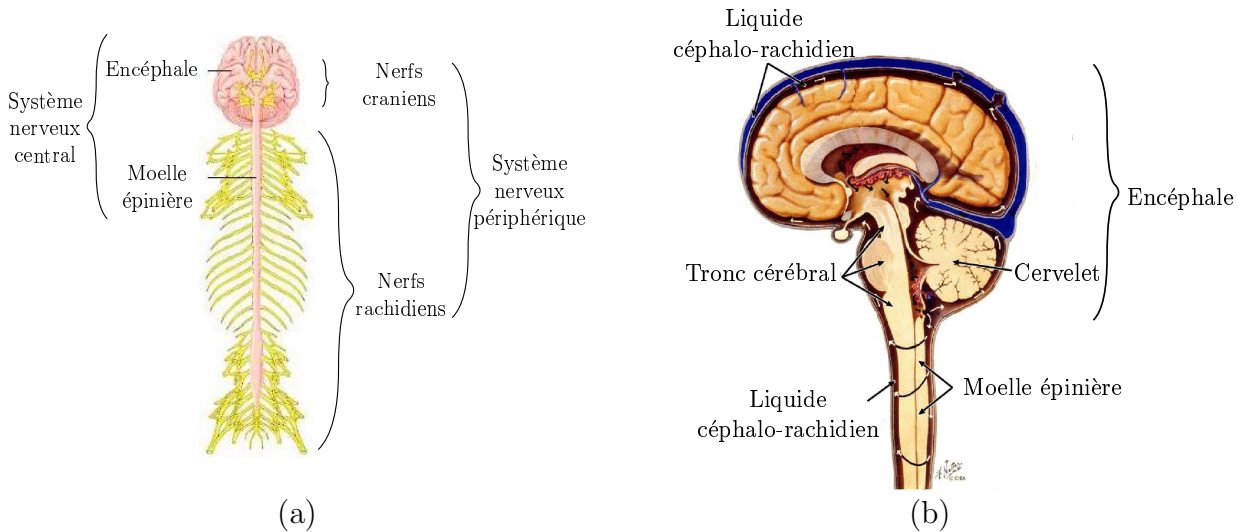


FIG. 1.1 – Anatomie du système nerveux. (a) Le système nerveux est composé du système nerveux central, en rose, et du système nerveux périphérique, en jaune (illustration d'après [mhhe.com/biosci/esp/2001\\_saladin/folder\\_structure/](http://mhhe.com/biosci/esp/2001_saladin/folder_structure/), modifiée). (b) Le système nerveux central (illustration d'après [167], modifiée).

### 1.1.2 Structures anatomo-fonctionnelles du système nerveux

D'un point de vue anatomique, le système nerveux se divise en deux grandes parties : le système nerveux central (SNC) et le système nerveux périphérique (SNP) (figure 1.1.a). Le SNC est composé de l'encéphale (cerveau, tronc cérébral) et de la moelle épinière, l'ensemble flottant dans un coussin protecteur : le liquide céphalo-rachidien (figure 1.1.b). Le SNP, quant à lui, est formé principalement de nerfs qui permettent de relier le SNC à tous les organes du corps humain. Les nerfs peuvent être issus de l'encéphale (nerfs crâniens) ou de la moelle épinière (nerfs rachidiens) (figure 1.1.a).

D'un point de vue fonctionnel, le SNC interprète les informations sensorielles et élabore une action sensée alors que le SNP peut être perçu comme l'interface ou comme la ligne de communication entre l'organisme et le SNC. Il comprend deux types de voies :

- la voie sensitive, ou afférente qui permet de transporter vers le SNC les informations provenant des récepteurs sensoriels ;
- la voie motrice, ou efférente, qui permet de transmettre aux organes effecteurs (muscles, glandes) les influx provenant du SNC (ces influx étant à l'origine du déclenchement d'une réponse motrice adaptée).

L'organisation fonctionnelle du système nerveux est résumée figure 1.2. Les récepteurs sensoriels du SNP créent à partir des différents stimuli une information sensorielle qui est délivrée par les voies sensibles au SNC. Ce dernier interprète l'information sensorielle et élabore l'action motrice à entreprendre. Le SNP achemine finalement aux organes effecteurs le message qui sera à l'origine de la réponse motrice.

Pour comprendre comment l'information est codée, comment elle se propage, et comment elle est analysée, il est nécessaire de s'intéresser à l'unique type de cellules qui compose le système nerveux : les neurones. Ces derniers sont les unités fonctionnelles du système nerveux. Agencés en réseaux, ils sont spécialisés pour recevoir, intégrer et transmettre l'information.

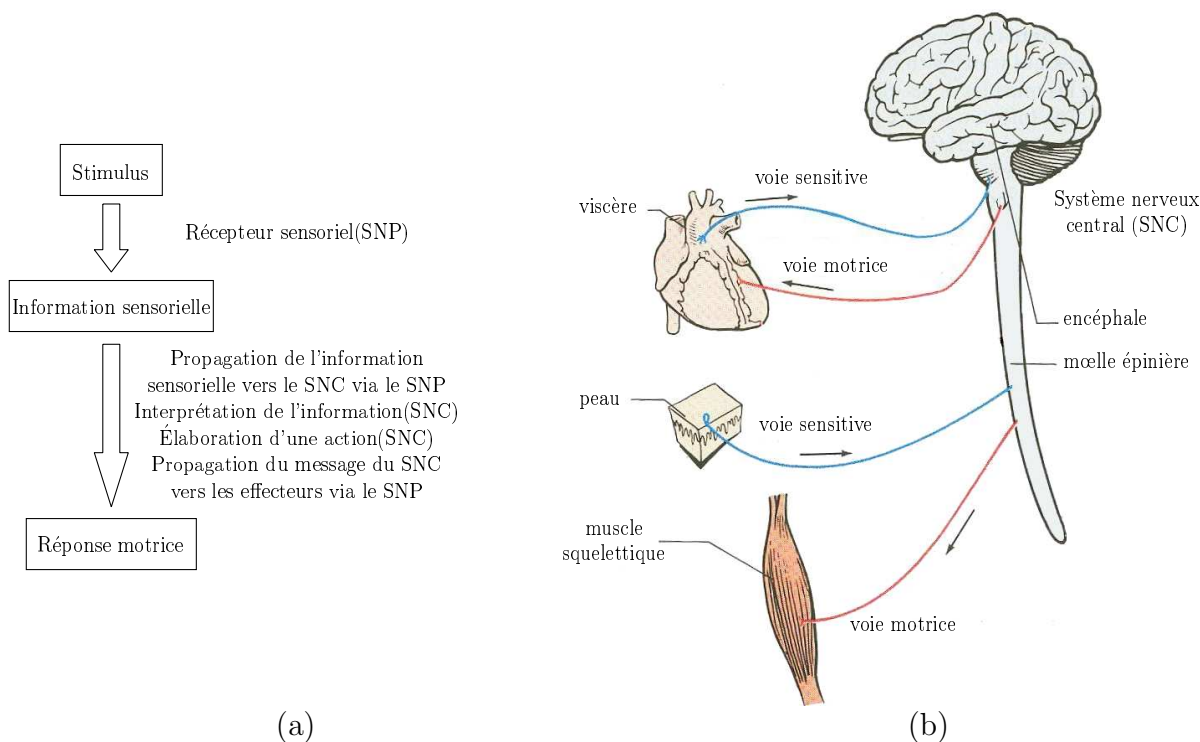


FIG. 1.2 – Organisation fonctionnelle du système nerveux. (a) Du stimulus au déclenchement d'une réponse motrice adaptée. (b) Organisation du système nerveux : en rouge, voie motrice, et en bleu, voie sensitive (illustration d'après [153], modifiée).

### 1.1.3 L'unité fonctionnelle du système nerveux : le neurone

D'un point de vue anatomique, les neurones comprennent un corps cellulaire<sup>1</sup>, dont sont issus un ou plusieurs prolongements. Il existe deux types de prolongements neuronaux : les dendrites et les axones (figure 1.3). La morphologie d'un neurone peut être très variable. Elle dépend essentiellement du nombre de prolongements issus du corps cellulaire. Cependant, la plupart des neurones possèdent trois structures fonctionnelles en commun : une structure réceptrice, les dendrites, une structure conductrice, les axones, et une structure sécrétrice, les terminaisons présynaptiques. L'information circule au sein du neurone et le long de l'axone sous forme d'une activité électrique le *potentiel d'action*, l'intensité du stimulus étant codée par la fréquence de transition de l'influx électrique. L'information est traduite au niveau des terminaisons présynaptiques (arborisation terminale) en intensité de libération d'une substance chimique, le neurotransmetteur. Le neurotransmetteur excite ou inhibe les neurones ou les cellules effectrices avec lesquels l'axone est en contact étroit. Par exemple, dans le cas d'une communication neurone à neurone, le neurotransmetteur se lie à des récepteurs spécifiques, sur le versant postsynaptique (dendrite) d'un autre neurone et induit un potentiel gradué dont le voltage est proportionnel à l'intensité du stimulus. Ce potentiel peut aboutir à la genèse de nouveaux potentiels d'action ou au contraire à leur inhibition. Les dendrites peuvent recevoir un très grand nombre de signaux provenant d'autres neurones grâce à l'immense surface qu'elles couvrent.

D'un point de vue fonctionnel, on distingue trois types de neurone :

<sup>1</sup>Le corps cellulaire est situé dans la plupart des cas à l'intérieur du SNC où il est protégé par les os du crâne ou par ceux de la colonne vertébrale. De plus, le neurone a besoin d'un tissu de soutien et d'environnement : la glie (cf. gaine de Myéline, figure 1.3).

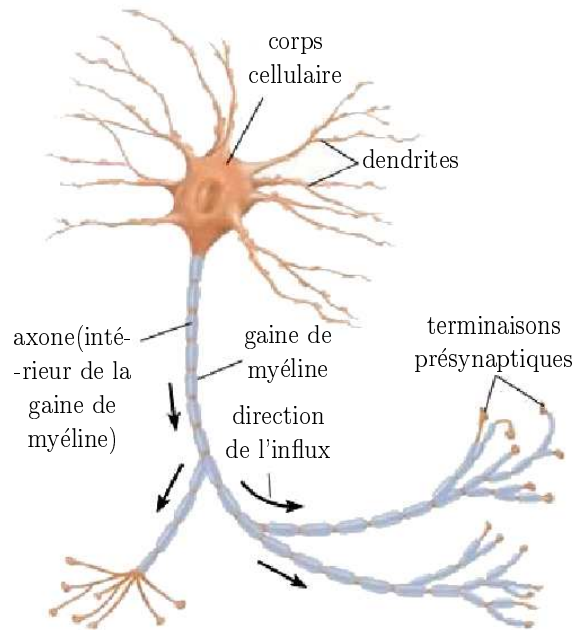


FIG. 1.3 – Structure d'un neurone (illustration d'après *driesen.com*, modifiée).

- Les neurones sensoriels captent les messages des récepteurs sensoriels et les communiquent au SNC.
- Les neurones moteurs conduisent la commande motrice du SNC aux muscles.
- Les interneurons sont situés entre les neurones sensoriels et les neurones moteurs. Ils représentent 99 % des neurones de l'organisme et on les retrouve en général dans le SNC.

Cependant, pour comprendre comment l'information sensorielle est interprétée et comment l'action à entreprendre est élaborée par le SNC, il ne faut pas considérer un unique neurone. En effet, le SNC a une organisation hiérarchique : les milliards de neurones du SNC sont répartis en groupes de neurones. Chacun des groupes rassemble l'information provenant des récepteurs ou d'autres groupes de neurones pour élaborer une nouvelle information. La plupart des groupes de neurones sont composés de milliers de neurones. Les capacités fonctionnelles des groupes sont déterminées principalement par la disposition des synapses dans les groupes de neurones, la synapse étant le lieu de connexion entre une terminaison axonale et une dendrite. Parmi ces groupes, on peut citer par exemple les réseaux divergents, convergents, parallèles postdécharges et réverbérants (figure 1.4). Dans le cas des réseaux divergents, un neurone déclenche des réponses dans un nombre toujours croissant de neurones. Par exemple, l'information provenant d'un récepteur sensoriel unique peut être transmise via la moelle épinière à diverses régions de l'encéphale.

#### 1.1.4 Les tissus cérébraux du SNC

Schématiquement, le SNC est formé par des amas de corps cellulaires de neurones (*substance grise*) interconnectés entre eux par l'intermédiaire des axones (*substance blanche*). La substance grise représente les centres de l'activité cérébrale alors que la substance blanche permet de relier ces différents centres entre eux.

Dans la moelle épinière ainsi que dans la majeure partie du tronc cérébral, la substance

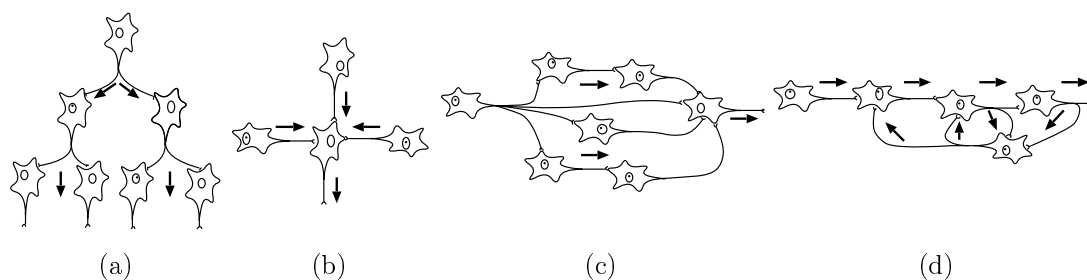


FIG. 1.4 – Illustration schématique d'un réseau divergent (a), convergent (b), parallèle postdécharge (c) et réverbérant (d) (illustration d'après [153], modifiée).

grise est répartie sur le centre alors que la substance blanche est placée en périphérie. Les hémisphères cérébraux et le cervelet présentent une structure beaucoup plus complexe. La substance grise des deux hémisphères cérébraux se retrouve dans le cortex et dans un ensemble de noyaux sous-corticaux (noyaux gris centraux, noyaux thalamiques, noyaux hypothalamiques, noyaux du cervelet). Le cortex peut être vu comme une sphère d'épaisseur relativement constante (autour de 1,7 mm) repliée sur elle-même de manière très intriquée, les bosses étant dénommées *gyri* et les creux *sulci*. Les deux hémisphères sont séparés partiellement par la fissure longitudinale. D'autres sillons et fissures divisent les hémisphères en lobes : le lobe frontal, le lobe pariétal, le lobe temporal, le lobe occipital et le lobe limbique (figure 1.5).

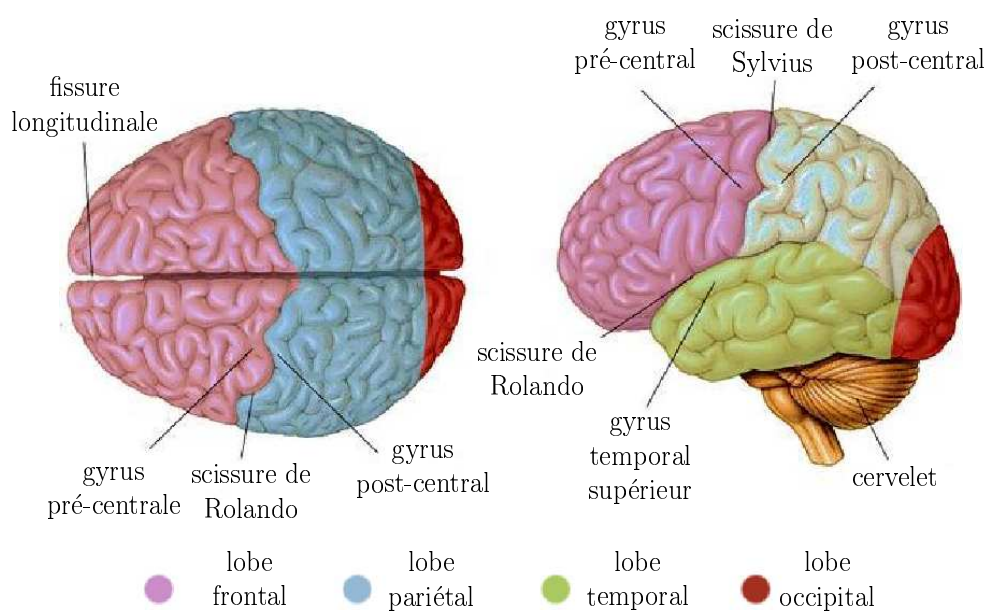
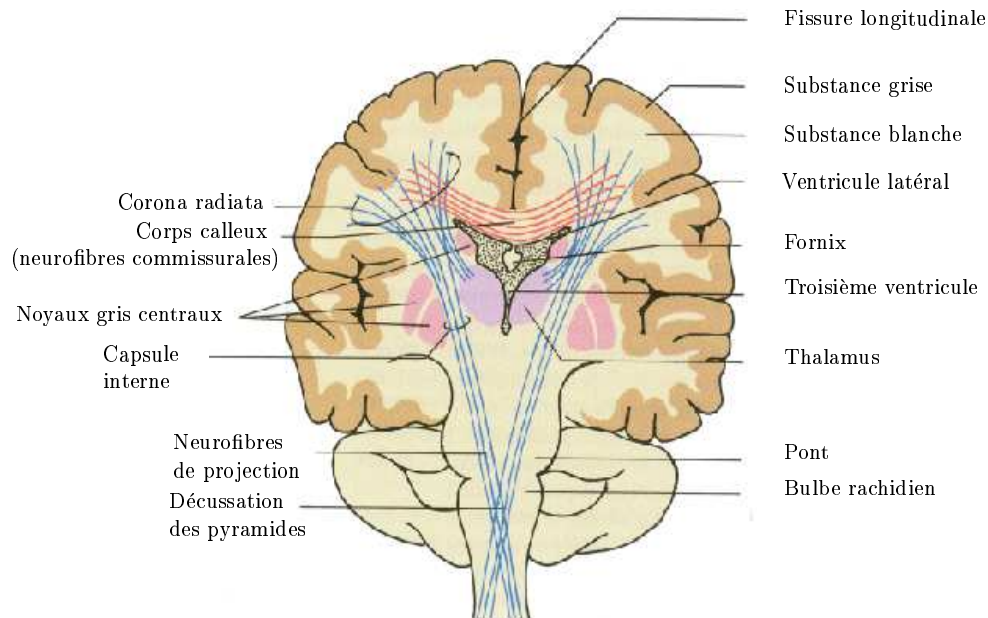


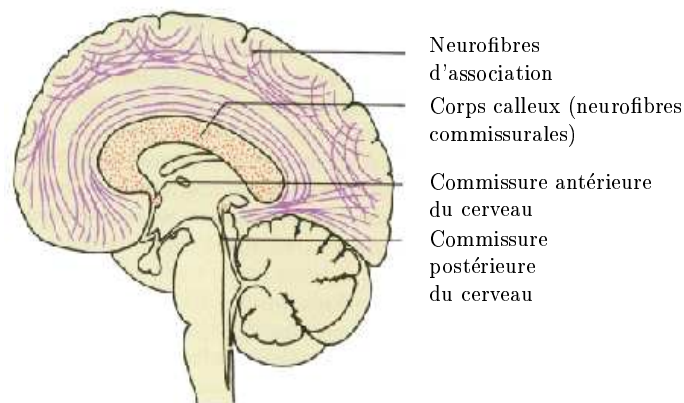
FIG. 1.5 – Les 5 lobes des hémisphères cérébraux, le lobe limbique n'est pas représenté (illustration d'après *driesen.com*, modifiée)

La substance blanche cérébrale relie les différents centres de l'activité cérébrale. Elle est constituée de neurofibres (les prolongements longs des neurones portent le nom de fibre nerveuse ou neurofibre) regroupées en faisceaux. Cet ensemble de faisceaux de fibres constitue des circuits d'information complexes. Suivant leur orientation, les neurofibres sont dites commissurales, d'association ou de projection (figure 1.6) :

- Les neurofibres commissurales (corps calleux et commissure antérieure du cerveau) forment les commissures qui relient les aires homologues des deux hémisphères et permettent ainsi leur coordination.
- Les neurofibres d'association transmettent les informations à l'intérieur d'un même hémisphère (liaison entre des lobes du cerveau ou entre des gyri adjacents).
- Les neurofibres de projection relient le cortex au reste du système nerveux.



(a)



(b)

FIG. 1.6 – Neurofibres composant la substance blanche. (a) Coupe sagittale médiane de l'hémisphère droit cérébral. (b) Coupe frontale de l'encéphale (illustration d'après [153], modifiée).

### 1.1.5 Les régions fonctionnelles du cortex

La substance grise se décompose en régions aux fonctions spécifiques bien localisées telles que le langage ou la vision. Le concept de régions fonctionnelles du cortex localisées anatomi-



quement fut mis au jour par le neurologue français Paul Broca au XIX<sup>e</sup> siècle. Il découvrit, à la mort de l'un de ses patients qui comprenait le langage mais ne pouvait pas parler, une petite lésion dans une région spécifique du cerveau. Il en déduisit que la production du langage devait être liée à cette région.

Aujourd'hui, nous savons que le cortex nous fournit entre autres nos facultés de perception, de communication, de mémorisation, de compréhension, de jugement et d'accomplissement des mouvements volontaires, ces facultés relevant du comportement conscient. Les deux hémisphères du cortex ne sont pas absolument égaux sur le plan fonctionnel. Il y a plutôt latéralisation, c'est-à-dire spécialisation du cortex de chaque hémisphère par rapport à certaines fonctions cérébrales. Le cortex de chacun des hémisphères cérébraux est par exemple essentiellement le siège de la perception sensorielle et de la régulation de la motricité volontaire du côté opposé du corps. La figure 1.7 présente les différentes régions fonctionnelles du cortex cérébral. Les numéros indiquent les aires définies par Brodmann <sup>2</sup>. Le cortex cérébral renferme trois types d'aires fonctionnelles :

- les aires motrices qui président à la fonction motrice volontaire ;
- les aires sensibles qui permettent les perceptions sensorielles ;
- les aires associatives qui comprennent toutes les aires corticales qui ne sont pas qualifiées par l'adjectif « primaire ».

Les aires sensibles primaires sont situées à proximité des aires associatives avec lesquelles elles communiquent (cf. neurofibres d'association, figure 1.6.a) Les aires associatives servent principalement à intégrer les diverses informations sensorielles afin d'envoyer des commandes motrices aux effecteurs musculaires et glandulaires. Ainsi, elles communiquent également entre elles et avec l'aire motrice. La figure 1.8 présente de manière plus détaillée l'organisation fonctionnelle de l'aire somesthésique et motrice primaire.

## 1.2 Modalités d'observation de l'activité cérébrale

Jusque dans les années 80, la cartographie des aires fonctionnelles se basait sur l'étude conjointe des lésions, du comportement et des performances intellectuelles. Les techniques d'imagerie fonctionnelle non traumatiques permettent aujourd'hui d'appréhender l'architecture spatio-temporelle des fonctions du cerveau. Cependant, il reste à donner une définition à l'activité cérébrale. La mesure de l'activité cérébrale peut être associée à l'activité neuronale, à l'activité métabolique ou à des mesures effectuées sur le système cérébrovasculaire dont notamment l'imagerie du volume ou du débit sanguin. Les liens entre ces différents types d'activité sont largement étudiés dans la littérature. Par exemple, il existerait une relation linéaire entre l'activité synaptique et le débit sanguin et une relation non linéaire entre le débit sanguin et le signal BOLD (Blood Oxygenation Level Dependent) [164]. L'effet BOLD est présenté section 1.2.3, page 17.

---

<sup>2</sup>L'anatomiste allemand Korbinian Brodmann a établi au début du XX<sup>e</sup> siècle une carte cérébrale basée sur les différences d'architecture cellulaire des différentes régions du cortex. Chaque région du cortex ayant la même organisation cellulaire a reçu un numéro allant de 1 à 52. L'intuition de Brodmann, qui s'est vue fréquemment confirmée par la suite, était qu'à une organisation anatomique donnée correspond une fonction particulière.

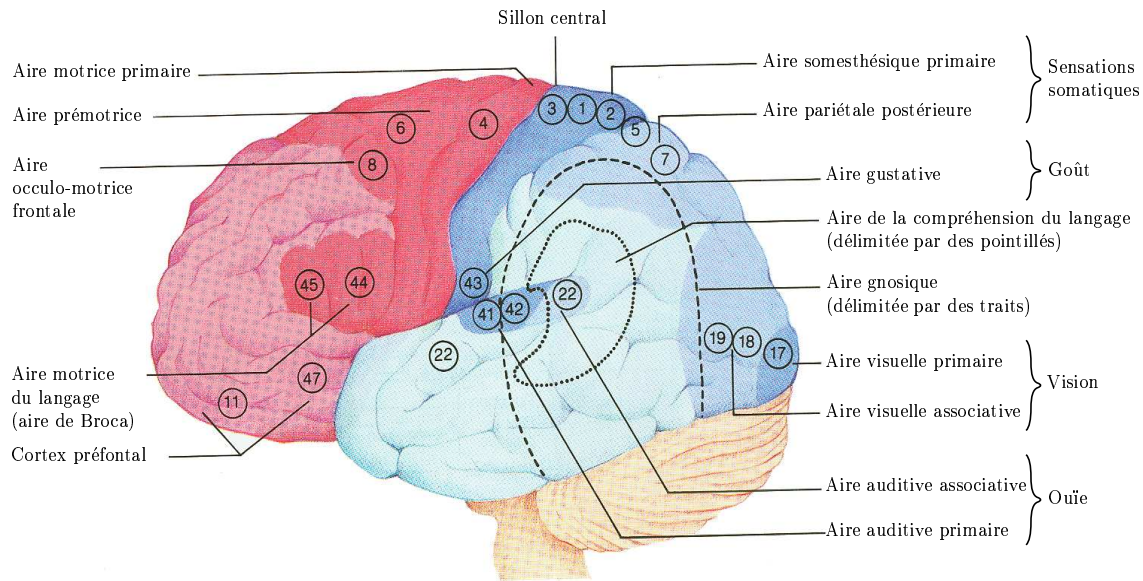


FIG. 1.7 – Aires fonctionnelles du cortex cérébral gauche (illustration d'après [153], modifiée).

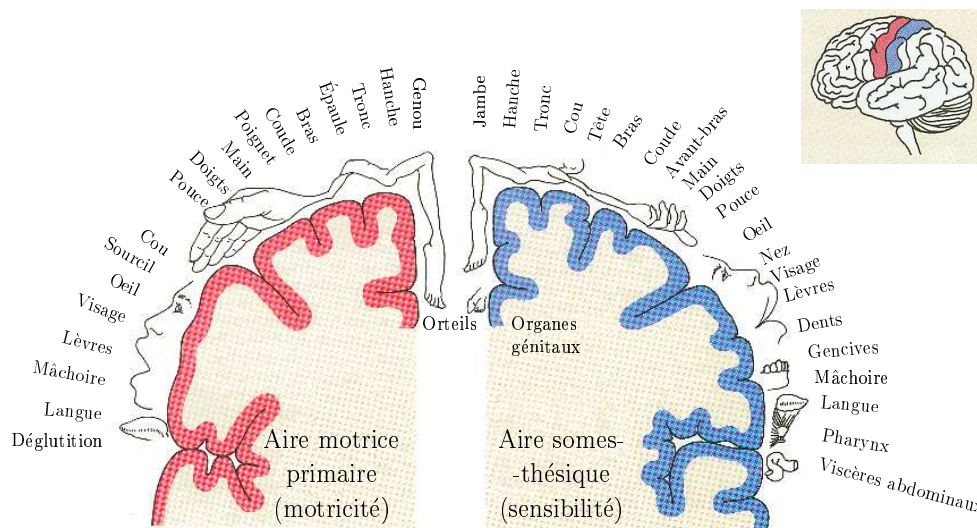


FIG. 1.8 – Aires sensibles et motrices du cortex cérébral gauche (illustration d'après [153], modifiée).

### 1.2.1 Mesure de l'activité neuronale

L'activité neuronale peut se mesurer en termes de transfert d'information entre les différents neurones. Les potentiels d'action donnent lieu à des champs quadripolaires non détectables de façon externe. Les potentiels gradués, quant à eux, engendrent un champ électromagnétique dipolaire détectable par électro ou magnéto encéphalographie.

L'électroencéphalographie (EEG) mesure le champ électrique, à l'aide d'électrodes placées sur le scalp du sujet alors que la magnétoencéphalographie (MEG) mesure le champ magnétique (figure 1.9). L'EEG et la MEG possèdent une très bonne résolution temporelle (de l'ordre de la milliseconde) mais une très mauvaise résolution spatiale (de l'ordre du cen-

timètre)<sup>3</sup>. De plus, seule l'activité corticale peut être observée.

La transmission du champ électromagnétique dans le cerveau, l'os du crâne et le scalp est plus simple en MEG qu'en EEG. Cependant, la MEG nécessite un appareillage beaucoup plus lourd et coûteux. Les deux méthodes restent néanmoins complémentaires dans la mesure où l'EEG est très sensible aux sources radiales (perpendiculaires à la surface du cortex) alors que la MEG s'intéresse aux sources tangentielles uniquement.



FIG. 1.9 – Système d'acquisition en EEG (a) et en MEG (b) (Illustration d'après [web.ccr.jussieu.fr/meg-center/media/cours%20Line/CoursLine01.pdf](http://web.ccr.jussieu.fr/meg-center/media/cours%20Line/CoursLine01.pdf))

### 1.2.2 Mesure de l'activité métabolique

L'activité neuronale se traduit par une consommation d'énergie sous forme d'adénosine triphosphate (ATP) nécessitant un apport continu de glucose et d'oxygène. La mesure locale du métabolisme du glucose est donc une mesure de l'activité cérébrale locale.

Les changements du métabolisme du glucose peuvent être observés par tomographie par émission de positons (TEP). Pour cela, on peut utiliser le fluoro-deoxy-glucose marqué au  $F^{18}$ . Le principe de la TEP consiste à introduire des molécules biologiques (molécules traceuses) marquées par des isotopes radioactifs produits dans un cyclotron. Ces derniers, chargés positivement réagissent avec les électrons du tissu biologique et émettent un rayonnement gamma détectable par une caméra à positons. On suit ainsi le trajet des molécules traceuses. Différentes fonctions de l'organisme peuvent être explorées selon le choix de la molécule traceuse.

### 1.2.3 Mesures basées sur le système cérébrovasculaire

L'existence d'un couplage entre l'activité cérébrale et le système cérébrovasculaire est maintenant établie. Cependant, les mécanismes qui sont à l'origine de ce couplage restent un sujet de controverse [104]. Il y a tout de même un consensus pour dire que l'activité d'une population neuronale engendre localement une augmentation (figure 1.10) :

1. du volume sanguin (vasodilatation) ;
2. de la vitesse du sang ;
3. du débit sanguin (conséquence des points 1. et 2.) ;

<sup>3</sup>La position des zones cérébrales actives est déterminée connaissant les mesures effectuées sur le champ magnétique ou électrique en différents endroits. La résolution de ce problème inverse nécessite de fortes hypothèses simplificatrices.

4. du taux d'oxygène dans le sang ;
5. de la consommation d'oxygène ;
6. du rapport oxyhémoglobine sur désoxyhémoglobine (effet BOLD, Blood Oxygenation Level Dependent). L'effet BOLD sera expliqué en détail par la suite.

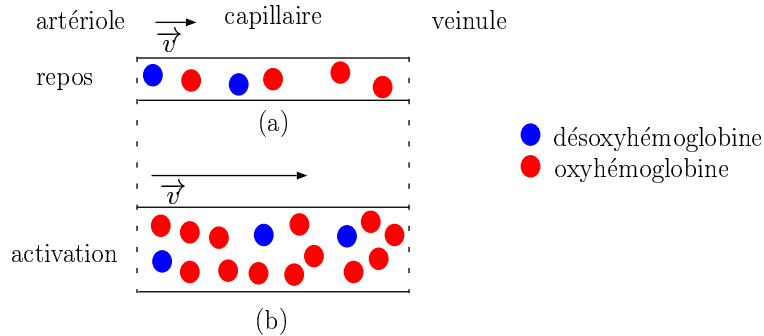


FIG. 1.10 – Représentation schématique du couplage entre activité et système cérébrovasculaire. (a) inactivité (b) activité.

La TEP permet par exemple d'établir une carte de quantification du débit sanguin cérébral en utilisant de l'eau marquée ( $H_2^{15}O$ ).

L'imagerie par résonance magnétique fonctionnelle (IRMf) permet d'estimer localement le volume sanguin, le débit sanguin et le rapport oxyhémoglobine sur désoxyhémoglobine (effet BOLD). L'IRMf par effet BOLD est la plus répandue. Elle est présentée en détail dans la prochaine section.

## 1.2.4 Comparaison

Les différentes méthodes peuvent être classées selon leur résolution temporelle et spatiale. Les résolutions temporelles sont de l'ordre de la milliseconde pour l'EEG et la MEG, de la seconde pour l'IRMf et enfin de la minute pour la TEP. Les résolutions spatiales sont de l'ordre du centimètre pour l'EEG et la MEG, de quelques millimètres pour la TEP et enfin du millimètre pour l'IRMf.

Les différences importantes (résolution temporelle et spatiale, phénomène observé) entre les méthodes basées sur la mesure du champ électromagnétique (EEG et MEG) et les méthodes basées sur des imageurs classiques (TEP et IRMf) font que ces méthodes sont souvent complémentaires justifiant des études basées par exemple sur l'enregistrement conjoint du signal EEG et du signal IRMf. Par la suite, nous allons uniquement comparer l'IRMf et la TEP.

La TEP a deux inconvénients majeurs. D'une part, les isotopes utilisés ont une durée de vie courte (environ 2 minutes pour l'oxygène  $^{15}O$  et de 100 minutes pour le fluor  $^{18}F$ ). Cela impose souvent leur fabrication sur le lieu même de l'étude, ceci nécessitant d'avoir un cyclotron. De plus, dans le cas de l'oxygène  $^{15}O$ , il est nécessaire d'effectuer plusieurs acquisitions (toutes les 15 minutes environ). D'autre part, la TEP a une faible résolution temporelle (l'acquisition d'une image dure plusieurs minutes) limitant son intérêt pour l'étude de nombreux processus cognitifs. Pour toutes ces raisons, l'IRMf a pris l'avantage ces dernières années sur la TEP. Son principal défaut provient du fait que les bases physiologiques sont encore mal connues. L'IRMf, contrairement à la TEP, ne permet pas une mesure quantitative des phénomènes observés.

Cette technique impose enfin des contraintes fortes liées à l'environnement d'utilisation (fort champ magnétique, bruit important).

## 1.3 L'IRMf par contraste BOLD

L'IRMf par contraste BOLD est la modalité IRMf d'observation la plus répandue aujourd'hui. En révélant l'activité cérébrale, elle permet de cartographier de manière précise les différentes fonctions cérébrales, c'est-à-dire de localiser les aires impliquées dans la réalisation d'un processus. Elle se base sur le fait que le signal IRMf par contraste BOLD augmente avec l'activité cérébrale. Nous nous concentrons tout d'abord sur les principes physiques de l'acquisition de l'image. Nous expliquons ensuite l'augmentation fonctionnelle du signal IRMf. Nous rentrons enfin plus en détail sur la mise en œuvre d'un protocole en IRMf.

### 1.3.1 Principe physique de l'IRM

L'imagerie par résonance magnétique (IRM) est fondée sur la résonance magnétique nucléaire (RMN). Ce phénomène est expliqué par la suite à travers l'exemple du noyau l'hydrogène, le signal RMN provenant principalement des atomes d'hydrogène présents dans le corps, sous forme de molécules d'eau dans les tissus.

#### 1.3.1.1 Principe physique de la résonance magnétique nucléaire

Les noyaux possèdent un moment magnétique propre : le *spin* noté  $j$ . Ce dernier est quantifié et ne peut prendre que deux valeurs distinctes dans le cas du noyau d'hydrogène :  $j$  vaut soit  $\frac{1}{2}$  (état 1) soit  $-\frac{1}{2}$  (état 2). A chacun de ces deux états est associée une énergie distincte  $E_1$  et  $E_2$  ( $E_1 < E_2$ ).

Lorsqu'un noyau est placé dans un champ magnétique  $B_o \vec{u}_z$ , son moment magnétique décrit un cercle à la fréquence de Larmor  $f_o = \gamma B_o$  (le rapport gyromagnétique  $\gamma$  dépend de la nature du noyau). Ce cercle est dans un plan orthogonal à l'axe  $\vec{u}_z$ . On dit que le moment magnétique a un mouvement de *précession* autour de l'axe  $\vec{u}_z$  (figure 1.11.a). Le

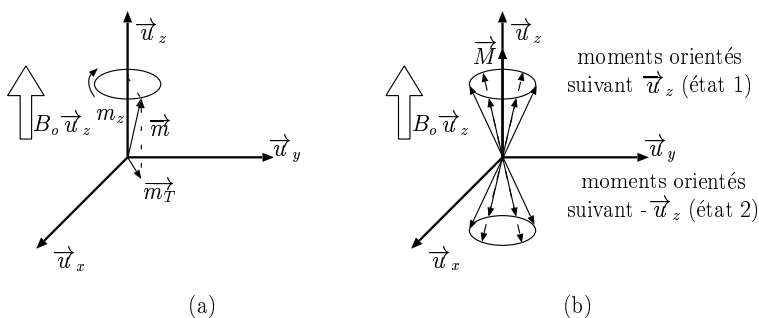


FIG. 1.11 – (a) Mouvement de précession du moment magnétique d'un noyau d'hydrogène placé dans le champ  $B_o \vec{u}_z$  (b) Ensemble de noyaux d'hydrogène placés dans le champ  $B_o \vec{u}_z$ . Chaque moment magnétique précesse autour de l'axe  $\vec{u}_z$ . Le moment magnétique total  $\vec{M}$  est la somme de tous les moments magnétiques et il est égal à  $\vec{M} = M_{zeq} \vec{u}_z$  (voir texte)

moment magnétique de chaque noyau  $\vec{m}$  peut se décomposer suivant deux projections (cf. figure 1.11.a), l'une suivant  $\vec{u}_z$  notée  $m_z$  (moment magnétique longitudinal) et l'autre dans

le plan  $(O; \vec{u}_x, \vec{u}_y)$  notée  $\vec{m}_T$  (moment magnétique transversal). Le moment magnétique longitudinal est, soit orienté suivant  $B_o \vec{u}_z$  (état 1) ou soit orienté suivant  $-\vec{u}_z$  (état 2). Enfin,  $\vec{m}_T$  décrit un mouvement circulaire.

Le moment magnétique macroscopique  $\vec{M}$  est la somme de tous les moments magnétiques. Comme le nombre de noyaux d'hydrogène dans l'état 1 ( $N_1$ ) est supérieur au nombre de noyaux d'hydrogène dans l'état 2 ( $N_2$ )<sup>4</sup>, le moment magnétique longitudinal est dans le même sens que  $\vec{u}_z$ . De plus,  $\vec{M}_T$  est statistiquement le vecteur nul car les moments magnétiques des différents noyaux d'hydrogène tournent indépendamment les uns des autres (on dit que les moments magnétiques sont déphasés sur le cône de précession). En définitive, le moment magnétique macroscopique  $\vec{M}$  est colinéaire et dans le même sens que  $\vec{u}_z$  (cf figure 1.11.b). Pour la suite, on note  $\vec{M} = M_{zeq} \vec{u}_z$ .

Il est possible de faire passer certains spins des noyaux d'hydrogène de l'état 1 à l'état 2 en les stimulant avec une impulsion d'énergie  $E_2 - E_1 = h\gamma|B_o|$  ou de fréquence  $f_o$  ( $E = h\nu$ ). En pratique, l'énergie est apportée à travers un champ magnétique  $B_1 \cos(\omega_o t) \vec{u}_x$  réalisé à l'aide d'une bobine parcourue par un courant alternatif. Si l'énergie est suffisante, on peut aboutir à la configuration pour laquelle  $N_1 = N_2$ . Dans ce cas, le moment magnétique longitudinal est nul. De plus, comme le champ alternatif rephase les moments magnétiques sur le cône de précession, on assiste à l'apparition d'un moment magnétique transverse tournant à la fréquence de Larmor (cf figure 1.12.b). On note  $\|\vec{M}_{T_{heq}}\|$  le module du champ magnétique transverse.

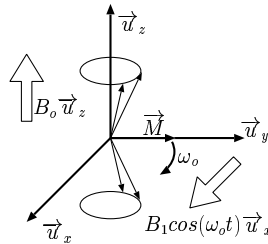


FIG. 1.12 – Condition de résonance. Le moment magnétique total transverse tourne dans le plan  $(O; \vec{u}_x, \vec{u}_y)$  à la fréquence de Larmor.

Lorsque l'application de  $B_1$  cesse, le moment macroscopique évolue spontanément, selon un phénomène de relaxation vers sa position initiale. La norme du moment magnétique longitudinal passe donc de 0 à  $M_{zeq}$  et la norme du moment magnétique transverse passe de  $\|\vec{M}_{T_{heq}}\|$  à 0. On peut montrer que le retour à l'équilibre s'effectue exponentiellement avec une constante de temps  $T_1$  (temps de relaxation spin-réseau ou longitudinal) pour  $M_z$  et avec une constante de temps  $T_2$  (temps de relaxation spin-spin ou transversal) pour  $\|\vec{M}_T\|$ .

La résonance magnétique nucléaire permet ainsi d'apporter des informations sur les temps de relaxation  $T_1$ ,  $T_2$  et sur la densité de protons. Ces informations permettent par exemple de discriminer les principaux tissus imagés (IRM anatomique).

$T_2$ , de par sa définition, est une « mesure » du temps que mettent les moments des spins pour se déphaser. Or, le déphasage s'accélère à cause des hétérogénéités du champ magnétique. Ainsi, on parle également du temps de relaxation  $T_2^*$  ( $T_2 > T_2^*$ ).

<sup>4</sup>La distribution de Boltzmann s'écrit :  $\frac{N_2}{N_1} = e^{-\frac{\Delta E}{kT}}$  où  $k$  est la constante de Boltzmann,  $T$  la température absolue du milieu contenant les spins,  $\Delta E$  est la différence d'énergie entre les deux niveaux, c'est-à-dire  $h\gamma B_o$ . Enfin,  $N_1 + N_2$  est fixé.

### 1.3.1.2 Formation d'une image

La formation d'une image nécessite la localisation spatiale du signal émis. La méthode la plus répandue est le codage spatial par transformée bi- (coupe par coupe) ou tri-dimensionnelle de Fourier. L'acquisition tri-dimensionnelle présente de nombreux avantages par rapport à une acquisition coupe par coupe (meilleur rapport signal sur bruit, coupes très fines jointives). Cependant, le temps d'acquisition est beaucoup plus important.

D'une manière générale, une image IRM est obtenue en appliquant une séquence de gradients de champ. Par exemple, dans le cas du codage par transformée bi-dimensionnelle de Fourier, la coupe est sélectionnée suivant le principe suivant : au lieu d'appliquer un champ magnétique uniforme, on applique un champ qui croît linéairement suivant l'axe  $z$  (champ de gradient). Ainsi, la fréquence de résonance dépendra de  $z$ . A chaque valeur de l'impulsion radio fréquence, seuls les noyaux d'hydrogène situés sur le plan  $z = cte$  rentrent en résonance. Ce principe permet donc de sélectionner les coupes les unes après les autres.

De manière générale, les différentes séquences sont construites de manière à privilégier la mesure du  $T_1$ ,  $T_2$ ,  $T_2^*$  ou de la densité de protons au détriment des autres.

Dans le cadre de l'IRMf, le temps d'acquisition de l'image doit être suffisamment court pour permettre d'étudier le fonctionnement dynamique du cerveau. Dans ce but, la séquence à trains d'échos de gradients (EPI, Echo Planar Imaging) a été développée, cette dernière permettant d'échantillonner le plan de Fourier en une seule impulsion radiofréquence. Pour palier les défauts de cette séquence (sensibilité aux hétérogénéités du champ et à la qualité des gradients de champ), la séquence à trains d'échos de spins RARE (Rapid Acquisition Relaxation Enhancement) a été introduite. L'acquisition d'une image RARE ou EPI dure en général entre 2 et 5 secondes alors que la taille d'un voxel est de l'ordre de 4x4x4 mm.

### 1.3.2 L'effet BOLD

Il est admis aujourd'hui que l'activité d'une population neuronale engendre localement une augmentation du rapport oxyhémoglobine sur désoxyhémoglobine (effet BOLD, Blood Oxygenation Level Dependent). Le transport de l'oxygène dans le sang est majoritairement (98%) assuré par l'hémoglobine, chaque molécule d'hémoglobine fixant 4 molécules d'oxygène. Lorsque l'hémoglobine est saturée en oxygène, elle est appelée oxyhémoglobine. Après avoir libéré l'oxygène dans les tissus, elle change de conformation et elle est appelée désoxyhémoglobine. L'effet BOLD s'oppose donc au principe de parcimonie dans la mesure où il implique le constat suivant : en situation d'activité, l'augmentation locale de la consommation d'oxygène est plus petite que l'augmentation de l'apport en oxygène.

L'oxygène contenu dans l'hémoglobine (oxyhémoglobine HbO<sub>2</sub>) étant diamagnétique, il a une susceptibilité magnétique faible et ne perturbe donc que très peu le champ magnétique. La désoxyhémoglobine, quant à elle est paramagnétique à cause des quatre électrons du fer non appariés et a ainsi une forte susceptibilité magnétique. La différence de susceptibilité entre les capillaires contenant de la désoxyhémoglobine et le milieu interstitiel induit un gradient local de champ (figure 1.13.c). La perturbation locale de champ entraîne une diminution des temps de relaxation  $T_2$  et  $T_2^*$  suivant deux processus complexes : la diffusion (pour  $T_2$  et  $T_2^*$ ) et le déphasage intra-voxel (pour  $T_2^*$  uniquement). L'augmentation fonctionnelle du rapport oxyhémoglobine sur désoxyhémoglobine induit une diminution de l'intensité des champs perturbateurs et donc un rallongement du temps de relaxation transversal  $T_2$  et  $T_2^*$  et donc une augmentation du signal IRMf. La figure 1.13 rappelle les points importants qui font que l'augmentation de l'activité neuronale entraîne l'augmentation du signal IRMf.

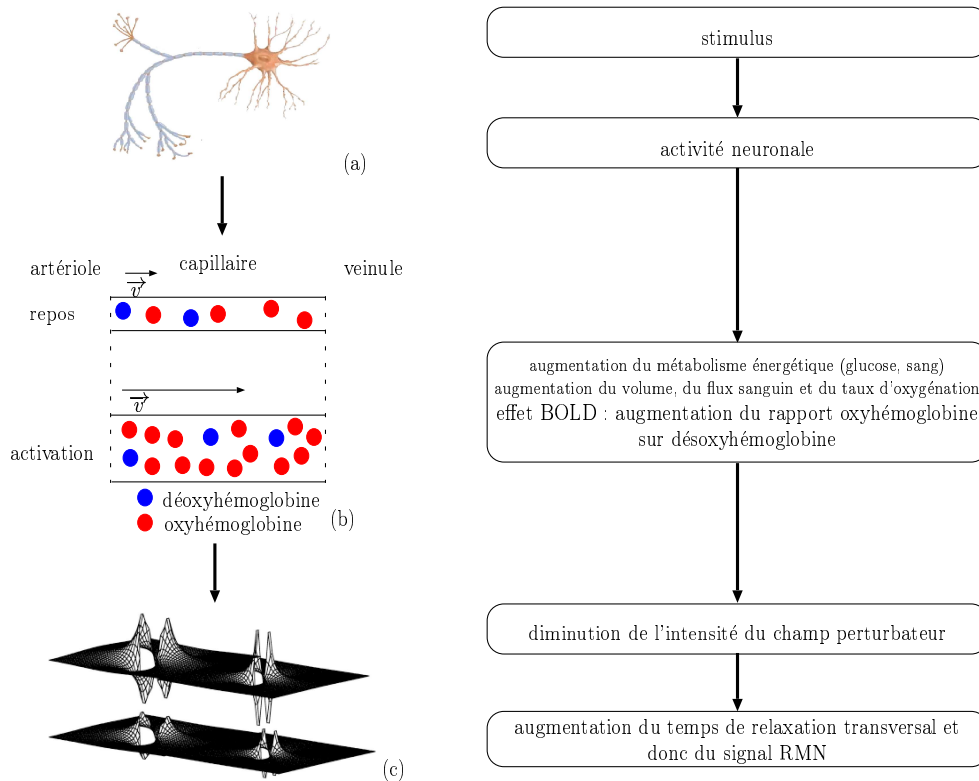


FIG. 1.13 – De l'augmentation de l'activité cérébrale à l'augmentation du signal IRM. (a) Le neurone (illustration d'après *driesen.com*, *modifiée*). (b) Représentation schématique du couplage entre activité cérébrale et système cérébrovasculaire. (c) Perturbations du champ magnétique statique induites autour de deux vaisseaux par des différences de susceptibilité magnétique entre les espaces intra- et extra-vasculaires. Les deux vaisseaux sont orientés verticalement, mais n'ont pas été dessinés explicitement. Haut : condition au repos. Bas : condition d'activation (illustration tirée de <http://www-u594.ujf-grenoble.fr/people/mdojat/papers/sfp03.pdf>).

### 1.3.3 Mise en œuvre : du paradigme d'activation à la carte d'activité cérébrale

Nous décrivons ici comment les principes développés précédemment sont pratiquement mis en œuvre pour cartographier l'activation cérébrale. D'abord et avant tout, le patient est placé au centre d'un scanner IRM. Durant toute la période de stimulation, il lui est demandé de ne pas ou peu bouger. La stimulation du patient se déroule alors suivant un protocole très précis, alternant phases de repos et phases d'activation, appelé communément *paradigme d'activation*. Simultanément à la stimulation, une séquence d'images en 3 dimensions du cerveau est acquise. On obtient une séquence temporelle d'images 3-D. A chaque voxel  $V_{ijk}$  du volume cérébral est associé un signal IRMf  $S_{ijk}$  composé de  $T$  échantillons  $\{S_{ijk}(1), S_{ijk}(2), \dots, S_{ijk}(T)\}$ ,  $T$  désignant la taille, en nombre d'images, de la séquence acquise.

De façon très générale, le processus (cognitif, moteur, sensoriel) d'intérêt est caractérisé par les tâches soumises au sujet pendant les phases de repos et d'activation. Une région cérébrale est dite impliquée dans le traitement du processus d'intérêt dès lors que les groupes neuronaux constituant la région sont « actifs » pendant les phases d'activation et « inactifs » pendant les



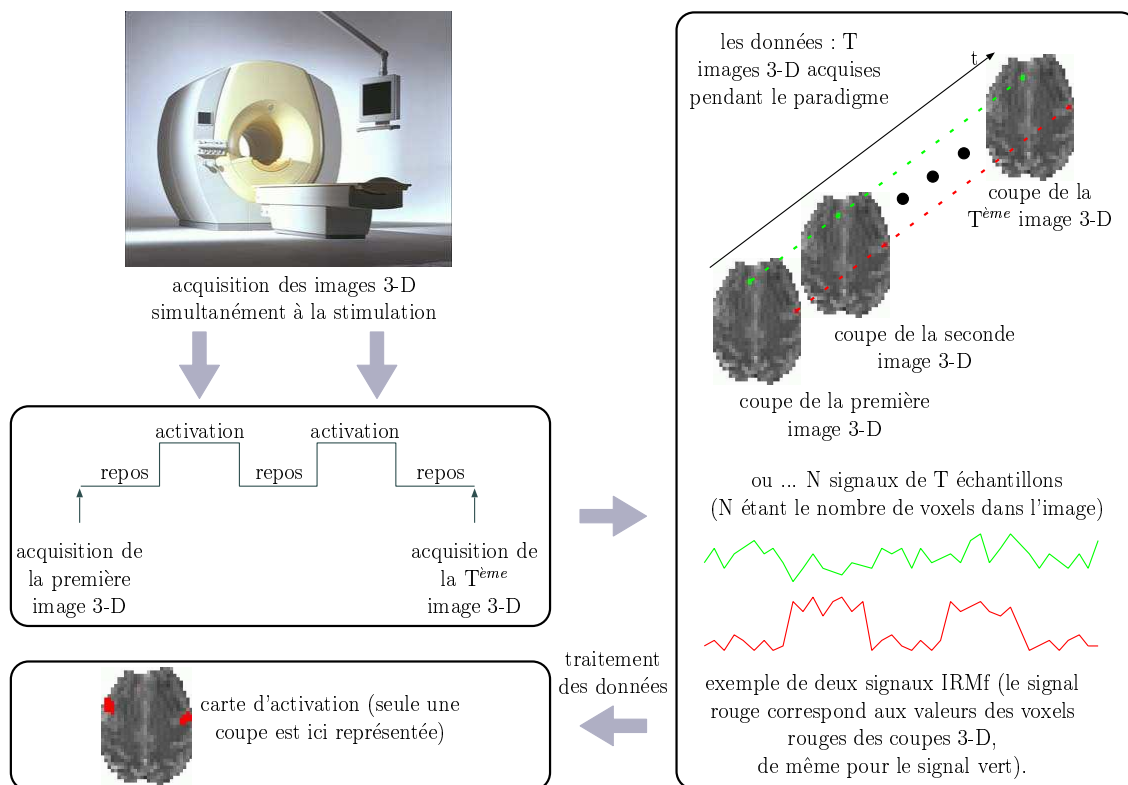


FIG. 1.14 – Fondements de l'IRMf. L'image d'un scanner IRM est représentée en haut à gauche (illustration d'après [www.ulb.ac.be/philo/phonolab/fmri.html](http://www.ulb.ac.be/philo/phonolab/fmri.html)). Le paradigme d'activation, représenté en dessous est décrit en terme de successions de tâches. A droite, sont représentées les données qui sont acquises pendant le protocole. Elles peuvent être, soit considérées comme une séquence d'images 3-D ( $T$  images de  $N$  voxels) ou comme  $N$  signaux de  $T$  échantillons. Le signal rouge étant corrélé avec le paradigme, ce signal est dit actif. Le voxel associé à ce dernier délimite une zone corticale participant au traitement de la tâche d'intérêt et non à celle de contrôle. La carte d'activation est représentée en bas à gauche (les zones associées au processus cognitif d'intérêt sont en rouge).

phases de repos. Le signal  $S_{ijk}$  est représentatif de l'activité du groupe neuronal correspondant au voxel  $V_{ijk}$ . Comme le signal IRMf augmente avec l'activité cérébrale, le signal  $S_{ijk}$ , et par conséquent la zone associée  $V_{ijk}$ , sont déclarés actifs dès lors que le signal  $S_{ijk}$  présente une forte corrélation avec le paradigme (cf. signal rouge de la figure 1.14). Des méthodes de traitement du signal permettent ainsi de localiser les aires impliquées dans la réalisation du processus d'intérêt à partir du paradigme et des signaux IRMf  $S_{ijk}$ . Ces méthodes fournissent une carte appelée indifféremment *carte d'activation* ou *carte d'activité*, sur laquelle figurent les régions cérébrales ayant participé au traitement du processus d'intérêt (voir figure 1.14).

### 1.3.4 Types de paradigmes d'activation cérébrale

Deux questions prévalent lors de l'établissement du paradigme d'activation : quelles sont les tâches que devra effectuer le patient ? Quand est-ce que le patient devra effectuer les différentes tâches ? Nous présentons les éléments de réponses à la seconde question, puis à la première.

### 1.3.4.1 Le déroulement temporel du paradigme d'activation

Il existe deux grands types de paradigme du point de vue du déroulement temporel : les paradigmes en blocs et les paradigmes événementiels.

Un paradigme en blocs est décrit en termes de succession d'états. Les conceptions en blocs isolent en effet différents processus cognitifs dans des périodes de temps distinctes (figure 1.15). Il s'agit entre autres de définir la durée des différents blocs. La durée d'une tâche (tâche de

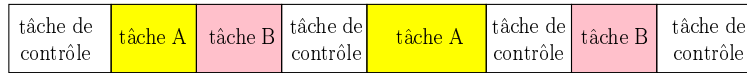


FIG. 1.15 – Exemple de paradigme en blocs avec deux tâches d'intérêt (A et B) et la tâche de contrôle

contrôle ou d'intérêt) est de l'ordre d'une trentaine de secondes (soit environ six à dix images acquises). Pour éviter des phénomènes d'anticipation du patient à un bloc de stimulation, les paradigmes en blocs périodiques sont rarement utilisés. Le choix de la durée se fait en partie à partir des caractéristiques fréquentielles du bruit et de la réponse hémodynamique<sup>5</sup> [87], la réponse hémodynamique étant la réponse impulsionnelle du système linéaire modélisant le signal IRMf.

Les paradigmes événementiels sont décrits en termes de succession de stimuli.

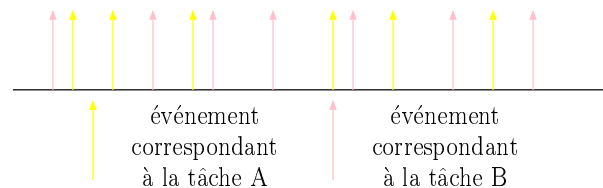


FIG. 1.16 – Exemple de paradigme événementiel avec deux types de stimuli

Les avantages des paradigmes événementiels ont largement été discutés dans la littérature [54][193][119]. Ils permettent entre autres :

- d'éliminer des phénomènes perturbatoires tels que les phénomènes d'habituation, d'anticipation et les effets liés à la mise en place d'une stratégie de traitement de la tâche par le patient. Ils permettent d'une manière générale d'isoler plus facilement les différents processus cognitifs ce qui facilite grandement le travail d'inférence sur le fonctionnement cérébral ;
- d'offrir une plus grande souplesse d'utilisation. De par le grand nombre d'événements, l'analyse peut se restreindre à une catégorie de réponses déterminée *a posteriori* selon la performance du sujet et du temps de réaction. Ils permettent également d'analyser des processus qui ne peuvent pas être maintenus dans le temps ou d'analyser des événements qui se passent de façon non prévisible, le patient indiquant par appui sur un bouton, l'instant d'apparition de ces derniers.

Cependant, le traitement des séquences d'images acquises avec des paradigmes événementiels nécessite une modélisation fine du signal attendu et des propriétés du bruit. De plus,

<sup>5</sup>Le bruit est très important aux basses fréquences (1/64 Hz et moins) alors que le maximum d'énergie de la réponse hémodynamique se trouve aux environs de 1/32 Hz.

la construction des paradigmes événementiels est complexe : il est nécessaire de définir le temps entre les différents stimuli noté ISI (interstimulus interval) ou SOA (stimulus onset asynchrony). La construction des paradigmes est ainsi devenue un sujet de recherche à part entière [119][45][105][146] [25][216].

Un paradigme est caractérisé par son pouvoir de détection, de caractérisation et sa prédictibilité. Son pouvoir de détection représente la capacité du paradigme à permettre une bonne détection des zones d'activité. Son pouvoir de caractérisation représente sa capacité à permettre une bonne estimation de la réponse hémodynamique. La prédictibilité du paradigme permet entre autres d'estimer les phénomènes d'habituation et d'anticipation du patient.

Dans le but de construire le paradigme, l'utilisateur définit les caractéristiques souhaitées du paradigme. Les méthodes de construction se placent ensuite toutes dans le cadre du modèle linéaire général (cf. partie II) mais elles se différencient :

- par les critères utilisés pour caractériser le paradigme ;
- par le nombre de tâches dans le paradigme ;
- par la modélisation du signal actif et du bruit (prise en compte des non linéarités du signal IRMF pour un ISI inférieur ou égal à 3 secondes) ;
- par l'algorithme d'optimisation utilisé.

Liu *et al.* [146] [144] ont démontré sous certaines conditions que les pouvoirs de détection et de caractérisation sont fortement liés. Par exemple, l'augmentation du pouvoir de détection entraîne une diminution du pouvoir de caractérisation.

Les paradigmes en blocs non périodiques ont un pouvoir de détection optimal mais un mauvais pouvoir de caractérisation. Les propriétés des paradigmes événementiels dépendent beaucoup de leur construction. Le choix aléatoire selon différentes densités de probabilité des différents ISIs permet d'améliorer le pouvoir de caractérisation mais diminue le pouvoir de détection. De plus, le choix aléatoire permet de diminuer les phénomènes liés à l'anticipation et à l'habituation.

Pour obtenir à la fois un fort pouvoir de détection et de caractérisation, le paradigme est construit selon un processus pseudo-aléatoire. Les paradigmes ainsi obtenus sont en général fortement non stationnaires : certaines parties sont dédiées à la caractérisation (ISI grand), d'autres à la détection (ISI petit  $\sim 3$  secondes) . On peut citer également l'existence de paradigmes mixtes (paradigmes en blocs mixés avec une séquence aléatoire ou pseudo-aléatoire de stimuli).

#### 1.3.4.2 Choix des tâches

La seconde difficulté consiste à définir les tâches que le patient doit effectuer. Ces dernières doivent permettre d'isoler l'activité liée au processus d'intérêt. Par la suite, on appellera *paire* un ensemble de deux tâches ou de deux stimuli et *processus* un mécanisme élémentaire induit par une tâche. Une tâche induit en général plusieurs processus : la tâche consistant à nommer un objet présenté sous forme d'une image peut induire un processus de traitement des formes, de traitement des couleurs, un processus sémantique, phonologique et d'articulation. La plupart des études sont basées sur le principe de la soustraction cognitive dont le principe est illustré figure 1.17. Le processus étudié est supposé être isolé en termes de différences entre deux tâches. Une tâche de contrôle induit différents processus alors qu'une seconde tâche d'intérêt induit les mêmes processus auxquels va cependant se rajouter le processus étudié. On conçoit donc une tâche A, dite de « contrôle », et une tâche B, impliquant le processus étudié et les processus de la tâche A. On s'intéresse alors aux variations A-B, c'est à dire aux variations entre les images acquises pendant la tâche A et celles acquises pendant la tâche

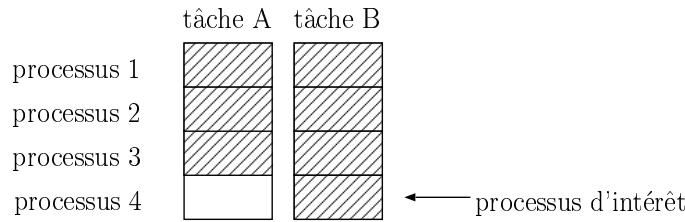


FIG. 1.17 – Soustraction cognitive : la tâche de contrôle (A) induit les processus 1,2 et 3. La tâche d'intérêt (B) induit les mêmes processus auxquels se rajoute le processus 4. Le processus 4 est le processus d'intérêt. L'hypothèse d'insertion pure stipule que l'apparition du processus 4 ne modifie pas la mise en œuvre des processus 1,2 et 3.

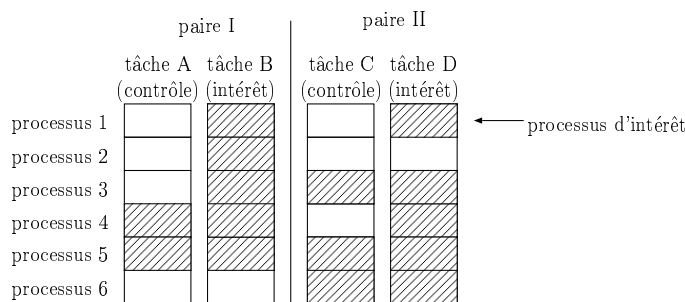


FIG. 1.18 – Conjonction cognitive : La paire I permet de mettre en évidence les processus 1-2 et 3 alors que la paire II met en évidence les processus 1 et 4. Ces deux paires ont en commun le processus 1 qui est le processus d'intérêt. Supposons que l'apparition du processus 2 modifie la mise en œuvre du processus 5. La zone corticale correspondant au processus 5, le processus 2 étant induit, sera détectée comme active au regard de la première expérience. Cependant, la paire II permet de ne pas considérer cette zone comme active. Ainsi, la conjonction cognitive permet de relâcher l'hypothèse d'insertion pure.

B. Si la définition de telles tâches n'est pas simple, les méthodes à base de la soustraction cognitive soulèvent encore deux problèmes. Tout d'abord, elles présupposent une hypothèse forte : l'*insertion pure*. Le rajout du processus d'intérêt est supposé n'avoir aucune influence sur les différents processus induits par la tâche de contrôle. Ensuite, il reste le problème des traitements implicites ou cachés. Le processus cognitif d'intérêt peut être implicitement déclenché par la tâche de contrôle. L'activité liée au processus d'intérêt est alors diminuée lors de l'analyse. De plus, d'autres processus cachés peuvent être déclenchés par la tâche B. Ainsi, l'activité observée ne sera pas liée uniquement au processus d'intérêt. Pour résoudre ce problème, deux schémas ont été proposés : l'analyse factorielle et la conjonction cognitive.

Les études conjonctives [182] se fondent sur plusieurs paires de tâches qui induisent toutes le processus d'intérêt pendant la tâche B et non pendant la tâche de contrôle. La conjonction cognitive recherche en définitive les points communs entre plusieurs différences. Elle permet de s'affranchir de l'hypothèse d'*insertion pure* car les différentes paires sont supposées indépendantes (les interactions entre les processus sont supposées s'exprimer différemment, figure 1.18). Cependant, il semblerait que les études conjonctives ne résolvent pas les problèmes posés par les études soustractives [34].

Les analyses factorielles se proposent quant à elles de mesurer les interactions que la sous-

traction cognitive ignore. Pour étudier deux processus P1 et P2 et leurs interactions, il est nécessaire de définir quatre tâches, une tâche impliquant P1 et pas P2, une tâche impliquant P1 et P2, une tâche impliquant P2 et pas P1, une tâche impliquant ni P1 ni P2. Sur l'exemple de la figure 1.19, les deux processus étudiés sont les processus de dénomination et de reconnaissance d'objet. La tâche A met en jeu les deux processus, la tâche B uniquement le processus de reconnaissance, la tâche C uniquement le processus de dénomination et la tâche D aucun des deux. Il est alors possible d'étudier l'influence du processus de reconnaissance sur le processus de dénomination ou d'étudier l'influence du processus de dénomination sur le processus de reconnaissance. Les études psychopharmacologiques sont enfin l'application typique des analyses factorielles : une paire de tâches (tâche de contrôle et tâche d'intérêt) est soumise à un patient avant et après l'administration d'un médicament [42]. Le terme d'interaction représente alors l'effet de la drogue sur la réponse physiologique à la tâche.

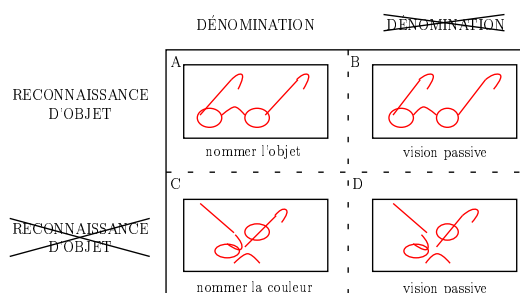


FIG. 1.19 – Exemple d'analyse factorielle (voir texte). (exemple tiré de [http://web.ccr.jussieu.fr/meg-center/media/megirmf2003/ProtocolesExpe\\_ALP-NG.pdf](http://web.ccr.jussieu.fr/meg-center/media/megirmf2003/ProtocolesExpe_ALP-NG.pdf)).

### 1.3.5 Caractéristiques des données recueillies

Les séquences de données IRMf acquises se composent généralement de plusieurs centaines d'images 3D. La fréquence d'échantillonnage est de l'ordre de la seconde (entre 2 et 5 secondes) et la résolution spatiale (dans une direction de l'espace) est de l'ordre de quelques millimètres (3-4 millimètres). Il convient de noter enfin que les différentes coupes 2D d'un volume cérébral 3D n'ont pas été acquises au même instant.

#### 1.3.5.1 Les différentes sources d'artéfacts

Les inévitables mouvements du sujet ont des conséquences très complexes sur l'obtention de l'image (les spins excités au niveau d'une première coupe peuvent être excités au niveau d'une seconde coupe à cause du mouvement) et sur celle du signal IRMf également (le même voxel n'est pas associé exactement à la même zone corticale tout au long de l'expérience).

La contribution majeure au bruit dans une série temporelle provient de fluctuations physiologiques (respiration et rythme cardiaque en particulier) auxquelles s'ajoute un bruit thermique lié au processus d'acquisition des images [130]. Dans une moindre mesure, on peut également citer des artéfacts liés à la construction de l'image IRM, des artéfacts relatifs au signal BOLD mesuré près des vaisseaux de taille importante et des artéfacts en relation avec des différences de susceptibilité magnétique importante à l'interface de différents tissus. Enfin, il existe une dérive de la valeur du champ principal de l'aimant au cours du temps. Ainsi, les sources de

bruit apparaissent sous la forme d'une dérive de la ligne de base (fluctuation de la moyenne du signal dans le temps) et sous la forme de fluctuations lentes ou rapides.

### 1.3.5.2 La variabilité des données acquises

Il existe tout d'abord une grande variabilité de la réponse à un stimulus (réponse hémodynamique). La réponse peut varier avec l'âge [192], d'un individu à l'autre [108] et d'une région à l'autre du cerveau [108]. Cette grande variabilité de signaux actifs rend la modélisation du signal IRMf actif complexe. A cela s'ajoute également une grande variabilité fonctionnelle. Par exemple, une même tâche n'est pas toujours traitée de la même manière chez tous les sujets. Ainsi, les zones impliquées dans le traitement peuvent être différentes. Cette variabilité fonctionnelle se retrouve également à un niveau individuel : le niveau d'attention du sujet, sa fatigue, les conditions expérimentales et son entraînement à exécuter la tâche peuvent conduire à ce qu'une même tâche soit traitée différemment par le même individu à deux moments différents [188].

Enfin, il existe également une forte variabilité anatomique. Les structures du cerveau sont présentes chez tous les individus mais elles montrent cependant une forte variabilité de forme. Par exemple, les sillons (secondaires) sont très variables d'un individu à l'autre, aussi bien au niveau de leur existence, de leur position que de leurs relations d'adjacence.

# Chapitre 2

## Méthodes de cartographie en IRMf cérébrale : état de l'art

Nous nous confrontons à différents problèmes lors de l'analyse des séquences d'images IRMf. Tout d'abord, la nature de l'information portée par le signal IRMf fait toujours débat. Ce signal présente également un faible rapport signal à bruit auquel s'ajoute une forte variabilité anatomo-fonctionnelle intra- et inter-individus. Enfin, de nombreux prétraitements appliqués aux données brutes tels que le recalage spatial et temporel des images, sont susceptibles de modifier le contenu informatif du signal original.

Dans un premier temps, nous présentons succinctement les prétraitements communément utilisés en cartographie IRMf cérébrale. Ensuite, nous présentons les deux grandes classes de méthodes utilisées en cartographie : les méthodes inférentielles et les méthodes exploratoires. Nous terminons le chapitre par une discussion autour du problème de validation des méthodes de cartographie.

### 2.1 Prétraitements

L'étape de prétraitement consiste à créer les signaux IRMf relatifs à chaque voxel de l'image à partir des  $T$  images de volume cérébral. Les principaux prétraitements sont consignés ci-après.

Tout d'abord, deux coupes d'une même image n'étant pas acquises au même instant, les différentes coupes peuvent être resynchronisées entre elles [110]. Cette étape peut s'effectuer soit avant, soit après le recalage des images selon le type d'entrelacement utilisé lors de l'acquisition.

Une seconde étape dite de recalage consiste à supprimer en partie les inévitables mouvements du sujet. En pratique, on cherche la transformation rigide qui minimise une fonction de coût de préférence robuste [76]. Des traitements plus sophistiqués [132] prennent en compte les fortes distorsions géométriques des images fonctionnelles dans la direction du gradient de phase (ces distorsions variant avec le temps).

Une troisième étape de normalisation spatiale peut être utilisée lors des études de groupe [8] [44]. Elle consiste à recaler les images EPI sur un atlas.

Enfin, une étape de débruitage se révèle indispensable. Elle vise à éliminer la dérive de ligne de base (fluctuation de la moyenne du signal dans le temps) [203] ainsi que certains effets physiologiques ne variant pas sur toute l'image mais variant avec le temps [149]. On retrouve alors principalement le bruit thermique et les fluctuations physiologiques. Des méthodes courantes

de filtrage ou de restauration des données sont alors utilisées [220] [206] [35][130].

La qualité des prétraitements a une influence importante sur les résultats finaux de cartographie [44] [220] [113]. Les prétraitements dépendent également de la méthode de cartographie utilisée et en particulier du modèle choisi pour définir les signaux IRMf. Ainsi, la suppression de la dérive de ligne de base n'est pas nécessaire si cette dernière fait partie intégrante du modèle décrivant le signal actif. De même, le recalage des images peut s'effectuer simultanément avec la détection de l'activation, rendant inutile l'étape de recalage dans les prétraitements [172].

## 2.2 Les méthodes inférentielles

### 2.2.1 La méthodologie SPM (Statistical Parametric Mapping)

L'abréviation SPM (Statistical Parametric Mapping) désigne aussi bien un logiciel, un processus d'analyse, ou des images (« carte SPM ») dont chaque voxel suit une loi connue sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale au voxel  $v$ . Cette approche a été développée par K. Friston *et al.* au Hammersmith Hospital à Londres il y a bientôt 15 ans pour l'analyse des images TEP [80]. SPM a évolué jusqu'à devenir la référence aussi bien en TEP qu'en IRMf. Une description détaillée de la méthodologie SPM est présentée dans [82]. Cette méthodologie est constituée de trois grandes étapes : étape de prétraitement, calcul de la carte statistique paramétrée et détection des activités fonctionnelles par seuillage de la carte statistique paramétrée. Au cœur de SPM, se trouve le modèle linéaire général. Le succès de SPM s'explique en partie par la grande richesse de ce modèle et par le cadre théorique qu'il fournit pour résoudre les problèmes d'inférence. L'utilisation du modèle linéaire général dans l'analyse de séquences IRMf est ainsi devenue très courante. C'est pourquoi, nous commençons tout d'abord par présenter le modèle linéaire général. Nous verrons ensuite les différentes variations qui existent autour de ce modèle.

#### 2.2.1.1 Le modèle linéaire général

Le modèle linéaire général exprime le signal IRMf cérébral comme une combinaison linéaire de composantes attendues (variables explicatives) auxquelles s'ajoute une erreur résiduelle. Précisément, le signal IRMf associé au voxel  $n$ ,  $\mathbf{X}_n = [X_n(1), X_n(2), \dots, X_n(T)]^T$ , est supposé être la somme pondérée de  $R$  signaux,  $\{\mathbf{g}_r(\cdot)\}_{r=1..R}$ , et d'un signal d'erreur,  $\boldsymbol{\epsilon}_n = [\epsilon_n(1), \epsilon_n(2), \dots, \epsilon_n(T)]^T$ . En notant  $b_n(r)$  la contribution de la  $r$ -ième variable explicative pour le voxel  $n$ , on obtient :

$$X_n(t) = \sum_{r=1}^R b_n(r)g_r(t) + \epsilon_n(t) \quad (2.1)$$

Les erreurs  $\boldsymbol{\epsilon}_n$  sont supposées indépendantes et identiquement distribuées selon la loi normale  $\mathcal{N}(0, \sigma_n^2)$ . Une première étape consiste tout d'abord à estimer le vecteur  $\mathbf{b}_n = [b_n(1), b_n(2), \dots, b_n(R)]^T$  selon la méthode des moindres carrés ou bien par la méthode du maximum de vraisemblance. Sous les hypothèses précédentes, on obtient dans les deux cas l'équation (2.2). Par commodité, l'ensemble des régresseurs  $\{g_r(t)\}_{1 \leq r \leq R, 1 \leq t \leq T}$  est regroupé dans une matrice  $\mathbf{G}$  appelée matrice de dessin. Une colonne de  $\mathbf{G}$  est représentative d'un signal explicatif de  $\mathbf{X}_n$ . On obtient :

$$\hat{\mathbf{b}}_n = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{X}_n \text{ avec } \hat{\mathbf{b}}_n \rightsquigarrow \mathcal{N}(\mathbf{b}_n, \sigma_n^2 (\mathbf{G}^T \mathbf{G})^{-1}) \quad (2.2)$$



On considère implicitement dans l'équation précédente que la matrice  $\mathbf{G}^T \mathbf{G}$  est inversible, c'est-à-dire que la matrice de dessin est de rang maximal. Conformément à l'équation (2.2), il apparaît que l'estimateur du vecteur  $\mathbf{b}_n$  est non biaisé et qu'une estimation de sa matrice de variance-covariance nécessite l'estimation de la variance résiduelle  $\sigma_n^2$ . L'équation (2.3) donne une estimation,  $\hat{\Sigma}_n$ , de la matrice de variance-covariance de l'estimateur de  $\mathbf{b}_n$  :

$$\hat{\Sigma}_n = \frac{\boldsymbol{\epsilon}_n^T \boldsymbol{\epsilon}_n}{T - R} (\mathbf{G}^T \mathbf{G})^{-1} \quad (2.3)$$

Le point clé de la méthode est donc de concevoir une matrice de dessin,  $\mathbf{G}$ , représentative des données analysées. Cette étape de conception est à la charge de l'utilisateur. La matrice de dessin  $\mathbf{G}$  se divise en composantes représentatives des perturbations et en composantes d'activation attendues. Par exemple, des sinusoides de basse fréquence peuvent être utilisées pour modéliser les fluctuations physiologiques. La composante d'activation attendue est quant à elle souvent considérée comme la sortie d'un système linéaire et invariant dans le temps (LIT) dont l'entrée correspond à un modèle déterministe du paradigme d'activation. Par exemple, dans le cadre des paradigmes en blocs, ce modèle se résume en une fonction carrée représentative de l'alternance des phases de contrôle (repos) et d'activation du paradigme. Le système LIT est entièrement caractérisé par sa réponse impulsionnelle  $h$  liant le signal de sortie  $y$  (composante d'activation attendue) à celui d'entrée  $x$  (représentation déterministe du paradigme d'activation) par la relation de convolution :

$$y(t) = \int_{-\infty}^{\infty} x(u)h(t - u)du = (x * h)(t) \quad (2.4)$$

Dans ce type d'approche, la réponse impulsionnelle  $h$  est représentative de la réponse hémodynamique cérébrale à un stimulus isolé.

En pratique, la matrice de dessin est définie par l'utilisateur. Il s'agit ensuite d'estimer la contribution de chaque variable explicative et la précision de ces estimations. Pour cela, on détermine  $\hat{\mathbf{b}}_n$  et  $\hat{\Sigma}_n$  au travers des équations (2.2) et (2.3). Il reste alors à définir le test statistique. Deux types de test peuvent être définis permettant d'obtenir les cartes statistiques paramétriques  $\{t\}$  ou  $\{f\}$ .

### 2.2.1.2 Les cartes paramétriques $\{t\}$

Les tests sont reliés à la notion de *contraste*. Un contraste  $\mathbf{c}$  est un vecteur permettant de définir une combinaison linéaire entre les composantes du vecteur  $\hat{\mathbf{b}}_n$ . Le vecteur  $\hat{\mathbf{b}}_n$  représentant la contribution de chaque variable explicative pour la synthèse du signal IRMf associé au voxel  $n$ , la notion de contraste permet de nous intéresser à l'effet de certaines conditions en particulier à travers le scalaire  $\mathbf{c}^T \hat{\mathbf{b}}_n$ . Un test statistique permettant de mesurer l'effet de certaines conditions est donc élaboré. Sous l'hypothèse nulle  $\mathcal{H}_0$  d'inactivité cérébrale au voxel  $n$ , il vient  $\mathbf{c}^T \mathbf{b}_n = 0$ , les conditions étudiées étant censées être nulles sous  $\mathcal{H}_0$ . Toujours sous  $\mathcal{H}_0$ ,  $\mathbf{c}^T \hat{\mathbf{b}}_n$  représente l'erreur d'estimation de  $\mathbf{c}^T \mathbf{b}_n$  tandis que  $\mathbf{c}^T \hat{\Sigma}_n \mathbf{c}$  représente la variance de l'estimateur de  $\mathbf{c}^T \mathbf{b}_n$ . Le test statistique peut alors s'écrire sous  $\mathcal{H}_0$  par :

$$s_n = \frac{\mathbf{c}^T \hat{\mathbf{b}}_n}{\sqrt{\mathbf{c}^T \hat{\Sigma}_n \mathbf{c}}} \rightsquigarrow t_{T-R} \text{ avec } t_d : \text{loi de Student à } d \text{ degrés de liberté} \quad (2.5)$$

A chaque voxel  $n$ , est associée la valeur  $s_n$  à partir de laquelle est déterminé un z-score ou de manière équivalente une p-valeur. On obtient alors une carte paramétrique  $\{t\}$ . Pour obtenir une carte d'activation, il convient finalement de seuiller la carte  $\{t\}$ .

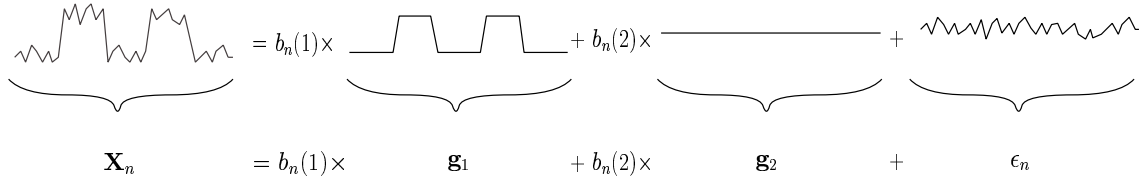


FIG. 2.1 – Exemple d'utilisation du modèle linéaire en IRMf : le signal IRMf (à gauche) est considéré comme étant la somme pondérée de deux signaux explicatifs ( $\mathbf{g}_1$  et  $\mathbf{g}_2$ ) et d'un terme d'erreur (à droite).

Pour comprendre cette approche, il est intéressant de considérer l'exemple suivant. On considère un paradigme en blocs constitué de deux tâches (une tâche de contrôle (repos) et une tâche d'intérêt). On appelle  $\{\tau_1\}$  l'ensemble des instants discrets associé à la tâche d'intérêt et  $\{\tau_2\}$  celui associé à la tâche de contrôle (repos). Par définition,  $\{\tau_1\}$  et  $\{\tau_2\}$  sont disjoints et leur union est égale à  $\{1, 2, \dots, T\}$ . On note  $n_i = \text{card}(\tau_i)$ ,  $i = 1, 2$ . On considère deux signaux explicatifs :  $g_1(t) = 1$  si  $t \in \{\tau_1\}$ , 0 sinon et  $\forall t$ ,  $g_2(t) = 1$ . (cf. Fig. 2.1). Le signal  $\mathbf{g}_1$  est un signal explicatif permettant de modéliser la réponse attendue aux stimuli<sup>1</sup>. Le signal  $\mathbf{g}_2$  modélise les composantes de non intérêt présentes à la fois pour les signaux actifs et les signaux inactifs. Ainsi, sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale au voxel  $n$ , le vecteur  $b_n(1)$  doit être nul. Le contraste  $\mathbf{c} = [1, 0]^T$  sera donc utilisé. On estime au préalable la contribution de chaque signal explicatif et la précision de cette estimation. Les équations (2.2) et (2.3) s'écrivent de la façon suivante :

$$\begin{aligned} \hat{\mathbf{b}}_n &= \frac{1}{n_1(n_1+n_2)-n_1^2} \begin{pmatrix} n_2 \sum_{t \in \{\tau_1\}} X_n(t) - n_1 \sum_{t \in \{\tau_2\}} X_n(t) \\ n_1 \sum_{t \in \{\tau_2\}} X_n(t) \end{pmatrix} \\ \hat{\Sigma}_n &= \frac{\boldsymbol{\epsilon}_n^T \boldsymbol{\epsilon}_n}{T-2} \frac{1}{n_1(n_1+n_2)-n_1^2} \begin{pmatrix} n_1 + n_2 & -n_1 \\ n_1 & n_1 \end{pmatrix} \end{aligned} \quad (2.6)$$

L'équation (2.5) s'écrit finalement pour le contraste  $\mathbf{c} = [1, 0]^T$  :

$$s_n = \frac{\sqrt{n_1(n_1+n_2)-n_1^2}}{\sqrt{n_1+n_2}} \frac{n_2 \sum_{t \in \{\tau_1\}} X_n(t) - n_1 \sum_{t \in \{\tau_2\}} X_n(t)}{\hat{\sigma}_n} \rightsquigarrow t_{T-2} \quad (2.7)$$

L'équation (2.7) permet de définir une carte statistique paramétrique  $\{t\}$  pouvant être transformée en carte de p-valeur ou de z-score.

### 2.2.1.3 Les cartes paramétriques $\{f\}$

Alors que les cartes  $\{t\}$  s'intéressent à l'effet de certaines conditions en particulier, les cartes  $\{f\}$  s'intéressent à l'effet de plusieurs conditions. Elles sont basées sur la norme au carré du vecteur  $\mathbf{\Gamma} \hat{\mathbf{b}}_n$ ,  $\mathbf{\Gamma}$  étant une matrice où chaque ligne représente un contraste :

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \dots \\ \mathbf{c}_l^T \end{pmatrix} \quad (2.8)$$

<sup>1</sup>Par simplicité, nous n'avons pas utilisé un modèle de réponse hémodynamique.

Sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale au voxel  $n$ , les facteurs d'intérêt représentés dans la matrice  $\mathbf{\Gamma}$  n'ont pas d'effet significatif, d'où  $\mathbf{\Gamma}\mathbf{b}_n = \vec{0}$ , ou de manière équivalente :  $(\mathbf{\Gamma}\mathbf{b}_n)^T\mathbf{\Gamma}\mathbf{b}_n = 0$ . Ainsi, sous  $\mathcal{H}_0$ ,  $(\mathbf{\Gamma}\hat{\mathbf{b}}_n)^T(\mathbf{\Gamma}\hat{\mathbf{b}}_n)$  représente l'erreur d'estimation de la norme au carré du vecteur  $\mathbf{\Gamma}\mathbf{b}_n$  alors que  $\mathbf{\Gamma}^T\hat{\Sigma}_n\mathbf{\Gamma}$  représente la variance de l'estimateur. Le test statistique s'écrit alors :

$$s_n = \frac{(\mathbf{\Gamma}\hat{\mathbf{b}}_n)^T(\mathbf{\Gamma}\hat{\mathbf{b}}_n)}{\mathbf{\Gamma}^T\hat{\Sigma}_n\mathbf{\Gamma}} \frac{d_2}{d_1} \rightsquigarrow F_{d_1, d_2} \quad (2.9)$$

où  $F_{d_1, d_2}$  représente une loi de Fisher de degrés de liberté  $d_1$  et  $d_2$ ,  $d_1$  et  $d_2$  étant respectivement le nombre de degrés de liberté du numérateur et du dénominateur. L'équation (2.9) conduit à définir une carte statistique paramétrique  $\{f\}$ . Cette carte peut être transformée secondairement en carte de p-valeur ou de z-score.

#### 2.2.1.4 Seuillage des cartes de p-valeur ou de z-score

Les cartes SPM ( $\{f\}$  ou  $\{t\}$ ) sont finalement interprétées dans le but de localiser les composantes dont on est « sûr » qu'elles soient significatives. Par la suite, on parlera indifféremment de p-valeur et de z-score dans la mesure où une p-valeur peut être calculée à partir d'un z-score, l'inverse étant vrai également. Une p-valeur est la probabilité de rejeter l'hypothèse  $\mathcal{H}_0$  (l'hypothèse d'inactivité cérébrale dans le cadre de notre application) alors que cette dernière est vraie. Une revue des différentes méthodes de seuillage est présentée dans [151] et [148].

Une première famille de méthodes définit le seuil de manière à contrôler le nombre de fausses alarmes. Si on seuille la carte avec une p-valeur  $p$ , on s'attend à obtenir en moyenne  $N \times p$  fausses alarmes où  $N$  représente le nombre de signaux inactifs. Une solution pour diminuer le nombre de fausses alarmes est de modifier le seuil. La correction de Bonferroni consiste ainsi à choisir le seuil de manière à obtenir en moyenne  $p$  fausses alarmes. Ainsi, plus  $N$  est important, plus le seuil devient petit augmentant ainsi le risque de non détection des signaux actifs. La correction de Bonferroni est aussi mal adaptée à la segmentation des cartes SPM dans la mesure où elle ne prend pas en compte la corrélation spatiale qui existe dans ces cartes. Ainsi, il y a une différence importante entre le nombre de tests « décorrélés » et le nombre de tests utilisés. Ce problème est connu sous le nom du problème de comparaison multiple.

Pour résoudre ce problème, les principales méthodes définissent le seuil de manière à contrôler la probabilité d'avoir une ou plusieurs fausses alarmes dans la carte. Par conséquent, il est nécessaire d'estimer la probabilité que la valeur maximale de la carte soit plus grande qu'un seuil donné, la carte étant représentative de l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale à tous les voxels  $v$ . Dit autrement, la distribution du maximum de la carte de z-score doit être déterminée sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale en tous les voxels  $v$ .

Dans ce but, Worsley [226] propose de modéliser la carte d'activité comme un champ aléatoire gaussien, permettant ainsi d'estimer la distribution désirée au travers de la caractéristique d'Euler moyenne d'une carte binaire. Cette caractéristique est définie comme étant le nombre de composantes connexes d'un ensemble auquel on soustrait le nombre de trous dans les composantes connexes. Au-dessus d'un certain seuil, les trous disparaissent et les composantes connexes sont indépendantes. La distribution de leur nombre est approchée par une loi de Poisson. La valeur moyenne de la caractéristique d'Euler peut alors être reliée avec le seuil utilisé pour obtenir la carte binaire et la régularité du champ gaussien (la régularité du champ est estimée à travers la largeur à mi-hauteur de la réponse impulsionnelle du système). On obtient ainsi une estimation de la probabilité que la valeur maximale de la carte soit plus grande qu'un seuil donné. Cette approche donne des résultats satisfaisants dès lors que les

cartes sont lissées. Ainsi, il est fortement recommandé de filtrer les données au préalable [227]. Il est proposé enfin sous les mêmes hypothèses des tests basés sur l'étendue spatiale des activations [178] ou des tests basés à la fois sur l'intensité et l'étendue spatiale [179]. L'utilisation des champs aléatoires gaussiens repose cependant sur des hypothèses difficilement vérifiables, dont notamment le caractère gaussien du champ aléatoire et la stationnarité de la carte SPM.

D'autres méthodes, basées sur des techniques de permutation, se concentrent sur la construction empirique de la distribution désirée (voir [168] pour une revue des différentes méthodes). Dans ce but, il est nécessaire de simuler, sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale en tous les voxels  $v$ , une nouvelle séquence d'images. La difficulté provient essentiellement de la prise en compte de l'autocorrélation du bruit et de la dérive de la ligne de base. Ainsi, la permutation se fait généralement à partir des résidus du modèle linéaire. A noter que lorsque la permutation est effectuée dans le plan des ondelettes [31], on se soustrait des problèmes d'autocorrélation du bruit, puisqu'on tire profit des propriétés de décorrélation des ondelettes. Il est enfin à noter que des comparaisons ont été effectuées entre les méthodes à base de champs aléatoires gaussiens et les méthodes à base de permutation. Les auteurs Nichols *et al.* [168] démontrent que les méthodes à base de permutation donnent des résultats plus satisfaisants. A noter que Raz *et al.* [190] proposent de considérer des paradigmes différents de celui utilisé lors de l'expérience pour estimer la distribution désirée.

Les méthodes précédentes n'autorisent pas les fausses alarmes. Le seuil en terme de z-score est donc élevé, augmentant ainsi le risque de ne pas détecter les signaux actifs. Dans ce but, Benjanini [21] propose par exemple de contrôler le nombre de fausses alarmes sur le nombre de voxels déclarés actifs. Il est à noter que la valeur du seuil dépend du nombre de voxels actifs. Ce travail est à rapprocher de la méthode proposée par Friston [89] dans laquelle ce dernier définit des cartes *a posteriori* dans le cadre bayésien. Dans le même cadre, nous pouvons citer les travaux d'Everitt [62] et leur extension à la prise en compte du voisinage spatial [109]. Ils calculent, après avoir estimé la distribution des z-scores sous l'hypothèse  $\mathcal{H}_0$  et  $\mathcal{H}_1$  ( $\mathcal{H}_1$  étant l'hypothèse d'activité), la probabilité *a posteriori* que le voxel soit actif (ou inactif). Cependant, ces deux approches ne semblent pas avoir eu de véritables échos dans la communauté IRMf bien qu'elles soient fondées sur une représentation « satisfaisante » pour l'esprit dans la mesure où les deux distributions sont modélisées. Ce manque de succès est peut-être dû en partie au caractère arbitraire du choix des différentes densités de probabilité.

## 2.2.2 De l'applicabilité des méthodes inférentielles en IRMf

### 2.2.2.1 Limitations

Trois points de modélisation sont discutés dans la littérature : l'autocorrélation temporelle du bruit, la non linéarité de la réponse et la variabilité de la réponse hémodynamique. Nous détaillons ici ces trois problèmes de modélisation. Dans la prochaine sous-section, nous abordons les solutions proposées.

Tout d'abord, alors que l'on a supposé les erreurs  $\epsilon_n$  indépendantes et identiquement distribuées selon la loi normale  $\mathcal{N}(0, \sigma_n^2)$ , le fait qu'il existe une corrélation temporelle du bruit est maintenant établi.

Ensuite, la composante d'activation attendue a été modélisée comme la sortie d'un système LIT. Cependant, Miller *et al.* [164] ont montré que si l'activité cérébrale est proportionnelle au flux sanguin, elle n'est pas une fonction linéaire de l'intensité de la stimulation, ni une fonction linéaire du signal BOLD. De plus, la non linéarité du signal BOLD n'est pas stationnaire spatialement [26]. Cette non linéarité est enfin grandement diminuée à haut champ (7 Teslas)

[175] laissant ainsi supposer que les propriétés de la matière blanche seraient à l'origine de la non linéarité. L'hypothèse de linéarité reste cependant valide dès lors que l'intervalle de temps séparant deux stimuli est supérieur à 4 secondes [99]. Vazquez arrive aux mêmes conclusions dans [210] où il montre que l'hypothèse de linéarité est d'autant moins valide que les périodes de stimulation sont courtes et que l'intensité du stimulus est faible.

Enfin, il existe une grande variabilité de la réponse hémodynamique. Cette dernière varie avec l'âge [192], d'un individu à l'autre [108] et d'une région à l'autre du cerveau [108]. Il est important dans les traitements d'utiliser une réponse hémodynamique adaptée aux données [108]. Nous voyons par la suite que cette réponse peut être estimée conjointement avec l'activation.

### 2.2.2.2 Solutions

En ce qui concerne la prise en compte de l'autocorrélation temporelle du bruit, on peut citer trois grandes familles de méthodes.

La première introduit une forte corrélation temporelle au travers d'un filtrage passe-bas  $k$ . L'équation (2.1) s'écrit désormais :

$$[k * X_n](t) = \sum_{r=1}^R b_n(r)[k * g_r](t) + [k * \epsilon_n](t) \quad (2.10)$$

La dépendance temporelle intrinsèque aux données (non connue) est alors négligée devant celle ajoutée. Des estimateurs biaisés de  $\mathbf{b}_n$  et de  $\Sigma_n$  ont été ensuite proposés. Cette méthode est très robuste et elle est utilisée dans le logiciel SPM99.

Les recherches, aujourd'hui, se tournent plutôt vers une seconde famille de méthodes, reposant sur le blanchiment du bruit. L'équation (2.1) s'écrit désormais :

$$\mathbf{A}_n \mathbf{X}_n = \sum_{r=1}^R b_n(r) \mathbf{A}_n \cdot \mathbf{g}_r + \mathbf{A}_n \cdot \epsilon_n \quad (2.11)$$

où  $\mathbf{A}_n$  est une matrice carrée de taille  $T \times T$  avec  $\mathbf{A}_n = \mathbf{V}_n^{-\frac{1}{2}}$  où  $\sigma_n^2 \mathbf{V}_n$  est la matrice de variance-covariance de  $\epsilon_n$ . Le modèle défini par l'équation (2.11) est un modèle linéaire dont les erreurs sont indépendantes et identiquement distribuées. Pour déterminer  $\mathbf{A}_n$ , les méthodes procèdent souvent de manière récursive. Le bruit est tout d'abord supposé blanc. Ensuite, connaissant  $\epsilon_n$ , la matrice  $\mathbf{A}_n$  peut être estimée. On obtient ainsi un nouveau modèle linéaire à partir duquel on peut réestimer  $\epsilon_n$  et ensuite  $\mathbf{A}_n$ . Les méthodes proposées dans la littérature se différencient par les méthodes d'estimation de la matrice de variance-covariance du bruit et par la modélisation du bruit (modèle auto-régressif d'ordre 1 [87][185] [32], modèle auto-régressif d'ordre  $p$  [228]).

Enfin, la dernière grande famille de méthodes visant à tenir compte de l'autocorrélation temporelle du bruit utilise les propriétés de décorrélation des ondelettes discrètes : bien que le signal soit fortement corrélé, les coefficients de l'ondelette peuvent être relativement peu corrélés, voire non corrélés sous certaines conditions. Tout d'abord, Bullmore *et al.* [31] proposent de construire un test statistique non biaisé à partir du test statistique classique qui ne prend pas en compte la corrélation temporelle. Ils estiment, dans ce but, la distribution de la statistique sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale à la place de considérer cette dernière comme une loi de Student (équation (2.5)). Pour cela, ils construisent des signaux représentatifs de l'inactivité cérébrale en utilisant des méthodes de permutation dans le plan

de l'ondelette. L'autocorrélation du signal original est en effet égale sous certaines conditions à l'autocorrélation du signal reconstruit, après permutation aléatoire des coefficients de l'ondelette <sup>2</sup>. Le travail de Fadili *et al.* [63] est plus ambitieux dans la mesure où l'estimation des paramètres du modèle linéaire s'effectue dans l'espace des ondelettes, profitant ainsi de la décorrélation des coefficients de l'ondelette. Cependant, l'inférence s'effectue encore dans l'espace temporel. Dans [161] et dans [152], l'inférence est désormais effectuée directement à partir des coefficients des ondelettes. A noter que contrairement à [161], l'approche de [152] n'est valide que pour des paradigmes périodiques dans la mesure où elle utilise la transformée de Fourier.

En ce qui concerne les non linéarités de la réponse, un modèle hémodynamique permettant de relier l'activité au signal BOLD a été défini dans [33]. Ce modèle semble être suffisamment riche pour permettre d'expliquer les non linéarités observées [88]. De plus, il peut s'exprimer en terme de séries de Volterra, formalisme qui avait été auparavant utilisé dans [86] (les séries de Volterra permettent de définir un système non linéaire invariant dans le temps). Dans [81], l'inférence bayésienne est utilisée à partir du modèle de [88] pour cartographier les zones d'activité.

Enfin, le dernier problème visant à prendre en compte la forte variabilité de la réponse hémodynamique nécessite tout d'abord de choisir un type de modélisation : modèle paramétrique ou non. Parmi les modèles paramétriques, on peut citer la loi de Poisson [85], la fonction Gamma [133], des modèles orientés physiologie [129], [201]. Le modèle le plus répandu (par exemple [99]) est celui utilisé dans le logiciel SPM, où :

$$h(t, \boldsymbol{\alpha}) = G(t, \frac{\alpha_0}{\alpha_2}, \alpha_2) - \frac{1}{\alpha_4} G(t, \frac{\alpha_1}{\alpha_3}, \alpha_3) \text{ avec } G(t, a, b) = \frac{1}{b^a \Gamma(a)} t^{a-1} e^{-\frac{t}{b}} \quad (2.12)$$

$\alpha_0$  représente le délai entre la stimulation et le maximum de la réponse,  $\alpha_1$  le délai entre la stimulation et le minimum de la réponse (« undershoot »),  $\alpha_2$  la dispersion de la réponse,  $\alpha_3$  la dispersion de l'« undershoot » et  $\alpha_4$  le rapport d'amplitude entre la réponse et l'« undershoot ».

Un intérêt particulier a été porté sur l'estimation des paramètres de délai. Pour prendre en compte d'éventuels retards d'activation, les dérivées temporelles de la réponse peuvent être considérées [83] et incorporées dans la matrice de dessin. Cependant, les résultats obtenus ne sont pas satisfaisants [141]. Aussi, d'autres méthodes d'estimation du retard ont été proposées [141] [196]. Liao *et al.* [141] proposent d'utiliser deux fonctions de base dans la matrice de dessin. Ces deux fonctions de base permettent d'inférer sur le retard d'activité. Saad *et al.* [196] ne se placent pas dans le cadre du modèle linéaire. Ils estiment le retard à partir de la fonction d'intercorrélacion entre le paradigme et le signal IRMf. Nous allons voir enfin dans le prochain paragraphe que l'inférence bayésienne est d'un grand intérêt pour prendre en compte les hyper-paramètres et donc pour prendre en compte la variabilité de la réponse hémodynamique.

---

<sup>2</sup>Il est à noter que la permutation dans le plan de l'ondelette a également d'autres applications dans le cadre de la cartographie IRMf. Par exemple, la permutation spatio-temporelle des données IRMf dans l'espace des ondelettes permet de déterminer si la corrélation qui existe entre deux régions du cerveau est due au hasard ou si la corrélation est significative. Ainsi, cette méthode a un grand intérêt pour l'étude de la connectivité fonctionnelle [29].

### 2.2.3 De nouvelles méthodologies

#### 2.2.3.1 Cadre d'inférence bayésien

Dans le cadre de la cartographie en IRMf, Friston [90] montre dans quel cas l'inférence bayésienne est identique à l'inférence classique. Cependant, lorsque l'on veut à la fois prendre en compte la variabilité de la réponse hémodynamique, la non linéarité de la réponse et l'information spatiale, l'inférence classique basée sur le maximum de vraisemblance trouve rapidement ses limites. Aussi, l'inférence bayésienne est préférée [95][224][154][40]. Elle est d'une grande aide pour prendre en compte un nombre important d'hyperparamètres dans la mesure où elle peut permettre de ne pas effectuer un choix particulier quant au jeu d'hyperparamètres [95][224]. Au contraire, à partir des *a priori*, il est possible de déterminer la distribution conjointe *a posteriori* des paramètres non connus (étant données les observations). Enfin, les cartes obtenues avec une approche bayésienne sont plus simples à analyser que celles obtenues avec une inférence classique dans la mesure où les premières se décrivent en termes d'*a posteriori*. Il est enfin à noter que les *a priori* sont également d'une grande aide dans [154] et dans [40] pour apporter des informations physiologiques sur la réponse hémodynamique ou dans [198] pour prendre en compte l'information spatiale.

Par exemple, Smith *et al.* [198] proposent d'utiliser une carte de labels  $\gamma_i$ ,  $\gamma_i$  étant le label au voxel  $i$  avec  $\gamma_i = 0$  en cas d'inactivité et  $\gamma_i = 1$  en cas d'activité. L'*a priori* est défini ensuite sur cette carte grâce au modèle d'Ising. Le signal IRMf est supposé être la somme pondérée d'une composante d'activation et de composantes représentatives des perturbations. On note  $\beta_i$  le coefficient pondérant la composante d'activation pour le voxel  $i$ . On a ainsi :  $\beta_i = 0$  si et seulement si  $\gamma_i=0$  et  $\beta_i \neq 0$  si et seulement si  $\gamma_i=1$ . La vraisemblance *a posteriori* que  $\gamma_i = 1$  (connaissant les données) est ensuite estimée par des méthodes de Monte-Carlo. L'analyse de la carte obtenue est ainsi simplifiée, un lien pouvant même être établi avec la notion de p-valeur classique [198].

Woolrich *et al.* [224] modélisent le bruit comme un processus auto-régressif spatio-temporel non séparable. La réponse hémodynamique est ensuite modélisée comme quatre demi-périodes de cosinus (huit paramètres). L'auteur estime grâce à des approches du type Monte-Carlo la densité de probabilité *a posteriori*  $f$  de la hauteur  $h$  de la réponse (connaissant les observations). En cas d'activité, on s'attend à ce que la hauteur  $h$  soit strictement positive et qu'elle soit nulle en cas d'inactivité. Un signal est donc déclaré actif dès lors que la vraisemblance  $P[h > 0] = \int_{h=0}^{+\infty} f$  est grande.

Højen-Sørensen *et al.* [111] utilisent une modélisation markovienne cachée. Pour chaque voxel, une variable caché binaire  $X_t$  est associée à chaque échantillon du signal. Cette dernière vaut 1 si la zone corticale associée au voxel  $v$  est active au temps  $t$ , 0 sinon. Le processus observable est lié au processus caché de la façon suivante : l'observation est définie comme étant la convolution du processus caché par la réponse hémodynamique à laquelle on ajoute des termes modélisant la dérive de ligne de base et une constante. La réponse hémodynamique ainsi que les autres termes sont considérés comme des hyperparamètres. La méthodologie d'inférence bayésienne est ensuite utilisée dans des buts de restauration du signal.

#### 2.2.3.2 Autres méthodes inférentielles

Il existe de très nombreuses méthodes de cartographie en IRMf cérébrale. Il est hors de propos ici de toutes les présenter. Néanmoins, parce que souvent cités dans la littérature, nous présentons succinctement le test de Kolmogorov-Smirnov, les méthodes fondées sur la théorie de l'information et les méthodes multi-variées inférentielles.

Le test de Kolmogorov-Smirnov est un test non paramétrique visant à tester l'identité de deux distributions empiriques. Les deux distributions sont obtenues à partir d'échantillons du signal (une première distribution correspond aux échantillons du signal pendant la tâche de repos et une autre correspond aux échantillons pendant la tâche d'intérêt). Ainsi, ce test permet de ne pas faire d'hypothèses *a priori* sur les propriétés du signal et du bruit [189]. Cependant, les résultats obtenus avec ce test sont rarement satisfaisants. Tout d'abord, l'estimation des distributions empiriques est difficile en raison du nombre restreint de points à disposition. Ensuite, la méthode est très sensible aux non stationnarités du signal [2].

Les méthodes fondées sur la théorie de l'information permettent également de ne pas faire d'hypothèses trop restrictives sur le signal IRMf actif. Par exemple, une estimation de l'information mutuelle entre le paradigme expérimental et le signal IRMf peut être considérée comme une mesure de l'activité locale à un voxel [209]. La mesure de l'entropie du signal IRMf est également utilisée dans [47]. Cette dernière est calculée pour différentes portions du signal IRMf, chacune des portions étant définie à partir du paradigme expérimental. Un score, représentatif de l'activité est ensuite déterminé à partir des différentes valeurs de l'entropie. Une méthode plus adaptée, car prenant directement en compte l'information temporelle, est proposée dans [72], l'entropie étant remplacée par le taux d'entropie.

Les méthodes multi-variées visent à ne plus traiter les voxels indépendamment les uns des autres mais à considérer les données dans leur ensemble. Dans ce cadre, le modèle linéaire général s'écrit sous la forme :

$$\mathbf{X}^T = \mathbf{G}\mathbf{B} + \mathbf{E} \quad (2.13)$$

où  $\mathbf{X}$ ,  $\mathbf{B}$ ,  $\mathbf{G}$  et  $\mathbf{E}$  sont des matrices de taille  $N \times T$ ,  $R \times N$ ,  $T \times R$  et  $T \times N$  respectivement. Chaque colonne de  $\mathbf{X}$  représente une image et chaque ligne un signal IRMf (on gardera par la suite la même notation jusqu'à la fin du chapitre). Désormais, une estimation aux moindres carrés globale est utilisée :

$$\mathbf{B} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{X}^T \quad (2.14)$$

Au lieu de tester les résidus de chaque voxel, il est nécessaire de tester la matrice des résidus. Cependant, les tests classiques sont difficilement utilisables dans la mesure où le nombre de variables (voxels) est supérieur au nombre d'observations (images). Aussi, la matrice des données est au préalable réduite par une méthode de réduction de données comme par exemple la CVA (Canonical Variate Analysis) [5]. Dans ce type d'approches, l'inférence est généralement effectuée sur l'ensemble du cerveau. Ceci permet par exemple de conclure qu'il y a une différence d'activation sur le volume cérébral quand deux tâches différentes sont soumises au patient. Cependant, il reste possible de définir des tests locaux à un voxel [229].

## 2.3 Les méthodes exploratoires

Les méthodes exploratoires se divisent en deux grandes familles, d'un côté les méthodes produisant un modèle générateur des données, telles que l'analyse en composantes principales (ACP), l'analyse en composantes indépendantes (ACI) et l'analyse de corrélation canonique (ACC), de l'autre les méthodes de clustering associant un modèle de réponse à chaque cluster. Nous présentons ces méthodes succinctement dans la mesure où nous avons privilégié une approche de cartographie de type inférentiel.



### 2.3.1 La décomposition en valeurs singulières

La décomposition en valeurs singulières (singular value decomposition (SVD)) vise à créer un nouvel espace de données de dimension plus faible que celle du jeu de données originelles. Cet espace est déterminé de manière à expliquer le maximum de la variance des données de départ. Un centrage des données  $\mathbf{X}$  est au préalable effectué de manière à ce que la somme de chaque colonne soit nulle. On a alors :

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (2.15)$$

où  $\mathbf{U}$  et  $\mathbf{V}$  sont des matrices orthonormales de taille  $N \times N$  et  $T \times T$  respectivement et  $\mathbf{\Lambda}$  est une matrice diagonale de taille  $N \times T$ . Chaque vecteur colonne de  $\mathbf{V}$  représente un parcours temporel ( $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_T]$ ) alors que chaque vecteur colonne de  $\mathbf{U}$  ( $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_N]$ ) représente une image. A chaque couple ou effet  $\{\mathbf{U}_i, \mathbf{V}_i\}_{i \in [1, \min(N, T)]}$  est associée la valeur propre  $\lambda_i = \Lambda(i, i)$  correspondant à la contribution de l'effet à la variance des données. Les données sont donc décomposées en une somme de différents effets ordonnés par les valeurs propres. En pratique, les composantes basses fréquences contenues dans les données sont préalablement filtrées [6] [91].

Cependant, les méthodes les plus répandues définissent un espace d'intérêt au travers de la matrice de dessin  $\mathbf{G}$  de taille  $R \times T$  (il ne s'agit donc plus véritablement de méthodes exploratoires). La SVD est effectuée sur des matrices calculées à partir de  $\mathbf{G}$ , de  $\mathbf{X}$  et éventuellement de la matrice de corrélation temporelle des données. On obtient alors les méthodes de CVA [84], des moindres carrés partiels [156] et du modèle linéaire général [229] [127]. On retrouve les mêmes méthodes de réduction de données que pour les méthodes inférentielles multivariées, la seule différence étant qu'ici, aucun test statistique n'est utilisé. Une fois les données réduites, ces dernières sont directement laissées à l'interprétation du spécialiste.

### 2.3.2 L'analyse en composantes indépendantes

L'analyse en composantes indépendantes (ACI) cherche à décomposer les données  $\mathbf{X}$  en composantes indépendantes soit spatialement soit temporellement. En pratique, l'indépendance spatiale est utilisée dans la mesure où le nombre de voxels dans une image est beaucoup plus important que le nombre  $T$  d'images de la séquence à analyser, rendant ainsi l'estimation de l'indépendance plus robuste (l'ACI temporelle est cependant utilisée dans [213]).

L'objectif est alors d'estimer  $K$  images indépendantes ou  $K$  sources ( $K < T$ ). Chaque image IRMf est une combinaison linéaire des  $K$  sources à laquelle se rajoute un terme d'erreur. On obtient ainsi :

$$\mathbf{X}^T = \mathbf{M}\mathbf{S} + \mathbf{E} \quad (2.16)$$

où  $\mathbf{X}$ ,  $\mathbf{M}$ ,  $\mathbf{S}$  et  $\mathbf{E}$  sont des matrices de taille  $N \times T$ ,  $T \times K$ ,  $K \times N$  et  $T \times N$ , respectivement, et de rang  $T$ ,  $K$ ,  $K$  et  $T - K$ . La matrice  $\mathbf{S}$  peut être considérée comme un ensemble de cartes d'activations associées avec les différents effets présents dans l'ensemble de données.

L'analyse en composantes indépendantes nécessite un critère d'indépendance permettant d'estimer  $\mathbf{M}$  et  $\mathbf{S}$  [61]. L'interprétation des résultats dont notamment la séparation des effets liés au signal et au bruit est souvent effectuée directement par le spécialiste même si des traitements sont proposés [73]. L'ACI est indiscutablement la méthode exploratoire la plus utilisée dans le cadre de l'IRMf [157] [214] [211] [56]. A noter que l'ACI, comme l'ACP peut être simplement utilisée dans le but d'enlever des composantes de non intérêt [206].

### 2.3.3 L'analyse de corrélation canonique

Les processus réels sont souvent autocorrélés. Lorsqu'ils sont additionnés les uns avec les autres, l'autocorrélation du signal final est souvent plus faible que celle des sources originales. A la place de chercher des sources indépendantes comme l'ACI, il est possible, à travers l'analyse de corrélation canonique, de déterminer les sources les plus autocorrélées possibles [27]. Comme dans l'approche ACI, il est possible d'effectuer une analyse temporelle ou une analyse spatiale [78]. Il semblerait que la méthode basée sur la corrélation canonique soit plus robuste que l'ACI [78]. Parmi ses autres avantages, on peut citer le fait que les différentes sources peuvent être classées et que l'algorithme d'optimisation n'est pas itératif. Néanmoins, comme dans le cas de l'ACI, le choix du nombre de sources à utiliser reste un problème non résolu.

### 2.3.4 Les méthodes de clustering

Les méthodes de clustering visent à obtenir une carte 3D définissant des régions. Toutes les zones d'une même région ont répondu de manière similaire au paradigme de stimulation en termes d'une métrique calculée à partir des signaux IRMf. Plusieurs types d'algorithme de clustering ont été utilisés dans le cadre de l'IRMf : l'algorithme des k-moyennes [14], des k-moyennes floues [64] [20] et le recuit déterministe [221]. La définition de la métrique utilisée est très variable : distance euclidienne dans l'espace d'origine [221], distance basée sur la corrélation [65] ou la métrique de Mahalanobis [101]. Le problème de la détermination du nombre de clusters est traité dans [65][165][200][10]. Les cartes auto-organisées de Kohonen sont aussi utilisées en IRMf comme méthode de clustering des données [39].

## 2.4 Bilan

Les méthodes de cartographie en IRMf sont aujourd'hui très nombreuses. Cependant, il est difficile de prendre du recul et de comparer les différentes méthodes entre elles. Faire la lumière sur les différentes méthodes de cartographie reste très complexe pour différentes raisons. *(i)* Tous les membres de la communauté ne traitent pas les mêmes données. Les images sont en effet gardées précieusement par les différentes équipes et elles sont rarement disponibles pour l'ensemble de la communauté. *(ii)* La vérité terrain n'est pas connue avec certitude. *(iii)* Il existe une grande variabilité des données (acquisition des images avec des scanners différents) et des paradigmes (paradigmes en blocs, événementiels). *(iv)* Les méthodes de cartographie possèdent en général trois étapes (prétraitement, calcul d'une carte de z-score sous l'hypothèse d'inactivité et seuillage de la carte). Toutes les méthodes de cartographie ne nécessitent pas les mêmes prétraitements et il est difficile de discerner l'influence des différentes étapes dans les résultats.

De nombreux logiciels permettent de faire de la cartographie en IRMf. Nous pouvons citer par exemple SPM<sup>3</sup>, brainVisa<sup>4</sup>, brainVoyager<sup>5</sup> et le logiciel développé dans notre équipe Medimax<sup>6</sup>. Cependant, les médecins et neurologues utilisent principalement tous le même logiciel : SPM. Ceci pourrait être un point très positif dans la mesure où cela permet de valider les algorithmes de ce logiciel et d'avoir une méthode de comparaison pour les nouvelles

<sup>3</sup><http://www.fil.ion.ucl.ac.uk/spm/>

<sup>4</sup><http://brainvisa.info/>

<sup>5</sup><http://www.brainvoyager.com/>

<sup>6</sup>[http://alsace.u-strasbg.fr/ipb/gitim/medimax/medimax\\_f.html](http://alsace.u-strasbg.fr/ipb/gitim/medimax/medimax_f.html)

---

méthodes développées. Par exemple, nous comparons par la suite nos approches de cartographie à SPM. Cependant, l'utilisation de ce logiciel reste délicate, le nombre de paramètres que doit rentrer l'utilisateur étant très important. Si les résultats ne sont pas satisfaisants, l'utilisateur peut à son gré changer la matrice de dessin, prendre en compte ou ne pas prendre en compte l'autocorrélation temporelle du bruit, jusqu'à obtenir les résultats souhaités. Cette trop grande liberté laissée à l'utilisateur apparaît surprenante conduisant à s'interroger sur la réelle significativité des résultats obtenus.



# Chapitre 3

## Approches markoviennes cachées : une synthèse

La théorie des modèles de Markov cachés (MMC) a pour origine les travaux du mathématicien russe Markov en analyse du langage par chaînes à états interconnectés, dites chaînes de Markov. C'est au début des années 70 que les modèles markoviens cachés ont véritablement connu leur essor en traitement du signal, essentiellement dans le cadre de la reconnaissance automatique de la parole. Ceci s'explique en partie par la découverte par Baum [17] d'une procédure générale d'ajustement des paramètres des modèles aux caractéristiques du signal à reconnaître et par Viterbi [215] d'un algorithme de décodage utilisé en théorie de l'information puis adapté par la suite au décodage des états cachés de la chaîne. Depuis lors, moyennant l'introduction quasi continue de nouveaux éléments de modélisation, le cadre théorique des modèles markoviens cachés s'est considérablement élargi tandis que les approches par MMC se sont généralisées à d'autres domaines d'application : traitement des images, biologie (analyse de séquences de gènes), météorologie, analyse financière... Aujourd'hui, les réseaux bayésiens dynamiques peuvent être perçus comme les héritiers des MMCs. La richesse de modélisation est beaucoup plus importante avec des réseaux bayésiens dynamiques dans la mesure où le contexte utilisé pour définir les observations peut incorporer plusieurs états cachés. Ils permettent également de modéliser des processus en interaction et donc de fusionner différents types d'information. Le nombre de variantes ou d'extensions des MMCs est tel qu'il nous est impossible de les aborder toutes, du moins dans le cadre de cette thèse. Par contre, nous nous concentrons sur les éléments de modélisation nécessaires à la thèse.

Dans la première section, nous présentons le modèle markovien caché standard (ou chaîne de Markov cachée), ainsi que les problèmes d'inférence et les algorithmes de résolution qui s'y rattachent. La seconde section aborde un élément de modélisation qui a contribué à enrichir le modèle standard, à savoir la modélisation explicite de la durée d'occupation des états cachés. La présentation de cette extension du MMC standard, appelée modèle semi-markovien caché (MSMC), est nécessaire afin de mieux appréhender les fondements théoriques du modèle semi-markovien caché de séquence d'événements développé en deuxième partie de cette thèse. La troisième section est consacrée aux réseaux bayésiens dynamiques lesquels peuvent être vus comme une généralisation des approches markoviennes cachées. Pour les présenter, il est commode de faire appel à la notion de graphe de dépendance. Les problèmes d'inférence et leur résolution y sont présentés dans ce cadre. Enfin, la dernière section de ce chapitre, intitulée « fusion markovienne de processus en interaction » discute des principales approches markoviennes cachées développées dans la littérature pour la modélisation et l'analyse de processus en interaction, l'objectif de cette section étant de cadrer la problématique d'alignement-fusion

multiséquence d'événements développée en partie III.

## 3.1 Le modèle markovien caché standard

### 3.1.1 Définition

Un modèle markovien caché (MMC) est un double processus stochastique  $(\mathbf{X}, \mathbf{O})$  où  $\mathbf{X} = (X_t)_{t=1}^T$  est une chaîne de Markov à temps discret, homogène d'ordre 1, et  $\mathbf{O} = (O_t)_{t=1}^T$  est le processus observable supposé indépendant conditionnellement à  $\mathbf{X}$ . Le modèle est dit caché car le processus  $\mathbf{X}$  ne peut être observé directement. Par contre, il peut être estimé indirectement par l'intermédiaire du processus observable  $\mathbf{O}$ . Par définition, on a :

$$\begin{cases} P[X_{t+1} = q_j | X_t = q_i, \mathbf{X}_1^{t-1}] = P[X_{t+1} = q_j | X_t = q_i] \\ P[\mathbf{O} | \mathbf{X}] = \prod_{t=1}^T P[O_t | X_t] \end{cases} \quad (3.1)$$

Une chaîne de Markov cachée est entièrement définie par la donnée :

- de l'espace d'état du modèle,  $\mathcal{Q} = \{q_j\}_{1 \leq j \leq N}$ , dans lequel le processus aléatoire  $\mathbf{X}$  prend ses valeurs. Le processus  $\mathbf{X}$  est supposé présenter dans chaque état  $q_j$  un ensemble de propriétés distinctes et mesurables.
- du vecteur de distribution initiale du modèle,  $\boldsymbol{\Pi} = (\pi_i)_{1 \leq i \leq N}$ , où  $\pi_i \triangleq P[X_1 = q_i]$ .
- de la matrice de transition du modèle,  $\mathbf{A} = (a_{ij})_{1 \leq i, j \leq N}$ , où  $a_{ij} \triangleq P[X_{t+1} = q_j | X_t = q_i]$ .
- de  $N$  lois d'observation,  $\mathbf{B} = \{b_j(\cdot)\}_{1 \leq j \leq N}$ , chacune définie par :  $P[O_t | X_t = q_j] = b_j(O_t)$ .

Les lois d'observations peuvent être discrètes, continues [187] ou semi-continues [114].

Par suite, un MMC est entièrement défini par la donnée du quadruplet  $\boldsymbol{\lambda} = \{\mathcal{Q}, \boldsymbol{\Pi}, \mathbf{A}, \mathbf{B}\}$ . Pour de plus amples détails, le lecteur pourra se référer au tutoriel classique de Rabiner [186].

### 3.1.2 Graphe d'états

Les sommets du graphe d'états représentent les états de la chaîne. Le graphe a une arête orientée de  $q_i$  vers  $q_j$  si et seulement si  $a_{ij} > 0$ . On associe alors à cette arête l'étiquette  $a_{ij}$ . Le graphe est donc représentatif des possibilités d'évolution séquentielle du processus modélisé. Sa topologie dépend directement des éléments non nuls de la matrice de transition. En particulier, un MMC dont la matrice  $\mathbf{A}$  est pleine (aucun terme nul) est dit ergodique : toutes les possibilités d'évolution sont permises (figure 3.1.a). Les MMCs dont la matrice de transition est triangulaire supérieure sont appelés modèles gauches-droits (figure 3.1.b). Ils sont particulièrement adaptés à la modélisation de processus temporels, le retour sur des états occupés antérieurement étant topologiquement prohibé.

### 3.1.3 Aspects algorithmiques

Les modèles markoviens cachés trouvent leur intérêt dans la résolution des trois problèmes fondamentaux que sont l'évaluation, le décodage et l'apprentissage [186].

#### 3.1.3.1 Evaluation

L'évaluation consiste à déterminer la vraisemblance d'une séquence d'observation  $\mathbf{O} = \{O_1, O_2, \dots, O_T\}$  connaissant les paramètres du modèle  $\boldsymbol{\lambda}$ . Elle s'effectue classiquement selon la procédure avant-arrière dont la complexité algorithmique est en  $O(N^2T)$  (équation (3.2)).

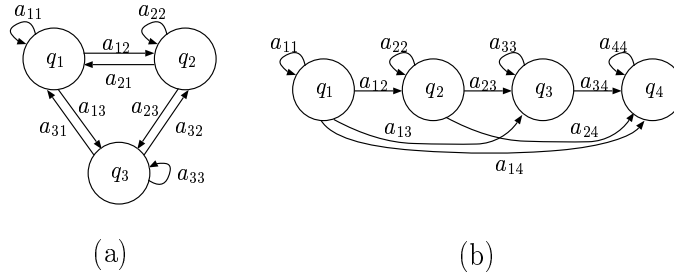


FIG. 3.1 – Graphe d'états d'une chaîne de Markov cachée. (a) Chaîne ergodique ( $N=3$ ) et (b) chaîne de topologie gauche-droite ( $N=4$ ).

On introduit les probabilités avant-arrière, ou « forward-backward »,  $\alpha_t(i)$  et  $\beta_t(i)$ , définies par  $\alpha_t(i) = P[\mathbf{O}_1^t, X_t = q_i | \boldsymbol{\lambda}]$  et par  $\beta_t(i) = P[\mathbf{O}_{t+1}^T | X_t = q_i, \boldsymbol{\lambda}]$ . Elles sont calculées récursivement grâce aux relations dites d'induction :

$$\begin{aligned} \alpha_{t+1}(j) &= \left( \sum_{i=1}^N \alpha_t(i) a_{ij} \right) b_j(O_{t+1}) \\ \beta_t(i) &= \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \end{aligned} \quad (3.2)$$

avec les conditions initiales  $\alpha_1(i) = \pi_i b_i(O_1)$  et  $\beta_T(i) = 1$  pour  $i = 1, \dots, N$ . Le calcul de  $P[\mathbf{O} | \boldsymbol{\lambda}]$  peut alors s'effectuer de  $T$  façons différentes :

$$(\forall t \in [1, T]), P[\mathbf{O} | \boldsymbol{\lambda}] = \sum_{i=1}^N \alpha_t(i) \beta_t(i) \quad (3.3)$$

En pratique, on utilise l'équation (3.3) avec  $t = T$  ( $\beta_T(i)=1$  pour tout  $i$ ) permettant ainsi de ne pas calculer les coefficients arrière. Cependant, ces derniers seront largement utilisés lors de la résolution du problème de l'apprentissage.

Pour terminer, il convient de souligner que le calcul des probabilités avant-arrière conduit à des dépassements inférieurs de capacité. Pour pallier ce phénomène, Levinson [139] propose de multiplier ces probabilités par un facteur d'échelle, variant selon  $t$ , et choisi heuristiquement. Par la suite, Devijver [55] a montré que cela peut revenir (selon le choix du coefficient) à considérer le problème en termes de probabilités conditionnelles et non en termes de probabilités jointes.

### 3.1.3.2 Décodage

Le décodage consiste à inférer à partir de la séquence observée  $\mathbf{O}$  la séquence d'états cachés  $\mathbf{q}^*$  qui en est à l'origine, l'estimation pouvant s'effectuer suivant différents critères. Dans un cadre bayésien, la séquence d'états cachée optimale  $\mathbf{q}^*$  minimise le risque de Bayes  $\mathcal{R}(\mathbf{q})$ , ce dernier étant défini comme l'espérance d'une fonction de coût conditionnellement aux observations :

$$\mathbf{q}^* = \arg \min_{\mathbf{q}} \mathcal{R}(\mathbf{q}) = \arg \min_{\mathbf{q}} E[C(\mathbf{x}, \mathbf{q}) | \mathbf{O}] \quad (3.4)$$

où  $\mathbf{x}$  est le vecteur sur lequel on intègre et  $C(\mathbf{x}, \mathbf{q})$  représente le coût de choisir  $\mathbf{q}$  alors que la solution est  $\mathbf{x}$ . Plusieurs estimateurs bayésiens correspondant à des fonctions de coût différentes peuvent être choisis. On peut citer par exemple l'estimateur du MAP (Maximum A Posteriori) ou du MPM (Mode des Marginales A Posteriori). On peut également se référer à [186] pour d'autres critères.

Le critère du MPM est associé à une fonction de coût qui pénalise la configuration  $\mathbf{q}$  au prorata du nombre de variables aléatoires mal étiquetées :  $C(\mathbf{x}, \mathbf{q}) = \sum_{t=1}^T (1 - \delta_{x_t}(q_t))$ . Ce critère revient finalement à considérer comme optimale la séquence d'états dont les états sont optimaux localement :

$$X_t^{MPM} = \arg \max_{1 \leq i \leq N} P[X_t = q_i | \mathbf{O}, \boldsymbol{\lambda}] = \arg \max_{1 \leq i \leq N} [\alpha_t(i) \beta_t(i)] \quad (3.5)$$

Si le critère MPM maximise le nombre attendu d'états correctement choisis dans la séquence d'états, la séquence d'états obtenue peut cependant être incompatible avec les possibilités d'évolution du modèle. En effet, certaines transitions entre états autorisées par décodage ne peuvent finalement l'être car interdits initialement dans la matrice de transition.

Le critère du MAP, quant à lui, est associé à la fonction de coût qui pénalise de la même façon toutes les configurations différentes de  $\mathbf{x}$  :  $C(\mathbf{x}, \mathbf{q}) = 1 - \delta_{\mathbf{x}}(\mathbf{q})$  où  $\delta_{\mathbf{x}}(\mathbf{q})$  est la masse de Dirac en  $\mathbf{x}$ . On obtient :

$$\mathbf{q}^{MAP} = \arg \min_{\mathbf{q}} \left[ \sum_{\mathbf{x}} C(\mathbf{x}, \mathbf{q}) P[\mathbf{x} | \mathbf{O}, \boldsymbol{\lambda}] \right] = \arg \max_{\mathbf{q}} P[\mathbf{q} | \mathbf{O}, \boldsymbol{\lambda}] \quad (3.6)$$

Le critère du MAP revient donc à choisir la séquence d'états la plus vraisemblance conditionnellement aux observations  $\mathbf{O}$  et au modèle  $\boldsymbol{\lambda}$ . Le décodage de la séquence d'états cachés  $\mathbf{q}^{MAP}$  s'effectue à l'aide de l'algorithme de Viterbi [215] [74] d'où l'appellation décodage de Viterbi. A l'origine, cet algorithme a été utilisé en programmation dynamique pour la recherche de chemins optimaux dans un graphe. L'algorithme définit la variable suivante :

$$\delta_t(j) = \max_{\mathbf{q}_1^{t-1}} P[\mathbf{X}_1^{t-1} = \mathbf{q}_1^{t-1}, X_t = q_j, \mathbf{O}_1^t | \boldsymbol{\lambda}] \quad (3.7)$$

Ces probabilités sont calculées récursivement par la relation :

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t) \quad (3.8)$$

avec pour conditions initiales  $\delta_1(j) = \pi_j b_j(O_1)$  pour  $j = 1 \dots N$ . L'obtention effective de  $\mathbf{q}^{MAP}$  requiert l'introduction d'un pointeur  $\Psi_t(j)$ , associé à  $\delta_t(j)$ , de sauvegarde de l'indice d'état  $i$  maximisant l'équation (3.8). Le chemin optimal  $\mathbf{q}^{MAP}$  est obtenu en fin de procédure par retour arrière sur le tableau de pointeurs en partant du pointeur  $\Psi_T(j^*)$  où  $j^* = \max_j \delta_T(j)$ .

### 3.1.3.3 Apprentissage

L'apprentissage d'un modèle de Markov caché consiste à estimer son vecteur paramètre  $\boldsymbol{\lambda}$  sur la base d'un ensemble de séquences d'observations appelé corpus d'apprentissage.

**3.1.3.3.1 L'algorithme de Baum-Welch** L'algorithme d'apprentissage le plus communément utilisé est l'algorithme de Baum-Welch [15], lequel est une instance de l'algorithme EM (Expectation-Maximisation) appliquée aux MMCs [51],[23]. Il se fonde sur le critère du Maximum de Vraisemblance (MV). Il cherche à maximiser la vraisemblance du corpus d'apprentissage supposé constitué de  $R$  séquences d'observations mutuellement indépendantes <sup>1</sup>  $\mathbf{O} = \{\mathbf{O}^1, \mathbf{O}^2, \dots, \mathbf{O}^R\}$  :

$$\boldsymbol{\lambda}^* = \arg \max_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda}) = \arg \max_{\boldsymbol{\lambda}} \prod_{r=1}^R P[\mathbf{O}^r | \boldsymbol{\lambda}] \quad (3.9)$$

<sup>1</sup>On pourra se référer à [140] si les données ne sont pas indépendantes.



La maximisation de la vraisemblance  $L(\boldsymbol{\lambda})$  ne pouvant être effectuée directement, l'algorithme de Baum-Welch introduit une fonction auxiliaire  $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$  définie comme suit :

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') = \sum_{r=1}^R \sum_{\mathbf{X}^r} P[\mathbf{X}^r | \mathbf{O}^r, \boldsymbol{\lambda}] \log P[\mathbf{X}^r, \mathbf{O}^r | \boldsymbol{\lambda}'] \quad (3.10)$$

On démontre en utilisant l'inégalité de Jensen la propriété suivante de la fonction auxiliaire :

$$\forall(\boldsymbol{\lambda}, \boldsymbol{\lambda}'), \quad Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') - Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}) \leq L(\boldsymbol{\lambda}') - L(\boldsymbol{\lambda}) \quad (3.11)$$

L'algorithme de Baum-Welch s'appuie sur cette propriété en construisant itérativement une suite de vecteurs paramètres  $\{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots\}$  vérifiant  $Q(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_{i+1}) \leq Q(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_i)$ . La suite  $\{L(\boldsymbol{\lambda}_1), L(\boldsymbol{\lambda}_2), \dots\}$  est donc une suite croissante. Elle converge vers un point stationnaire, ce dernier pouvant être un point selle ou un maximum local [51]. La construction de la suite de vecteurs paramètres s'effectue par l'application directe des formules de réestimation des paramètres, résumées dans l'expression suivante :  $\boldsymbol{\lambda}_{i+1} = BW(\boldsymbol{\lambda}_i)$ . Pour obtenir ces dernières, il suffit de remarquer que la fonction auxiliaire de l'équation (3.10) peut s'écrire comme la somme de trois fonctions indépendantes optimisables séparément :

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') = \sum_{r=1}^R \sum_{i=1}^N P[X_1^r = q_i | \mathbf{O}^r, \boldsymbol{\lambda}] \log(\pi'_i) + \sum_{r=1}^R \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T^r-1} P[X_t^r = q_i, x_{t+1}^r = q_j | \mathbf{O}^r, \boldsymbol{\lambda}] \log(a'_{ij}) + \sum_{r=1}^R \sum_{i=1}^N \sum_{t=1}^{T^r} P[X_t^r = q_i | \mathbf{O}^r, \boldsymbol{\lambda}] \log(b'_i(O_t)) \quad (3.12)$$

En notant  $\xi_t^r(i, j) = P[X_t^r = q_i, X_{t+1}^r = q_j | \mathbf{O}^r, \boldsymbol{\lambda}]$  et  $\gamma_t^r(i) = P[X_t^r = q_i | \mathbf{O}^r, \boldsymbol{\lambda}]$ , on obtient, dans le cas de lois d'observation gaussiennes, les formules de réestimation suivantes :

$$\left\{ \begin{array}{l} \pi'_i = \frac{1}{R} \sum_{r=1}^R \gamma_1^r(i) \\ a'_{ij} = \frac{\sum_{r=1}^R \sum_{t=1}^{T^r-1} \xi_t^r(i, j)}{\sum_{r=1}^R \sum_{t=1}^{T^r-1} \gamma_t^r(i)} \\ \mu'_i = \frac{\sum_{r=1}^R \sum_{t=1}^{T^r} \gamma_t^r(i) O_t^r}{\sum_{r=1}^R \sum_{t=1}^{T^r} \gamma_t^r(i)}, \quad \Sigma'_i = \frac{\sum_{r=1}^R \sum_{t=1}^{T^r} \gamma_t^r(i) (O_t^r - \mu'_i)(O_t^r - \mu'_i)^T}{\sum_{r=1}^R \sum_{t=1}^{T^r} \gamma_t^r(i)} \end{array} \right. \quad (3.13)$$

Le calcul de  $\xi$  et de  $\gamma$  s'effectue enfin à partir des probabilités avant et arrière :

$$\begin{aligned} \xi_t^r(i, j) &= \frac{\alpha_t^r(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P[\mathbf{O}^r | \boldsymbol{\lambda}]} \\ \gamma_t(i) &= \sum_{j=1}^N \xi_t(i, j) \text{ si } t = 1 \dots T-1 \text{ et } \gamma_T(i) = \frac{\alpha_T(i)}{P[\mathbf{O}^r | \boldsymbol{\lambda}]} \end{aligned} \quad (3.14)$$

On peut remarquer que les formules de réestimation de l'équation (3.13) correspondent aux estimateurs courants du MV (à condition de pondérer les observations par leur poids).

L'algorithme de Baum-Welch converge vers un maximum local. Aussi, une bonne initialisation est nécessaire à une convergence rapide et à l'obtention finale d'un « bon » vecteur de paramètres. Par conséquent, il est courant d'utiliser une autre procédure qui fournit à l'algorithme de Baum-Welch une bonne estimation des paramètres. Cette procédure est présentée ci-après. Si cette dernière est encore une procédure itérative, elle a l'avantage d'être robuste et de dépendre peu de son initialisation.

**3.1.3.3.2 La procédure de Viterbi** La procédure de Viterbi (segmental k-means algorithm en anglais) maximise la probabilité des séquences d'observations de la base d'apprentissage uniquement le long des chemins optimaux obtenus par décodage de Viterbi [187] :

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} \prod_{r=1}^R P[\mathbf{O}^r, \mathbf{X}_{\boldsymbol{\lambda}}^{r, MAP} | \boldsymbol{\lambda}] \quad (3.15)$$

où  $\mathbf{X}_\lambda^{rMAP}$  est le chemin optimal obtenu à partir de la  $r^{\text{ème}}$  séquence d'observation par décodage de Viterbi et avec le modèle  $\lambda$ . La fonction auxiliaire s'écrit dorénavant :

$$Q(\lambda, \lambda') = \sum_{r=1}^R \log P[\mathbf{X}_\lambda^{rMAP}, \mathbf{O}^r | \lambda'] \quad (3.16)$$

En pratique, à partir d'un modèle initial  $\lambda_1$ , on détermine les  $R$  chemins optimaux (un pour chaque séquence d'observations du corpus). Pour l'ensemble des chemins cachés, on note  $n_{i|j}$  le nombre de transitions de l'état  $i$  à l'état  $j$ ,  $n_i$  le nombre de chemins pour lequel le premier état caché est  $q_i$  et enfin  $\mathbf{o}_i = \{o_{i1}, o_{i2}, \dots, o_{im_i}\}$  l'ensemble des observations effectuées pendant que le modèle était dans l'état caché  $n_i$ . La maximisation de la fonction auxiliaire permet d'obtenir les équations de réestimation suivantes :

$$\begin{cases} \pi'_i = \frac{n_i}{R} \\ a'_{ij} = \frac{n_{i|j}}{\sum_{j=1}^N n_{i|j}} \\ \mu'_i = \frac{\sum_{k=1}^{m_i} o_{ik}}{n_i}, \quad \Sigma'_i = \frac{\sum_{k=1}^{m_i} (o_{ik} - \mu'_i)(o_{ik} - \mu'_i)^T}{n_i} \end{cases} \quad (3.17)$$

On peut remarquer que les formules de réestimation de l'équation (3.17) correspondent aux estimateurs courants du MV sans aucune pondération des observations.

**3.1.3.3.3 Autres procédures d'apprentissage** Le critère du MV est inadapté dans certaines situations, notamment dans le cadre d'une application de reconnaissance multi-modèles. Supposons que l'on ait deux classes de signaux notées A et B. La classe d'un signal inconnu  $\mathbf{O}$  est estimée à partir de l'*a posteriori* du modèle : si  $P[\lambda_A | \mathbf{O}] > P[\lambda_B | \mathbf{O}]$ , le signal est dit de classe A sinon il est dit de classe B. Disposant d'une base d'apprentissage pour chaque classe de signal  $\mathbf{O}^A$  et  $\mathbf{O}^B$ , l'objectif est de déterminer les paramètres des modèles  $\lambda_A$  et  $\lambda_B$ . Ici, on ne cherche pas à modéliser et à caractériser finement les signaux si bien que le critère du MV n'est pas adapté. Au contraire, on cherche plutôt à discriminer les différentes classes entre elles. On peut faire ici une analogie avec les méthodes de réduction de données, certaines d'entre elles cherchant à caractériser les données (analyse en composantes principales par exemple), les autres à les discriminer (analyse discriminante linéaire par exemple).

Il est donc parfois recommandé de faire l'apprentissage sous des critères discriminants tels que minimiser le taux d'erreur de classification sur l'espace d'apprentissage (critère MCE : minimum Classification Error) [197][147][11][123][120] ou utiliser son approximation à travers le critère MMI (Maximum Mutual Information) [12] [223]. Dans le cadre des MMCs, l'algorithme de Baum-Welch étendu au critère MMI est introduit dans [100] pour le cas des densités d'observation discrètes et dans [171] pour le cas des densités d'observation continues.

Un autre défaut des procédures d'apprentissage concerne le caractère markovien prêté *a priori* au signal d'entrée. Celui-ci peut se révéler non fondé *a posteriori*, conduisant inmanquablement à une procédure de réestimation des paramètres sous optimale. L'approche MDI (Minimum Discriminant Information Approach) [58][59] résout ce problème en minimisant l'information de discrimination entre la source, génératrice du signal, et le MMC qui la modélise, sans toutefois exiger de la source qu'elle soit impérativement à caractère markovien. Cette approche est une généralisation des approches MV et MMI.

## 3.2 Le modèle semi-markovien caché

### 3.2.1 Origine

Le modèle markovien caché standard fait l'hypothèse que le processus observable est constitué d'une suite de segments de mesures stationnaires et de durées variables, un état caché de la chaîne étant associé à chaque segment de mesures. Le comportement stationnaire du processus observable sur un segment est modélisé par une loi d'observation  $b_i(\cdot)_{1 \leq i \leq N}$ . L'ordre d'apparition des segments sur l'axe temporel est lié aux probabilités de transitions  $\{a_{ij}\}_{1 \leq i, j \leq N, i \neq j}$ . Enfin, la modélisation de la durée d'un segment est effectuée par les probabilités de bouclage sur états  $\{a_{ii}\}_{1 \leq i \leq N}$ . Il en découle que la durée d'un segment suit une loi géométrique. En effet, le processus caché étant dans l'état  $q_i$  à l'instant  $t$ , la probabilité qu'il y reste encore  $d$  fois est donnée par :

$$p_i(d) = P[\mathbf{X}_{t+1}^{t+d-1} = q_i, X_{t+d} \neq q_i | X_t = q_i] = a_{ii}^{d-1}(1 - a_{ii}) \quad (3.18)$$

La modélisation de la loi de durée par une loi géométrique est souvent incompatible avec la nature réelle des processus à modéliser. Les modèles semi-markovien cachés (MSMC) [195][138] permettent de résoudre ce problème.

### 3.2.2 Définition

Un MSMC est défini sur la base d'un MMC standard auquel est imposée la contrainte de nullité de ses probabilités de bouclage  $\{a_{ii}\}_{1 \leq i \leq N}$  au profit d'une modélisation explicite des temps de séjour dans chaque état. La loi de durée explicite peut être discrète [195] ou continue [138]. On notera par la suite :

$$p_i(d) = P[\mathbf{X}_{t+1}^{t+d-1} = q_i, X_{t+d} \neq q_i | X_{t-1} \neq q_i, X_t = q_i] \quad (3.19)$$

Une fois le processus caché entré dans l'état  $q_i$ , la loi de durée  $p_i(\cdot)$  permet de modéliser le nombre d'observations générées par le modèle ou de manière équivalente, la durée d'occupation de l'état  $q_i$ . La propriété de Markov d'ordre 1 n'est plus vérifiée à chaque instant mais uniquement lorsque le processus caché transite d'un état  $q_i$  vers un état  $q_j$ ,  $j$  différent de  $i$ , justifiant l'appellation « semi-markovien » du modèle. Si  $D$  représente la durée maximale d'occupation d'un même état, alors, la chaîne semi-markovienne est markovienne d'ordre  $D$  :

$$P[X_t | \mathbf{X}_1^{t-1}] = P[X_t | \mathbf{X}_{t-D}^{t-1}] \quad (3.20)$$

En pratique, cela revient à compléter le vecteur de paramètres  $\boldsymbol{\lambda}$  d'un cinquième sous-groupe de paramètres supplémentaires,  $\mathbf{D}$ , relatifs aux lois de durées explicites  $\{p_i(\cdot)\}_{1 \leq i \leq N}$ . Un MSMC est alors entièrement défini par la donnée du vecteur  $\boldsymbol{\lambda} = \{\boldsymbol{Q}, \boldsymbol{\Pi}, \mathbf{A}, \mathbf{B}, \mathbf{D}\}$ .

### 3.2.3 Graphe d'états

Le graphe d'états d'un MSMC est identique à celui d'un MMC, l'unique différence étant que le graphe d'états d'un MSMC ne possède pas d'arête reliant le même sommet ( $a_{ii}=0$ ). Tout modèle markovien d'ordre  $D$  pouvant se ramener à un modèle markovien d'ordre 1 par démultiplication de ses états, il est commode de représenter le graphe d'états d'un MSMC sous sa forme markovienne standard équivalente (figure 3.2). Chaque état  $q_j$  d'un MSMC est alors remplacé par  $D_j$  états markoviens standards  $\{q_j^d\}_{1 \leq d \leq D_j}$ , où  $D_j$  représente la durée maximale

d'occupation de l'état  $q_j$ . Ces états sont connectés en cascade de façon à former une ligne à retard, dans laquelle le modèle s'y trouve piégé pour y émettre  $d$  observations successives à condition d'être entré dans l'état semi-markovien  $q_j$  par l'état markovien  $q_j^d$ .

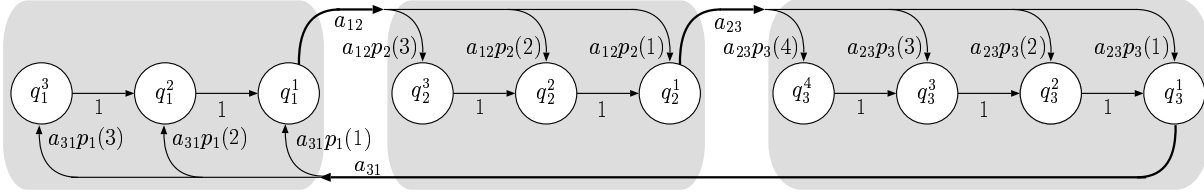


FIG. 3.2 – Représentation markovienne standard équivalente d'un MSMC composé de trois états semi-markoviens  $q_1$ ,  $q_2$  et  $q_3$  (en gris) de durée d'occupation maximale  $D_1=3$ ,  $D_2=3$  et  $D_3=4$  respectivement, et d'une matrice de transition avec trois éléments non nuls ( $a_{12}, a_{23}, a_{31}$ ).

### 3.2.4 Aspects algorithmiques

Les trois problèmes fondamentaux que sont l'évaluation, le décodage et l'apprentissage se posent et se résolvent de façon similaire dans le cadre des MSMCs en utilisant des versions modifiées des algorithmes présentés dans le cadre des MMCs standards [186]. Pour éviter les problèmes numériques de dépassements inférieurs de capacité, une procédure de normalisation est présentée également dans [102]. Enfin, si on considère que pour chaque état  $D_j=D$ , alors le coût calculatoire des « nouvelles » probabilités avant-arrière est approximativement multiplié par  $\frac{D^2}{2}$  par rapport au cas standard [186].

## 3.3 Graphes de dépendance

Pour répondre au large éventail de leurs applications possibles en traitement du signal et de l'image, les modèles markoviens cachés se sont largement complexifiés et diversifiés au fil des années. Pour en faciliter la présentation, leur représentation sous forme de graphe de dépendance se révèle très utile. Cette représentation est privilégiée par la suite dès lors que l'on s'intéresse aux aspects de modélisation de processus aléatoires, et en particulier de processus aléatoires en interaction (partie III).

### 3.3.1 Définitions

Les graphes de dépendance sont une représentation compacte des relations de dépendance entre différentes variables aléatoires. Ils permettent de comprendre rapidement les liens entre les différents processus aléatoires et d'adapter les différents algorithmes utilisés, dans le cadre des graphes. On rappelle que si  $\mathbf{A}$ ,  $\mathbf{B}$  et  $\mathbf{C}$  sont trois ensembles de variables aléatoires, on dit, par définition, que  $\mathbf{A}$  et  $\mathbf{B}$  sont indépendants conditionnellement à (ou sachant)  $\mathbf{C}$  si et seulement si  $P[\mathbf{A}, \mathbf{B} | \mathbf{C}] = P[\mathbf{A} | \mathbf{C}] P[\mathbf{B} | \mathbf{C}]$ . Cette relation d'indépendance conditionnelle est notée  $\mathbf{A} \perp \mathbf{B} | \mathbf{C}$ .

### 3.3.1.1 Graphe de dépendance non orienté

Un graphe de dépendance non orienté est un graphe dont les nœuds représentent les variables aléatoires alors que les arêtes, par leur absence, représentent les relations d'indépendance conditionnelle entre les variables. Par définition, un processus aléatoire  $\mathbf{X}$  défini sur un graphe de dépendance non orienté est un champ de Markov. Il vérifie ainsi la propriété de Markov globale pour tout ensemble de nœuds  $\mathbf{A}, \mathbf{B}$  et  $\mathbf{C}$ . Si  $\mathbf{C}$  sépare  $\mathbf{A}$  et  $\mathbf{B}$  (tous les chemins entre un nœud de  $\mathbf{A}$  et un nœud de  $\mathbf{B}$  passent obligatoirement par un nœud de  $\mathbf{C}$ ) alors,  $\mathbf{A}$  et  $\mathbf{B}$  sont indépendants conditionnellement à  $\mathbf{C}$ . Il existe également une propriété de Markov locale : pour chaque nœud  $i$  du graphe, la variable aléatoire  $X_i$  est indépendante de toutes les autres variables connaissant ses voisins, les nœuds étant voisins s'ils sont reliés par une arête. Théoriquement, un système de voisinage peut être défini par les arêtes du graphe dans la mesure où l'on doit simplement vérifier les deux conditions suivantes : si  $X_i$  est voisin de  $X_j$ , alors  $X_j$  est voisin de  $X_i$  et  $X_i$  n'appartient pas à son propre voisinage.

Finalement, la distribution conjointe de  $\mathbf{X}$  s'écrit sous la forme d'un champ de Gibbs à condition que  $P[\mathbf{X}]$  soit strictement positive pour tout  $\mathbf{X}$  (théorème de Hammersley-Clifford) :

$$P[\mathbf{X}] = \frac{1}{Z} \prod_{C \in \mathcal{C}} \Psi_C(\mathbf{X}_C) \quad (3.21)$$

$\mathcal{C}$  désigne l'ensemble des cliques maximales<sup>2</sup>,  $\Psi_C(\cdot)$  une fonction potentielle<sup>3</sup> à valeur réelle positive sur la clique  $C$ ,  $\mathbf{X}_C$  la restriction de  $\mathbf{X}$  à  $C$  et,  $Z$  un facteur de normalisation. Un exemple de graphe de dépendance non orienté est présenté figure 3.3 :

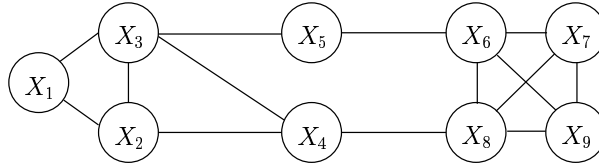


FIG. 3.3 – Graphe de dépendance d'un champ de Markov. L'ensemble des voisins de  $X_3$  est  $\{X_1, X_2, X_4, X_5\}$ . Une des propriétés globales de Markov s'écrit :  $\{X_1, X_2, X_3\} \perp \{X_6, X_7, X_8, X_9\} | \{X_4, X_5\}$  alors que l'une des propriétés locales s'écrit :  $\{X_5\} \perp \{X_1, X_2, X_3, X_4, X_6, X_7, X_8, X_9\} | \{X_3, X_6\}$ . On a enfin :  $P[\mathbf{X}] = \frac{1}{Z} \Psi_{123}(X_1, X_2, X_3) \Psi_{234}(X_2, X_3, X_4) \Psi_{48}(X_4, X_8) \Psi_{35}(X_3, X_5) \Psi_{56}(X_5, X_6) \Psi_{6789}(X_6, X_7, X_8, X_9)$ .

### 3.3.1.2 Graphe de dépendance orienté

Un graphe de dépendance orienté est appelé également *réseau bayésien*. Il s'agit d'un graphe dont les nœuds représentent les variables aléatoires et dont les arêtes orientées représentent les relations de dépendance entre les différentes variables aléatoires. Un arc de  $A$  vers  $B$  signifie que  $A$  a tendance à causer  $B$ . Les graphes de dépendance orientés sont donc acycliques (en prenant en compte l'orientation des arêtes, il n'y a pas de chemin fermé). Il est commode de

<sup>2</sup>Une clique correspond soit à un nœud, soit à un sous-ensemble du graphe tel que tous les sites soient mutuellement voisins. Une clique est dite maximale si aucune autre clique ne la contient.

<sup>3</sup>Généralement, la fonction potentielle désigne le terme apparaissant dans la forme exponentielle  $P[\mathbf{X}] = \frac{1}{Z} \exp^{-\sum_{C \in \mathcal{C}} V_C(\mathbf{X}_C)}$ .

distinguer, pour un nœud donné  $X_i$ , ses parents  $parent(i)$  (nœuds qui « pointent » vers lui), ses enfants (nœuds vers lesquels il pointe) et sa descendance (enfants, enfants des enfants...).

Pour comprendre les relations de dépendance qu'exprime un graphe de dépendance orienté, on considère tout d'abord l'exemple décrit à la figure 3.4. La pluie ou le fait que le jardinier arrose le jardin aboutissent aux mêmes effets : la terre devient grasse. Ainsi, on a  $P[J = 1|G =$

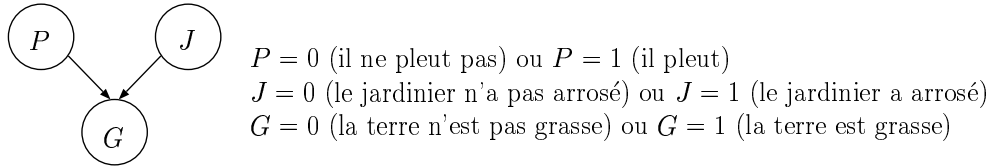


FIG. 3.4 – Voisinage et réseau bayésien

$P = 1] < P[J = 1|G = 1]$  car dans le premier cas, le fait que le terrain soit gras a déjà été expliqué par le fait qu'il pleuve. En définitive,  $P$  et  $J$  ne sont pas indépendants sachant  $G$ . Les parents d'un même enfant ne sont donc pas mutuellement indépendants sachant cet enfant. Les relations de dépendance peuvent s'exprimer de deux façons différentes.

Tout d'abord, un nœud est conditionnellement indépendant de ses non-descendants étant donné ses parents. Par exemple, pour le réseau bayésien de la figure 3.5.a., on a  $X_1 \perp X_2$ ,  $X_2 \perp \{X_1, X_3\}$  ou  $X_4 \perp \{X_1, X_2, X_3\} | \{X_2, X_3\}$ . Le graphe n'ayant pas de cycle, il est possible d'indicer les nœuds de manière à ce qu'il n'y ait aucun descendant de  $X_i$  dans  $\mathbf{X}_1^{i-1}$  et que tous les parents de  $X_i$  soient dans  $\mathbf{X}_1^{i-1}$ . Il découle de la propriété précédente que  $P[X_i | \mathbf{X}_1^{i-1}] = P[X_i | parent(X_i)]$  dans la mesure où l'on a :  $\mathbf{A} \perp \mathbf{B} | \mathbf{C} \Rightarrow P[\mathbf{A} | \mathbf{B}, \mathbf{C}] = P[\mathbf{A} | \mathbf{C}]$ . On obtient en définitive ( $K$  représentant dans l'équation ci-dessous le nombre de composantes de  $\mathbf{X}$ ) :

$$P[\mathbf{X}] = \prod_{i=1}^K P[X_i | parent(i)] \quad (3.22)$$

Une seconde façon pour exprimer les relations de dépendance dans un graphe orienté est donnée ci-après : un nœud est indépendant de tous les autres nœuds du graphe sachant ses voisins, les voisins du nœud étant ses parents, ses enfants et les autres parents de ses enfants. Par exemple, pour l'exemple de la figure 3.5, on a :  $X_3 \perp \{X_1, X_2, X_4, X_5\} | \{X_1, X_2, X_4\}$ . On déduit de cette propriété qu'un graphe orienté peut se présenter de manière équivalente sous forme d'un graphe non orienté (l'inverse n'étant pas toujours vrai [52]). On convertit un graphe orienté en un graphe non orienté en deux étapes : (i) toutes les arêtes orientées deviennent non orientées et (ii) on rajoute des arêtes entre les parents (non liés) qui ont des enfants en commun. Par exemple, le graphe de dépendance orienté de la figure 3.5.a peut également se représenter sous le graphe non orienté de la figure 3.5.b.

### 3.3.2 Les réseaux bayésiens dynamiques

Les réseaux bayésiens dynamiques incorporent l'idée d'évolution temporelle. Ils élargissent ainsi les réseaux bayésiens en modélisant des probabilités de distribution sur une collection de variables aléatoires  $\mathbf{X}^1, \dots, \mathbf{X}^t$  qui évoluent dans le temps. En pratique, chacun des  $\mathbf{X}^i$  représente un réseau bayésien de même topologie. Ainsi, un réseau bayésien dynamique est défini par un réseau bayésien et par des dépendances temporelles entre les différents réseaux,

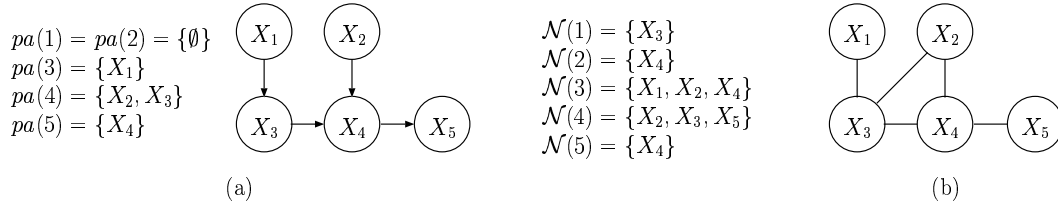


FIG. 3.5 – (a) Graphe de dépendance d'un graphe orienté (b) Graphe de dépendance non orienté modélisant les dépendances du graphe représenté en (a). Tout processus défini sur le graphe orienté (a) peut également être représenté par un champ de Markov sur le graphe non orienté (b). Pour cela, il suffit de vérifier les conditions suivantes :  $\Psi_{13}(X_1, X_3) = P[X_1]P[X_3|X_1]$ ,  $\Psi_{234}(X_2, X_3, X_4) = P[X_2]P[X_4|X_2, X_3]$  et  $\Psi_{45}(X_4, X_5) = P[X_5|X_4]$ . En effet, on a  $P[\mathbf{X}] = P[X_1]P[X_2]P[X_3|X_1]P[X_4|X_2, X_3]P[X_5|X_4]$  pour (a) et  $P[\mathbf{X}] = \Psi_{13}[X_1, X_3]\Psi_{234}[X_2, X_3, X_4]\Psi_{45}P[X_4, X_5]$  pour (b)

ces dépendances ne variant pas avec le temps. En général, on suppose le modèle markovien d'ordre 1. En supposant les indices bien arrangés, on obtient alors :

$$P[\mathbf{X}^t | \mathbf{X}^1, \dots, \mathbf{X}^{t-1}] = P[\mathbf{X}^t | \mathbf{X}^{t-1}] = \prod_{i=1}^K P[X_i^t | parent(X_i^t)] \quad (3.23)$$

$K$  étant le nombre de variables aléatoires de  $\mathbf{X}^t$  ( $\forall t$ ),  $X_i^t$  le  $i^{\text{ème}}$  nœud au temps  $t$  et  $parent(X_i^t)$  les nœuds parents de  $X_i^t$  dans le graphe (ces nœuds peuvent être associés au temps  $t$  ou  $t - 1$  en supposant le modèle markovien d'ordre 1). La figure 3.6 présente le graphe de dépendance d'un réseau bayésien dynamique. Il apparaît clairement que le réseau bayésien dynamique est une suite de réseaux bayésiens, chaque réseau bayésien étant entouré par un rectangle. On voit également les liens de dépendance entre les différents réseaux. On peut enfin remarquer qu'un réseau bayésien dynamique est également un réseau bayésien

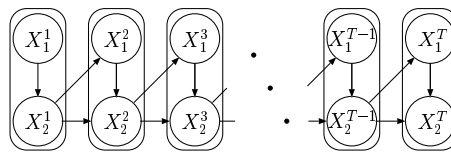


FIG. 3.6 – Graphe de dépendance d'un réseau bayésien dynamique. Un réseau bayésien dynamique est une collection de réseaux bayésiens de même topologie. Sur la figure, chaque bloc de deux variables aléatoires représente le « même » réseau bayésien. Les arêtes sortant des blocs représentent les dépendances temporelles.

Un autre exemple de réseau bayésien dynamique est la chaîne de Markov cachée (modèle standard défini au premier chapitre). Avec les mêmes notations que celles utilisées pour la présentation du modèle markovien standard, on obtient le graphe de dépendance de la figure 3.7. A noter que les nœuds des observations apparaissent en gris. Nous utiliserons cette convention par la suite.

Après avoir défini ce qu'est un graphe de dépendance, il nous reste maintenant à résoudre les problèmes d'inférence dans ce cadre général.

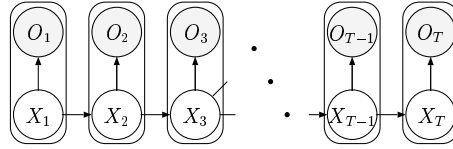


FIG. 3.7 – Graphe de dépendance du modèle de Markov caché standard

### 3.3.3 Problèmes d'inférence : évaluation et décodage

Les problèmes d'inférence sont présentés dans le cadre des champs de Markov (graphe non orienté) dans la mesure où l'on peut transformer un graphe orienté en un graphe non orienté. On suppose que  $\mathbf{X}$  est le processus caché (prenant ses valeurs dans un espace fini et discret) et que  $\mathbf{O}$  est le processus observable. L'évaluation consiste à calculer la vraisemblance  $P[\mathbf{O}]$  alors que le décodage consiste le plus souvent à déterminer le chemin caché au sens du MAP :

$$\begin{cases} P(\mathbf{O}) = \sum_{\mathbf{X}} P[\mathbf{O}, \mathbf{X}] & (i) \\ \mathbf{X}^{MAP} = \arg \max_{\mathbf{X}} P[\mathbf{O}, \mathbf{X}] & (ii) \end{cases} \quad (3.24)$$

Pour une étude exhaustive des méthodes de résolution des problèmes de l'évaluation et du décodage, le lecteur peut se référer à [103]. Nous insistons ici sur les méthodes les plus utilisées en pratique, en commençant par les méthodes conduisant aux solutions exactes des équations (3.24.i) et (3.24.ii), pour terminer sur les méthodes visant à approcher ces dernières. Par la suite, nous utilisons pour résoudre les problèmes d'inférence des versions adaptées des algorithmes présentés dans le cadre des MMCs standards. Ainsi, cette section n'est pas essentielle pour la compréhension du travail de thèse. Il nous semble néanmoins intéressant de présenter comment les problèmes d'inférence se résolvent dans un cadre général.

#### 3.3.3.1 Calcul exact

Le processus observable n'intervient pas dans la sommation de l'équation (3.24.i). Par conséquent, nous voyons simplement comment il est possible de calculer  $P[X_i]$ . Le calcul de  $P[X_i]$  nécessite en effet de marginaliser sur toutes les variables aléatoires sauf  $X_i$ . Le même principe permet de calculer  $P[X_i, \mathbf{O}]$  et donc  $P[\mathbf{O}]$  après une dernière étape de marginalisation suivant  $X_i$ . Enfin, nous présentons très brièvement le décodage au sens du MAP dans la mesure où les problèmes de l'estimation et du décodage se résolvent de manière analogue.

Il suffit pour résoudre le problème de l'évaluation d'effectuer une sommation (équation (3.24.i)) ! Cependant, des algorithmes ont été mis en place de manière à factoriser la somme de l'équation (3.24.i). Ce problème de factorisation est résolu dans un cadre très général par Aji [3]. Les algorithmes présentés ici ne sont qu'une illustration de [3].

**3.3.3.1.1 Algorithme du passage de message** L'algorithme du passage de message est utilisé pour résoudre les problèmes d'évaluation et de décodage dès lors que le graphe est un arbre [150]. Il généralise ainsi les algorithmes d'inférence (évaluation et décodage de Viterbi) qui existent par exemple pour les chaînes de Markov cachées et le quadarbre. L'algorithme du passage de message peut prendre des formes très variées.

Pour des raisons de simplicité, nous commençons par présenter l'algorithme du passage de message en utilisant l'exemple donné par [150] et illustré figure 3.8.a. Cet exemple est aussi général ou particulier que la résolution du problème de l'évaluation sur un arbre. Cependant,



en raison de sa simplicité, on a une compréhension intuitive de l'algorithme. Pour des aspects théoriques, le lecteur peut se référer à [3]. Chaque nœud de l'arbre représente désormais un individu. Chaque individu a le droit de parler uniquement à ses voisins (son père ou ses enfants). A la fin des différentes discussions, chaque individu doit connaître le nombre de personnes dans l'arbre. L'algorithme est le suivant :

- 1 Si un individu a 1 voisin, il lui dit « 1 ».
- 2 S'il a reçu  $N - 1$  messages ( $N$  étant le nombre ses voisins), il identifie le voisin qui ne lui a pas envoyé de messages et lui dit la somme de tous les nombres qu'il a reçus additionné de 1.
- 3 S'il a reçu un message de tous ses voisins, le nombre de personnes dans l'arbre correspond alors à la somme de tous les nombres qu'il a reçus additionné de 1. A chacun des voisins à qui il n'a pas parlé, il dit le nombre d'individus diminué de  $v$  où  $v$  est le nombre que lui avait donné ledit voisin.

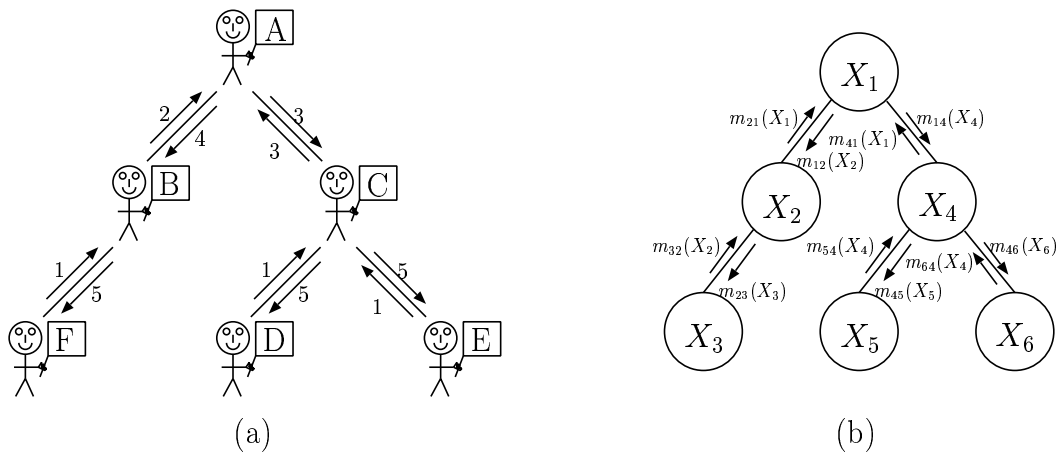


FIG. 3.8 – Exemple de l'algorithme du passage de message dans deux applications différentes.

Un individu reçoit et envoie un message à chacun de ces voisins. Ainsi, il y a  $2(K - 1)$  messages envoyés,  $K$  étant le nombre d'individus. L'ordre dans lequel les messages sont générés n'est pas unique, un individu pouvant parler dès lors qu'il en a la possibilité.

On considère désormais un arbre formé de variables aléatoires notées  $X_i$  (graphe de dépendance, figure 3.8.b). L'algorithme du passage de message prend alors la forme suivante : le nœud  $i$  envoie au nœud  $j$  un message  $m_{ij}$  constitué de  $N$  réels (où  $N$  est le nombre d'états cachés) dès lors qu'il lui est possible de calculer ce dernier. Si  $\mathcal{N}(i)$  représente les voisins du nœud  $i$ , on a :

$$m_{ij}(X_j) = \sum_{X_i} \psi_{ij}(X_i, X_j) \prod_{k \in \mathcal{N}(i) \setminus j} m_{ki}(X_i) \text{ pour } X_j = q_1, \dots, q_N \quad (3.25)$$

Une fois qu'un nœud  $X_i$  a reçu tous les messages de voisins, il peut déterminer la vraisemblance  $P[X_i]$  en utilisant la relation :

$$P[X_i] = \frac{1}{Z} \prod_{k \in \mathcal{N}(i)} m_{ki}(X_i) \quad (3.26)$$

Pour résoudre le problème de l'évaluation, il suffit de prendre le premier  $i$  pour lequel on connaît  $P[X_i]$  et de marginaliser suivant  $X_j$ . On retrouve ici le fait que le calcul des  $\alpha$  est suffisant

pour résoudre le problème de l'évaluation dans le cadre du modèle markovien caché standard. Cependant, le calcul complet est nécessaire pour la résolution du problème de l'apprentissage ou du décodage au sens du MPM.

Nous nous proposons ici de calculer  $P[X_4]$  et de vérifier que nous retrouvons bien  $P[X_4] = \frac{1}{Z} m_{54}(X_4) m_{64}(X_4) m_{14}(X_4)$ .

$$\begin{aligned}
P[X_4] &= \frac{1}{Z} \sum_{X_1, X_2, X_3, X_5, X_6} \psi_{12}(X_1, X_2) \psi_{23}(X_2, X_3) \psi_{14}(X_1, X_4) \\
&\quad \psi_{45}(X_4, X_5) \psi_{46}(X_4, X_6) \\
&= \frac{1}{Z} (\sum_{X_5} \psi_{45}(X_4, X_5)) \times (\sum_{X_6} \psi_{46}(X_4, X_6)) \times \\
&\quad (\sum_{X_1} \psi_{14}(X_1, X_4) \sum_{X_2} \psi_{12}(X_1, X_2) \sum_{X_3} \psi_{23}(X_2, X_3)) \quad (3.27) \\
&= \frac{1}{Z} m_{54}(X_4) m_{64}(X_4) (\sum_{X_1} \psi_{14}(X_1, X_4) \sum_{X_2} \psi_{12}(X_1, X_2) m_{32}(X_2)) \\
&= \frac{1}{Z} m_{54}(X_4) m_{64}(X_4) (\sum_{X_1} \psi_{14}(X_1, X_4) m_{21}(X_1)) \\
&= \frac{1}{Z} m_{54}(X_4) m_{64}(X_4) m_{14}(X_4)
\end{aligned}$$

L'inférence pour les graphes en arbre se fait donc de manière exacte avec des algorithmes à faible coût calculatoire (envoi de  $2(K-1)$  messages,  $K$  étant le nombre de variables aléatoires, chaque message étant exprimé sous la forme de l'équation (3.25)). Ceci explique que beaucoup de modèles markoviens ont une structure en forme d'arbre. Enfin, la détermination du chemin optimal au sens du MAP fonctionne suivant le même principe, la sommation étant remplacée par une maximisation.

**3.3.3.1.2 Algorithme de l'arbre de jonction** L'algorithme de l'arbre de jonction permet de marginaliser suivant un graphe quelconque [134] et non plus obligatoirement sur un arbre. Son principe consiste néanmoins à construire un arbre (de jonction) à partir duquel on peut utiliser l'algorithme vu précédemment. Par exemple, l'arbre présenté figure 3.9.b est un arbre de jonction du graphe de dépendance de la figure 3.9.a. Les nœuds de l'arbre correspondent à un ensemble de variables aléatoires du processus à analyser. Par exemple, l'arbre de jonction a trois nœuds et le premier nœud contient les variables aléatoires  $\{X_1, X_2, X_3\}$ . L'arbre n'est pas un arbre quelconque, il s'agit d'un arbre de jonction : les nœuds qui sont sur l'unique chemin entre les nœuds  $C_i$  et  $C_j$  contiennent les variables aléatoires  $C_i \cap C_j$ . Ainsi, nous vérifions sur notre exemple que le nœud  $C_2$  contient les variables aléatoires  $C_1 \cap C_3 = X_3$ . A une arête de l'arbre reliant  $C_i$  et  $C_j$  est associée un séparateur  $S_{ij}$  constitué des variables aléatoires  $C_i \cap C_j$ . L'arbre de jonction est construit de manière à avoir la relation suivante :

$$P[\mathbf{X}] = \frac{\prod_i P[C_i]}{\prod_{ij} P[S_{ij}]} \quad (3.28)$$

D'après l'équation (3.28), on a :  $P[\mathbf{X}] = \frac{P[X_1, X_2, X_3] P[X_2, X_3, X_4] P[X_3, X_4, X_5]}{P[X_2, X_3] p[X_3, X_4]}$ . En effet,  $(X_2, X_3, X_4)$  sépare  $X_1$  et  $X_5$ .

Les algorithmes vus précédemment dans le cas des arbres sont utilisés sur l'arbre de jonction. Les messages étant désormais les suivants :

$$m_{ij}(\mathbf{X}_{S_{ij}}) = \sum_{C_i \setminus S_{ij}} \Psi_{C_i}(\mathbf{X}_{C_i}) \prod_{k \in \mathcal{N}(i) \setminus j} m_{ki}(\mathbf{X}_{S_{ki}}) \quad (3.29)$$

pour tous les  $\mathbf{X}_{S_{ij}}$  possibles

où  $\mathcal{N}(i)$  sont les cliques voisines (ou nœuds voisins) de  $C_i$  dans l'arbre de jonction. Le message  $m_{ij}(\mathbf{X}_{S_{ij}})$  comporte  $N^{|S_{ij}|}$  réels où  $N$  est le nombre d'états cachés supposé identique pour chaque variable aléatoire de  $S_{ij}$ . On obtient finalement :

$$p(\mathbf{X}_{C_i}) = \frac{1}{Z} \prod_{k \in \mathcal{N}(i)} m_{ki}(\mathbf{X}_{S_{ki}}) \quad (3.30)$$

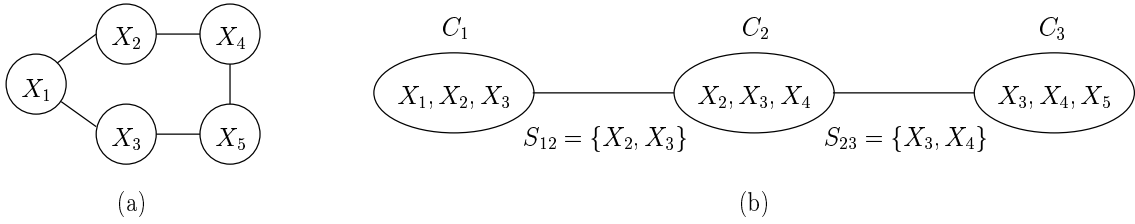


FIG. 3.9 – Algorithme de l’arbre de jonction. (a) Graphe de dépendance quelconque. (b) Un arbre de jonction du graphe de dépendance illustré en (a).

La marginalisation sur un unique nœud peut enfin être obtenue par une étape de marginalisation supplémentaire. La séquence d’états cachés optimale au sens du MAP peut également être obtenue à partir de l’arbre de jonction [46]. Il reste désormais à définir comment l’arbre de jonction est construit.

On commence tout d’abord par définir un ordre de marginalisation  $(X_1, X_2, X_3, X_4, X_5)$  pour l’exemple de la figure 3.10 (on cherche à déterminer l’arbre de jonction du graphe de dépendance illustré figure 3.10.a). En marginalisant suivant  $X_1$ , on obtient le graphe de dépendance de la figure 3.10.b [181]<sup>4</sup>. En effet, en sommant par rapport à  $X_1$ , on obtient :

$$P[X_2, \dots, X_5] = \psi_{24}(X_2, X_4) \times \psi_{45}(X_4, X_5) \times \psi_{35}(X_3, X_5) \times \underbrace{\psi_{23}(X_2, X_3)}_{\sum_{X_1} \psi_{12}(X_1, X_2) \psi_{13}(X_1, X_3)} \quad (3.31)$$

Ensuite, en marginalisant par rapport à  $X_2$ , on obtient le graphe de la figure 3.10.c. Une fois l’ensemble des transformations sur les graphes terminé, on ajoute au graphe initial les arêtes nouvellement créées par les différentes opérations (figure 3.10.f). Le graphe obtenu est un graphe triangulé. Tout processus défini sur le graphe d’origine peut être défini sur le graphe triangulé dans la mesure où l’on a simplement ajouté des arêtes. De plus, en marginalisant avec le même ordre que précédemment, aucune nouvelle arête n’est désormais créée. L’arbre de jonction est calculé à partir de ce graphe triangulé. A noter que tout graphe triangulé de la figure 3.10.a ne permet pas de construire un arbre de jonction.

On construit ensuite le graphe dont les nœuds correspondent aux cliques maximales du graphe triangulé et dont les arêtes relient des cliques qui ont au moins une variable aléatoire en commun. On associe alors à l’arête un poids correspondant au nombre de variables aléatoires en commun (figure 3.10.g). L’arbre de jonction est alors défini comme étant un arbre de recouvrement de poids maximal du graphe précédemment construit<sup>5</sup>.

La principale difficulté consiste à déterminer l’ordre d’élimination des différents nœuds (ou l’ordre de marginalisation) de manière à ce que les cliques de l’arbre de jonction soient les plus petites possibles [135] [204][137], la complexité algorithmique étant exponentielle par rapport à la taille de la plus grande clique. Cependant, dès lors que le nombre d’arêtes et le nombre de nœuds est important [219], cet algorithme ne peut pas être utilisé en pratique : aussi utilise-t-on des méthodes approximatives comme les méthodes de Monte-Carlo ou les méthodes variationnelles.

<sup>4</sup>Si  $X$  est markovien sur un graphe  $G$ , alors la restriction  $X_A$  de  $X$  au sous-ensemble de nœuds  $A$  est markovien ( $p(X_A) = \sum_{X_{\bar{A}}} P[X_A, X_{\bar{A}}]$  où  $\bar{A} = S \setminus A$ ,  $S$  étant l’ensemble des variables aléatoires de  $G$ ).  $X_A$  est markovien sur un graphe tel que deux sites de  $A$  sont voisins s’ils sont voisins dans  $G$  ou s’ils appartiennent dans  $G$  au voisinage d’une même composante connexe de  $\bar{A}$ .

<sup>5</sup>Un arbre de recouvrement maximal d’un graphe est un arbre contenant chaque sommet du graphe construit de manière à maximiser la somme cumulée des différents poids des arêtes

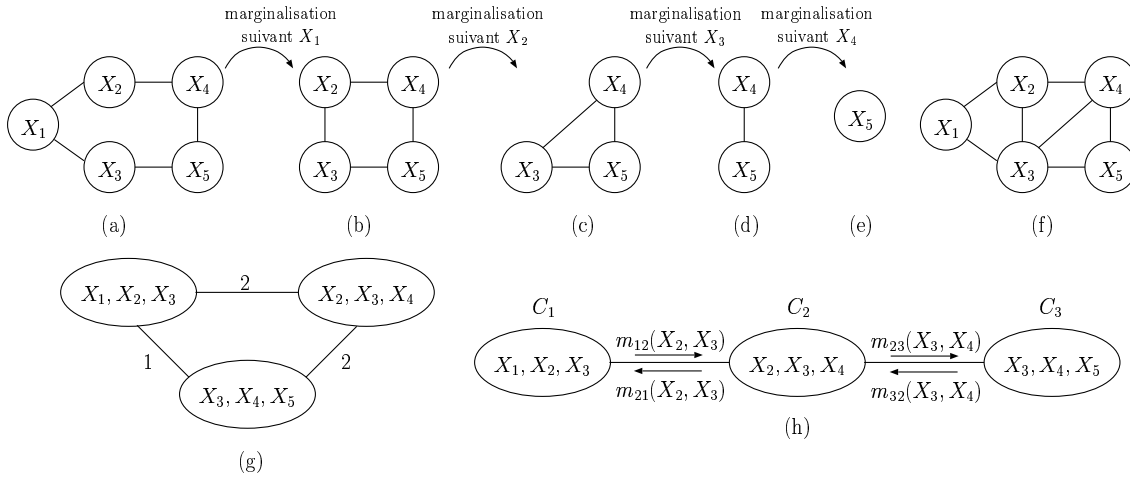


FIG. 3.10 – Résumé des grandes étapes pour construire l'arbre de jonction (voir texte). L'arbre décrit en (h) est un arbre de jonction du graphe de dépendance illustré en (a)

### 3.3.3.2 Méthodes approximatives

**3.3.3.2.1 Les méthodes de Monte-Carlo** Les méthodes de Monte-Carlo font l'approximation suivante :

$$E[f(\mathbf{X})] = \int f(\mathbf{X})p(\mathbf{X})d\mathbf{X} \approx \frac{1}{K} \sum_{t=1}^K f(\mathbf{X}^t) \text{ où } \mathbf{X}^t \rightsquigarrow p(\mathbf{X}) \quad (3.32)$$

Dans notre application, les méthodes de Monte-Carlo permettent de résoudre le problème d'estimation (prendre  $f(\mathbf{X}) = P[\mathbf{O}|\mathbf{X}]$  dans l'équation (3.32)) et de décodage sous le critère MPM (il suffit d'associer à la variable aléatoire cachée  $X_i$  l'état caché apparaissant le plus fréquemment dans la suite des échantillons pour le site  $i$ ). Dans le cas où le critère MAP est choisi, il est possible d'associer aux méthodes de Monte-Carlo un algorithme de recuit simulé [106] [94]. L'intérêt de cet algorithme est d'autoriser, au début, lorsque la température est élevée, le choix temporaire de configurations qui diminuent la vraisemblance dans le but de sortir des maxima locaux. Ensuite, en diminuant progressivement la température, le recuit simulé va évoluer vers un comportement déterministe refusant les configurations qui diminuent la vraisemblance. Pour palier au coût calculatoire de ces algorithmes, l'algorithme ICM (Iterated Conditioned Modes) [22] a été introduit. Il correspond au recuit simulé à température nulle : en chaque site  $i$ , on choisit l'état caché  $X_i$  le plus vraisemblable conditionnellement à l'état caché de ses voisins (et à l'observation  $O_i$ ). L'algorithme converge en un nombre fini d'itérations vers un maximum local qui dépendra de l'initialisation et de la stratégie de balayage des sites.

Quand la dimension de la loi  $p(\cdot)$  est importante, la principale difficulté réside dans la simulation de la loi  $p(\cdot)$  : on a alors recours aux algorithmes de Monte Carlo par chaînes de Markov . Il s'agit de méthodes permettant de produire une chaîne de Markov ergodique admettant  $p(\cdot)$  pour loi stationnaire. A l'équilibre, c'est-à-dire après un régime transitoire de  $M$  itérations, cet algorithme génère des échantillons sous la loi  $p(\cdot)$  et on réalise l'approximation  $E[f(\mathbf{X})] \approx \frac{1}{K-M} \sum_{t=M+1}^K f(\mathbf{X}^t)$  où  $\mathbf{X}^t$  est la  $t^{\text{ème}}$  observation générée par la chaîne. Les deux principaux algorithmes permettant de produire de telles chaînes sont l'algorithme de Metropolis-Hasting [159] et l'échantillonneur de Gibbs [94]. Nous allons présenter brièvement ces derniers.

L'algorithme de Metropolis-Hasting consiste à :

- générer  $\mathbf{Y}$  selon  $q(\cdot|\mathbf{X}^t)$  ( $q$  est appelée loi instrumentale) ;
- accepter  $\mathbf{Y}$  ( $\mathbf{X}^{t+1} \stackrel{\Delta}{=} \mathbf{Y}$ ) avec la probabilité  $\min\left(1, \frac{p(\mathbf{Y})q(\mathbf{X}^t|\mathbf{Y})}{p(\mathbf{X}^t)q(\mathbf{X}^t|\mathbf{Y})}\right)$  et en cas de rejet,  $\mathbf{X}^{t+1} \stackrel{\Delta}{=} \mathbf{X}^t$ .

La chaîne de Markov générée selon l'algorithme de Metropolis-Hasting admet  $p$  pour loi stationnaire pour une loi  $q(\cdot)$  quelconque. Cependant, le choix de  $q$  est guidé par le compromis suivant. Le taux d'acceptation doit être élevé pour que l'algorithme converge rapidement (temps de transitoire faible) mais pas trop pour que l'algorithme puisse décrire le support de  $p(\cdot)$  en un temps raisonnable. En pratique,  $\mathbf{Y}$  est souvent déterminé à partir de  $\mathbf{X}$  en choisissant un site au hasard (loi uniforme) et en tirant un nouvel état caché selon une loi uniforme. Le calcul de  $\frac{p(\mathbf{Y})}{p(\mathbf{X})}$  se réduit donc à :

$$\frac{p(\mathbf{Y})}{p(\mathbf{X})} = \frac{\sum_{C \in \mathcal{C}_{X,Y}} \psi_C(\mathbf{Y}_C)}{\sum_{C \in \mathcal{C}_{X,Y}} \psi_C(\mathbf{X}_C)} \quad (3.33)$$

où  $\mathcal{C}_{X,Y}$  représente l'ensemble des cliques maximales où apparaissent les sites pour lesquelles  $\mathbf{Y}$  et  $\mathbf{X}$  n'ont pas les mêmes états cachés.

L'échantillonneur de Gibbs est quant à lui un algorithme dans lequel les composantes de la variable d'intérêt sont mises à jour une par une, soit dans l'ordre lexicographique, soit par tirage aléatoire. On pose alors  $X_i^t = X_i^{t-1}$  pour tous les sites  $i$  exceptés pour la composante  $k$  qui va être mise à jour selon la loi :  $X_k^t \sim p(X_k^t | X_1^t, X_2^t, \dots, X_{k-1}^t, X_{k+1}^t, \dots, X_K^t)$  où  $K$  est le nombre de composantes de  $\mathbf{X}$ . Le calcul de  $p(X_k^t | X_1^t, X_2^t, \dots, X_{k-1}^t, X_{k+1}^t, \dots, X_K^t)$  peut se faire très rapidement dans le cadre des modèles markoviens :

$$p(X_k^t | X_1^t, X_2^t, \dots, X_{k-1}^t, X_{k+1}^t, \dots, X_K^t) \propto \prod_{C \in \mathcal{C}_k} \Psi_C(\mathbf{X}_C^t) \quad (3.34)$$

$$\text{avec } \sum_{i=1}^N p(X_k^t = q_i | X_1^t, X_2^t, \dots, X_{k-1}^t, X_{k+1}^t, \dots, X_K^t) = 1$$

où  $\mathcal{C}_k$  représente l'ensemble des cliques maximales qui contiennent le site  $k$ .

**3.3.3.2.2 Les méthodes variationnelles** Les méthodes variationnelles sont des méthodes d'approximation qui transforment un problème convexe ou non en un problème plus simple dans le sens où les degrés de liberté du problème sont découplés. Le découplage est réalisé par l'introduction de nouveaux paramètres, les paramètres variationnels.

Dans le cadre des modèles graphiques, on introduit une famille de distributions  $\mathbf{R} = \{q(\mathbf{X}|\boldsymbol{\lambda})\}$  paramétrées par  $\boldsymbol{\lambda}$  (les paramètres variationnels). Le graphe d'inférence représentant  $\mathbf{R}$  est en général obtenu en supprimant des arêtes dans le graphe du processus d'intérêt. La distribution d'intérêt va être approchée par une distribution particulière  $q(\cdot)$  de  $\mathbf{R}$ .  $q(\cdot)$  est choisie de manière à minimiser la divergence de Kullback-Leibler  $D(q||p)$  :

$$\lambda^* = \operatorname{argmin}_{\boldsymbol{\lambda}} \sum_{\{\mathbf{x}\}} q(\mathbf{X}|\boldsymbol{\lambda}) \ln \frac{q(\mathbf{X}|\boldsymbol{\lambda})}{p(\mathbf{X})} \quad (3.35)$$

Ce critère fournit la meilleure borne inférieure de la vraisemblance dans la famille d'approximation  $\mathbf{R}$  et il se justifie également au travers de la théorie de la convexité duale [118]. Une approximation connue sous le nom d'approximation du champ moyen consiste à supprimer tous les liens entre les variables aléatoires. On a alors :  $q(\mathbf{X}|\boldsymbol{\lambda}) = \prod_i q_i(X_i)$ . Pour plus de détails, on pourra se référer à [118].

### 3.3.4 Problèmes d'inférence : apprentissage

L'apprentissage consiste à estimer le vecteur paramètre du modèle  $\lambda$  sur la base d'un ensemble de séquences d'observations appelé corpus d'apprentissage.

#### 3.3.4.1 L'algorithme EM

L'algorithme EM (Expectation Maximization) est un algorithme itératif utilisé pour le calcul des paramètres au sens du MV quand les observations sont ou peuvent être vues comme étant des données incomplètes [51]. L'algorithme de Baum-Welch présenté dans le cadre du modèle markovien caché s'avère être une instance particulière de l'algorithme EM. L'algorithme EM procède de façon itérative jusqu'à convergence :

- *étape E (Expectation)* : on commence par calculer la fonction auxiliaire  $Q(\lambda^{p-1}, \lambda)$ ,  $\lambda^{p-1}$  provenant de l'itération précédente :

$$\begin{aligned} Q(\lambda^{p-1}, \lambda) &\triangleq E[\log P[\mathbf{X}, \mathbf{Y}|\lambda]|\mathbf{Y}, \lambda^{p-1}] \\ &= \int_{\mathbf{X}} f(\mathbf{X}|\mathbf{Y}, \lambda^{p-1}) \log P[\mathbf{X}, \mathbf{Y}|\lambda] d\mathbf{X} \end{aligned} \quad (3.36)$$

- *étape M (Maximization)* : maximisation de la fonction auxiliaire

$$\lambda^p = \arg \max_{\lambda} Q(\lambda^{p-1}, \lambda) \quad (3.37)$$

L'algorithme EM possède des inconvénients majeurs : l'estimée finale dépend fortement de l'initialisation et l'algorithme peut se piéger dans un maximum local ou un point selle. Afin de tenter de résoudre ces problèmes, de nouvelles méthodes d'apprentissage, dont certaines sont des variantes de l'algorithme EM, ont été proposées. Nous les présentons maintenant.

#### 3.3.4.2 Autres méthodes d'apprentissages

**3.3.4.2.1 Variantes de l'EM** Une première famille de variantes est celle des algorithmes EM stochastiques. Elle comprend le SEM (stochastic EM) [36][155] et le MCEM (Monte-Carlo EM) [217]. Leur point commun est d'introduire une étape stochastique pour le calcul de la fonction auxiliaire.

Le SEM simule uniquement un échantillon  $\mathbf{X}$  à partir de  $f(\mathbf{X}|\mathbf{Y}, \lambda^{p-1})$  et remet à jour les paramètres à partir des données observées et manquantes simulées. La phase stochastique du SEM permet d'éviter les maxima locaux et les points instables (type point selle) de l'EM. Néanmoins, il n'est pas assuré de converger. C'est pour cette raison que le SAEM (Stochastic Approximation EM) [37] a été conçu. Cet algorithme passe progressivement d'un comportement de type SEM à un comportement de type EM, garantissant ainsi la convergence.

Le MCEM approche la fonction auxiliaire en générant un ensemble de  $M$  échantillons  $\mathbf{X}^1, \dots, \mathbf{X}^M$  selon la loi  $f(\mathbf{X}|\mathbf{Y}, \lambda^{p-1})$ . Si  $M = 1$ , on retrouve le SEM alors que pour  $M$  très grand, on s'approche de l'algorithme EM standard (loi des grands nombres). Une stratégie possible consiste à faire varier  $M$  au cours des différentes itérations [217]. En partant avec  $M = 1$  et en augmentant ce dernier progressivement, l'algorithme adopte au départ un comportement stochastique et devient progressivement déterministe.

Il existe encore une variante de l'algorithme EM appelée GEM (Generalized EM). L'étape de maximisation est simplifiée. Elle consiste à chercher  $\lambda^{p+1}$  tel que  $Q(\lambda^p, \lambda^{p+1}) \geq Q(\lambda^p, \lambda^p)$ . Cette variante est assurée également de converger [50].

**3.3.4.2.2 L'algorithme ICE** L'algorithme ICE (Iterative Conditional Estimation) est une méthode très générale d'estimation des paramètres pour des données partiellement observées [177]. Elle s'appuie sur l'hypothèse qu'il existe au moins un estimateur  $\hat{\lambda}(\mathbf{X}, \mathbf{O})$  défini sur les données complètes  $(\mathbf{X}, \mathbf{O})$ . L'objectif de l'ICE est alors de l'approcher au mieux par une fonction des seules données observées. La démarche de la procédure est encore une fois itérative :  $\lambda^{p+1} = E(\hat{\lambda}(\mathbf{X}, \mathbf{O})|\mathbf{O}, \lambda^p)$ . Pour comprendre les liens qui existent entre l'ICE et l'EM, on pourra se référer à [48] et [49].

En pratique,  $E(\hat{\lambda}(\mathbf{X}, \mathbf{O})|\mathbf{O}, \lambda^p) = \sum_{\mathbf{X}} \hat{\lambda}(\mathbf{X}, \mathbf{O})P[\mathbf{X}|\mathbf{O}, \lambda^p]$  permet de minimiser l'erreur quadratique sur les paramètres du modèle. Ainsi, la paramétrisation du modèle a une influence directe sur le résultat. De plus, les paramètres du modèle sont une somme pondérée de paramètres obtenus à partir de différents jeux de données (états cachés et observations). Ce choix ne semble pas forcément être judicieux pour tous les estimateurs, notamment pour des estimateurs robustes. Ces deux problèmes ne semblent pas être évoqués dans la littérature.

**3.3.4.2.3 Segmentation courante** Ces méthodes consistent à résoudre simultanément les problèmes d'apprentissage et de segmentation (décodage des états cachés). Elles fonctionnent de manière itérative. Tout d'abord, à partir du modèle  $\lambda^p$ , une méthode de segmentation fournit une carte de segmentation  $\mathbf{X}^p$ . Un modèle  $\lambda^{p+1}$  est ensuite estimé à partir des données complètes  $(\mathbf{X}^p, \mathbf{O})$ . Un exemple de segmentation courante a été présenté lors de la résolution des problèmes d'apprentissage dans le cas des chaînes de Markov cachées : l'algorithme d'apprentissage de Viterbi (sous-section 3.1.3.3.2 de la page 47). La segmentation se fait au sens du MAP et l'estimation des paramètres au sens du MV.

## 3.4 Fusion markovienne de processus en interaction

### 3.4.1 Introduction

Dans le cadre du développement d'une méthode de cartographie robuste, nous sommes amenés dans la troisième partie à considérer les signaux IRMf dans le voisinage d'un voxel  $v$  pour estimer l'activité au voxel  $v$ . De par la multiplicité des séquences à traiter, ce problème relève ainsi de la fusion de données. Différents signaux IRMf apportent en effet des informations complémentaires et/ou redondantes sur le processus d'activation. Le cadre formel retenu pour développer les méthodes est de type probabiliste markovien. Pour apporter une solution à notre problème, différentes solutions peuvent être envisagées, les deux solutions qui viennent immédiatement à l'esprit étant les suivantes :

- utiliser une chaîne de Markov par signal IRMf, chaque chaîne de Markov pouvant coopérer avec ses voisines. On arrive ainsi à la notion de processus qui interagissent les uns avec les autres.
- utiliser une unique chaîne de Markov et complexifier le processus observable de manière à prendre en compte l'information du voisinage.

De nombreux systèmes sont caractérisés par des processus qui interagissent les uns avec les autres. Par exemple, dans le domaine de la reconnaissance audiovisuelle de la parole, un premier processus est de nature visuelle (forme et position des lèvres) alors que le second est de nature auditive (signal acoustique). La lecture sur les lèvres permet alors, surtout dans des environnements bruités, d'améliorer les performances de reconnaissance. On peut citer également l'exemple de la reconnaissance d'actions complexes comme la langue des signes. Un premier processus correspond ici à la position de la main gauche alors que le second correspond

à celle de la main droite.

D'une manière générale, le cadre de fusion, dans ce contexte, peut être caractérisé entre autres par le nombre de processus cachés et par le nombre de processus observables. Il est hors de question ici de faire un état de l'art sur les différentes méthodes de fusion. Nous nous concentrons uniquement sur les approches où la fusion de donnée se fait au travers d'un modèle markovien modélisant les processus.

### 3.4.2 Cadre formel

Nous nous intéressons ici aux deux cas de figure les plus souvent rencontrés dans la littérature.

Le premier cas de figure consiste à modéliser  $N$  processus aléatoires en interaction superposant leurs observations dans un même et unique canal. Formellement,  $N$  séquences d'états cachés  $\{\mathbf{X}^i\}_{1 \leq i \leq N}$  sont considérées pour la partie cachée alors que le processus observable se résume en une seule et unique séquence d'observations  $\mathbf{O}$ . Généralement, les  $N + 1$  séquences sont échantillonnées aux mêmes instants discrets  $t \in [1, T]$ . Ainsi, pour chaque instant  $t$ , on dispose de  $N$  états cachés  $X_t^1, X_t^2, \dots, X_t^N$  et d'une unique observation  $O_t$ . On peut ainsi résoudre le problème de séparation de source où des signaux indépendants sont sommés à l'intérieur d'un même canal d'observation. Ce modèle peut également être vu comme un modèle multi-couches caché. Les différentes couches permettent de prendre en compte la complexité du processus caché sans pour autant augmenter de façon importante l'espace d'état. Une couche peut par exemple apporter une information contextuelle de haut niveau aux couches inférieures. Il peut sembler que nous nous éloignons quelque peu de notre application avec ce premier cas dans la mesure où nous avons ici une unique séquence d'observations. Néanmoins, il est également intéressant de voir comment les différentes chaînes coopèrent, des observations pouvant par la suite être ajoutées.

Le deuxième cas de figure consiste à modéliser  $N$  processus aléatoires en interaction observés au travers de  $N$  canaux d'observation distincts, chaque processus possédant son propre canal d'observation. Formellement, on dispose alors de  $N$  séquences d'observations  $\{\mathbf{O}^i\}_{1 \leq i \leq N}$  apportant des informations redondantes et/ou complémentaires sur une même réalité.

### 3.4.3 Approches de modélisation

D'un point de vue théorique, un modèle markovien caché standard peut permettre de modéliser l'ensemble des cas de figure mentionnés ci-dessus. Il suffit pour ce faire de prendre comme espace d'état le produit cartésien des espaces d'états des  $N$  processus cachés considérés. Ainsi, si  $N_i$  est le nombre d'états cachés caractérisant le processus  $i$ , le modèle standard devrait être finalement formé de  $\prod_{i=1}^N N_i$  états pour prendre en compte les éventuelles interactions entre processus. Ce produit cartésien est cependant rarement une solution satisfaisante pour plusieurs raisons : (i) le coût calculatoire devient très important, (ii) le nombre de paramètres est si important que des risques de sur-représentation du modèle ou des données sont à craindre, (iii) les données nécessaires à l'apprentissage sont peu nombreuses au regard du nombre de paramètres à estimer. Une solution consiste alors à fusionner certains états, l'interaction entre les différents processus n'étant plus clairement explicitée. Il peut par ailleurs apparaître des problèmes de sous-représentation du modèle si des états pourtant bien distincts sont fusionnés. Par conséquent, il devient souvent préférable d'établir explicitement des liens de dépendance.

Les modèles mentionnés ci-après sont souvent apparus dans la littérature anglo-saxonne. Leur appellation ne possédant pas pour la plupart une traduction immédiate et explicite en



français, nous utilisons indifféremment l'appellation anglo-saxonne ou l'appellation française pour désigner certains d'entre eux.

### 3.4.3.1 Ensemble de processus en interaction avec un unique canal d'observation

Le modèle de la figure 3.11.a est un modèle markovien caché factoriel (FHMM : factorial Hidden Markov Model). Ce modèle peut permettre de résoudre le problème de séparation de sources. Cependant, comme on l'a déjà vu, ce modèle peut être perçu comme un modèle markovien multi-couches. Il a par exemple été utilisé par [30] dans le cadre de la reconnaissance de la parole avec deux couches, chaque couche étant en fait représentative d'une partie des données observées.

Le second modèle (figure 3.11.b) est un modèle markovien caché à arbre de décision (Hidden Markov Decision Tree) où la couche d'entrée n'est pas représentée. Pour ce modèle, des liens de dépendance à travers les états cachés sont explicitement rajoutés. Les processus cachés sont dépendants et interagissent entre eux. Ce modèle a notamment été utilisé par [70] avec deux couches pour modéliser la parole d'un conducteur sous tension. Il est à noter que les résultats obtenus par [30] et par [70] sont moins bons que ceux obtenus par une approche classique (modèle markovien standard par exemple). Les modèles sont ainsi mal adaptés à l'application. Par conséquent, le choix du modèle s'avère très délicat. Lorsque le choix du modèle est guidé par une connaissance *a priori* forte, le modèle a cependant plus de chance d'être bien adapté à son application.

Par exemple, le modèle de Markov caché à deux niveaux centiseconde introduit par Meziane [163] a une justification claire : une couche représente un niveau acoustique alors que le second représente le niveau phonétique. Ce modèle permet en définitive de tenir compte de la prosodie de la parole. On voit également apparaître avec ce modèle une notion de hiérarchie.

Le modèle défini par [71] (modèle markovien caché hiérarchique ou hierarchical HMM en anglais) permet de modéliser des structures ou des dépendances avec différentes échelles de longueur ou de temps. En pratique, les états du HHMM génèrent soit une observation soit un ensemble d'observations. Plus précisément, la génération de plusieurs observations par un même état correspond à l'appel par cet état d'un sous-HHMM, les différents appels pouvant être récursifs. Une première application de ce modèle est donnée en reconnaissance automatique de la parole [124]. Au niveau le plus élevé se trouve la notion de mots, cette notion faisant ensuite appel à un niveau inférieur représentant les phonèmes, les phonèmes faisant enfin appel à un niveau inférieur représentant les sous-phonèmes. Il est à noter que cette modélisation permet de prendre facilement en compte les différentes façons de prononcer un même mot.

### 3.4.3.2 Ensemble de processus en interaction avec un canal d'observation par processus

Les approches de modélisation proposées dans ce cadre peuvent être classées selon deux critères : la (dé)synchronisation relative des processus en interaction et la nature des interactions existant entre les processus.

La figure 3.12.a représente un LHMM (Linked Hidden Markov Model). Ce modèle est utilisé pour exprimer des interactions de nature synchrone symétrique. Ce type d'interaction peut être illustré dans le cas de deux joueurs de tennis dont la position de chacun représente un processus à modéliser. En pareille situation, il est improbable de voir les deux joueurs au filet en même temps.

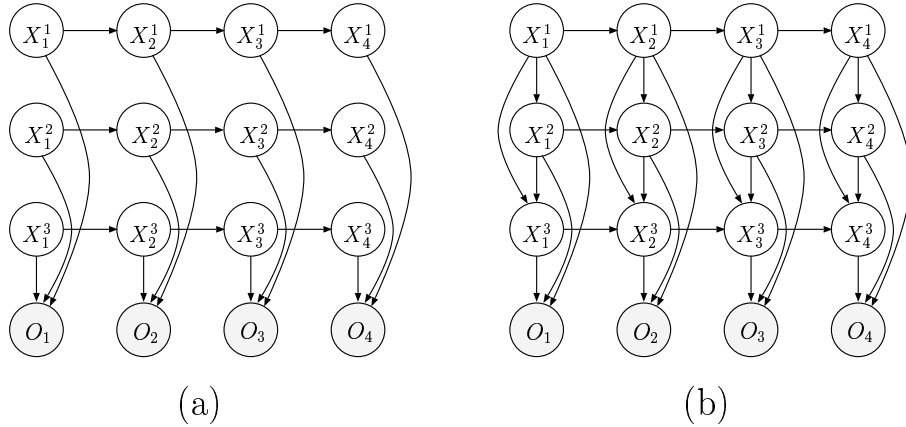


FIG. 3.11 – Graphe de dépendance d'un (a) FHMM (Factorial Hidden Markov Model) et d'un (b) HMDT (Hidden Markov Decision Tree)

Le modèle parallèle (ou maître-esclave) représenté sur la figure 3.12.c permet de modéliser un processus (le processus maître  $\mathbf{X}^1$ ) qui conditionne un autre processus de manière synchrone. Ce modèle a été utilisé dans [7] pour fusionner les données acoustiques et articulatoires. Le processus maître correspondant à la configuration des lèvres (lèvres ouvertes, fermées ou semi-ouvertes) détermine le processus esclave qui définit les unités acoustiques élémentaires (pseudo-diphones) [7].

Dans le cas où les différents processus interagissent de façon non synchrone, des arêtes doivent permettre de faire le lien entre des variables aléatoires d'un canal différent n'apparaissant pas au même instant (figure 3.12.b). Le modèle de la figure 3.12.b (Coupled hidden Markov model, CHMM) permet de modéliser des processus interagissant de façon asynchrone et éventuellement de façon causale. Par exemple, la montée d'un joueur de tennis au filet a tendance à faire reculer l'adversaire au fond du cours. Les deux modèles (LHMM et CHMM) ont été utilisés dans [28] pour reconnaître les mouvements de bras d'un homme pratiquant un art martial chinois (chaque position d'un bras est représentée par un processus). Le CHMM a montré de très bonnes performances contrairement au LHMM, démontrant ainsi la nécessité de prendre en compte les asynchronismes dans le cadre de cette application.

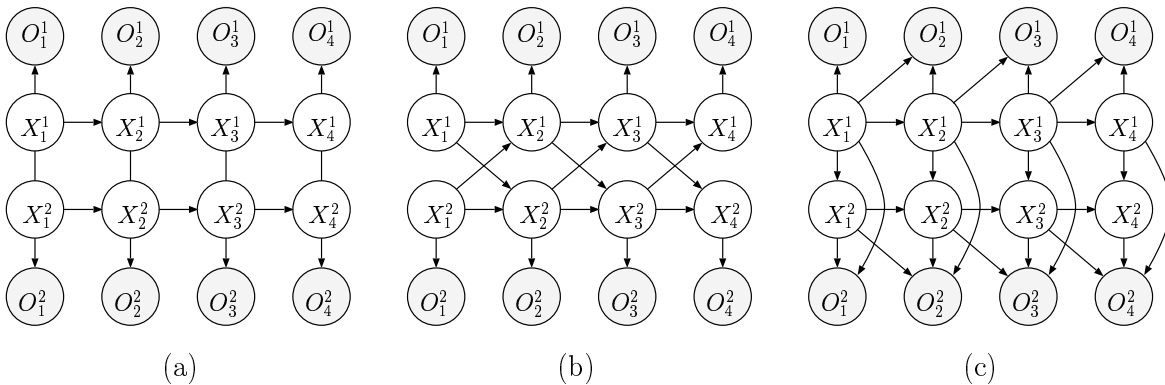


FIG. 3.12 – Graphe de dépendance d'un (a) LHMM (Linked Hidden Markov Model), d'un (b) Coupled Hidden Markov Model (CHMM) et (c) du modèle maître-esclave ou modèle parallèle.

Certains modèles permettent d'analyser les interactions à un niveau supérieur. Le modèle présenté figure 3.13.a s'appelle un mélange de MMCs [199]. Pour chaque valeur de la variable cachée  $X$ , on obtient un MMC particulier. Cette variable cachée permet de définir une variable contextuelle ou explicative de plus haut niveau. Ce modèle aurait pu être présenté dans la section précédente (unicité de la séquence d'observations). Cependant, si nous présentons ce modèle, c'est surtout pour présenter sa généralisation dont le principe consiste à rajouter une dimension : le temps par exemple (figure 3.13.b). Le modèle résultant porte plusieurs noms dans la littérature : hierarchical mixture of Markov chains [112], embedded HMM [166] pseudo-2D HMM [131]. Ces modèles sont par exemple utilisés pour représenter des images, dans le domaine de la reconnaissance de visage par exemple [166]. Conformément à la figure 3.13.b, les variables aléatoires  $X_1^i, X_2^i, X_3^i$  et  $X_4^i$  représentent les états cachés associés à chaque pixel de la  $i$ -ème ligne et la variable cachée  $X^i$  représente l'état caché associé à la  $i$ -ème ligne. Les dépendances dans la direction verticale sont ensuite assurées par les dépendances entre les états cachés  $X^i$ .

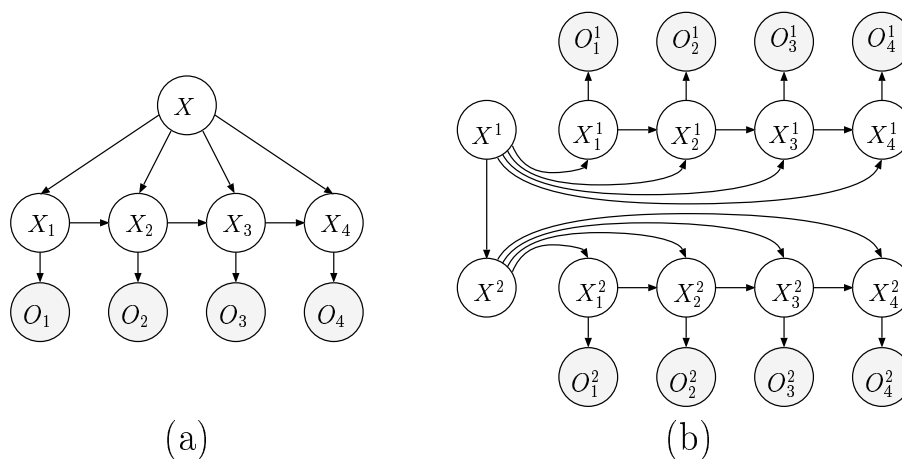


FIG. 3.13 – Graphe de dépendance d'un (a) mélange de MMC et (b) d'un mélange hiérarchique de MMC.



# Conclusion partielle

Cette première partie a permis de délimiter le cadre de travail. La présentation de l'IRMf a mis en évidence la complexité de la technique d'observation et de sa cible, le cerveau.

Les méthodes actuelles de cartographie IRMf se répartissent en deux grandes classes : les méthodes inférentielles et les méthodes exploratoires. Les méthodes inférentielles nécessitent souvent de faire des hypothèses restrictives sur le signal actif. Ceci est une limitation importante quand on pense à la forte variabilité du signal actif en termes de forme de la réponse hémodynamique et de chronologie d'activation cérébrale. Au contraire, les méthodes exploratoires ne font pas d'hypothèses *a priori* sur le signal actif. En revanche, l'interprétation des résultats reste difficile. Les méthodes de cartographie en IRMf sont finalement très nombreuses et les problèmes de l'évaluation et de la comparaison restent délicats.

Le travail présenté dans cette thèse vise à développer et à valider de nouvelles méthodes de cartographie IRMf cérébrale qui prennent en compte la grande variabilité du signal IRMf actif. De manière à ce que les résultats restent facilement interprétables, les approches privilégiées sont de nature inférentielle, et plus précisément de type modèles de Markov cachés. Nous avons donc présenté le formalisme markovien caché ou semi-markovien caché dans lequel s'inscrit l'approche de cartographie proposée en partie II. Nous avons également présenté des modèles markoviens permettant de fusionner des informations provenant de processus en interaction. Ces concepts sont abordés dans la partie III.

En définitive, les fondements théoriques nécessaires à la présentation des approches de modélisation proposées ont été présentés.



## Deuxième partie

**Le modèle semi-markovien caché de  
séquence d'événements : application  
en cartographie IRMf cérébrale**





# Introduction

Les méthodes actuelles de cartographie IRMf se répartissent en deux grandes classes : les méthodes inférentielles et les méthodes exploratoires. Les méthodes inférentielles nécessitent de faire des hypothèses restrictives sur la forme et sur la chronologie du signal IRMf actif. Les méthodes exploratoires, quant à elles, fournissent des résultats dont l'interprétation reste délicate.

Le travail présenté dans cette partie vise à développer et à valider une nouvelle méthode de cartographie IRMf cérébrale prenant en compte la grande variabilité du signal IRMf actif. Il convient notamment de réduire le nombre d'hypothèses *a priori* sur la forme de la réponse hémodynamique et sur la chronologie d'activation du signal actif. En pratique, un modèle statistique potentiellement représentatif d'une très grande variété de signaux actifs est appris à partir des données grâce aux capacités d'apprentissage des modèles markoviens cachés. L'approche étant largement guidée par les données, elle est également non supervisée. Les résultats sont par ailleurs faciles à interpréter dans la mesure où nous restons dans un cadre inférentiel.

Dans un premier temps, nous abordons le principe de modélisation. Nous présentons ensuite de façon détaillée le modèle markovien utilisé : le modèle semi-markovien caché de séquence d'événements (MSMCSE). Nous voyons ensuite comment cette représentation peut s'appliquer à la cartographie en IRMf cérébrale. Nous présentons et nous discutons enfin les résultats obtenus.



# Chapitre 4

## Principes et méthodes

L'objectif est de développer une méthode de cartographie en IRMf cérébrale qui prenne en compte la variabilité des signaux actifs, cette variabilité pouvant par exemple être liée à la forme de la réponse hémodynamique, à une éventuelle désynchronisation de la réponse par rapport à la stimulation, ou à une activation non maintenue tout au long du protocole. Dans ce cadre, l'approche que nous privilégions est de nature inférentielle, et plus précisément à base de modèles de Markov cachés. Rarement utilisée dans le contexte de la détection d'activation cérébrale, la modélisation markovienne cachée apparaît *a priori* bien adaptée à l'analyse du signal IRMf. Il s'agit en effet d'analyser un processus *caché*, le processus d'activation neuronale, au travers d'un processus aléatoire *observable* qui lui est corrélé : le signal IRMf cérébral. Nous constatons toutefois que les modèles de Markov cachés classiques sont surtout adaptés à la représentation de processus aléatoires de nature segmentale (c'est-à-dire stationnaires par morceaux). La nature événementielle et non-stationnaire des séquences IRMf nous amènera donc à considérer des modèles non-standards. Après un bref état de l'art sur les représentations envisagées dans la littérature, nous présentons notre propre approche originale : le modèle semi-markovien caché de séquences d'événements et la problématique sous-jacente.

### 4.1 Modélisation markovienne et signal IRMf

De par l'hypothèse de stationnarité par morceaux du processus observable sous-jacent aux modèles de Markov cachés (MMCs), ces modèles s'avèrent particulièrement adaptés à l'analyse de processus aléatoires de nature segmentale. Cependant, nombre de processus, rencontrés notamment en génie biomédical se présentent sous la forme d'une séquence d'événements localisés dans le temps, à caractère le plus souvent non stationnaire, et noyés dans du bruit ou toute autre activité de fond : ondes ECG (P, QRS, T) en électrocardiographie, pointes-ondes en électroencéphalographie, ruptures de dynamique en IRMf. Ainsi, les performances d'analyse de processus de nature événementielle par MMCs standards s'en trouvent dégradées. L'obtention de segments d'observations à la fois stationnaires sur le support temporel de ces événements et discriminants entre classes d'événements se révèle difficile, voire impossible, parce qu'incompatible avec la nature événementielle de ces processus.

Pour relâcher l'hypothèse de stationnarité par morceaux du processus observable, différentes approches ont été proposées dans la littérature. Nous ne faisons pas ici une étude exhaustive de ces approches mais nous citons les plus connues. Nous présentons notamment le modèle markovien caché auto-régressif, le modèle markovien caché segmental et un modèle présenté dans le cadre de la modélisation du signal IRMf.

### 4.1.1 Modèle markovien caché auto-régressif

Contrairement au modèle standard dont le graphe de dépendance est rappelé figure 4.1.b, le modèle markovien caché auto-régressif introduit explicitement des liens de dépendance entre les observations. Plus précisément, l'observation  $O_t$  est indépendante des observations et des états cachés conditionnellement aux observations  $O_{t-1}, O_{t+1}$  et aux états cachés  $X_t, X_{t+1}$ . Le graphe de dépendance d'un modèle markovien caché auto-régressif est représenté figure 4.1.a.

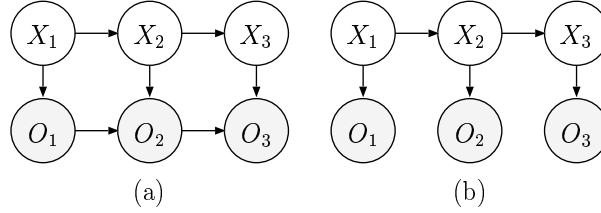


FIG. 4.1 – Graphe de dépendance d'un MMC auto-régressif (a) et d'un MMC (b).

Une relation de dépendance linéaire gaussienne [126] [218] [107] est souvent utilisée. Elle s'exprime de la façon suivante :

$$P[O_t | X_t = q_i, O_{t-1}] = \mathcal{N}(O_t; W_i O_{t-1} + \mu_i, \Sigma_i) \quad (4.1)$$

où  $W_i$  est la matrice de régression,  $X_t$  étant dans l'état  $q_i$  et où  $\mathcal{N}(O; \mu, \Sigma)$  désigne la vraisemblance de l'observation  $O$  sous l'hypothèse que le processus aléatoire associé à l'observation suit une loi normale de moyenne  $\mu$  et de matrice de variance-covariance  $\Sigma$ .

Ce modèle porte plusieurs noms dans la littérature (correlation HMM, conditionally Gaussian HMM, switching regression model, switching Markov model). Bilmes [24] généralise les modèles auto-régressifs en introduisant de nouvelles relations de dépendance entre les différentes observations. Les relations de dépendance peuvent varier avec le temps, ne sont pas obligatoirement linéaires et sont apprises automatiquement grâce à une approche discriminante.

Il faut noter que le terme « auto-régressif » fait référence à deux modèles différents. Le second est celui discuté dans [180][121][186]. Son graphe de dépendance est celui du modèle standard (figure 4.1.b). Cependant, la loi d'observation est construite sous l'hypothèse que la restriction du processus observable à un état caché est un processus auto-régressif. Ainsi, ce modèle permet également de relâcher l'hypothèse de stationnarité par morceaux du processus observable.

### 4.1.2 Modèle markovien caché segmental

Le principe consiste à découper le signal en un nombre  $\tau$  de segments. Le segment numéro  $i$ ,  $\mathcal{S}_i$ , a une longueur  $l_i$ . Il est associé à un unique état caché, l'état  $X_i$ . Ainsi,  $l_i$  observations sont générées pour un même état. A un état caché, est donc associée une loi d'observation plus complexe que dans le cas standard. Elle stipule le nombre d'échantillons générés et leur valeur. Le graphe de dépendance de ce modèle est présenté figure 4.2. On a :

$$P[\tau, \mathbf{X}_1^\tau, \mathbf{I}_1^\tau, \mathbf{O}_1^\tau] = P[\tau] P[X_1 | \tau] \prod_{i=2}^{\tau} P[X_i | X_{i-1}, \tau] \prod_{i=1}^{\tau} \left( P[l_i | X_i] P[\mathbf{O}_{\mathcal{S}_i} | l_i, X_i] \right) \quad (4.2)$$

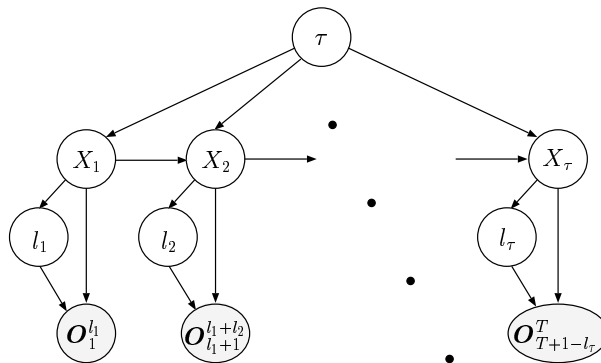


FIG. 4.2 – Graphe de dépendance d’un modèle markovien caché segmental. Le graphe de dépendance représenté est en fait simplifié car il ne prend pas en compte toutes les relations de dépendance. En effet, on a :  $\sum_{i=1}^{\tau} l_{\tau} = T$ .

où  $\mathcal{O}_{\{\mathcal{S}_i\}}$  est la restriction des observations au  $i$ -ème segment d’observation. La définition de la loi  $P[\mathcal{O}_{\{\mathcal{S}_i\}} | l_i, X_i]$  permet de modéliser des comportements quelconques sur le segment d’observation (voir [98] pour un exemple). On remarque enfin que les modèles segmentaux sont une généralisation des modèles semi-markoviens. On pourra se référer à [194][93][98][173] pour plus de détails.

### 4.1.3 Modèle proposé dans le cadre de l’IRMf

Nous avons déjà cité les travaux de Højen-Sørensen [111] dans le cadre des méthodes de cartographie en IRMf. Pour prendre en compte les propriétés du signal IRMf, Højen-Sørensen propose de complexifier la relation entre le processus caché et le processus observable. Une variable caché  $X_t$  binaire est associée à chaque échantillon du signal. La variable  $X_t$  vaut 1 si la zone corticale associée au voxel  $v$  est active au temps  $t$ , 0 sinon. Le processus observable est ensuite lié au processus caché de la façon suivante : l’observation est définie comme étant la convolution du processus caché par la réponse hémodynamique auquel on ajoute des termes modélisant la dérive de ligne de base, une constante et un terme d’erreur. Ainsi, ce modèle est relativement adapté pour résoudre des problèmes de déconvolution.

## 4.2 Solution proposée

On peut légitimement penser que le modèle markovien segmental et le modèle proposé par Højen-Sørensen permettent de modéliser convenablement le signal IRMf. Néanmoins, la modélisation reste très proche du signal et il est à craindre qu’il soit difficile avec de tels modèles de prendre en compte une grande variété de signaux actifs en termes de chronologie d’activation et de forme de la réponse. Le modèle proposé par Højen-Sørensen se place notamment dans le cadre du modèle linéaire et ne prend pas en compte les non linéarités de la réponse. Ces approches ne semblent donc pas des solutions complètement satisfaisantes pour atteindre notre objectif.

### 4.2.1 Un nouveau modèle markovien

Pour réduire le nombre d'hypothèses *a priori* et faire ainsi de sorte que le modèle soit potentiellement représentatif d'un espace de signaux actifs plus important, nous proposons d'analyser le signal IRMf uniquement à travers ses ruptures de dynamique (RDs). Les RDs peuvent correspondre à une fausse alarme du système de détection, à un début de réponse hémodynamique (DRH) d'une tâche de non-intérêt mais également à un DRH de la tâche d'intérêt. Analyser le signal IRMf à travers ses non stationnarités, à savoir les RDs, permet tout d'abord de limiter le nombre d'hypothèses sur la forme de la réponse à un bloc de stimulation ou à un stimulus. Ensuite, si le signal tout entier est difficilement modélisable par un MMC, une modélisation statistique par chaîne de Markov cachée de la distribution temporelle des événements supposés d'intérêt reste toujours pertinente.

En définitive, nous proposons d'adapter le formalisme markovien caché en le plaçant d'emblée dans un cadre de détection et d'analyse de séquences d'événements. En pratique, une étape de prétraitement détecte et caractérise les événements supposés d'intérêt composant le processus événementiel à analyser. Le modèle markovien se limite ensuite à l'analyse de cette séquence d'événements. La redéfinition nécessaire du contenu de la séquence d'observations conduit à développer un nouveau type de MMC pour l'analyser : le modèle semi-markovien caché de séquence d'événements (MSMCSE). Conçu sur la base d'un modèle semi-markovien caché (MSMC), le MSMCSE se révèle mieux adapté à l'analyse de processus aléatoires de nature événementielle plutôt que segmentale. Nous présentons ce modèle en détail dans le chapitre suivant.

### 4.2.2 Principe d'alignement

La méthode de cartographie repose sur l'analyse, par un modèle markovien, des RDs détectées dans le signal IRMf. Le MSMCSE modélise entre autres la distribution temporelle des événements détectés. Cependant, la répartition de ces événements sur l'axe des temps n'est pas quelconque et elle est très fortement contrainte par le paradigme d'activation. En cas d'activité pour un paradigme en blocs, on s'attend à ce que les RDs détectées dans le signal IRMf actif s'alignent, dans une certaine mesure, sur les transitions « off-on » du paradigme. Ce couplage entre les RDs détectées dans le signal IRMf et les transitions « off-on » du paradigme est illustré sur la figure 4.3. Pour un paradigme événementiel, les RDs doivent s'aligner dans une certaine mesure sur les stimuli eux-mêmes.

L'analyse du signal IRMf à travers le couplage entre les RDs détectées dans le signal IRMf et les transitions « off-on » du paradigme a tout d'abord été mis en œuvre dans [207] : le taux d'activité cérébrale est évalué à partir du degré d'alignement temporel entre la séquence de RDs détectées à partir du signal IRMf et les transitions associées au paradigme d'activité. Le degré d'alignement est estimé à travers le coût nécessaire pour transformer une séquence d'événements en une autre (alignement par programmation dynamique). Cette approche a permis de montrer que l'information contenue dans le signal IRMf est en grande partie contenue dans ses non-stationnarités. Cependant, l'approche présentée dans [207] souffre de l'introduction dans l'algorithme de différentes fonctions de coût heuristiques. Aussi, la reformulation de ce couplage dans le cadre statistique des MSMCSEs [66][67][68] est d'un grand intérêt. La modélisation du couplage soulève toutefois un nombre important de problèmes. Ces derniers sont présentés dans la section suivante.

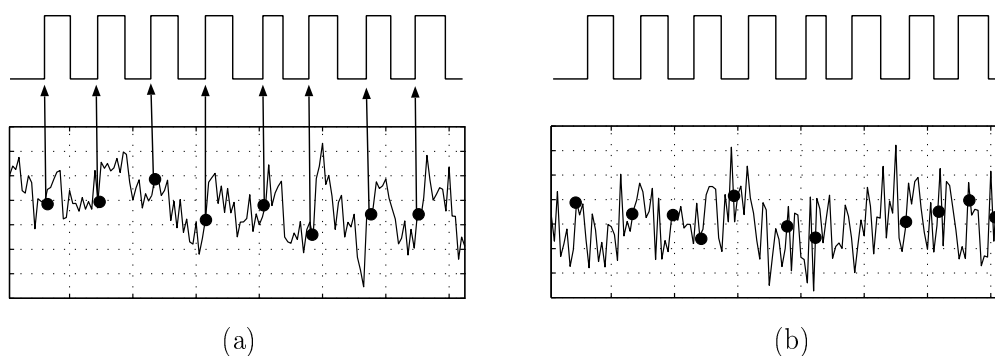


FIG. 4.3 – Principe d’alignement. (a) les RDs (rond noir) détectées dans le signal actif s’alignent relativement bien avec les transitions « off-on » du paradigme. (b) les RDs détectées dans un signal inactif ne s’alignent pas avec les transitions « off-on » du paradigme.

## 4.3 Problématiques

### 4.3.1 Que faut-il modéliser ?

#### 4.3.1.1 Le modèle biologique

Pour un signal IRMf, nous faisons l’hypothèse tout à fait justifiée d’un point de vue biologique qu’une transition du paradigme donne naissance à un DRH au plus. Supposons qu’un observateur idéal transforme (sans aucune fausse alarme, sans aucune non détection et sans aucune erreur sur l’instant de détection) le signal IRMf en une séquence de DRH. Faire de la cartographie à partir de la séquence de DRHs reste un problème non trivial. Trois problèmes illustrés sur la figure 4.4 apparaissent en effet :

- Certaines transitions du paradigme peuvent ne pas donner naissance à un DRH. En effet, le patient ne répond pas obligatoirement à tous les stimuli du paradigme.
- Une désynchronisation peut exister entre la séquence de DRH et la séquence de transitions du paradigme. En effet, il peut y avoir des retards d’activation, voire des anticipations.
- Tous les DRHs ne peuvent pas être associés avec une transition du paradigme dans la mesure où tous les DRHs détectées ne correspondent pas obligatoirement à la tâche d’intérêt.

En résumé, il est nécessaire d’avoir un modèle biologique permettant de caractériser l’activité cérébrale que l’on cherche à détecter ou un modèle biologique permettant de discriminer l’activité cérébrale recherchée de toute autre activité de fond.

#### 4.3.1.2 Le modèle du détecteur

Dans notre démarche, l’observateur idéal est remplacé par une étape de prétraitement des données qui fournit une séquence de RDs pour chaque signal IRMf. Les RDs sont détectées dans le signal IRMf au moyen d’une analyse multirésolution par ondelettes [4]. On fait l’hypothèse relativement naturelle qu’un DRH peut être associé au maximum à une RD et qu’une RD peut être associée au maximum à un DRH. A cause de l’étape de détection, on peut citer trois nouveaux problèmes illustrés également sur la figure 4.4 :

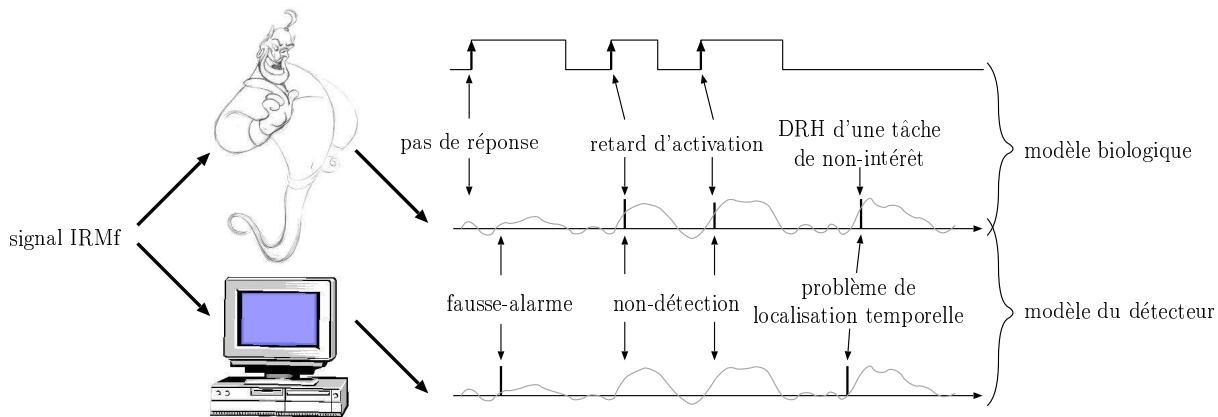


FIG. 4.4 – Modèle biologique et modèle du détecteur. On suppose tout d’abord qu’un observateur idéal transforme de manière exacte le signal IRMf en une séquence de DRH. Le modèle biologique doit permettre de modéliser trois problèmes (désynchronisation entre la séquence de DRH et la séquence de transitions du paradigme, pas de réponse à tous les stimuli et existence de DRHs ne correspondant pas à la tâche d’intérêt). Lorsque l’étape de détection n’est plus parfaite, le modèle du détecteur doit faire face à trois nouveaux problèmes (problème de non détection, de fausse alarme et de localisation temporelle)

- Une RD peut ne pas être associée à un DRH. Il s’agit d’une fausse alarme du système de détection.
- A cause des problèmes de non détection du détecteur, tous les DRHs ne peuvent pas être associés avec une RD.
- Des problèmes de localisation temporelle de la RD par rapport au DRH correspondant peuvent enfin exister.

Ainsi, il est nécessaire d’avoir un modèle permettant de définir les propriétés du détecteur.

#### 4.3.1.3 Un modèle multicouche ?

Il serait *a priori* intéressant de posséder deux modèles. Un premier modèle représenterait l’étape de préparation des données (probabilité de détection, probabilité de fausses alarmes, mauvaise localisation temporelle des détections) alors qu’un second modèle serait lié à la caractérisation de l’activité cérébrale ou à la discrimination entre l’activité cérébrale recherchée et tout autre activité de fond. En définitive, on découplerait ainsi les aspects liés à la mesure de ceux liés au fonctionnement du cerveau. On retrouve ainsi des aspects du filtrage de Kalman ou du filtrage particulière puisque l’on aurait un modèle « physique » et un modèle « d’observation ».

Cependant, notre choix ne s’est pas tourné vers une telle démarche. Caractériser les performances de l’étape de détection est très délicat en raison de la forte variabilité des données (variabilité de la réponse hémodynamique et du rapport signal sur bruit en fonction des tâches et des zones corticales notamment). La définition du modèle biologique serait encore plus complexe ! Enfin, avec une telle solution, il y a un risque certain que le modèle soit incapable de traiter de façon univoque les données, parce que sur-représentées (espace paramètre trop grand, problème mal posé).

Nous avons finalement préféré utiliser un unique modèle. Ce modèle est représentatif à la fois de l’activité cérébrale et des caractéristiques du détecteur utilisé pour transformer le



signal IRMf en une séquence de RDs. Ainsi, par la suite, nous ne parlons plus que de RDs. On fait l'hypothèse relativement naturelle qu'une transition « off-on » du paradigme peut être associée au maximum à une RD et qu'une RD peut être associée au maximum à une transition du paradigme. On retrouve enfin les trois problèmes déjà mentionnés :

- il peut exister un problème de synchronisation entre la RD et la transition du paradigme avec laquelle elle est associée (origine biologique ou/et problème de localisation temporelle lors de l'étape de détection) ;
- toutes les transitions « off-on » du paradigme ne peuvent pas être associées avec une RD (origine biologique – aucune réponse du patient – ou non détection du système de détection) ;
- toutes les RDs ne peuvent pas être associées avec une transition « off-on » du paradigme (origine biologique – DRH de non-intérêt – ou fausse alarme du système de détection).

### 4.3.2 Intérêt de la modélisation markovienne cachée

Les trois problèmes d'inférence ont un intérêt certain dans le cadre de notre application. Ils sont illustrés sur la figure 4.5. L'activité cérébrale locale est mesurée en terme de vraisemblance de la séquence de RDs connaissant le modèle d'activité (résolution du problème de l'évaluation). La résolution du problème de décodage permet de dire, pour chaque RD s'il s'agit d'une fausse alarme ou d'une bonne détection, chaque bonne détection pouvant être associée avec une transition du paradigme. Enfin, le modèle d'activité peut être appris à partir d'un ensemble de signaux actifs en résolvant le problème de l'apprentissage.

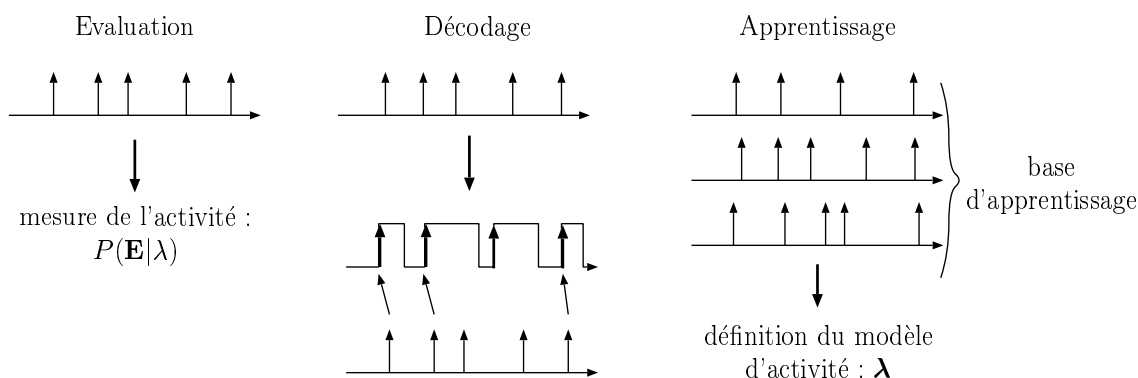


FIG. 4.5 – Problèmes d'inférence et cartographie en IRMf cérébrale. La résolution du problème de l'évaluation permet de mesurer l'activité cérébrale locale, celle du décodage permet d'apporter une information sémantique et celle de l'apprentissage de définir un modèle représentatif d'un corpus d'apprentissage.



# Chapitre 5

## Le modèle semi-markovien caché de séquence d'événements

Le modèle semi-markovien caché de séquence d'événements (MSMCSE) a été initialement développé par L. Thoraval pour l'analyse du signal électrocardiographique [208]. Ce modèle se révèle bien adapté à l'analyse de processus aléatoires de nature événementielle. Dans ce modèle, le processus observable est issu d'une étape de détection et de caractérisation des événements d'intérêt composant le processus d'entrée à analyser. Le MSMCSE modélise ensuite la distribution statistique des événements détectés sur l'axe des temps.

Dans un premier temps, nous présentons la notion de séquence d'événements et de séquence d'observations d'un MSMCSE, la seconde séquence se construisant à partir de la première. Ensuite, les différents éléments d'un MSMCSE sont présentés ainsi que le graphe d'états du modèle. Nous terminons ce chapitre en traitant, comme à l'accoutumée, de la résolution des problèmes d'inférence associés.

### 5.1 Séquence d'événements, séquence d'observations

Dans une approche par MSMCSE, une étape de prétraitement détecte et caractérise les événements supposés d'intérêt composant le processus événementiel à analyser. A chaque instant de détection  $t_l$ , une observation, encore appelée événement,  $e_l$ , est produite.  $\mathbf{E} = e_1 e_2 \dots e_L$  désigne la séquence des  $L$  événements détectés. Les événements  $e_1$  et  $e_L$  sont deux événements fictifs ajoutés artificiellement aux extrémités de la séquence  $\mathbf{E}$  pour des raisons de modélisation des durées inter-événements. On note par  $\tau_E = \{t_1 = 1, t_2, \dots, t_L = T\}$  l'ensemble des instants de détection. Par construction, on a :  $t_1 = 1 < t_2 < \dots < t_L = T$ . De plus,  $\tau_E$  est inclus dans  $\tau_O$ ,  $\tau_O$  étant l'ensemble des instants d'observation du processus à analyser ( $\tau_O = \{1, 2, \dots, T\}$ ).

La séquence d'observation  $\mathbf{O} = O_1 O_2 \dots O_T$  est déduite de la séquence  $\mathbf{E}$  par insertion d'un événement *null*, noté  $\emptyset$ , dans  $\mathbf{E}$  pour chaque  $t \in \tau_{\bar{E}} = \tau_O - \tau_E$ . L'écriture  $O_t = \emptyset$  signifie qu'un trou d'observation, résultat d'une non détection, apparaît à l'instant  $t$ . A titre d'exemple, si  $T = 20$  et  $\tau_E = \{1, 3, 6, 12, 15, 20\}$ , alors la séquence d'événements détectés est  $\mathbf{E} = e_1 e_2 e_3 e_4 e_5 e_6$ , laquelle conduit à la séquence d'observations  $\mathbf{O} = e_1 \emptyset e_2 \emptyset \emptyset e_3 \emptyset \emptyset \emptyset \emptyset e_4 \emptyset \emptyset e_5 \emptyset \emptyset \emptyset e_6$ .

## 5.2 Eléments d'un MSMCSE

Le modèle semi-markovien caché de séquence d'événements (MSMCSE) est un double processus stochastique constitué d'un processus caché  $\mathbf{X}$  et d'un processus observable  $\mathbf{O}$ . Le processus caché  $\mathbf{X}$  modélise la distribution statistique des événements détectés et des trous d'observation le long de l'axe temporel. Le processus observable représente, quant à lui, les caractéristiques à court terme des événements détectés.

### 5.2.1 Espace d'états

En raison de la nature orientée détection du prétraitement appliqué au processus à analyser, toute séquence d'observations d'un MSMCSE se voit potentiellement constituée d'événements vrais positifs (*evp*) caractérisant le processus, ainsi que d'événements faux positifs (*efp*) et de trous d'observation (*null*). Si  $C - 2$  représente le nombre de classes d'*evp* caractérisant le processus modélisé, alors l'espace d'état  $\mathcal{Q}$  du MSMCSE résulte de l'union de deux sous-espaces :

$$\mathcal{Q} = \mathcal{Q}_{evp} \cup \mathcal{Q}_{efp/null} \quad (5.1)$$

où :

- l'espace d'état  $\mathcal{Q}_{evp} = \{q_i\}_{1 \leq i \leq C}$  est constitué d'états markoviens cachés standards  $q_i$  produisant des *evps* aux instants  $t_l \in \tau_E$ .  
Les états de départ et d'arrivée du processus caché,  $q_1$  et  $q_C$  modélisent les *evps* fictifs  $e_1$  et  $e_L$  respectivement. En d'autres termes, on a  $X_1 = q_1$ ,  $X_T = q_C$  et  $X_t \notin \{q_1, q_C\}$  pour  $t = 2 \dots T - 1$ .
- L'espace d'état  $\mathcal{Q}_{efp/null} = \{q_{ij}\}_{1 \leq i < C, 1 < j \leq C}$  est formé, quant à lui, d'états semi-markoviens cachés  $q_{ij}$  produisant des *efps* aux instants  $t_l \in \tau_E$  ou des *null* aux instants  $t \in \tau_{\bar{E}}$ . Dans cette configuration, un état  $q_{ij}$  est inséré entre deux états *evp*  $q_i$  et  $q_j$ . Cet état modélise la sous-séquence d'observations de la forme  $\mathbf{O}_{t_k+1}^{t_k-1}$  insérée entre deux événements d'intérêt successifs  $o_{t_k} = e_k$  et  $o_{t_l} = e_l$  d'états cachés  $q_i$  et  $q_j$ , respectivement. Les distributions de probabilité associées à  $q_{ij}$  (loi de durée et loi d'observation) sont caractéristiques de la paire d'états  $q_i$  et  $q_j$  dans laquelle  $q_{ij}$  est inséré. Par définition, l'état semi-markovien  $q_{ij}$  résulte de l'agrégation de deux états :  $q_{ij} = \{q_{ij}^o, q_{ij}^\emptyset\}$  où  $q_{ij}^o$  modélise les *efps* et  $q_{ij}^\emptyset$  les trous d'observation. Dans la suite, l'état semi-markovien  $q_{ij}$  sera nommé également état agrégé.

### 5.2.2 Matrice de transition

La matrice de transition d'un MSMCSE est définie sur le sous-espace d'états  $\mathcal{Q}_{evp}$ . En effet, de par la structure du modèle, étant dans l'état agrégé  $q_{ij}$ , le dernier et le prochain état *evp* visités avant et après  $q_{ij}$  par le processus caché sont nécessairement l'état *evp*  $q_i$  et l'état *evp*  $q_j$ , respectivement. De plus, en raison de la spécificité des états  $q_1$  et  $q_C$ , la matrice  $\mathbf{A} = \{a_{ij}\}$  est définie pour  $1 \leq i < C$ ,  $1 < j \leq C$  avec pour terme général :

$$a_{ij} \triangleq P[X_{t+1} = q_j \text{ ou } q_{ij} | X_t = q_i] \quad (5.2)$$

Nous notons ici que la matrice  $\mathbf{A}$  est de taille  $C - 1 \times C - 1$

### 5.2.3 Lois de durée

On désigne par  $\mathbf{D} = \{p_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$  l'ensemble des lois de durée inter-événements. La loi  $p_{ij}(\cdot)$  représente non seulement la durée d'occupation de l'état  $q_{ij}(\cdot)$  mais aussi la loi de durée entre deux événements d'intérêts  $q_i$  et  $q_j$ . Elle est définie par :

$$p_{ij}(t_l - t_k) = P[t_l - t_k | X_{t_l} = q_j, \mathbf{X}_{t_k+1}^{t_l-1} = q_{ij}, X_{t_k} = q_i] \quad (5.3)$$

### 5.2.4 Lois d'observation

L'ensemble des lois d'observation,  $\mathbf{B} = \{b_j(\cdot)\}_{1 \leq j \leq C}$ , associé aux états *evp* est défini par :

$$b_j(O_t) = P[O_t | X_t = q_j] \quad (5.4)$$

Par définition,  $b_1(O_1 = e_1) = 1$  et  $b_C(O_T = e_L) = 1$ .

La distribution de probabilité d'observation associée à  $q_{ij}$ ,  $b_{ij}(\cdot)$ , est construite sur un processus de Bernoulli de paramètre  $e_{ij}$ . La probabilité  $e_{ij}$  est représentative du taux d'émission d'*efps* quand le processus caché est dans l'état  $q_{ij}$ . Cette probabilité modélise ainsi le taux moyen de fausses alarmes entre les deux états *evps*  $q_i$  et  $q_j$  :

$$e_{ij} = P[X_t = q_{ij}^o | X_t = q_{ij}] \quad (5.5)$$

On note  $\mathbf{E}$  la matrice de probabilité d'émission :  $\mathbf{E} = \{e_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$ . Comme  $P[O_t = \emptyset | X_t = q_{ij}^o] = 1$  et  $P[O_t = \emptyset | X_t = q_{ij}] = 0$ , la distribution de probabilité d'observation associée à  $q_{ij}$  s'écrit finalement :

$$\begin{aligned} P[O_t | X_t = q_{ij}] &= P[O_t | X_t = q_{ij}^o] \cdot e_{ij} \cdot \mathbf{1}_{\tau_E}(t) + (1 - e_{ij}) \cdot \mathbf{1}_{\tau_{\bar{E}}}(t) \\ &= b_{ij}(O_t) \cdot e_{ij} \cdot \mathbf{1}_{\tau_E}(t) + (1 - e_{ij}) \cdot \mathbf{1}_{\tau_{\bar{E}}}(t) \\ &= b_{ij}^+(O_t) \end{aligned} \quad (5.6)$$

où  $\mathbf{1}_X(\cdot)$  est la fonction indicatrice vérifiant  $\mathbf{1}_X(t) = 1$  si  $t \in X$ , 0 sinon. On note  $\mathbf{B}' = \{b_{ij}(\cdot)\}_{1 \leq i < C, 1 < j \leq C}$  l'ensemble des lois d'observations associées aux états  $q_{ij}^o$ . La probabilité de la sous-séquence  $\mathbf{O}_{t_k+1}^{t_l-1}$  peut finalement s'écrire en considérant les observations indépendantes conditionnellement aux états :

$$\begin{aligned} P[\mathbf{O}_{t_k+1}^{t_l-1} | \mathbf{X}_{t_k+1}^{t_l-1} = q_{ij}] &= \prod_{\tau=t_k+1}^{t_l-1} b_{ij}^+(O_\tau) \\ &= b_{ij}^+(\mathbf{O}_{t_k+1}^{t_l-1}) \end{aligned} \quad (5.7)$$

Un MSMCSE est donc finalement entièrement caractérisé par la donnée du vecteur paramètre :  $\boldsymbol{\lambda} = \{\mathbf{Q}, \mathbf{A}, \mathbf{B}, \mathbf{B}', \mathbf{D}, \mathbf{E}\}$ .

## 5.3 Graphe d'états d'un MSMCSE

Un exemple de graphe d'états est représenté figure 5.1.b. Les cercles représentent les états *evp* et les rectangles bicolores placés entre les états représentent les états agrégés (*efp/null*). On rappelle en effet qu'un état agrégé  $q_{ij}$  est par construction toujours inséré entre les états *evp*  $q_i$  et  $q_j$ .

La représentation markovienne standard équivalente d'un MSMCSE peut être obtenue en démultipliant chaque état agrégé  $q_{ij}$  en deux lignes à retard, comme le montre la figure 5.1.c. L'état semi-markovien  $q_{ij}$  peut être vu en effet comme deux lignes à retard composées d'états

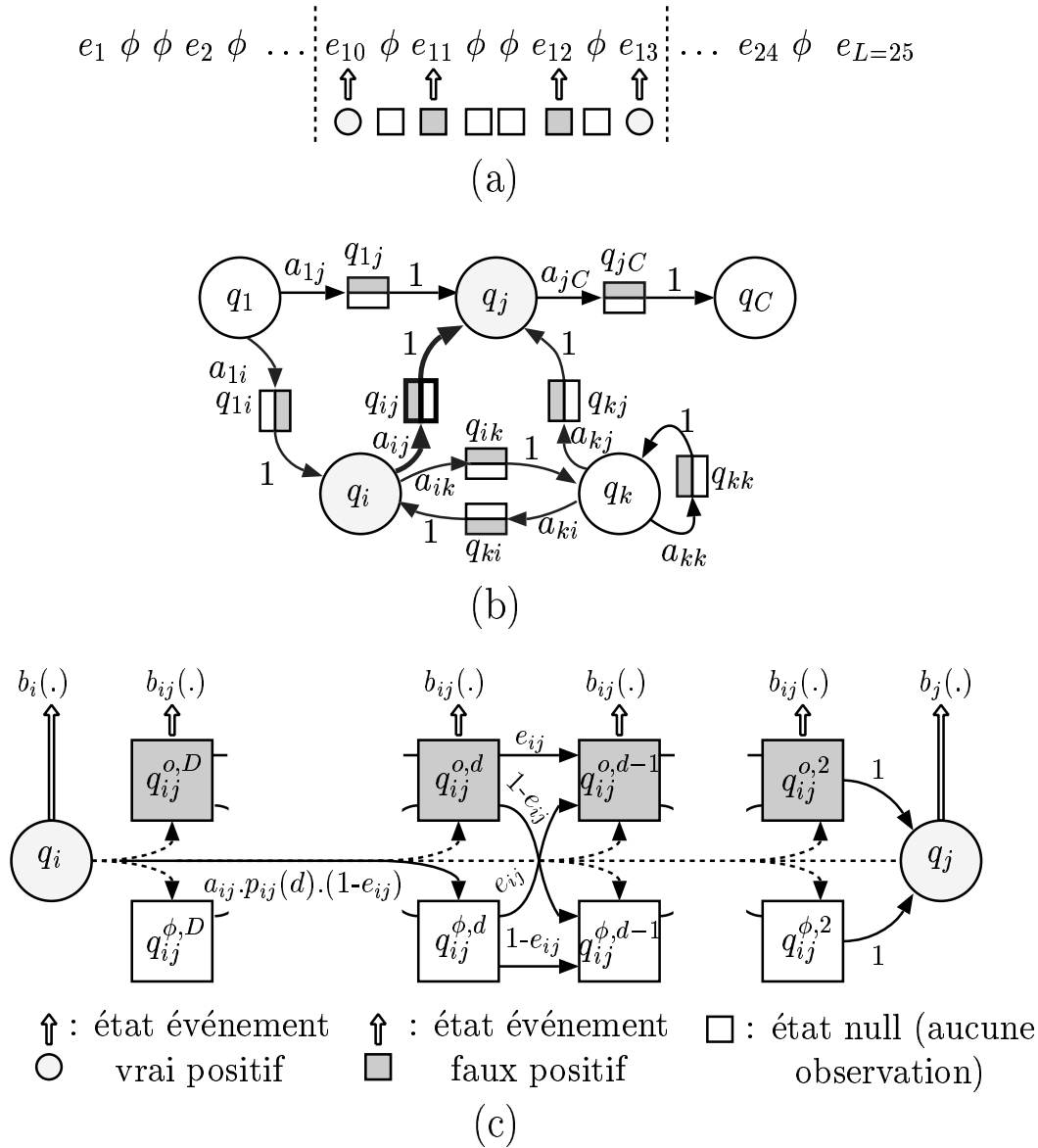


FIG. 5.1 – Exemple de graphe d'état d'un MSMCSE. (a) Structure d'une séquence d'observation d'un MSMCSE avec les états cachés associés. Observer la sous-séquence d'observations revient à occuper successivement les états  $q_i$ ,  $q_{ij}^{\phi,7}$ ,  $q_{ij}^{\phi,6}$ ,  $q_{ij}^{\phi,5}$ ,  $q_{ij}^{\phi,4}$ ,  $q_{ij}^{\phi,3}$ ,  $q_{ij}^{\phi,2}$  et  $q_j$ . En raison de l'étape de détection, une séquence d'observation est constituée de trous d'observation et d'événements détectés qui peuvent être soit des bonnes détections ( $e_{10}$  et  $e_{13}$ ), soit des fausses alarmes ( $e_{11}$  et  $e_{12}$ ). (b) Représentation compacte du graphe d'état d'un MSMCSE. (c) Représentation markovienne standard équivalente sous forme de graphe d'état de l'état semi-markovien  $q_{ij}$  intercalé entre les états markoviens  $q_i$  et  $q_j$ .

markoviens en parallèle entre lesquelles le processus caché transite suivant les probabilités  $e_{ij}$  et  $1 - e_{ij}$ . Si  $D$  dénote la durée maximale autorisée entre deux états *evp* successifs, la première ligne est composée de  $D - 1$  états markoviens standards  $q_{ij}^{o,d}$ ,  $d = 2, \dots, D$ , avec pour densité de probabilité d'observation associée  $b_{ij}(\cdot)$ . La seconde est également composée de  $D - 1$  états markoviens standards  $q_{ij}^{o,d}$ ,  $d = 2, \dots, D$ , mais sans aucune densité de probabilité associée ou plutôt  $P[X_t = \emptyset | X_t = q_{ij}^{o,d}] = 1$ . Par exemple, observer la sous-séquence d'observations de la figure 5.1.a revient à occuper successivement les états  $q_i$ ,  $q_{ij}^{o,7}$ ,  $q_{ij}^{o,6}$ ,  $q_{ij}^{o,5}$ ,  $q_{ij}^{o,4}$ ,  $q_{ij}^{o,3}$ ,  $q_{ij}^{o,2}$  et  $q_j$ .

Un MSMC peut enfin être vu comme une instance particulière des MSMCSEs. En posant  $e_{ij} = 1$ ,  $b_{ij}(\cdot) \equiv b_j(\cdot)$  et  $p_{ij}(\cdot) \equiv p_j(\cdot)$  dans la figure 5.1.c, on retrouve la représentation standard équivalente d'un état semi-markovien donnée par Levinson [138].

## 5.4 Résolution des problèmes d'inférence

Les trois problèmes fondamentaux que sont l'évaluation, l'apprentissage et le décodage classiquement rencontrés dans le cadre des modèles markoviens ou semi-markoviens cachés se posent et se résolvent de manière analogue dans le cas des MSMCSEs. Leur résolution conduit à des versions modifiées de l'algorithme Forward–Backward, de l'algorithme de Baum–Welch et de l'algorithme de Viterbi, respectivement.

### 5.4.1 Evaluation

On rappelle que  $\tau_E = \{t_1 = 1, t_2, \dots, t_L = T\}$  est l'ensemble des instants de détection ( $t_1 = 1 < t_2 < \dots < t_L = T$ ). On notera  $d_{kl} = t_l - t_k$ ,  $l > k$ , la durée écoulée entre les événements  $e_k$  et  $e_l$ . La vraisemblance  $P[\mathbf{O}|\boldsymbol{\lambda}]$  est déterminée grâce aux variables avant et arrière définies pour la suite d'instant de détection  $t_l$  par :

$$\alpha_{t_l}(j) = P[\mathbf{O}_1^{t_l}, X_{t_l} = q_j | \boldsymbol{\lambda}] \quad (5.8)$$

$$\beta_{t_l}(i) = P[\mathbf{O}_{t_l+1}^T | X_{t_l} = q_i, \boldsymbol{\lambda}] \quad (5.9)$$

$\alpha_{t_l}(j)$  peut être déterminé itérativement. Pour  $l = 2, \dots, L$ ,  $j = 2, \dots, C$  :

$$\alpha_{t_l}(j) = \sum_{i=1}^{C-1} \sum_{\substack{k=1, \\ 1 \leq d_{kl} \leq D}}^{l-1} \alpha_{t_k}(i) \cdot a_{ij} \cdot p_{ij}(d_{kl}) \cdot b_{ij}^+(\mathbf{O}_{t_k+1}^{t_l-1}) \cdot b_j(O_{t_l}) \quad (5.10)$$

avec les conditions initiales  $\alpha_{t_1=1}(i) = 1$ , si  $i = 1$ , 0 sinon. On a enfin  $P[\mathbf{O}|\boldsymbol{\lambda}] = \alpha_{t_L}(C)$ . Les variables arrières  $\beta_{t_l}(i)$  peuvent également être obtenues de manière itérative. Pour  $l = L - 1, \dots, 1$ ,  $i = 1, \dots, C - 1$  :

$$\beta_{t_l}(i) = \sum_{j=2}^C \sum_{\substack{m=l+1, \\ 1 \leq d_{lm} \leq D}}^L \beta_{t_m}(j) \cdot a_{ij} \cdot p_{ij}(d_{lm}) \cdot b_{ij}^+(\mathbf{O}_{t_l+1}^{t_m-1}) \cdot b_j(O_{t_m}) \quad (5.11)$$

avec les conditions initiales  $\beta_{t_L}(j) = 1$  si  $j = C$ , 0 sinon.

Il est à noter que la charge de calcul ainsi que les ressources mémoires sont sensiblement diminuées par rapport à une approche par MSMC dès lors que  $L \ll T$ .

### 5.4.2 Décodage

La meilleure séquence d'états  $\mathbf{q}^{MAP}$  est sélectionnée parmi toutes les séquences d'états possibles de manière à maximiser la probabilité jointe  $P[\mathbf{O}, \mathbf{q}|\boldsymbol{\lambda}]$  :

$$\mathbf{q}^{MAP} = \arg \max_{\mathbf{q}} P[\mathbf{O}, \mathbf{q}|\boldsymbol{\lambda}] \quad (5.12)$$

La maximisation est obtenue en utilisant la variable  $\delta_{t_l}(j)$  définie par :

$$\delta_{t_l}(j) = \max_{\mathbf{q}_1^{t_l-1}} P[\mathbf{O}_1^{t_l}, \mathbf{q}_1^{t_l-1}, X_{t_l} = q_j | \boldsymbol{\lambda}], \quad (5.13)$$

Comme les variables avant et arrière, la variable  $\delta_{t_l}(j)$  peut être calculée de manière itérative après avoir substitué dans l'équation (5.10) les «  $\alpha$  » par des «  $\delta$  » et la double somme par une double maximisation.

Une variable additionnelle  $\psi(j, l)$  est requise de manière à sauvegarder la paire d'indices  $(i, k)$  maximisant l'équation (5.10) ainsi modifiée. La meilleure séquence d'états  $\mathbf{q}^{MAP}$  est obtenue ensuite en deux étapes. On commence par déterminer la sous-séquence de  $\mathbf{q}^{MAP}$  formée uniquement des événements *evp* par retour-arrière sur le tableau de pointeurs  $\psi(j, l)$ . Ensuite, l'obtention de  $\mathbf{q}^{MAP}$  à partir de la sous-séquence juste déterminée est immédiate en raison de la structure du modèle.

### 5.4.3 Apprentissage

Nous présentons ici succinctement l'extension de l'algorithme de Baum-Welch dans le cadre des MSMCSEs. On considère un corpus d'apprentissage  $\mathbf{O} = \{\mathbf{O}^1, \mathbf{O}^2, \dots, \mathbf{O}^V\}$  constitué de  $V$  séquences d'observation supposées mutuellement indépendantes. La fonction auxiliaire  $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$  est définie comme suit :

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') = \sum_{v=1}^V \sum_{\mathbf{X}^v} P[\mathbf{X}^v | \mathbf{O}^v, \boldsymbol{\lambda}] \log P[\mathbf{X}^v, \mathbf{O}^v | \boldsymbol{\lambda}'] \quad (5.14)$$

Dans le cadre des MSMCSEs, la fonction auxiliaire  $Q(\cdot, \cdot)$  peut s'exprimer de la manière suivante :

$$\begin{aligned} Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') &= \sum_{v=1}^V \sum_{i=1}^{C-1} \sum_{j=2}^C \sum_{k=1}^{L^v-1} \gamma_{ij}^{1v}(k) \log(a'_{ij}) \\ &+ \sum_{v=1}^V \sum_{i=2}^{C-1} \sum_{k=2}^{L^v-1} \gamma_i^{3v}(k) \log(b'_i(O_{t_k}^v)) \\ &+ \sum_{v=1}^V \sum_{i=1}^{C-1} \sum_{j=2}^C \sum_{k=2}^{L^v-1} \gamma_{ij}^{4v}(t_k) \log(b'_{ij}(O_{t_k}^v)) \\ &+ \sum_{v=1}^V \sum_{i=1}^{C-1} \sum_{j=2}^C \left( \sum_{k=2}^{L^v-1} \gamma_{ij}^{4v}(t_k) \log(e'_{ij}) + \sum_{t \in \tau_{\mathbb{E}}^v} \gamma_{ij}^{4v}(t) \log(1 - e'_{ij}) \right) \\ &+ \sum_{v=1}^V \sum_{i=1}^{C-1} \sum_{j=2}^C \sum_{k=1}^{L^v-1} \sum_{l=k+1}^{L^v} \gamma_{ij}^{2v}(k, l) \log(p'_{ij}(t_l - t_k)) \end{aligned} \quad (5.15)$$

où :

$$\begin{cases} \gamma_{ij}^{1v}(k) = P[X_{t_k}^v = q_i, X_{t_{k+1}}^v = q_{ij} \text{ ou } q_j | \mathbf{O}^v, \boldsymbol{\lambda}] \\ \gamma_{ij}^{2v}(k, l) = P[X_{t_k}^v = q_i, \mathbf{X}_{t_k}^{t_{k+1}^v-1} = q_{ij}, X_{t_l}^v = q_j | \mathbf{O}^v, \boldsymbol{\lambda}] \\ \gamma_i^{3v}(k) = P[X_{t_k}^v = q_i | \mathbf{O}^v, \boldsymbol{\lambda}] \\ \gamma_{ij}^{4v}(t) = P[X_t^v = q_{ij} | \mathbf{O}^v, \boldsymbol{\lambda}] \end{cases} \quad (5.16)$$

Conformément à l'équation (5.15), la fonction auxiliaire  $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$  se décompose donc en une somme de cinq termes, chacun maximisable séparément. Comme dans le cas standard, la



maximisation peut s'effectuer en utilisant des méthodes d'optimisation basées sur les multiplicateurs de Lagrange. Les formules de réestimation sont reportées ci-après dans le cas où les densités de probabilité sont toutes des lois gaussiennes à une dimension. On utilise les notations suivantes :  $b_i \sim \mathcal{N}(\mu_i, \sigma_i)$ ,  $b_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$  et  $p_{ij} \sim \mathcal{N}(\mu d_{ij}, \sigma d_{ij})$ . Enfin, les paramètres du modèle  $\lambda'$  sont tous munis d'un « ' ».

$$\left\{ \begin{array}{l} \mu'_i = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_i^{3v}(k) O_t^v}{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_i^{3v}(k)}, \quad \sigma'_i = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_i^{3v}(k) (O_t^v - \mu'_i)^2}{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_i^{3v}(k)} \\ \mu_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{4v}(t_k^v) O_t^v}{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{4v}(t_k^v)}, \quad \sigma'_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{4v}(t_k^v) (O_t^v - \mu'_{ij})^2}{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{4v}(t_k^v)} \\ \mu d'_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v-1} \sum_{l=k+1}^{L^v} \gamma_{ij}^{2v}(k,l) d_{kl}}{\sum_{v=1}^V \sum_{k=1}^{L^v-1} \sum_{l=k+1}^{L^v} \gamma_{ij}^{2v}(k,l)}, \quad \sigma d'_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v-1} \sum_{l=k+1}^{L^v} \gamma_{ij}^{2v}(k,l) (d_{kl} - \mu' d_{ij})^2}{\sum_{v=1}^V \sum_{k=1}^{L^v-1} \sum_{l=k+1}^{L^v} \gamma_{ij}^{2v}(k,l)} \\ a'_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{1v}(k)}{\sum_{v=1}^V \sum_{k=1}^{L^v-1} \sum_{l=2}^C \gamma_{il}^{1v}(k)} \\ e'_{ij} = \frac{\sum_{v=1}^V \sum_{k=1}^{L^v} \gamma_{ij}^{4v}(t_k^v)}{\sum_{v=1}^V \sum_{t=1}^{T^v} \gamma_{ij}^{4v}(t)} \end{array} \right. \quad (5.17)$$

avec :

$$\left\{ \begin{array}{l} \gamma_{ij}^{1v}(k) = \frac{\alpha_{t_k}(i) \left( \sum_{l=k+1}^L p_{ij}(d_{kl}^v) b_j(O_{t_l}^v) \beta_{t_l}^v(j) b_{ij}^+ \left( \mathbf{O}_{t_k+1}^{v, t_l-1} \right) \right)}{P[\mathbf{O}^v | \lambda]} \\ \gamma_{ij}^{2v}(k, l) = \frac{\alpha_{t_k}^v(i) \beta_{t_l}^v(j) b_{ij}^+ \left( \mathbf{O}_{t_k+1}^{v, t_l-1} \right)}{P[\mathbf{O}^v | \lambda]} \\ \gamma_i^{3v}(k) = \frac{\alpha_{t_k}^v(i) \beta_{t_k}^v(i)}{P[\mathbf{O}^v | \lambda]} \\ \gamma_{ij}^{4v}(t) = \frac{\sum_{k=1, t_k < t}^{L^v-1} \sum_{l=k+1, t_l > t}^{L^v} \alpha_{t_k}^v(i) \beta_{t_l}^v(j) a_{ij} p_{ij}(t_l - t_k) b_j(O_{t_l}^v) b_{ij}^+ \left( \mathbf{O}_{t_k+1}^{v, t_l-1} \right)}{P[\mathbf{O}^v | \lambda]} \end{array} \right. \quad (5.18)$$

Nous n'avons pas mis les détails des calculs dans la mesure où l'équation (5.17) tombe sous le sens. Il s'agit tout simplement, comme dans le cas standard des estimateurs au sens du MV pondérés par les probabilités *a posteriori* qui conviennent.



# Chapitre 6

## Application en cartographie IRMf cérébrale

Nous présentons dans ce chapitre l'application du MSMCSE à la cartographie en IRMf cérébrale.

### 6.1 Présentation générale

Les données IRMf  $\mathcal{F}$  sont composées d'un ensemble de signaux IRMf  $f^v(\cdot)$ , chaque signal IRMf étant représentatif d'un voxel différent  $v$ . Le protocole de stimulation considéré est un paradigme en blocs à deux états (un état de contrôle et un état d'intérêt). On note  $P - 2$  le nombre de transitions « off-on » dans le paradigme et  $t'_i$  l'instant d'occurrence de la  $(i-1)$ -ème transition. Pour les problèmes de modélisation des lois de durée, on introduit deux transitions fictives au début et à la fin de la séquence :  $t'_1 = 1$  et  $t'_P = T$ .

Pour rappel, notre méthode de cartographie se base sur le principe, qu'en cas d'activité, il existe un couplage entre les RDs détectées dans le signal IRMf et les transitions « off-on » du paradigme (paradigme en blocs), ou un couplage entre les RDs détectées dans le signal IRMf et les différents stimuli du paradigme (paradigme événementiel). Dans cette partie, nous ne parlons toutefois que des paradigmes en blocs. Les résultats obtenus avec des paradigmes événementiels ne sont en effet pas satisfaisants en raison des limitations de la méthode que nous décrivons également dans ce chapitre. Ces limitations seront complètement levées dans la troisième partie de cette recherche, dans laquelle nous considérons de nouveaux éléments de modélisation.

Le synoptique de traitement est représenté à la figure 6.1. La première étape est une étape de prétraitement des données. Pour chaque voxel, la séquence d'observation  $\mathbf{O}^v$  est obtenue par prétraitement de  $f^v(\cdot)$  (section 6.2). Dans une seconde étape, les états cachés et la topologie du MSMCSE sont définis à partir du paradigme de stimulation (section 6.3). Enfin, dans une dernière étape, on cherche conjointement à estimer les paramètres du modèle et à cartographier les différentes zones d'activité (section 6.4). Finalement, la dernière section présente les nouvelles fonctionnalités que permet d'obtenir la modélisation markovienne cachée en IRMf cérébrale, dans l'analyse des retards d'activation, des modes d'activation et dans l'analyse de la réponse hémodynamique.

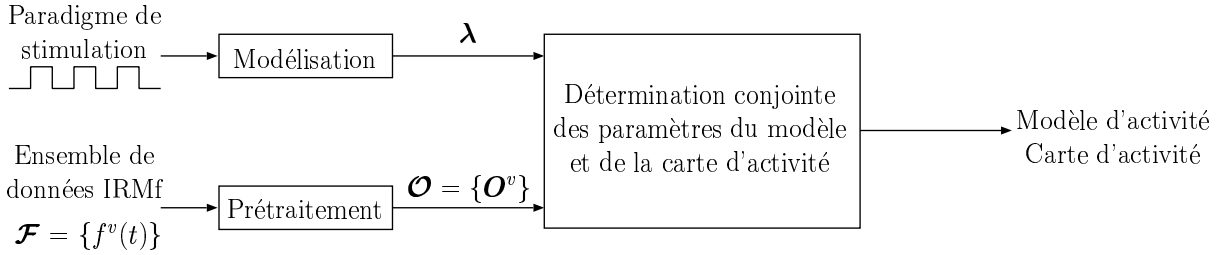


FIG. 6.1 – Synoptique de traitement

## 6.2 Séquence d'événements IRMf, séquence d'observations

L'étape de prétraitement transforme l'ensemble des données IRMf  $\mathcal{F} = \{f^v(.)\}$  en un ensemble de séquences d'observation  $\mathcal{O} = \{\mathcal{O}^v\}$ , chaque observation  $\mathcal{O}^v$  étant une observation du MSMCSE. En d'autres termes, il s'agit de détecter et de caractériser les événements supposés d'intérêt de chaque signal IRMf.

### 6.2.1 Détection des événements

Un événement correspond à une augmentation du signal IRMf. Cette augmentation est liée à la réponse hémodynamique et s'étend temporellement entre un éventuel minimum de la réponse (« under-shoot ») et le maximum de la réponse. Le motif ainsi défini possède approximativement en son centre un point d'inflexion dont le temps d'occurrence est associé au temps d'occurrence de l'événement.

La capacité de la transformée en ondelettes continue (TOC) à détecter les non-stationnarités dans le domaine temps-échelle est ici utilisée [4]. L'ondelette est choisie complexe de manière à bénéficier de l'information de la phase de la TOC du signal IRMf. Des singularités symétriques du signal analysé sont associées à des valeurs particulières de la phase de sa TOC [4]. Plus précisément, un point d'inflexion positif au temps  $t_l^v$  est associé pour des résolutions hautes au passage à «  $+\pi/2$  » de la phase. L'instant de détection d'un événement au voxel  $v$  est donc obtenu en cherchant à une haute résolution d'analyse  $s_h$ , l'instant  $t_l^v$  pour lequel la phase de la TOC du signal passe par la valeur de «  $+\pi/2$  ». Une telle situation est décrite figure 6.2 pour un signal actif et un signal inactif. Les courbes décrites par les passages par «  $+\pi/2$  » de la phase le long des échelles sont appelées *empreintes*.

Le choix de l'échelle  $s_h$  est important car elle contrôle le taux de détection. En pratique, la résolution  $s_h$  est choisie de manière à ce que  $k \times (P - 2)$  RDs soient détectées en moyenne par signal IRMf sur l'ensemble des données IRMf ( $k$  est initialement fixé à 1.5). Si l'échelle de basse résolution  $s_l$  n'a aucune influence sur le taux de détection, elle doit être suffisamment petite de manière à fournir des empreintes significatives (les empreintes sont utilisées par la suite pour caractériser l'événement). Dans nos expériences, l'échelle  $s_l$  est choisie de manière à ce que  $0.5 \times (P - 2)$  RDs soient détectées en moyenne par signal IRMf. Finalement, le nombre de décompositions d'échelles est fixé à 20.

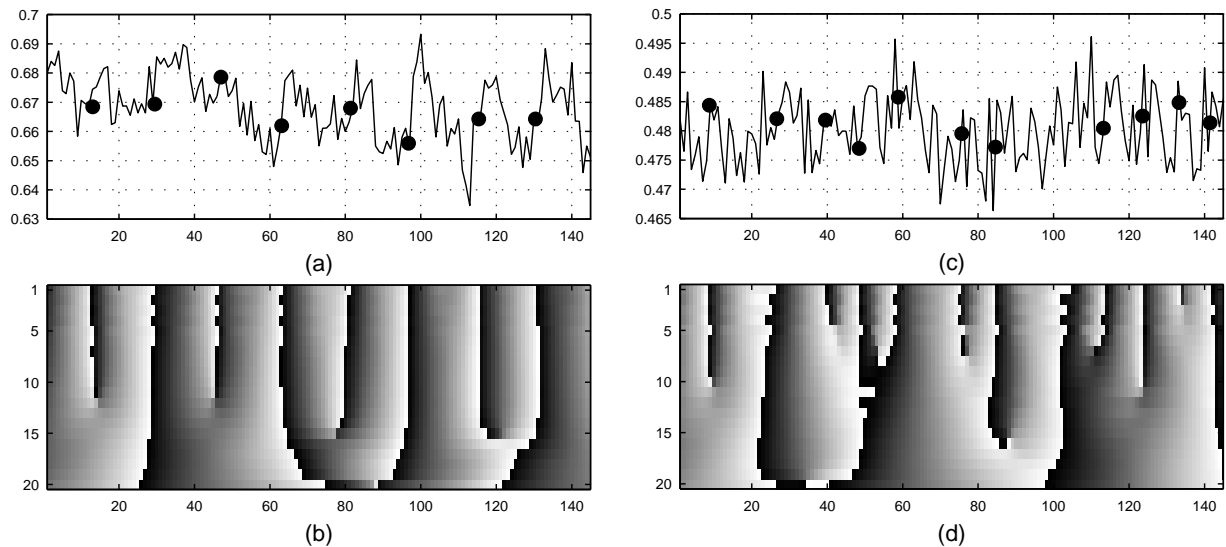


FIG. 6.2 – Prétraitement du signal IRMf. Signal IRMf (a) actif et non actif (c). (b), (d) Plans de phase de la TOC correspondant aux signaux (a) et (c). Le temps et le nombre d'échelles de la TOC sont reportés sur l'axe  $X$  et  $Y$ , respectivement. Les échelles hautes et basses de résolution  $s_h$  et  $s_l$  correspondent aux indices 1 et 20, respectivement. Les passages par «  $+\pi/2$  » de la phase correspondent au passage du blanc au noir. La détection de RDs est effectuée en détectant à l'échelle 1 le passage par «  $+\pi/2$  » de la phase. Les temps de détection des RDs sont indiqués par des points noirs placés sur les signaux.

### 6.2.2 Caractérisation des événements

Notre objectif n'est pas de caractériser au mieux les événements d'intérêt mais plutôt de déterminer des mesures permettant de discriminer au mieux les bonnes détections des fausses alarmes. Dans ce but, nous avons tout d'abord défini différentes mesures. Une base d'apprentissage constituée de mesures associées à des bonnes détections et à des fausses alarmes a été construite avec l'aide d'un spécialiste. Nous avons ensuite effectué une analyse discriminante linéaire de Fisher [57] pour atteindre notre objectif.

Plus précisément, à chaque événement détecté à l'instant  $t_l^v$ , un vecteur de caractéristiques,  $\mathbf{f}_l^v$ , de 24 composantes est calculé :  $\mathbf{f}_l^v = \{\mathbf{f}_{1,l}^v, \mathbf{f}_{2,l}^v, \dots, \mathbf{f}_{24,l}^v\}$ .

- 14 composantes représentent le comportement du signal  $f^v(\cdot)$  autour du temps de détection  $t_l^v$  :
  - 5 composantes sont les échantillons d'une version centrée, réduite de  $f^v(\cdot)$ .
  - 5 autres composantes représentent 5 échantillons d'une version centrée, réduite de  $f^v(\cdot)$  préalablement filtrée (filtrage gaussien).
  - 4 composantes sont les produits de corrélation entre des versions filtrées et non filtrées de  $f^v(\cdot)$  avec un échelon unité, seul le signal autour de  $t_l^v$  étant considéré.
- 6 coefficients décrivent les maxima du module de la TOC du signal IRMf au temps  $t_l^v$  (les 5 modules les plus importants et la fréquence pour laquelle le module de la TOC est maximal).
- Les 4 dernières composantes sont calculées à partir de l'empreinte associée à l'événement détecté. 3 mesures concernent la longueur de l'empreinte et une dernière mesure concerne sa déviation temporelle à travers les échelles.

Ensuite, le vecteur caractéristique  $\mathbf{f}_l^v$  est réduit en une unique mesure scalaire  $e_l^v = \mathbf{c}^T \mathbf{f}_l^v$

de manière à discriminer au mieux les bonnes détections des fausses alarmes. Dans notre approche, le vecteur  $\mathbf{c}$  a été estimé une unique fois avec une analyse discriminante linéaire de Fisher [57].

Pour construire la base d'apprentissage, la procédure semi-automatique décrite ci-dessous a été utilisée à partir de l'ensemble des données IRMf à disposition (119 séquences d'images avec 3 paradigmes différents).

- 1 déterminer les cartes de t-test. Après seuillages des différentes cartes, sélectionner avec l'aide d'un expert les régions actives ;
- 2 détecter, pour chaque signal déclaré actif les RDs et calculer le vecteur caractéristique correspondant ;
- 3 classer les différents vecteurs caractéristiques avec l'algorithme illustré figure 6.3. Si l'instant de détection de l'événement est à proximité de l'instant de transition « off-on » du paradigme, l'événement est classé comme étant un *evp*. Au contraire, si l'instant de détection est éloigné de chacun des instants de transition du paradigme, l'événement est considéré comme étant un *efp*. Enfin, pour des positions intermédiaires, l'événement n'est pas classé.

Plusieurs bases de données ont été créées en faisant varier  $v, w, x, y$  (voir figure 6.3) et le seuil du t-test. Il ressort que le vecteur  $\mathbf{c}$  ne varie que très peu en fonction des différentes bases de donnée. Il apparaît également que les caractéristiques permettant de discriminer au mieux les *evps* des *efps* sont la longueur des empreintes. En d'autres termes, un *evp* est détecté sur une large bande de fréquence, contrairement aux fausses alarmes. On retrouve dans une certaine mesure le constat qui a été effectué dans les approches basées sur le « scale-space » linéaire pour l'analyse d'images : les structures qui existent sur un grand intervalle d'échelle sont souvent les structures d'intérêt ! En définitive, la caractérisation des données est fondée sur des mesures permettant de discriminer au mieux les bonnes détections des fausses alarmes. On s'abstient ainsi de faire un nombre important d'hypothèses quant à la forme de la réponse hémodynamique.

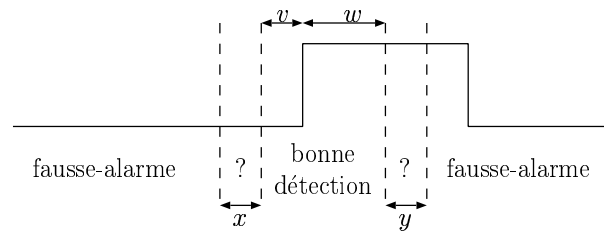


FIG. 6.3 – Méthode de classification des différents événements détectés

## 6.3 Modélisation par MSMCSE de l'activité cérébrale

### 6.3.1 Principe

La procédure est développée pour un paradigme en blocs à  $P$  blocs dont 2 sont fictifs. La modélisation est basée sur le principe d'alignement que nous rappelons ici brièvement. En cas d'activité, les RDs détectées dans le signal IRMf doivent s'aligner dans une certaine mesure sur les transitions « off-on » du paradigme. Ce couplage entre les RDs détectées dans le signal

IRMf et les transitions du paradigme est modélisé au travers du MSMCSE. Ce dernier modélise entre autres la distribution temporelle des événements d'intérêt.

On suppose qu'il existe un principe de correspondance 1-1 entre un événement d'intérêt et une transition « off-on » du paradigme. La modélisation du processus caché se fait alors à partir de la séquence de transitions du paradigme alors que la modélisation de la chronologie d'activation se fait à partir des temps d'occurrence associés aux transitions du paradigme (figure 6.4). Le principe de causalité devrait imposer que la RD prenne place après la transition « off-on » du paradigme avec laquelle elle est associée. Cependant, l'incertitude liée au temps de détection et le phénomène d'anticipation des patients nous incitent à ne pas considérer ce principe. Le modèle complet est spécifié de la façon suivante :

- *Etats cachés* ( $\mathcal{Q}$ ) : un état *evp*  $q_i$  est associé à chaque transition « off-on » du paradigme. L'indice  $i$  est représentatif de l'ordre d'apparition de la transition dans le paradigme créant une bijection entre les états *evp* et un bloc de stimulation. Ainsi, on a  $C = P$  états *evp* dans l'espace d'états du modèle.
- *Matrice de transition* ( $\mathbf{A}$ ) : pour prendre en compte l'évolution temporelle du processus, la topologie de la chaîne est de type gauche-droite, avec pour contrainte supplémentaire  $a_{ii} = 0, \forall i$ , rendant ainsi la matrice  $\mathbf{A}$  rectangulaire supérieure. Pour limiter l'espace de recherche et pour éviter de déclarer actif un signal IRMf  $f^v(\cdot)$  qui répond à un nombre limité de blocs de stimulation, on ajoute la contrainte que  $a_{ij} = 0$  si  $j > i + \Delta$  avec  $\Delta \sim \frac{P}{2}$ . A noter que si  $\Delta = 1$ ,  $f^v(\cdot)$  doit répondre à tous les blocs de stimulation.
- *Lois de durée et d'observation* ( $\mathbf{B}, \mathbf{B}', \mathbf{D}$ ) : toutes les densités de probabilité sont modélisées par une loi gaussienne. Ce choix est essentiellement guidé par la simplicité d'utilisation de la loi gaussienne. On note  $\mu d_{ij}$  et  $\sigma d_{ij}$  la moyenne et l'écart-type associés à la loi de durée  $p_{ij}(\cdot)$ .

Un exemple de modèle est représenté figure 6.4 avec  $\Delta = 2$  et un paradigme de stimulation

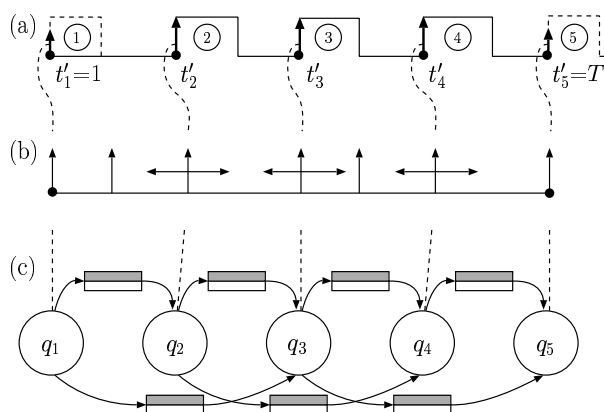


FIG. 6.4 – Principe de modélisation. (a) Paradigme de stimulation avec  $P = 5$  blocs, les blocs 1 et 5 étant des blocs fictifs. Les flèches verticales indiquent le début du stimulus associé à chaque bloc avec le temps d'occurrence  $t'_p$ . (b) Représentation schématique de la séquence de RDs. Les flèches horizontales soulignent l'imprécision de l'instant de détection des RDs. Enfin, certaines RDs ne sont pas associées avec des transitions du paradigme : il s'agit de fausses alarmes. (c) MSMCSE d'activation cérébrale avec les états *evp* (cercle) et les états agrégés (rectangles). Les lignes verticales en pointillé montrent le lien étroit qui peut exister entre le premier stimulus du bloc, une RD et un état *evp* du modèle.

pour lequel  $P = 5$ .

Finalement, pour réduire le nombre de paramètres à estimer, le concept de liaison de paramètres [186] est utilisé.

## 6.3.2 Paramètres du modèle et liaison de paramètres

### 6.3.2.1 La matrice de transition

En raison de la spécificité des états fictifs, la matrice  $\mathbf{A}$  est de taille  $C - 1 \times C - 1$ . Le terme associé à la  $i$ -ème ligne et  $j$ -ème colonne correspond donc à la probabilité  $a_{i,j+1}$ .

Soit  $\mathbf{a}_i$  la  $i$ -ème colonne de  $\mathbf{A}$  et  $\mathbf{0}_i$  un vecteur colonne de  $i$  zéros. Alors, on a :  $\mathbf{a}_1 = \{a_{12}, \dots, a_{1\Delta+1}, \mathbf{0}_{P-\Delta-1}\}$ . L'ensemble  $\{a_{12}, \dots, a_{1\Delta+1}\}$  tient compte spécifiquement de l'installation de l'activité induite par le protocole expérimental (aucune liaison de paramètres n'est effectuée). Par exemple, il est possible que le patient ne comprenne pas immédiatement la tâche à effectuer ou que des zones cérébrales ne soient actives qu'à partir du milieu du protocole de stimulation (phénomène d'apprentissage).

Une fois que le sujet a répondu à un bloc de stimulation, un autre ensemble de probabilités est utilisé :  $\{a_1, \dots, a_\Delta\}$ . On pose :

$$a_{k,k+i} = a_i \text{ si } k \neq 1 \text{ et } k+i \neq C \quad (6.1)$$

L'ensemble  $\{a_1, \dots, a_\Delta\}$  modélise ainsi la poursuite de l'activité. La probabilité de passer de l'état  $q_k$  à l'état  $q_{k+i}$  ( $k \neq 1$  et  $k+i \neq C$ ) ne dépend donc pas de  $k$ . Cette hypothèse n'est pas forcément justifiée d'un point de vue de l'application mais elle a l'avantage de réduire de façon importante le nombre de paramètres. Il est possible ainsi d'effectuer l'apprentissage dans de bonnes conditions même quand le cardinal du corpus d'apprentissage est très faible. Il peut parfois être de l'ordre de 100 dans notre application. De plus, l'estimation de la matrice de transition est très sensible à la taille du nombre d'échantillons dans la mesure où la probabilité de sauter des états est très faible. Estimer les probabilités associées à des sauts d'états est donc problématique et la liaison de paramètres s'avère donc être une solution satisfaisante.

Comme  $\mathbf{A}$  est une matrice stochastique, on aboutit finalement à la matrice définie par :  $\mathbf{a}_i = \{\mathbf{0}_{i-1}, a_1, \dots, a_\Delta, \mathbf{0}_{P-\Delta-i}\}$ , pour  $i = 2, \dots, P-\Delta$ , et  $\mathbf{a}_i = \{\mathbf{0}_{i-1}, a_1, \dots, a_{P-i-1}, a_{P-i} + \dots + a_\Delta\}$ , pour  $i = P - \Delta + 1, \dots, P - 1$ . Par exemple, si  $\Delta = 4$  et  $C = 7$ , la matrice de transition s'écrit :

$$\mathbf{A} = \begin{pmatrix} a_{12} & a_{13} & a_{14} & a_{15} & 0 & 0 \\ 0 & a_1 & a_2 & a_3 & a_4 & 0 \\ 0 & 0 & a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & 0 & a_1 & a_2 & a_3 + a_4 \\ 0 & 0 & 0 & 0 & a_1 & a_2 + a_3 + a_4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.2)$$

En pratique, cette matrice permet de modéliser convenablement des signaux actifs sur tout le support temporel ou juste sur la fin du support temporel. Cependant, cette matrice est mal adaptée pour modéliser des signaux qui ne sont actifs qu'au début de la séquence (phénomènes d'apprentissage, ou inattention du patient à la fin du protocole). Par exemple, si le patient n'a répondu qu'aux 5 premiers blocs de stimulation, alors l'état caché passe directement de l'état  $evp$   $q_5$  à l'état fictif final  $q_7$ . A cette trajectoire, est associée la probabilité  $a_3 + a_4$ . Ainsi, la probabilité  $a_3 + a_4$  sera grande, rendant le modèle peu contraint : on autorise à tort et avec une probabilité importante des sauts entre l'état  $a_1$  et l'état  $a_4$  par exemple. La liaison de paramètres est à l'origine de ce problème. Néanmoins, nous avons trouvé plus judicieux



de mal modéliser certains types de signaux plutôt que d'avoir des problèmes importants au niveau de l'apprentissage.

### 6.3.2.2 Les lois de durée

En présence d'un retard d'activité locale  $\mu d_0$ , on suppose que la séquence des RDs doit idéalement s'aligner sur la séquence de transition du paradigme translatée de  $\mu d_0$ . Par conséquent, les moyennes :

- $\mu d_{1j}$  sont fixées à  $t'_j + \mu d_0 - 1$  ;
- $\mu d_{iP}$  sont fixées à  $T - t'_i + \mu d_0$  ;
- $\mu d_{ij}$  ( $i \neq 1$  et  $j \neq P$ ) sont fixées à  $\mu d_{ij} = t'_j - t'_i$ .

Pour faciliter l'apprentissage, on considère enfin que les densités de probabilité  $p_{ij}(\cdot)$  partagent la même variance  $\sigma d_0$  si  $i = 1$  ou  $j = P$ ,  $\sigma d_1$  sinon.

En définitive, le paradigme étant donné, il y a 3 paramètres à régler :  $\mu d_0$ ,  $\sigma d_0$  et  $\sigma d_1$ . La moyenne  $\mu d_0$  représente le retard d'activité moyen et  $\sigma d_0$  la variabilité associée. La variabilité des retards d'activité entre des *evps* non fictifs est représentée par l'écart-type :  $\sigma d_1$ . Un exemple est représenté sur la figure 6.5 avec  $\Delta = 2$  et  $C = 5$ .

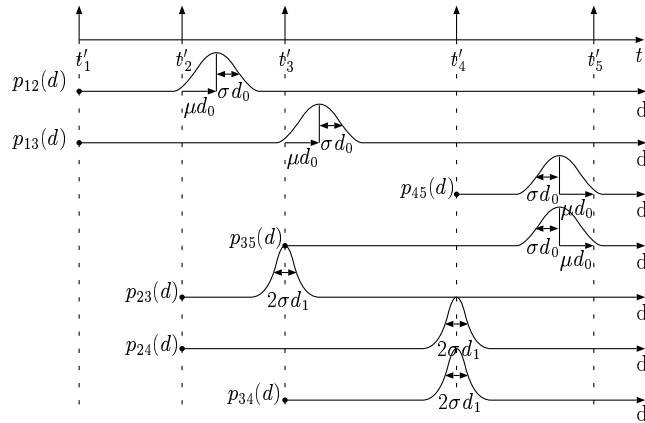


FIG. 6.5 – Lois de durée avec  $\Delta = 2$  et  $C = 5$ . En haut, les transitions du paradigme et en bas les différentes lois de durée.

### 6.3.2.3 Les lois d'observation

Une distinction statistique est uniquement effectuée entre les états *evp* et *efp* mais non à l'intérieur des classes. Ainsi, seulement 5 paramètres sont utilisés pour les lois d'observation, c'est-à-dire les moyennes  $\mu_{tpe}$  et  $\mu_{fpe}$ , les variances  $\sigma_{tpe}^2$  et  $\sigma_{fpe}^2$  et le taux de fausses alarmes  $e$ . Le taux de fausses alarmes est partagé entre toutes les lois d'observation dans la mesure où le nombre de fausses alarmes peut s'avérer très faible pour un signal actif. La liaison de paramètres rend ainsi son estimation plus fiable. Le fait d'utiliser uniquement les paramètres  $\mu_{tpe}$ ,  $\mu_{fpe}$ ,  $\sigma_{tpe}^2$  et  $\sigma_{fpe}^2$  se comprend enfin aisément eu égard à la méthode utilisée pour définir les mesures associées à un événement (voir section 6.2.2, page 91). Ces mesures ont été déterminées de manière à discriminer une bonne détection d'une fausse alarme et non de manière à différencier deux bonnes détections. En définitive, on a :

$$\begin{aligned} P[O_t | X_t = q_i] &= \mathcal{N}(O_t; \mu_{tpe}, \sigma_{tpe}) \\ P[O_t | X_t = q_{ij}] &= e \times \mathcal{N}(O_t; \mu_{fpe}, \sigma_{fpe}) \mathbb{1}_{\tau_E}(t) + (1 - e) \mathbb{1}_{\tau_{\bar{E}}}(t) \end{aligned} \quad (6.3)$$

## 6.4 Apprentissage et cartographie non supervisée de signaux d'activation cérébrale

### 6.4.1 Principe général

L'apprentissage non supervisé des paramètres du modèle et la cartographie sont effectuées de manière conjointe. La méthode se base sur les deux constats suivants : (i) si nous avons à notre disposition le modèle d'activité, une méthode de cartographie pourrait être utilisée pour déterminer la carte d'activation  $\mathcal{A}_i$  ou de manière équivalente l'ensemble  $\mathcal{F}_i$  des signaux déclarés actifs par le modèle d'activité. (ii) Si nous avons à notre disposition la carte d'activation, les signaux déclarés actifs pourraient servir de base d'apprentissage pour estimer les paramètres du modèle.

On cherche en définitive un modèle d'activité  $\lambda_i$  qui déclare actif un ensemble de signaux  $\mathcal{F}_i$  qui, une fois utilisé comme corpus d'apprentissage fournit comme modèle d'activité  $\lambda_i$ . La procédure consiste à construire itérativement et conjointement une suite de modèles d'activation,  $\{\lambda_1, \lambda_2, \dots\}$ , et une suite d'ensembles de signaux IRMf actifs,  $\{\mathcal{F}_1, \mathcal{F}_2, \dots\}$ , ou de manière équivalente, une série de carte d'activations  $\{\mathcal{A}_1, \mathcal{A}_2, \dots\}$ . Composés de tous les signaux déclarés actifs par  $\lambda_i$ ,  $\mathcal{F}_i$  est utilisé comme base d'apprentissage pour  $\lambda_{i+1}$ . Les trois séries convergent respectivement vers  $\lambda_{opt}$ ,  $\mathcal{F}_{opt}$ , et  $\mathcal{A}_{opt}$  dans le sens où l'apprentissage à partir de  $\mathcal{F}_{opt}$  fournit  $\lambda_{opt}$ , qui, à son tour, déclare actif l'ensemble  $\mathcal{F}_{opt}$ . A la fin de la procédure,  $\mathcal{A}_{opt}$  définit la carte d'activation finale alors que le modèle  $\lambda_{opt}$  peut être considéré comme statistiquement représentatif de la réponse neuronale induite par le protocole d'activation et de l'étape de prétraitement.

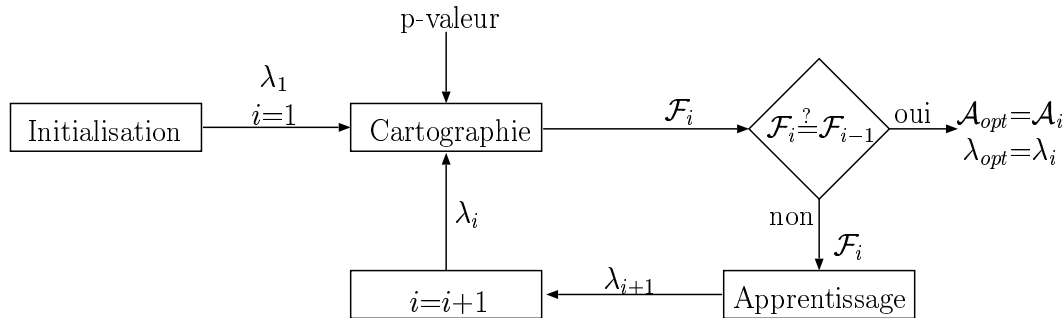


FIG. 6.6 – Détermination conjointe des paramètres du modèle et de la carte d'activité

L'algorithme de cartographie est résumé à la figure 6.6. Il est nécessaire d'éclaircir trois points :

- Comment obtenir un ensemble de signaux actifs à partir d'un modèle d'activité (problème de cartographie) ?
- Comment obtenir un modèle d'activité à partir d'un ensemble de signaux actifs (problème d'apprentissage) ?
- Comment définir le modèle initial  $\lambda_1$  ?

Nous allons répondre par la suite à ces trois questions.

### 6.4.2 Initialisation

Le modèle  $\lambda_1$  est choisi très peu contraint. En effet, il doit être potentiellement représentatif d'une très grande variété de signaux actifs. Si le modèle est trop contraint, l'algorithme risque de converger trop tôt et de ne pas être représentatif de la variabilité des signaux actifs présente dans les données. Les paramètres du modèle initial  $\lambda_1$  sont déterminés à partir de l'information temporelle du paradigme d'activation (temps d'occurrence des transitions) et à partir de statistiques réalisées sur l'ensemble  $\mathcal{O}$  de toutes les séquences d'observation. Les paramètres du modèle initial  $\lambda_1$  sont décrits ci-après :

- *matrice de transition* :
  - $a_{12} = a_{13} = \dots = a_{1\Delta+1} = \frac{1}{\Delta}$
  - $a_1 = a_2 = \dots = a_\Delta = \frac{1}{\Delta}$
- *Lois de durée* :
  - $\mu d_0 = 0$
  - $\sigma d_0 = \frac{1}{2} \cdot \inf_{(2 \leq p \leq P)} (t'_p - t'_{p-1})$ ,
  - $\sigma d_1 = 0.5$
- *Lois d'observation* : On note  $\mu$  et  $\sigma^2$  la moyenne et la variance des mesures associées aux événements  $\{e_l^v\}$  relatifs à l'ensemble d'observation  $\mathcal{O}$ . Soit  $\bar{L}$  le nombre moyen d'événements détectés par signal IRMf sur l'ensemble des observations ayant au moins  $P$  événements. On pose :
  - $\mu_{tpe} = \mu_{fpe} = \mu$
  - $\sigma_{tpe} = \sigma_{fpe} = \sigma$
  - $e = \frac{\bar{L}-P}{T-P}$

### 6.4.3 Cartographie

Il s'agit de déterminer les signaux actifs à partir d'un modèle d'activité  $\lambda$ . Pour chaque voxel  $v$ , la vraisemblance  $l^v = P[\mathcal{O}^v | \lambda]$  est déterminée. On obtient ainsi une carte de vraisemblance  $\mathcal{L} \equiv \{l^v\}$ . Cette carte est ensuite transformée en une carte de p-valeur  $\mathcal{P}_i \equiv \{p^v\}$  qu'il suffit de seuiller. On rappelle que la p-valeur associée à  $l^v$  est définie par  $p^v = \int_{l^v}^{+\infty} P[l | \mathcal{H}_0] dl$  où  $P[l | \mathcal{H}_0]$  représente la densité de probabilité de la vraisemblance  $l$  sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale. Le seuillage de la carte de p-valeur permet d'obtenir finalement la carte d'activité (pour une p-valeur donnée) ou de manière équivalente l'ensemble des signaux déclarés actifs par le modèle  $\lambda$ .

La principale difficulté réside dans le fait que la distribution  $P[l | \mathcal{H}_0]$  est inconnue. Aussi, cette distribution est inférée expérimentalement en générant aléatoirement des séquences d'événements représentatives de  $\mathcal{H}_0$  à partir de l'ensemble des séquences d'événements intracrâniennes. En d'autres termes, une séquence représentative de  $\mathcal{H}_0$  est obtenue par une procédure aléatoire qui mélange entre elles les différentes séquences d'événements intracrâniennes. Cette procédure est présentée à l'algorithme 6.1 et elle est illustrée figure 6.7.

Bien que cette procédure n'ait aucune justification biologique, elle sera validée par la suite d'un point de vue expérimental. Elle possède également des propriétés intéressantes : (i) elle conserve par construction un nombre important de propriétés « locales » (elle a une mémoire d'ordre 1 du point de vue des événements). Par exemple, si deux ruptures consécutives sont au minimum séparées de 4 échantillons dans l'ensemble de la base  $\mathcal{O}$ , alors, cette propriété sera conservée également pour la séquence d'observation créée. (ii) La probabilité pour que  $t$  appartienne à l'ensemble des instants émetteurs de la séquence créée est égale au nombre de séquences d'événements de  $\mathcal{O}$  qui ont  $t$  pour instant émetteur divisé par le nombre de

**Algorithme 6.1** Synthèse d'une séquence d'observation représentative de  $\mathcal{H}_O$  à partir de l'ensemble des séquences intracrâniennes  $\mathcal{O}$

\*\*\*Initialisation\*\*\*

-  $l = 1, t_l = 1$ , créer  $e_1$

\*\*\*Boucle\*\*\*

**Tant que**  $t_l \neq T$  **Faire**

- Déterminer l'ensemble des séquences d'observation  $S$  ( $S \subset \mathcal{O}$ ) telles qu'il existe  $m \in [1, L^v]$  tel que  $t_m^v = t_l$ .

- Choisir aléatoirement selon une loi uniforme une séquence parmi l'ensemble  $S$ . On note  $\mathcal{O}^x$  la séquence choisie et  $m$  l'indice pour lequel  $t_m^x = t_l$

-  $l = l + 1, t_l = t_{m+1}^x, e_l = \mathcal{O}_{t_{m+1}^x}^x$

**Fin Tant que**

-  $L = l$ .

séquences d'événements de  $\mathcal{O}$  (démonstration par récurrence sur le temps). (iii) L'espérance du nombre d'événements de la séquence créée correspond au nombre moyen d'événements contenu dans les séquences de  $\mathcal{O}$  (conséquence du point précédent).

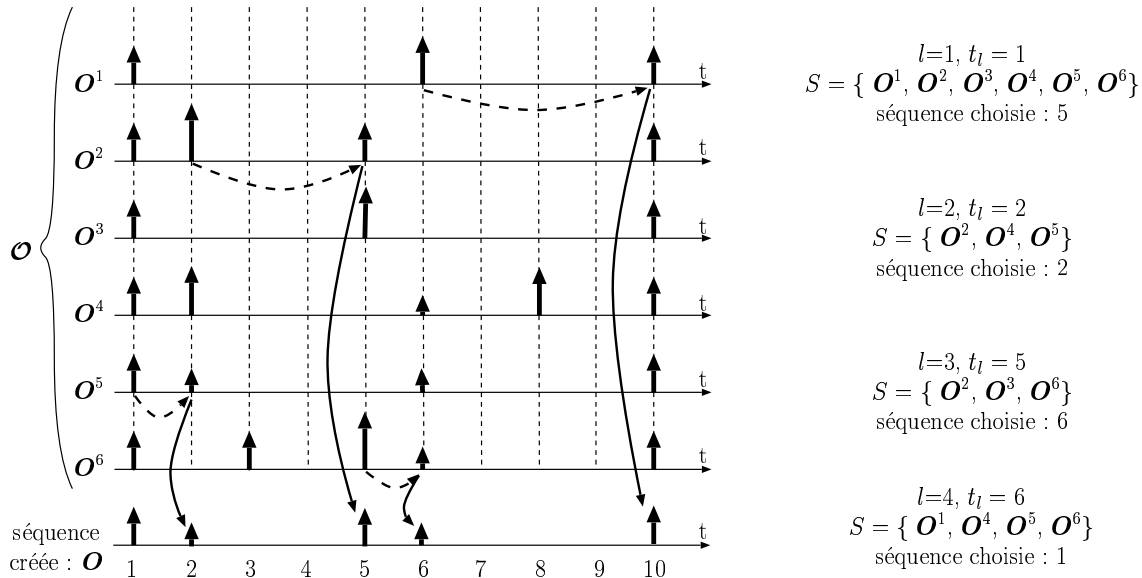


FIG. 6.7 – Illustration du principe de la procédure de mélange pour la création d'une séquence d'observation  $\mathcal{O}$  représentative de  $\mathcal{H}_O$  à partir de  $\mathcal{O}$ .

#### 6.4.4 Algorithme d'apprentissage

Etant donné un ensemble de signaux actifs, il est nécessaire de déterminer un modèle d'activité. Le critère du maximum de vraisemblance étant utilisé, il suffit d'utiliser l'algorithme de Baum-Welch décrit section 5.4.3, page 86. Il reste cependant à répondre à deux questions : quel modèle initial injecter dans l'algorithme de Baum-Welch ? Dans quelles mesures les formules de réestimation sont-elles affectées par la procédure de liaison des paramètres ?

Dans le cas où  $i > 1$  (voir figure 6.6), le modèle  $\lambda_i$  sert d'initialisation à l'algorithme de Baum-Welch permettant l'estimation de  $\lambda_{i+1}$ . Dans le cas où  $i = 1$ , le modèle  $\lambda_1$  sert

d'initialisation à une procédure d'apprentissage de Viterbi dont le résultat sert à initialiser convenablement l'algorithme de Baum-Welch.

Enfin, la liaison de paramètres modifie les formules de réestimation. L'équation (5.15) de la page 86 est toujours vraie. Les formules de réestimation des paramètres  $\{a_{1,i}\}_{2 \leq i \leq \Delta+1}$ ,  $\mu_{tpe}$ ,  $\sigma_{tpe}$ ,  $\mu_{fpe}$ ,  $\sigma_{fpe}$ ,  $\mu d_0$ ,  $\sigma d_0$  et  $\sigma d_1$  peuvent être obtenues simplement comme dans le cas standard. Cependant, il n'est pas possible de trouver une forme analytique pour la réestimation des paramètres  $\{a_i\}_{1 \leq i \leq \Delta}$ . Aussi, une méthode numérique d'optimisation est utilisée (gradient).

## 6.5 Décodage des signaux d'activation : vers de nouvelles fonctionnalités d'analyse

Le décodage de Viterbi permet d'obtenir la séquence d'états cachés optimale pour chaque voxel actif et donc en particulier la séquence d'états *evp*. Ainsi, on obtient une séquence de temps d'occurrence des *evps* :  $\tau_{tpe}^{v,MAP} = \tau_1^v \tau_2^v \dots \tau_{R^v}^v$  et la séquence des numéros des blocs de stimulation associés :  $i_{tpe}^{v,MAP} = i_{\tau_1^v} i_{\tau_2^v} \dots i_{\tau_{R^v}^v}$ . Cette information peut être exploitée avec profit pour fournir des informations sur la façon dont les régions actives ont répondu au stimulus.

### 6.5.1 Analyse des retards d'activation

A partir de  $\tau_{tpe}^{v,MAP}$  et  $i_{tpe}^{v,MAP}$ , il est possible d'estimer le retard d'activation pour le voxel  $v$ , c'est-à-dire la durée moyenne séparant une RD et la transition « off-on » avec laquelle elle est associée. On peut par exemple utiliser l'estimateur robuste suivant :

$$r = Med \left( \{ \tau_i^v - t'_{i_{\tau_i^v}} \}_{2 \leq i \leq R^v-1} \right) \quad (6.4)$$

### 6.5.2 Analyse des modes d'activation

Pour une région active, il est possible de compter le nombre de fois que chaque état *evp* a été visité dans les séquences d'état  $\mathbf{q}_{tpe}^{v,MAP}$  relatives à la région. On obtient ainsi le taux de réponse associé à chaque bloc.

Les modes d'activation permettent d'évaluer l'activité tout au long de l'expérience et donc d'étudier notamment les phénomènes liés à l'apprentissage. Plus généralement, les différents modes d'activation sont l'expression de la dynamique de recrutement des différents réseaux neurologiques impliqués dans la tâche.

Une illustration des modes d'activation est donnée dans le prochain chapitre, section 7.2.2.3, page 115.

### 6.5.3 Analyse de la réponse hémodynamique

Une analyse morphologique de la réponse à un bloc de stimulation peut enfin être effectuée. Cette analyse peut être réalisée à travers le temps, l'espace et les sujets. Par exemple, une réponse hémodynamique moyenne peut être estimée pour une région active et pour un bloc de stimulation donné  $H$ . Pour cela, on considère l'ensemble des signaux IRMf  $f^v(\cdot)$  relatifs à la région qui ont répondu au bloc  $H$  ( $i_{\tau^v} = H$ ). Chaque signal est alors recalé temporellement grâce à l'instant d'occurrence de l'événement d'intérêt associé au bloc  $H$ . Les signaux peuvent ensuite être sommés, fournissant ainsi la réponse moyenne au bloc.

Une illustration est donnée dans le prochain chapitre, section 7.2.2.4, page 115.



# Chapitre 7

## Résultats expérimentaux

Dans ce chapitre, nous évaluons les performances de détection d'activation de la méthode de cartographie par MSMCSE à la fois sur données IRMf synthétiques et sur données IRMf réelles. En contrôlant la « vérité terrain », l'évaluation sur données synthétiques permet de quantifier les performances de détection de la méthode MSMCSE et de les comparer à celles d'autres méthodes de cartographie. L'évaluation sur données réelles reste quant à elle incontournable pour valider l'approche en matière de détection d'activation, objectif majeur de la méthode proposée, mais également en matière d'information ajoutée quant à la caractérisation des phénomènes d'activation détectés. Parce que largement utilisée par la communauté IRMf, la méthode SPM (Statistical Parametric Mapping) est utilisée comme méthode de comparaison.

### 7.1 Evaluation sur données synthétiques

#### 7.1.1 Données

Six séquences synthétiques d'images IRMf 3-D, notées  $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$  ont été construites en additionnant, dans des zones définies comme actives, un signal d'activation synthétique au bruit. Dans chaque séquence, quatre cubes, de taille  $6 \times 6 \times 6$  voxels, ont été définis comme zones actives. Ces zones sont notées suivant l'ordre lexicographique  $Z_1$ ,  $Z_2$ ,  $Z_3$  et  $Z_4$ , respectivement (figure 7.1).

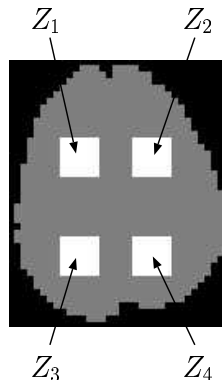


FIG. 7.1 – Masque d'activité définissant les zones actives ( $Z_1$ ,  $Z_2$ ,  $Z_3$  et  $Z_4$ ) et les zones inactives.

De façon plus formelle, pour chacune des six séquences considérées, si  $\mathbf{X}_n$  désigne le signal IRMf au voxel  $n$ , alors :

$$\mathbf{X}_n = \mathbf{X}_{1n} + \alpha_n \mathbf{X}_{2n} \quad (7.1)$$

Les signaux  $\mathbf{X}_n$ ,  $\mathbf{X}_{1n}$  et  $\mathbf{X}_{2n}$  sont de taille égale, à savoir  $T = 150$  :

- $\mathbf{X}_{1n}$  représente le bruit observé au voxel  $n$  dans la séquence  $\mathcal{F}_0$ . Cette séquence a été acquise avec un imageur IRM Bruker S200 2 Teslas (antenne de volume, EPI, orientation axiale, 32 coupes,  $64 \times 64$  pixels, volume voxel =  $4 \times 4 \times 4mm$ , TE=43ms, TR=5s). Pendant l'acquisition, le sujet n'a été soumis à aucun paradigme de stimulation. Il lui a été demandé au préalable de ne pas bouger et de rester le plus calme possible. Parce qu'acquise sous l'hypothèse d'inactivité cérébrale, la séquence  $\mathcal{F}_0$  peut être considérée comme constituée de bruit seul. Les images ont ensuite été recalées sur la première image de la séquence [169].
- $\mathbf{X}_{2n}$  représente le signal synthétique d'activation ajouté au bruit dans les zones définies comme actives (cas  $\alpha_n \neq 0$ ). Hors de ces zones, aucun signal d'activation n'est ajouté au bruit (cas  $\alpha_n = 0$ ). Le signal d'activation dépend de la séquence IRMf synthétique considérée, et de la zone active  $Z_i$ ,  $i = 1, 2, 3, 4$ , considérée dans cette séquence. Il résulte, à part pour la séquence  $\mathcal{F}_3$ , de la convolution d'une fonction déterministe  $x_i(\cdot)$  avec un modèle de réponse hémodynamique  $h_i(\cdot)$  :

$$\mathbf{X}_{2n}(t) = (x_i * h_i)(t) \quad (7.2)$$

La fonction  $x_i(\cdot)$  est soit égale à une fonction de base  $x_0(\cdot)$  (cas de  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$ ) soit obtenue par transformation temporelle et/ou morphologique de cette fonction (cas de  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$ , voir figure 7.2). La fonction  $x_0(\cdot)$  est représentative du paradigme d'activation. Ce dernier est de type bloc, périodique, et formé de 8 périodes de repos ( $x_0(t) = 0$ ) alternant avec 7 périodes d'activation ( $x_0(t) = 1$ ), de 10 échantillons chacune. La réponse hémodynamique  $h_i(\cdot)$  utilisée dans la plupart des cas est celle proposée par le logiciel SPM (équation (2.12) de la page 36), notée  $h_{SPM}(\cdot)$ , à l'exception de la séquence  $\mathcal{F}_3$  où le bloc de stimulation est directement remplacé par la réponse à un bloc, la réponse à un bloc ayant été estimée sur données réelles avec la méthode présentée précédemment (figure 7.2).

- $\alpha_n$  représente un coefficient d'activité égal à 0 en dehors des zones actives. Pour chaque signal actif, le coefficient  $\alpha_n$  est calculé à partir de la spécification du rapport signal sur bruit  $m$  de l'expérience, défini comme le rapport de l'amplitude de la réponse sur l'écart-type du bruit [170]. Ce dernier est estimé sur le bruit auquel on a au préalable soustrait la dérive de ligne de base.

Suivant ce principe, les caractéristiques de chaque séquence synthétique sont fonctions du jeu de paramètres  $\{m, (x_i.t), h_i(\cdot)\}_{1 \leq i \leq 4}$  utilisés pour la créer. Ainsi :

- $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$  illustrent un cas d'application idéal où ce qui observé comme signal d'activation est conforme en tout point à ce qui est attendu. Pour cela, les fonctions d'excitation  $x_i(\cdot)$  sont toutes égales à la fonction de base  $x_0(\cdot)$ . De même, les réponses hémodynamiques  $h_i(\cdot)$  sont toutes égales au modèle de réponse  $h_{SPM}(\cdot)$ . Seul le rapport signal à bruit  $m$  varie, avec  $m = 2$  pour  $\mathcal{F}_{1a}$  et  $m = 1$  pour  $\mathcal{F}_{1b}$ <sup>1</sup>. La figure 7.3.b représente deux signaux actifs (pour le même voxel) des séquences  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$ .

<sup>1</sup>D'après notre expérience et pour comparaison sur signaux réels, le rapport signal sur bruit  $m$  s'étend pour des paradigmes en bloc entre  $m=2$  et  $m=1$  selon la nature du processus cognitif, les processus moteurs étant beaucoup plus simples à détecter. Pour les paradigmes événementiels, une valeur réaliste s'étend entre  $m=1$  et  $m=0.25$ .



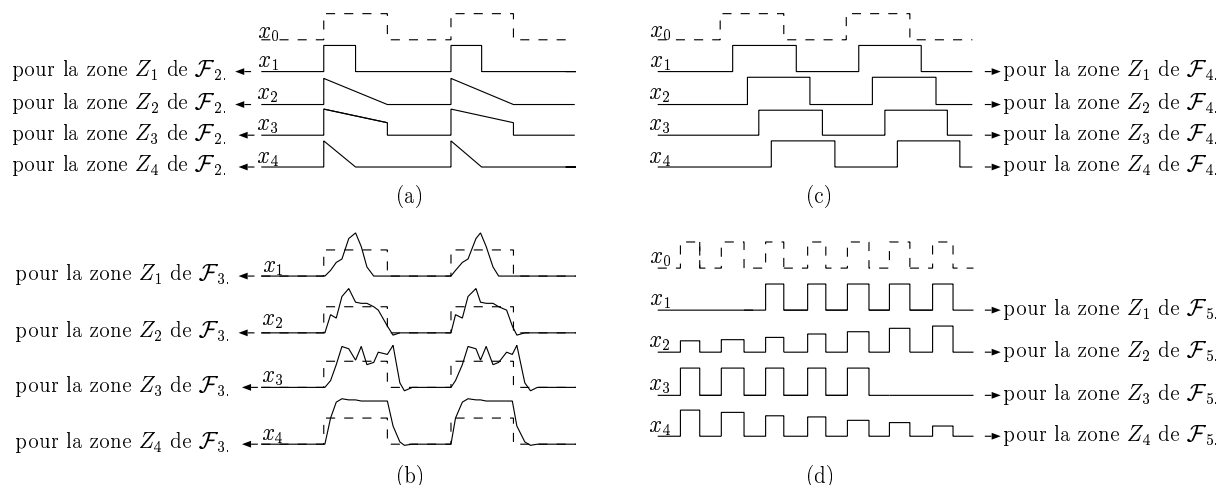


FIG. 7.2 – Construction des signaux synthétiques (voir texte).

- $\mathcal{F}_{2a}$  illustre différents cas de non maintien de la réponse hémodynamique dans un bloc de stimulation. Pour cela, les quatre fonctions d'excitation  $x_i(\cdot)$ ,  $i = 1, \dots, 4$  de la figure 7.2.a sont utilisées. Par ailleurs,  $h_i(\cdot) = h_{SPM}(\cdot)$  pour  $i = 1, \dots, 4$ , et  $m = 2$  pour les quatre zones actives de la séquence.
- $\mathcal{F}_{3a}$  est représentative de différents cas de variabilité de forme de la réponse hémodynamique à un bloc de stimulation. Ainsi, quatre formes de réponse à un bloc de stimulation ont été extraites à partir de données réelles et sont directement « superposées » à la fonction d'excitation  $x_0(\cdot)$  (figure 7.2.b). Le rapport signal à bruit  $m$  a été fixé à 2 pour les quatre zones actives de la séquence.
- $\mathcal{F}_{4a}$  a été conçue pour illustrer différents cas de retard d'activation sur un bloc de stimulation. Les fonctions d'excitation  $x_1(\cdot)$ ,  $x_2(\cdot)$ ,  $x_3(\cdot)$  et  $x_4(\cdot)$  sont des versions retardées (retard de 2, 4, 6 et 8 échantillons respectivement, figure 7.2.c) de la fonction de base  $x_0(\cdot)$  représentative du paradigme d'activation. Par ailleurs,  $h_i(\cdot) = h_{SPM}(\cdot)$  pour  $i = 1, \dots, 4$ , et  $m = 2$  pour les quatre zones actives de la séquence.
- $\mathcal{F}_{5a}$  illustre enfin différents cas de phénomènes d'habituation et d'apprentissage, avec possibilités d'absence de réponse en fin ou en début de stimulation, ou de baisse d'amplitude de la réponse hémodynamique en un ou plusieurs blocs. Pour cela, les quatre fonctions  $x_i(\cdot)$  ont été obtenues par transformation de la fonction  $x_0(\cdot)$  comme indiqué figure 7.2.d. Enfin,  $h_i(\cdot) = h_{SPM}(\cdot)$  pour  $i = 1, \dots, 4$ , et  $m = 2$  pour les quatre zones actives de la séquence.

Pour terminer, soulignons qu'un filtrage spatial gaussien isotrope (largeur à mi-hauteur de 8 mm) et un filtrage spatial temporel ( $\sigma = \sqrt{8}$  secondes) ont été appliqués sur chaque séquence synthétique avant cartographie.

## 7.1.2 Méthodologie d'évaluation

### 7.1.2.1 Courbes COR, $\alpha$ et $\beta$

L'algorithme de cartographie, après convergence, fournit une carte de p-valeurs que nous ne seuillons pas. Une courbe COR (caractéristiques opérationnelles des récepteurs) est calculée à partir de la carte de p-valeurs et du masque d'activation. En raison du filtrage spatial, une bande large de 2 voxels entourant les zones actives n'est pas considérée lors de la construction

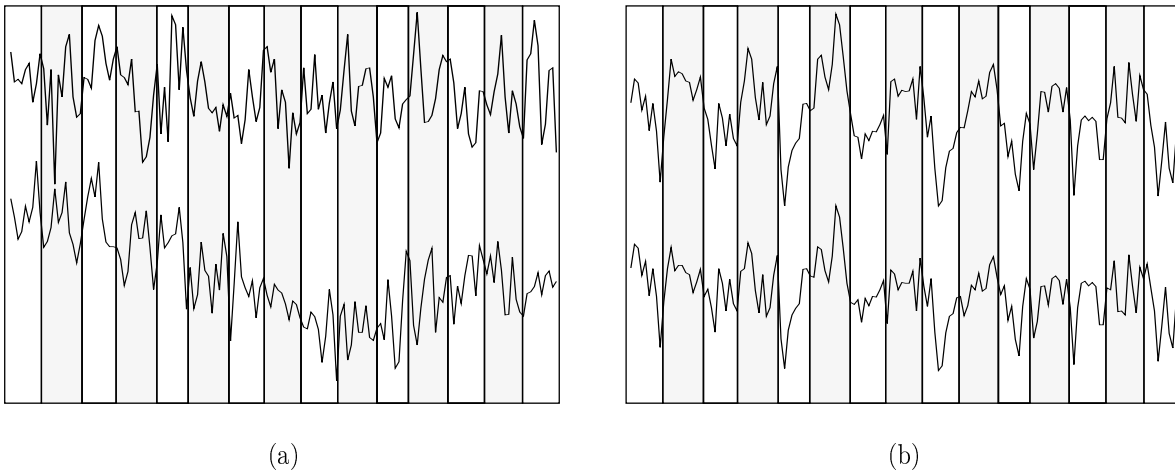


FIG. 7.3 – (a) exemple de signaux inactifs (b) exemple de signaux actifs de l'ensemble  $\mathcal{F}_{1a}$  (haut) et  $\mathcal{F}_{1b}$  (bas) pour le même voxel. Les zones grisées correspondent à des échantillons associés à l'état « on » alors que les autres sont associés à l'état « off » du paradigme.

des courbes COR (cf. figure 7.4).

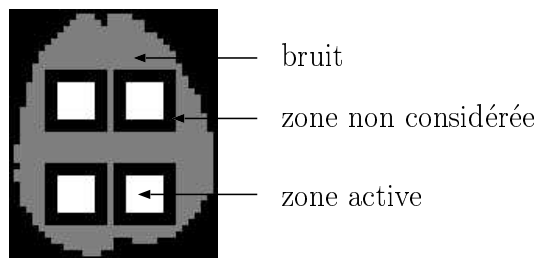
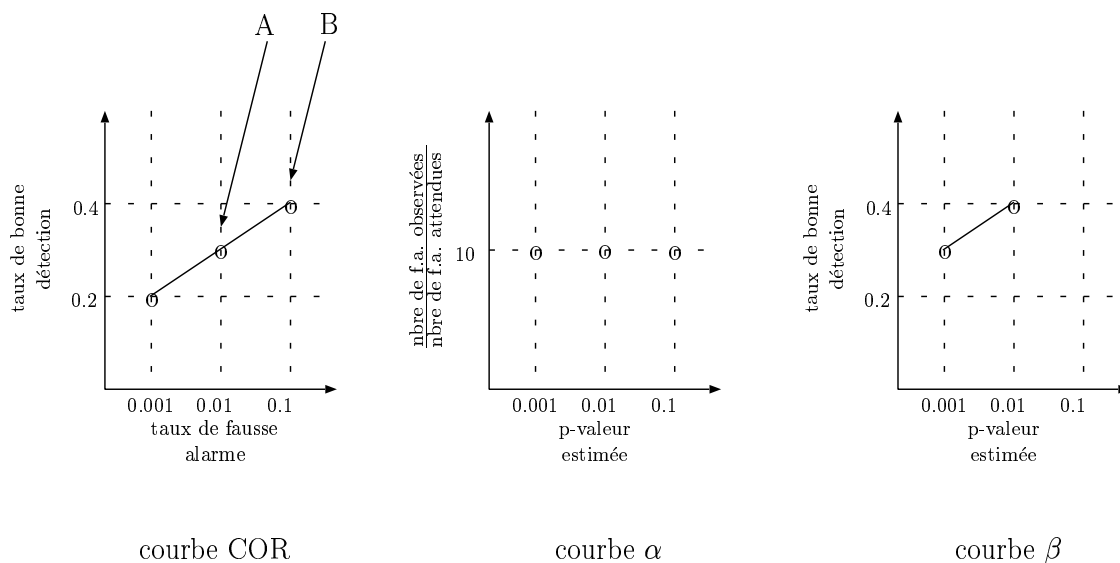


FIG. 7.4 – Masque utilisé pour l'analyse des résultats.

La courbe COR n'est pas suffisante pour caractériser la qualité des résultats. En effet, la courbe COR se base uniquement sur le classement des différents voxels les uns par rapport aux autres. Par exemple, dans le cadre de notre application, le calcul de la courbe COR à partir de la carte de vraisemblance permet d'obtenir les mêmes résultats que son calcul à partir de la carte de p-valeur puisque quand la vraisemblance augmente, la p-valeur diminue. La courbe COR crée un lien entre la probabilité de détection et la probabilité de fausse alarme (PFA). Cependant, lorsqu'on utilise l'algorithme dans des conditions réelles, le médecin ne seuille pas la carte à partir de la PFA (qu'il ne connaît pas) mais à partir d'une estimation de cette dernière : la p-valeur. Aussi, des problèmes importants apparaissent dès lors que la p-valeur est mal estimée.

Considérons en effet l'exemple de la figure 7.5. On obtient un taux de bonne détection de 0,3 pour une PFA de 0.01 (courbe COR). Lorsque l'on seuille ensuite la carte avec une p-valeur de 0.01, on s'attend, si les p-valeurs estimées sont correctes, à obtenir également une PFA de 0.01 (on rappelle que pour une p-valeur  $p$  et un nombre  $N$  de signaux inactifs, on s'attend à observer  $Np$  fausses alarmes). Ainsi, on espère se placer sur le point A de la courbe COR de la figure 7.5. Cependant, si la p-valeur associée aux différents voxels est mal estimée, il peut apparaître des problèmes de localisation sur la courbe COR. Il est donc

FIG. 7.5 – Définition des courbes COR,  $\alpha$  et  $\beta$ .

nécessaire d'effectuer un lien entre la PFA et la p-valeur. On introduit ainsi la courbe  $\alpha$  qui à une p-valeur fixée, associe le nombre de fausses alarmes observées divisé par le nombre de fausses alarmes attendues pour cette p-valeur. Dans notre exemple, nous voyons que pour une p-valeur  $p$  de 0.01, le nombre de fausses alarmes observées divisé par le nombre de fausses alarmes attendues est de 10 (bien que ce rapport puisse apparaître important, il reste réaliste dans le cadre de notre application). Ainsi, en seuillant avec une p-valeur de 0.01, on obtient une PFA de  $10 \times 0.01 = 0.1$ . On se place ainsi sur le point B de la courbe COR et non sur le point A ! La courbe  $\alpha$  permet ainsi de regarder avec quelle précision il est possible de se positionner sur la courbe COR.

La courbe  $\alpha$  et la courbe COR permettent en définitive de caractériser la qualité des résultats. Cependant, nous allons tracer également une autre courbe. Cette dernière peut être déterminée de manière déterministe à partir des courbes  $\alpha$  et COR. Il s'agit de tracer le pourcentage de détection obtenu lorsque l'on seuille en fonction de la p-valeur estimée (on appellera par la suite, cette courbe la courbe  $\beta$ ). Même si cette dernière apporte une information redondante par rapport aux deux autres courbes, la courbe  $\beta$  montre le résultat obtenu lorsque les tests sont effectués à l'aveugle et qu'il n'est pas possible de faire le lien entre la PFA et la p-valeur. Par exemple, en seuillant avec une p-valeur estimée de 0.01, on a vu que l'on se plaçait sur le point B de la courbe COR auquel est associé le taux de détection de 0.4. Ainsi, la courbe  $\beta$  de la figure 7.5 passe par le point de coordonnée (0.01 ; 0.4).

En définitive, les résultats sur données synthétiques sont caractérisés par la courbe COR et les courbes  $\alpha$  et  $\beta$ .

### 7.1.2.2 Méthode de comparaison utilisée : SPM

Comme mentionné dans l'introduction du chapitre, les résultats de cartographie de la méthode MSMCSE sont comparés à ceux obtenus par la méthode SPM, version SPM2. Pour une évaluation plus juste des résultats, deux implémentations de SPM ont été mises en œuvre.

L'implémentation SPMA correspond à l'utilisation courante du logiciel alors que SPMb correspond à un cas idéal où l'on injecte dans la matrice de dessin les composantes d'activation

des différents signaux actifs. Dans le cas des séries  $\mathcal{F}_1$ , le cas SPMa et SPMb sont identiques car la composante d'activation injectée dans les données correspond exactement à la composante d'activation donnée par défaut par le logiciel SPM.

Dans la première implémentation SPMa, on utilise un unique régresseur d'intérêt, à savoir le modèle de réponse hémodynamique proposé par ce logiciel convolué avec le paradigme d'activation où plus précisément le signal  $x_0(\cdot)$ .

La seconde implémentation, SPMb, correspond au cas idéal où la vérité terrain est connue et exploitée. Pour l'analyse des séquences  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$ , il est nécessaire pour chaque série d'utiliser quatre régresseurs correspondants aux composantes d'activation utilisées pour la création des séquences. Dans ce cas, le signal IRMf  $\mathbf{X}_n$  est supposé être la somme pondérée de quatre signaux d'intérêt :  $\mathbf{s}_1$ ,  $\mathbf{s}_2$ ,  $\mathbf{s}_3$  et  $\mathbf{s}_4$  et d'une composante constante représentative des perturbations :  $\boldsymbol{\mu}$  (voir équation (7.3)). Plus précisément, les signaux  $\mathbf{s}_1$ ,  $\mathbf{s}_2$ ,  $\mathbf{s}_3$  et  $\mathbf{s}_4$  correspondent au signal  $\mathbf{X}_{2n}$  de l'équation (13.1) pour les quatre zones distinctes du masque d'activité.

$$\mathbf{X}_n = b_n(1)\mathbf{s}_1 + b_n(2)\mathbf{s}_2 + b_n(3)\mathbf{s}_3 + b_n(4)\mathbf{s}_4 + b_n(5)\boldsymbol{\mu} + \boldsymbol{\epsilon} \quad (7.3)$$

En cas d'activité, on s'attend à obtenir un unique terme  $b_n(i)$  ( $i = 1 \dots 4$ ) différent de 0. Il est alors nécessaire d'effectuer 4 tests distincts fournissant finalement 4 cartes d'activité distinctes. Dans le but d'obtenir une unique carte d'activité, nous avons effectué un t-test sur la valeur  $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$  et un f-test sur la nullité des coefficients  $\alpha_1, \alpha_2, \alpha_3$  et  $\alpha_4$ . Il s'est avéré que les résultats sont meilleurs avec un t-test. Une raison pouvant expliquer ce fait est que l'on perd l'information de signe avec le f-test. Aussi, par la suite, nous ne présentons que les résultats obtenus avec le t-test.

## 7.1.3 Résultats

### 7.1.3.1 Procédure d'apprentissage-cartographie

L'algorithme de cartographie est un algorithme itératif. Aussi est-il important de s'intéresser à ses propriétés de convergence. D'un point de vue théorique, sous l'hypothèse que le modèle appris lors de la passe  $i$  (voir figure 6.6 de la page 96) ne dépende pas du point initial donné à l'algorithme de Baum-Welch, on démontre facilement que l'algorithme converge ou qu'il oscille entre  $n_0$  différentes cartes d'activation. Par exemple avec  $n_0 = 2$ ,  $\boldsymbol{\lambda}_{opt1}$  fournit la carte d'activité  $\mathcal{A}_{opt1}$  alors que l'apprentissage sur les voxels actifs de  $\mathcal{A}_{opt1}$  fournit le modèle  $\boldsymbol{\lambda}_{opt2}$ . Le modèle  $\boldsymbol{\lambda}_{opt2}$  permet d'obtenir la carte d'activité  $\mathcal{A}_{opt2}$ , qui par apprentissage permet d'obtenir le modèle  $\boldsymbol{\lambda}_{opt1}$ . En pratique, il est très rare d'observer un cas où  $n_0 > 1$ . De plus, dans un tel cas, les différentes cartes d'activité obtenues ne diffèrent que de quelques voxels. L'algorithme converge dans la plupart des cas et le nombre de passes pour atteindre la convergence est relativement faible (de l'ordre d'une dizaine).

Les figures 7.6.a et b illustrent un exemple de convergence. La figure 7.6.a représente la log-vraisemblance moyenne déterminée sur les différentes bases d'apprentissage pour chaque itération de l'algorithme EM. La log-vraisemblance moyenne doit théoriquement augmenter à l'intérieur de chaque passe. Il se peut cependant qu'elle diminue légèrement en raison de contraintes imposées sur certains coefficients (non nullité des écarts-types par exemple). Dans l'exemple, huit passes sont nécessaires à la convergence de l'algorithme. La figure 7.6.b montre la composition de chacune des bases d'apprentissage en termes de bonnes détections et de fausses alarmes. On voit très nettement l'intérêt de la procédure d'apprentissage dans la mesure où les résultats obtenus après convergence sont meilleurs que ceux obtenus avec la base  $\mathcal{F}_1$  (environ 95 % de bonnes détections contre environ 80 %). Le nombre de fausses alarmes est

enfin relativement cohérent par rapport au nombre attendu. On voit par la suite que la p-valeur est relativement fiable comparativement à celle obtenue avec SPM.

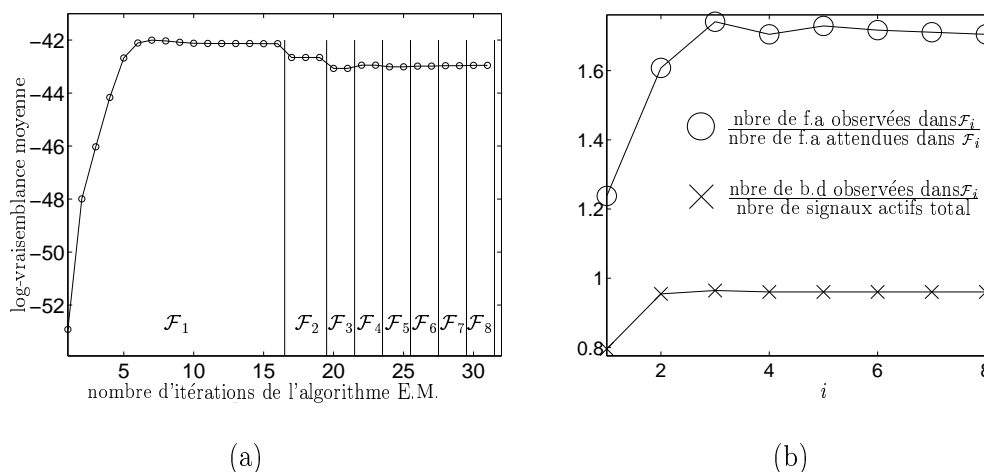


FIG. 7.6 – Convergence de la procédure d'apprentissage-cartographie sur données synthétiques. (a) log-vraisemblance moyenne déterminée sur les différentes bases d'apprentissage pour chaque itération de l'algorithme EM. (b) Composition des différentes bases d'apprentissage  $\mathcal{F}_i$  au cours des différentes passes  $i$  de l'algorithme.

### 7.1.3.2 Cartes d'activité

**7.1.3.2.1 Influence du bruit** Nous commençons par présenter les résultats obtenus sur les séquences  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$ . Nous rappelons que les fonctions d'excitation  $x_i(\cdot)$  sont toutes égales à la fonction de base  $x_0(\cdot)$ . De même, les réponses hémodynamiques  $h_i(\cdot)$  sont toutes égales au modèle de réponse  $h_{SPM}(\cdot)$ . Seul le rapport signal à bruit  $m$  varie, avec  $m = 2$  pour  $\mathcal{F}_{1a}$  et  $m = 1$  pour  $\mathcal{F}_{1b}$ . Nous rappelons enfin que dans le cas des séries  $\mathcal{F}_1$ , le cas SPMa et SPMb sont identiques car la composante d'activation injectée dans les données correspond exactement à la composante d'activation insérée par défaut dans la matrice de dessin par le logiciel SPM. Les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$  sont représentés sur les figures 7.7 et 7.8. Les cartes obtenues avec la méthode proposée sont présentées sur la première colonne. Celles obtenues avec SPMa sont représentées sur la seconde colonne. Les cartes de la figure 7.7 sont seuillées pour une PFA de 0.005 alors que les cartes de la figure 7.8 sont seuillées pour une p-valeur estimée de 0.005. Pour toutes les cartes, le blanc représente une bonne détection, le gris clair une fausse alarme ou une non détection. Le gris foncé représente les signaux non actifs légitimement non détectés. Enfin, le noir représente soit l'extérieur de la coupe, soit des détections dans la zone qui entoure les zones actives. La troisième colonne de la figure 7.7 correspond aux courbes COR alors que les troisièmes et quatrièmes colonnes de la figure 7.8 correspondent respectivement aux courbes  $\beta$  et  $\alpha$ .

Les résultats obtenus avec la séquence  $\mathcal{F}_{1a}$  (première ligne de la figures 7.7) valident la stratégie d'ensemble visant à analyser le signal IRMf à travers ses non-stationnarités et plus précisément le principe de couplage entre la séquence de RDs et la séquence de transitions « off-on » du paradigme. Cependant, les résultats se dégradent très rapidement dès lors que le rapport signal sur bruit diminue. Pour s'en convaincre, il suffit de regarder les résultats

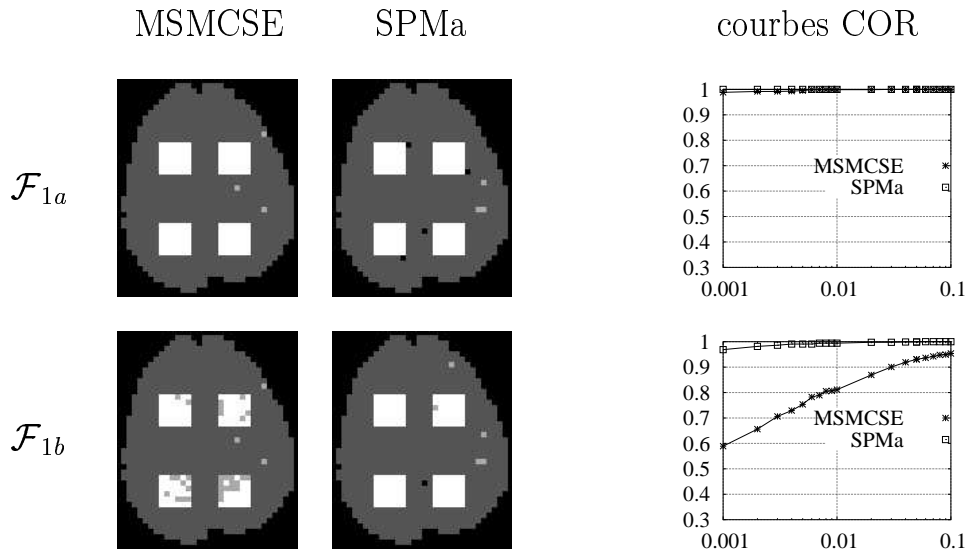


FIG. 7.7 – Résultats sur données synthétiques (voir texte)

obtenus avec la séquence  $\mathcal{F}_{1b}$  (seconde ligne des images 7.7). Nous tentons par la suite d'en cerner les causes et d'y apporter une solution.

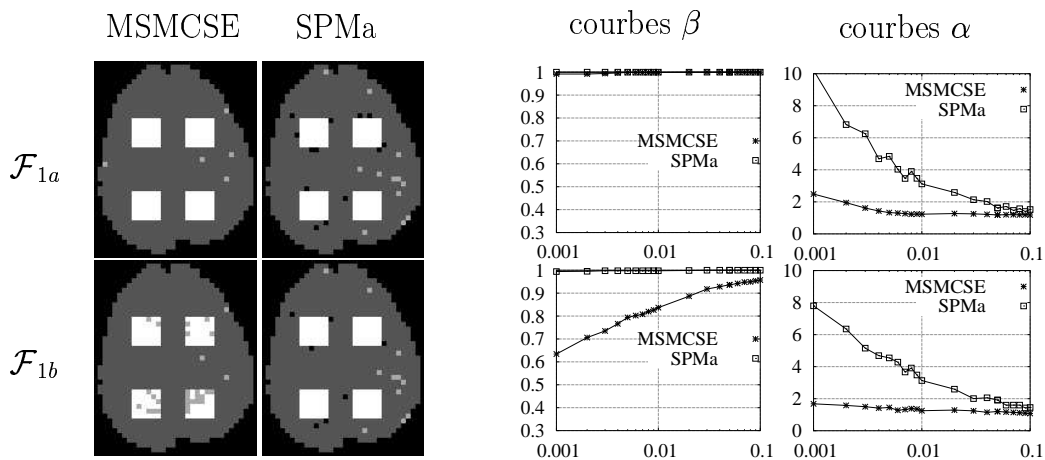


FIG. 7.8 – Résultats sur données synthétiques (voir texte)

A partir des résultats de la figure 7.8, on peut tout d'abord noter que la méthode proposée permet d'estimer plus fidèlement les p-valeurs que SPM (cf. courbes  $\alpha$  de la figure 7.8). Ceci valide la méthode proposée pour estimer la densité de probabilité de la vraisemblance sous l'hypothèse  $\mathcal{H}_0$  d'inactivité, densité de probabilité qui est utilisée pour associer une p-valeur à une vraisemblance. Au contraire, la p-valeur obtenue avec SPM est très biaisée en raison d'hypothèses non fondées comme le caractère gaussien du bruit. Les imagerie de la figure 7.8 correspondent à un cas d'application réel où l'on seuille pour une p-valeur estimée de 0.005. Il est à noter que les résultats obtenus avec notre méthode surpassent SPM quand le bruit est peu important (cas  $\mathcal{F}_{1a}$ ) : les zones actives sont bien détectées pour les deux méthodes. Cependant, SPMa fournit un nombre beaucoup plus important de fausses alarmes.

**7.1.3.2 Variabilité du signal actif** Nous rappelons (cf. figure 7.2 de la page 103) que  $\mathcal{F}_{2a}$  illustre différents cas de non maintien de la réponse hémodynamique dans un bloc de stimulation, que  $\mathcal{F}_{3a}$  est représentative de différents cas de variabilité de forme de la réponse hémodynamique à un bloc de stimulation, que  $\mathcal{F}_{4a}$  a été conçue pour illustrer différents cas de retard d'activation sur un bloc de stimulation et enfin que  $\mathcal{F}_{5a}$  illustre différents cas de phénomènes d'habituation et d'apprentissage. Nous rappelons enfin que l'implémentation SPMa correspond à l'utilisation courante du logiciel alors que SPMb correspond à un cas idéal où l'on injecte dans la matrice de dessin les composantes d'activation des différents signaux actifs.

Les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$  sont représentés sur les figures 7.9 et 7.10 avec la même méthode qu'utilisée précédemment à l'exception qu'une nouvelle colonne est utilisée pour insérer les imageries obtenues avec SPMb.

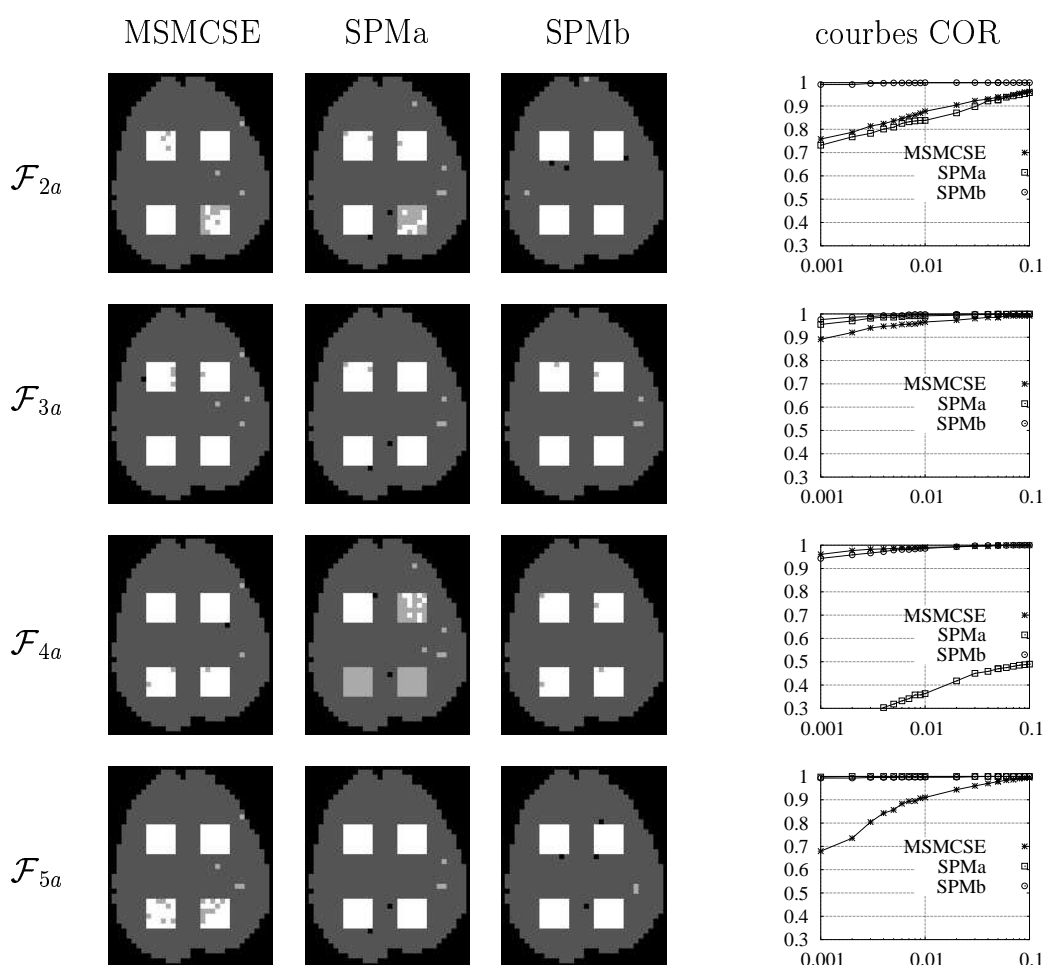


FIG. 7.9 – Résultats sur données synthétiques (voir texte)

Les résultats obtenus en termes de courbe COR sur la séquence  $\mathcal{F}_{2a}$  (influence du maintien de la réponse) montrent que les résultats obtenus avec SPMb sont très bons alors que la méthode proposée est approximativement au même niveau que SPMa. Les très bons résultats de SPMb ne sont pas surprenants dans la mesure où SPMb correspond à un cas idéal où l'on injecte dans la matrice de dessin les composantes d'activation des différents signaux actifs. Cependant, dès lors que la composante d'activation incorporée dans les données est différente de

celle injectée dans la matrice de dessin, les résultats se dégradent (cas SPMa). Ceci est relativement visible pour la zone  $Z_4$ . La méthode proposée voit également ses résultats se dégrader et spécialement pour la zone  $Z_4$ . En pratique, ceci est dû au fait que les RDs des signaux IRMf correspondants sont difficiles à détecter. Nous avons en effet observé que dès lors que les résultats de l'étape de prétraitement se dégradent, par exemple en raison de la diminution du rapport signal sur bruit (cas précédent), d'une activité non maintenue pendant le bloc de stimulation (cas présent), d'une réponse hémodynamique moins symétrique (prochain cas), nous observons une dégradation des résultats. Il convient par la suite d'analyser les causes de cette grande sensibilité à l'étape de prétraitement et d'apporter des solutions.

Les résultats obtenus sur la séquence  $\mathcal{F}_{3a}$  (variabilité de la réponse hémodynamique, seconde ligne) montrent la robustesse des différentes méthodes à la variabilité de la réponse hémodynamique. Ainsi, contrairement à ce qui peut être dit dans la littérature, la méthode SPM est très robuste à la variabilité de la réponse hémodynamique. L'implémentation SPMa donne en effet de très bons résultats même si un écart très important existe entre le signal d'activation injecté dans les données et celui inséré dans la matrice de dessin.

Les résultats obtenus sur la séquence  $\mathcal{F}_{4a}$  (introduction de retard d'activation, troisième ligne) montrent tout le bien fondé de la méthode proposée. Les très bons résultats de SPMb ne sont pas surprenants dans la mesure où SPMb correspond à un cas idéal de l'utilisation du logiciel SPM. Cependant, SPMa donne des résultats très mauvais, les zones avec un retard important n'étant pas détectées. Au contraire, notre méthode détecte automatiquement et de manière non supervisée l'ensemble des signaux actifs. La méthode proposée se révèle ainsi très robuste au retard d'activité contrairement à SPMa.

Les résultats obtenus sur la séquence  $\mathcal{F}_{5a}$  (introduction de variabilité liée à des phénomènes d'apprentissage, quatrième ligne) montrent qu'à la fois SPMa et SPMb donnent des résultats très satisfaisants. Au contraire, la faiblesse de notre modélisation concernant la matrice de transition apparaît ici clairement. La méthode est bien adaptée pour modéliser une activité qui augmente au cours du temps (les zones  $Z_1$  et  $Z_2$  sont bien détectées) alors qu'elle est mal adaptée à modéliser une activité qui diminue au cours du temps (des non-détections apparaissent au niveau des zones  $Z_3$  et  $Z_4$ ). Il convient par la suite d'apporter une solution à ce problème.

La figure 7.10 met encore une fois en évidence le fait que les p-valeurs sont relativement bien estimées avec la méthode MSMCSE contrairement à SPM (aussi bien SPMa que SPMb). Les imagerie de la figure 7.10 correspondent à un cas d'application réel où l'on seuil pour une p-valeur estimée de 0.005. A la vue des imagerie, les résultats obtenus avec notre méthode surpassent à la fois SPMa et SPMb pour la séquence  $\mathcal{F}_{4a}$  (introduction de retard d'activation, troisième ligne de la figure 7.10). L'implémentation SPMa ne détecte pas toutes les zones d'activité alors que SPMb obtient un très grand nombre de fausses alarmes.

## 7.2 Evaluation sur données réelles

### 7.2.1 Données

Dans le cadre de notre collaboration avec l'Institut de Physique Biologique, nous disposons d'une base de 119 séquences d'images IRMf 3-D acquises à partir de 17 sujets sains, âgés de 18 à 39 ans, à qui il a été demandé d'accomplir des tâches de traitement linguistique suivant trois paradigmes en blocs de type contrôle-activation. Les images des patients ont été acquises avec un imageur S200 Bruker 2 Teslas (antenne de volume, EPI, orientation axiale, 32 coupes,



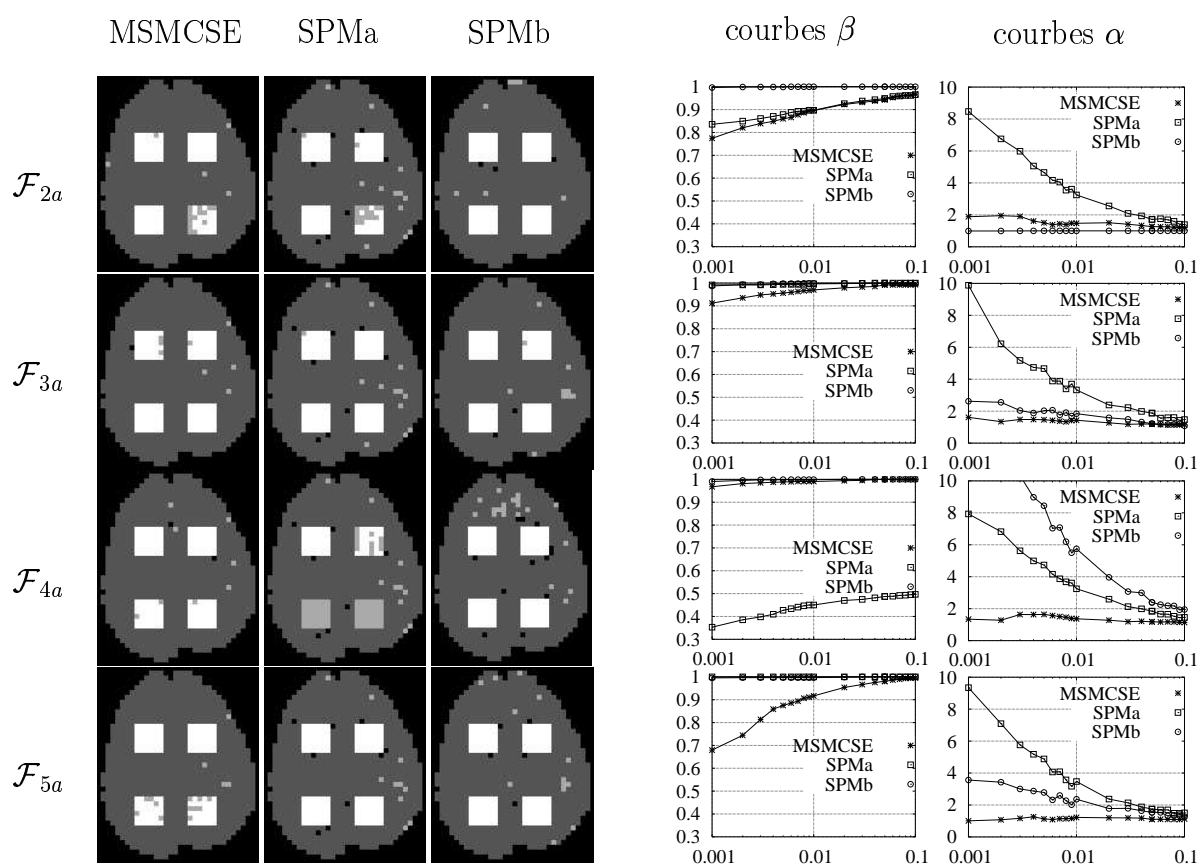


FIG. 7.10 – Résultats sur données synthétiques (voir texte)

64 x 64 pixels, volume voxel = 4 x 4 x 4 mm, TE=43ms, TR=5s). Les images ont toutes été recalées sur la première image de la série [169], la série étant composée de 145 images. Un filtrage spatial gaussien isotrope (largeur à mi-hauteur de 8 mm) et un filtrage temporel ( $\sigma=\sqrt{8}$  secondes) sont ensuite appliqués.

Trois paradigmes ont été conçus pour déterminer les zones corticales impliquées dans la recherche de mots et dans le traitement lexicologique auditif et visuel [160]. Les stimuli verbaux auditifs et visuels sont des noms communs du dictionnaire français, sélectionnés pour leur caractère concret et leur imagerie associée. Les sujets écoutent les stimuli à l'aide d'écouteurs de haute qualité. Les stimuli visuels ainsi que les instructions liées aux tâches sont présentés au travers de lunettes spécifiques, de marque Avotec.

Le premier paradigme, nommé VERBE, vise à identifier les aires corticales impliquées dans le traitement lexical, précisément dans le processus de sélection syntaxique/sémantique d'un mot. Le paradigme consiste en l'alternance d'une tâche de contrôle et d'une tâche d'activation pendant lesquelles des mots sont présentés au sujet sous forme auditive à la cadence de 40 mots par minute. Dans le même temps, il est demandé au sujet de suivre les instructions projetées visuellement. Durant la tâche de contrôle, il est demandé au sujet de répéter mentalement le mot qu'il entend. Durant la tâche d'activation, il lui est demandé de penser à un verbe, sans le vocaliser, en relation avec le nom entendu.

Le second paradigme, nommé AUDIO, vise à évaluer le processus de traitement auditif de mots. Il est demandé au sujet d'écouter des occurrences de mots sans accomplir de tâche active, si ce n'est de prêter un maximum d'attention aux mots concrets qui lui sont présentés, de manière à les reconnaître dans une liste, écrite, de 80 mots. En période d'activation, le sujet écoute une série de noms concrets à la cadence d'un nom par seconde. En période de contrôle, le sujet écoute passivement ces noms prononcés à l'envers, à la même cadence.

Le troisième et dernier paradigme, nommé VISUEL, consiste en une tâche de décision visuelle lexicale. Des chaînes de 4 à 7 caractères, d'une durée de 250 ms chacune, sont présentées au sujet. En période d'activation, des chaînes constituées de caractères prononçables sont présentées au sujet à la cadence d'une chaîne toutes les 2 secondes. Il est demandé au sujet de décider, par simple pression sur une souris à deux boutons, s'il s'agit oui ou non d'un mot réel ou d'un pseudo-mot. En période de contrôle, les chaînes constituées de caractères illisibles sont présentées au sujet. Il lui est alors demandé de répondre par oui ou par non s'il existe une large barre verticale dans la chaîne présentée. Les réponses du sujet ainsi que ses temps de réaction sont enregistrés automatiquement pour analyse ultérieure.

## 7.2.2 Résultats

### 7.2.2.1 Procédure d'apprentissage-cartographie

Les mêmes remarques que dans le cas synthétique peuvent être émises. Dans la très grande majorité des cas, l'algorithme converge. Dans le cas contraire, la différence entre les cartes successives est faible. La convergence de l'algorithme est également rapide, une dizaine de passes étant suffisante. Un exemple de convergence est illustré sur la figure 7.11. Il est à noter que le comportement exprimé par la figure 7.11.b, à savoir que le cardinal de l'ensemble  $\mathcal{F}_i$  augmente au fur et à mesure des passes  $i$ , est constant. Ce comportement est représentatif de l'apprentissage de l'activité par le modèle.

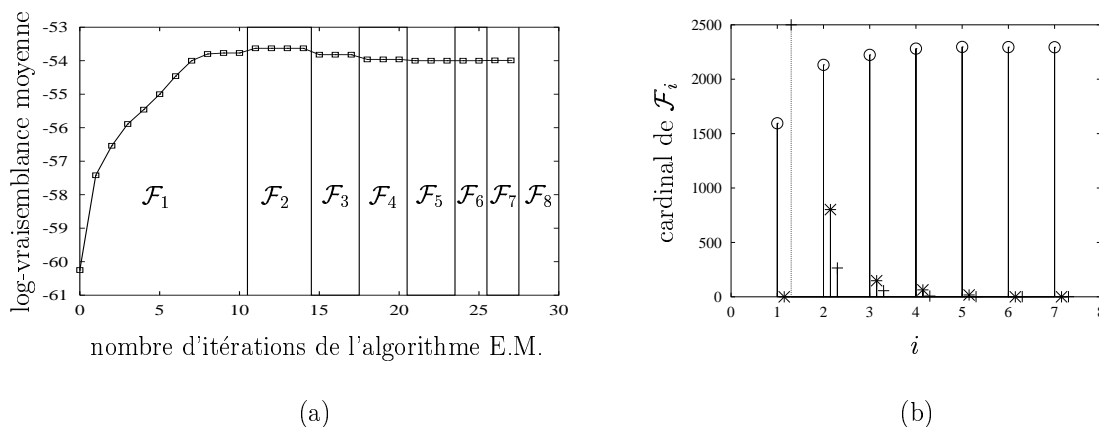


FIG. 7.11 – Convergence de la procédure d’apprentissage-cartographie sur un cas réel. (a) Log-vraisemblance moyenne déterminée sur les différentes bases d’apprentissage pour chaque itération de l’algorithme EM. (b) Cardinal des différentes bases d’apprentissage  $\mathcal{F}_i$ . Le rond représente le cardinal de la base  $\mathcal{F}_i$ , l’étoile représente le nombre de nouveaux voxels qui ont été introduits dans  $\mathcal{F}_i$  par rapport à  $\mathcal{F}_{i-1}$  et la croix représente le nombre de voxels de  $\mathcal{F}_{i-1}$  qui n’apparaissent plus dans  $\mathcal{F}_i$ . Le nombre de voxels intracrâniens est de l’ordre de 20000.

### 7.2.2.2 Cartes d’activité et retard d’activation

Des exemples de cartes d’activation sont représentés figure 7.12.a,b et c pour les protocoles VERBE, AUDIO, et VISUEL, respectivement. Les cartes d’activation obtenues avec la méthode SPM sont également insérées pour comparaison.

Chaque groupe de six cartes se compose tout d’abord d’une carte d’activation basée sur la méthode MSMCSE (en haut à gauche), de la carte de retard d’activation associée (en haut à droite). La carte au milieu à gauche représente uniquement les zones actives synchrones (retard d’activation estimé compris entre -2 et 2 échantillons) alors que la carte en bas à gauche représente les zones déphasées de  $\pi$  (les paradigmes sont périodiques avec une durée de l’état de repos étant égale à celle de l’état d’activation). Les cartes obtenues avec SPM sont représentées au milieu à droite et en bas à droite. Pour la carte au milieu, on a cherché à détecter les signaux en phase alors que pour la carte en bas, on a cherché à détecter les signaux déphasés de  $\pi$ . Pour ce faire, le traitement a été lancé deux fois en utilisant une exécution standard du logiciel SPM2 (correspondant à SPMa) pour deux paradigmes différents, le paradigme originel et le paradigme originel décalé d’une demi-période. Les cartes SPM ont été seuillées en utilisant les champs aléatoires gaussiens et une p-valeur corrigée de 0.05 alors que les cartes obtenues avec la méthode proposées ont été obtenues pour une p-valeur de 0.01. Il est à noter qu’il est difficile de comparer les résultats en termes d’extension spatiale des zones d’activité dans la mesure où des p-valeurs différentes sont utilisées pour le seuillage. De plus, nous avons vu dans la partie précédente qu’il pouvait y avoir des erreurs très importantes d’estimation des p-valeurs pour le logiciel SPM2.

Les résultats obtenus avec la méthode proposée et le logiciel SPM présentent toutefois des activations dans les mêmes secteurs corticaux. Cependant, la méthode proposée présente l’avantage d’être non-supervisée, les signaux déphasés ayant été détectés sans intervention de l’utilisateur.

Plus généralement, les cartes d’activation obtenues par le MSMCSE ont été comparées avec celles obtenues par le logiciel SPM sur les 119 séquences d’événements. En raison de la difficulté

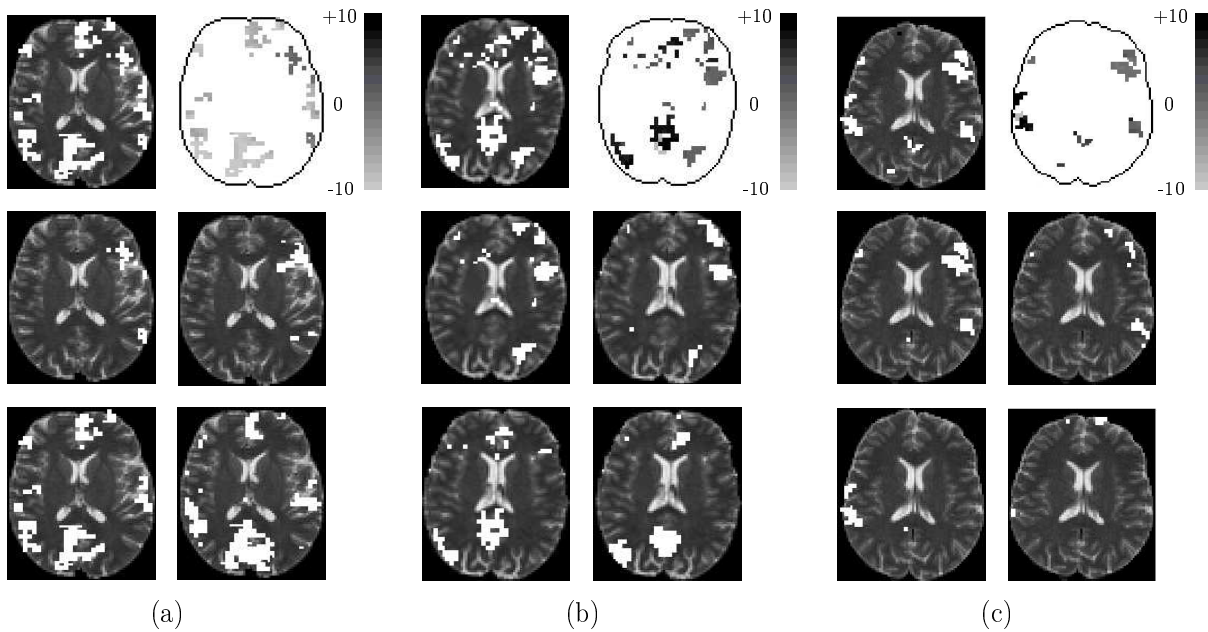


FIG. 7.12 – Cartes d'activation obtenues avec la méthode proposée et SPM pour trois paradigmes différents (de gauche à droite, VERBE, AUDIO, VISUEL). Chaque groupe de six cartes se compose tout d'abord d'une carte d'activation basée sur la méthode MSMCSE (en haut à gauche), de la carte de retard d'activation associée (en haut à droite). La carte au milieu à gauche représente uniquement les zones actives synchrones alors que la carte en bas à gauche représente les zones déphasées de  $\pi$ . Les cartes de SPM sont représentées au milieu à droite et en bas à droite. Pour la carte au milieu, on cherche à détecter les signaux en phase alors que pour la carte en bas, on a cherché à détecter les signaux déphasés de  $\pi$ .

à analyser les résultats et du temps que cela demande, il n’y a pas eu de validation poussée. Néanmoins, nous avons noté dans un nombre limité de cas des discordances. L’explication réside dans le fait que la proportion de signaux actifs retardés était beaucoup plus importante que la proportion de signaux actifs en phase. Puisque statistiquement sous-représentés, les signaux actifs en phase n’ont pas été détectés <sup>2</sup>.

Enfin, différentes informations supplémentaires peuvent être estimées à partir de la méthode proposée.

### 7.2.2.3 Modes d’activation

Pour une région active, il est possible de compter le nombre de fois que chaque état  $evp$  a été visité dans les séquences d’état  $\mathbf{q}_{tpe}^{v,MAP}$  relatives à la région. On obtient ainsi le taux de réponse associé à chaque bloc. Les modes d’activation permettent ainsi d’évaluer l’activité tout au long de l’expérience et donc d’étudier les phénomènes liés à l’apprentissage.

Par exemple, les modes d’activation de la figure 7.13 ont été obtenus pour le protocole AUDIO, et pour la même région active du cerveau de trois sujets. Avant comparaison, chaque histogramme de la figure 7.13 a été normalisée par le nombre  $V$  de voxels activés dans la région. A noter que le premier et le dernier taux de réponse n’ont aucun sens physiologique puisqu’ils correspondent à l’occupation systématique des états  $evp$  fictifs  $q_1$  et  $q_{10}$ . L’histogramme de la

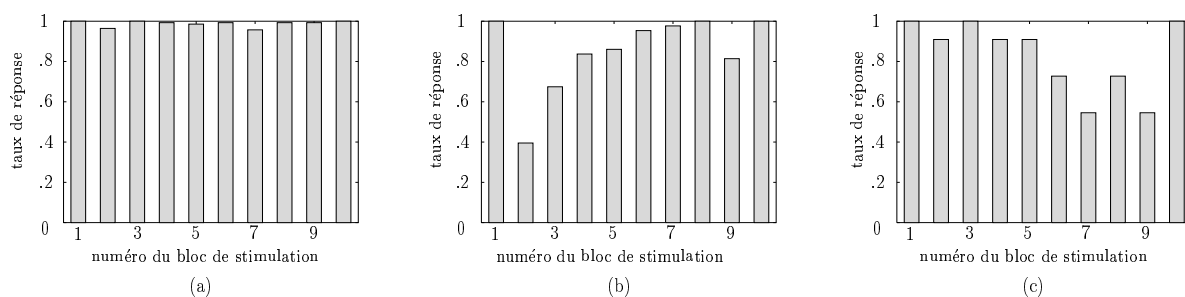


FIG. 7.13 – Analyse des modes d’activation.

figure 7.13.a indique un niveau régulier d’activation. Cette activité soutenue pourrait indiquer que la participation de cette aire du cerveau est essentielle au traitement cognitif impliqué dans la tâche. L’activation exprimée par la figure 7.13.b semble émerger progressivement comme si le recrutement de cette aire dépendait de l’état d’activation d’autres composantes du réseau neurologique. Enfin, le mode d’activation de la figure 7.13.c indiquerait que l’aire corticale n’est pas essentielle à la poursuite de la tâche. Il peut s’agir par exemple d’une aire liée à l’apprentissage de la tâche. Plus généralement, les différents modes d’activation sont l’expression de la dynamique de recrutement des différents réseaux neurologiques impliqués dans la tâche.

### 7.2.2.4 Réponse hémodynamique cérébrale

La méthode permet également d’estimer la forme de la réponse à un bloc de stimulation pour une région d’intérêt. La figure 7.14 montre trois formes de réponses moyennes de trois sujets différents pour la même région active (paradigme AUDIO, bloc de stimulation numéro 4). Il apparaît très nettement une grande variabilité de la réponse. La réponse représentée

<sup>2</sup>Ce problème de sous-représentativité de certains signaux actifs sera résolu dans la prochaine approche

sur la figure 7.14.a est soutenue pendant le bloc d'activation alors que la réponse de la figure 7.14.c ne l'est pas.

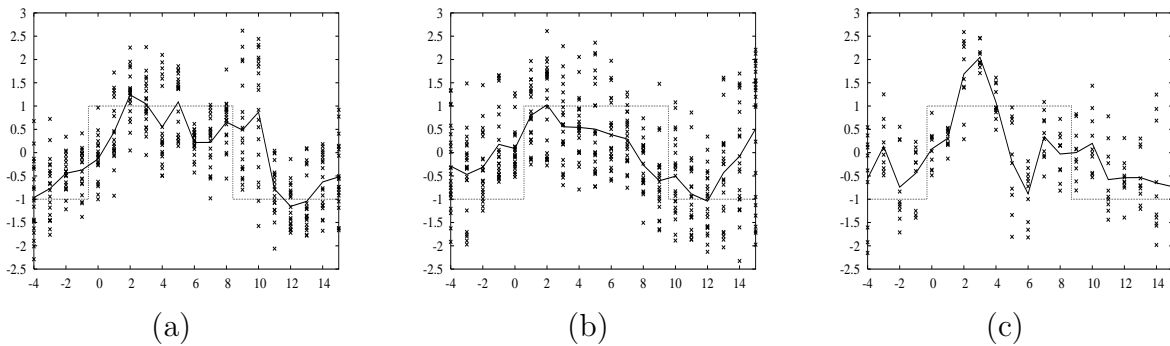


FIG. 7.14 – Analyse de la forme de la réponse hémodynamique : trois exemples de réponse à un bloc de stimulation. Avant tracé, les réponses au bloc ont été alignées sur l'instant 0 grâce à l'instant de détection de l'événement d'intérêt. Les croix représentent l'ensemble des signaux IRMf de la zone d'intérêt. Enfin, le bloc du paradigme est placé de manière à représenter le retard (ou l'avance) moyen d'activation.

# Chapitre 8

## Discussion

Ce chapitre est dans le prolongement direct du chapitre précédent. Il discute de l'influence de certains paramètres clés de la chaîne de traitement sur les résultats de détection de RDs et/ou de cartographie sur données synthétiques et réelles. A partir de ces études, nous ouvrons une discussion, à valeur de bilan, sur les avantages et les limites de l'approche de cartographie par MSMCSE.

### 8.1 Influence des paramètres sur données synthétiques

#### 8.1.1 Etape de détection

L'étape de détection est caractérisée, dans notre application, par le paramètre  $k$ . Ce paramètre détermine l'échelle de résolution la plus élevée,  $s_h$ , de la décomposition en ondelettes appliquée au signal IRMf. Cette échelle contrôle à son tour le taux de RDs détectées dans le signal IRMf. Rappelons également que  $k$  correspond au nombre moyen de RDs détectées par signal IRMf divisé par le nombre de transitions « off-on » présentes dans le paradigme d'activation, transitions fictives exclues. Etant donné que le traitement ne s'effectue qu'au travers des RDs détectées dans le signal IRMf, il est légitime de quantifier la performance des détecteurs indépendamment de l'algorithme de cartographie.

Dans un premier temps, nous avons examiné le nombre moyen de RDs détectées par signal IRMf pour les valeurs  $k = 1, 1.25, 1.5, 1.75$  et  $2$ . Les séquences synthétiques considérées sont  $\mathcal{F}_{1a}, \mathcal{F}_{1b}$ . Rappelons ici que ces séquences ont été construites à partir d'un paradigme de stimulation de type blocs comptant 7 transitions « off-on ». Afin de quantifier les performances de l'étape de détection, cinq mesures ont été utilisées : le nombre moyen de RDs détectées par signal actif,  $\mu_{nA}$ , le nombre moyen de RDs détectées par signal inactif,  $\mu_{nI}$ , les écarts-types associés,  $\sigma_{nA}$  et  $\sigma_{nI}$ , ainsi que la mesure discriminante  $|\mu_{nA} - \mu_{nI}| / (\sigma_{nA} + \sigma_{nI})$ . Les mesures obtenues pour les cinq seuils de détection considérés sont consignées dans le tableau 8.1. Elles montrent clairement que le pouvoir de discrimination basé sur le nombre de détections varie de façon très importante avec le seuil de détection  $k$ , la valeur 1.5 pour ce seuil conduisant à la meilleure discrimination. Le pouvoir de discrimination diminue enfin dès lors que le rapport signal à bruit diminue.

Cette première étude reste insuffisante. Elle s'appuie sur un simple comptage des RDs détectées sans prendre en compte l'information temporelle liée aux instants de détection des RDs. Nous avons donc cherché dans une seconde étude à quantifier les performances de l'étape de détection en termes de bonnes détections et de fausses alarmes. Sont considérées comme

	1	1.25	1.5	1.75	2
$\mathcal{F}_{1a}$	(6.98, 0.19)	(7.06, 0.26)	(7.43, 0.71)	(8.58, 1.26)	(10.68, 1.81)
	(6.98, 1.06)	(8.81, 1.15)	(10.63, 1.23)	(12.4, 1.34)	(14.14, 1.46)
	0	1.24	1.64	1.46	1.05
$\mathcal{F}_{1b}$	(6.97, 0.49)	(7.61, 0.8)	(8.85, 1.23)	(10.64, 1.55)	(12.87, 1.7)
	(6.98, 1.06)	(8.78, 1.15)	(10.56, 1.23)	(12.3, 1.34)	(14.03, 1.45)
	0	0.59	0.69	0.57	0.36

TAB. 8.1 – Nombre moyen de RDs détectées par signal IRMf actif et inactif pour les séquences synthétiques  $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$  et 5 détecteurs  $k = 1, 1.25, 1.5, 1.75$  et 2. A chaque cellule du tableau sont associées 5 mesures, avec dans l'ordre de lecture,  $\mu_{nA}$ ,  $\sigma_{nA}$ ,  $\mu_{nI}$ ,  $\sigma_{nI}$  et la mesure discriminante.

bonnes détections les RDs détectées dans un signal actif dont l'instant de détection s'inscrit dans une fenêtre temporelle de largeur  $2d$  centrée autour d'une des 7 transitions « off-on » du paradigme (figure 8.1). Sont considérées comme fausses alarmes les RDs détectées dans un signal inactif et remplissant le critère de fenêtrage précédent. De façon similaire à la première étude, cinq autres mesures ont été utilisées : le nombre moyen de bonnes détections par signal IRMf actif,  $\mu_{dA}$ , le nombre moyen de fausses alarmes (pouvant être considérées comme des bonnes détections) par signal IRMf inactif,  $\mu_{dI}$ , les écarts-types associés,  $\sigma_{dA}$  et  $\sigma_{dI}$ , enfin la mesure discriminante  $|\mu_{dA}-\mu_{dI}|/(\sigma_{dA}+\sigma_{dI})$ . Les mesures ont été effectuées pour deux tailles de

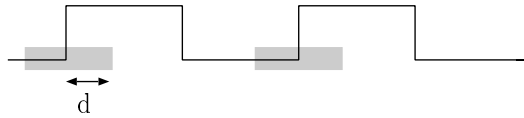


FIG. 8.1 – Définition des bonnes détections pour des signaux actifs ou des fausses alarmes pouvant être considérées à tort comme des bonnes détections pour des signaux inactifs : ces derniers tombent dans une fenêtre temporelle autour de l'instant de transition « off-on » du paradigme. Cette zone est représentée en gris sur la figure.

fenêtre différentes, à savoir  $d = 1$  et  $d = 2$ , et pour les cinq seuils de détection considérés plus haut. Les valeurs obtenues sont consignées dans le tableau 8.2.

	1	1.25	1.5	1.75	2
$\mathcal{F}_{1a}$	(5.68, 1.17)	(4.99, 1.24)	(4.63, 1.31)	(4.73, 1.31)	(4.72, 1.28)
	(0.93, 0.88)	(1.19, 0.97)	(1.47, 1.06)	(1.76, 1.12)	(2.02, 1.16)
	2.31	1.71	1.33	1.22	1.1
	(6.75, 0.53)	(6.51, 0.74)	(6.25, 0.91)	(6.33, 0.88)	(6.39, 0.81)
	(1.56, 1.08)	(2.02, 1.14)	(2.45, 1.22)	(2.9, 1.26)	(3.34, 1.29)
$\mathcal{F}_{1b}$	3.22	2.38	1.78	1.6	1.45
$\mathcal{F}_{1b}$	(4.08, 1.44)	(3.57, 1.41)	(3.36, 1.39)	(3.48, 1.39)	(3.59, 1.35)
	(0.93, 0.88)	(1.19, 0.97)	(1.46, 1.06)	(1.74, 1.11)	(2.01, 1.16)
	1.35	1	0.77	0.69	0.62
	(5.65, 1.14)	(5.18, 1.25)	(4.98, 1.31)	(5.08, 1.25)	(5.26, 1.19)
	(1.56, 1.08)	(2.01, 1.14)	(2.44, 1.22)	(2.88, 1.26)	(3.31, 0.49)
	1.84	1.32	1	0.87	1.16

TAB. 8.2 – Caractéristique de l'étape de prétraitement en termes de bonnes détections (pour les signaux actifs) et de fausses alarmes pouvant être considérées comme des bonnes détections par le modèle (pour les signaux inactifs). Les différentes mesures sont calculées pour les deux séries et les cinq détecteurs. A chaque cellule du tableau sont associées 10 mesures, avec dans l'ordre de lecture,  $\mu_{1I}$ ,  $\sigma_{1I}$ ,  $\mu_{1A}$ ,  $\sigma_{1A}$ , la mesure discriminante associée ( $d=1$ ) et  $\mu_{2I}$ ,  $\sigma_{2I}$ ,  $\mu_{2A}$ ,  $\sigma_{2A}$  et la mesure discriminante associée ( $d=2$ ).

Le pouvoir de discrimination diminue lorsque  $k$  augmente. En effet, augmenter le taux de détection entraîne une augmentation du nombre de fausses alarmes pouvant être considérées



comme de bonnes détections sans pour autant augmenter le nombre de bonnes détections. Il apparaît également que le pouvoir de discrimination obtenu en prenant en compte l'information temporelle n'est guère plus important que celui obtenu en prenant en compte l'information liée au nombre uniquement. Il est à noter que nous prenons en compte ces deux types d'information dans le modèle. En effet, le nombre de bonnes détections est associé indirectement à la matrice de transition alors que le nombre de détections total est lié à la matrice de transition (bonnes détections) et au taux de fausse alarme.

En conclusion, pour discriminer au mieux les signaux actifs des signaux inactifs, les deux études menées sur données synthétiques conduisent à s'orienter vers un seuil de détection  $k$  relativement faible, de l'ordre de 1.25. En effet, le seuil de détection  $k = 1$  permet une très bonne discrimination en utilisant l'information de localisation temporelle mais une très mauvaise discrimination en utilisant l'information de nombre de RDs.

### 8.1.2 Influence des paramètres sur les résultats de cartographie

Trois paramètres ont une influence importante sur les résultats de cartographie. Il s'agit de :

- la  $p$ -valeur utilisée pour seuiller les cartes d'activation. (cf. figure 6.6 de la page 96). Sa valeur détermine en partie le contenu de la base d'apprentissage ;
- le taux de détection permettant de déterminer la résolution de l'ondelette analysante ;
- les paramètres du modèle initial  $\lambda_1$  (cf. figure 6.6).

En pratique, le choix du modèle initial n'a que peu d'influence sur le résultat de cartographie dès lors que ce dernier n'est pas trop contraint (écarts-types grands par exemple). En effet, le modèle initial doit potentiellement être représentatif d'une très grande variété de signaux actifs. Il est donc hors de question de le contraindre à représenter une unique catégorie de signaux actifs.

Nous évaluons maintenant l'influence de la  $p$ -valeur et du paramètre  $k$  sur les résultats de cartographie. Les séquences d'images IRMf synthétiques ont été traitées pour trois  $p$ -valeurs distinctes ( $p_1=0.01$ ,  $p_2=0.005$  et  $p_3=0.001$ ) et 5 seuils de détection ( $k_1=1$ ,  $k_2=1.25$ ,  $k_3=1.5$ ,  $k_4=1.75$  et  $k_5=2$ ). Les séquences considérées sont  $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$ . Une carte de  $p$ -valeurs a été calculée pour chaque couple  $(k_i, p_j)$  et chaque séquence synthétique. Les valeurs en 0.01, 0.005 et 0.001 des courbes COR,  $\alpha$  et  $\beta$  associées à cette carte ont été retenues pour analyse. L'ensemble de ces valeurs est consigné dans le tableau 8.3.

Nous analysons tout d'abord plus spécifiquement les résultats obtenus avec les séquences  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$  puisque nous avons étudié sur ces deux séries la qualité de l'étape de prétraitement. Il convient de noter que de très bons résultats sont obtenus avec la séquence  $\mathcal{F}_{1a}$ . Pour la séquence  $\mathcal{F}_{1b}$ , conformément aux résultats obtenus précédemment sur le pouvoir discriminant des différents détecteurs, on observe de meilleurs résultats pour des faibles valeurs de  $k$ , à l'exception du taux de détection  $k_1$ , lequel conduit aux résultats les plus faibles. Ceci met en valeur le fait que le modèle statistique utilise à la fois l'information de localisation temporelle et l'information de nombre de RDs pour discriminer les signaux actifs des signaux inactifs.

Si l'on classe les détecteurs en termes de performance de cartographie, on obtient du plus performant au moins performant  $k_2$ ,  $k_3$ ,  $k_4$ ,  $k_5$  et  $k_1$ . Cependant, ce classement n'est pas toujours vrai. On obtient par exemple de meilleurs résultats avec le taux de détection  $k_1$  qu'avec le taux de détection  $k_5$  sur la séquence  $\mathcal{F}_{3a}$ . En définitive, cette différence est en grande partie expliquée par le fait que la performance intrinsèque des détecteurs en termes de discrimination entre les signaux actifs et les signaux inactifs dépend de la séquence d'images. Nous notons également que des taux de détection inférieurs à 1 (non présenté ici) permettent

	$\mathcal{F}_{1,a}$	$\mathcal{F}_{1,b}$	$\mathcal{F}_{2,a}$	$\mathcal{F}_{3,a}$	$\mathcal{F}_{4,a}$	$\mathcal{F}_{5,a}$
$(k_1, p_1)$	98, 1.8, 99	46, 1.6, 54	73, 2, 80	85, 1.5, 89	97, 1.4, 98	61, 1.7, 74
	98, 1.7, 98	36, 2.1, 47	63, 2.1, 73	79, 1.6, 83	95, 1.9, 97	47, 1.5, 56
	93, 1.9, 95	14, 2.6, 26	40, 2.5, 50	63, 2.2, 69	84, 2.7, 92	31, 1.8, 36
$(k_1, p_2)$	99, 1.3, 99	59, 1.3, 65	80, 1.5, 83	90, 1.5, 92	98, 1.2, 98	80, 1.1, 80
	99, 1.4, 99	50, 1.6, 57	73, 1.8, 79	86, 1.6, 89	96, 1.3, 97	64, 1, 64
	97, 1.9, 97	35, 2.8, 41	56, 2.1, 63	73, 2, 77	89, 0.8, 88	40, 1.4, 41
$(k_1, p_3)$	100, 1.2, 100	14, 1.6, 22	89, 1.1, 89	93, 1.2, 94	98, 1.1, 98	89, 0.8, 86
	100, 1.5, 100	6, 1.7, 11	86, 1, 86	90, 1.3, 91	98, 1.2, 98	80, 0.8, 76
	98, 1.1, 98	2, 3, 3	67, 1.5, 74	79, 2.6, 85	93, 1.3, 94	50, 0.6, 42
$(k_2, p_1)$	99, 1.7, 99	68, 2.2, 76	81, 2, 84	92, 2.1, 95	98, 1.6, 99	78, 1.4, 83
	99, 1.9, 99	58, 2.2, 69	75, 1.9, 81	90, 2.4, 92	98, 1.5, 98	71, 1.3, 74
	98, 2.8, 98	48, 2.8, 53	65, 0.7, 62	83, 3.2, 86	94, 1.7, 96	56, 2.4, 64
$(k_2, p_2)$	100, 1.7, 100	72, 1.7, 78	84, 1.8, 88	95, 1.7, 96	99, 1.8, 99	87, 1.9, 91
	99, 1.7, 99	65, 1.9, 72	81, 2.3, 85	93, 2.1, 95	98, 2, 99	77, 1.8, 86
	98, 1.2, 99	47, 1.1, 51	69, 2.9, 77	87, 3, 90	96, 2, 97	64, 1.8, 67
$(k_2, p_3)$	100, 1.2, 100	81, 1.3, 84	88, 1.5, 90	97, 1.4, 97	99, 1.4, 99	91, 1.2, 92
	100, 1.3, 100	75, 1.5, 79	84, 1.5, 86	95, 1.5, 96	99, 1.6, 99	86, 1.1, 88
	99, 2.5, 99	59, 1.7, 63	76, 1.9, 77	89, 1.6, 91	96, 1.3, 97	68, 1, 68
$(k_3, p_1)$	99, 2.3, 99	64, 2.2, 72	71, 2.1, 76	83, 2, 88	98, 2.6, 99	82, 2.2, 89
	98, 2.5, 99	55, 2.5, 66	64, 2.4, 72	80, 2.6, 85	97, 3.5, 99	76, 2.9, 85
	96, 2.9, 98	41, 2.7, 49	58, 5.5, 65	71, 4.3, 78	94, 5.8, 97	56, 4.3, 74
$(k_3, p_2)$	99, 2.2, 100	66, 2, 73	77, 2.1, 82	85, 2.3, 90	98, 2.3, 99	86, 2.3, 92
	99, 2.6, 99	59, 2.8, 70	72, 2.3, 79	82, 3, 88	97, 3, 99	80, 2.8, 89
	97, 3.2, 98	41, 3.4, 52	61, 2.8, 68	74, 1.8, 76	95, 3.97	61, 3.6, 76
$(k_3, p_3)$	100, 1.9, 100	73, 1.7, 79	84, 1.6, 86	91, 1.7, 93	99, 2.1, 99	90, 1.8, 93
	99, 2.1, 100	67, 1.9, 72	80, 1.8, 83	89, 2.4, 92	98, 2, 99	85, 2.2, 91
	99, 2.9, 99	49, 2.8, 60	69, 2.1, 74	80, 3.3, 87	95, 2.2, 97	72, 2.2, 77
$(k_4, p_1)$	98, 2.7, 99	53, 2.4, 63	60, 3, 69	80, 2.9, 86	98, 2.4, 99	80, 2.7, 88
	98, 2.7, 99	47, 3, 57	56, 3.9, 65	76, 3.1, 82	96, 2.8, 98	74, 2.6, 83
	95, 3.6, 97	34, 1.9, 38	47, 4, 53	70, 3.8, 74	94, 3.7, 96	67, 3.7, 72
$(k_4, p_2)$	99, 2.2, 99	60, 2, 69	69, 2.2, 76	85, 2.1, 88	98, 2.2, 99	86, 2.7, 92
	98, 2.6, 99	53, 2.3, 62	63, 2.8, 72	80, 1.9, 85	97, 2.6, 99	80, 2.9, 89
	97, 3.4, 98	42, 1.9, 46	56, 4.1, 62	74, 2, 76	94, 3.2, 96	73, 5.1, 80
$(k_4, p_3)$	99, 2, 100	69, 1.9, 77	83, 1.5, 85	88, 2.1, 92	99, 2.3, 99	89, 1.8, 93
	99, 2.5, 99	61, 2.5, 73	78, 1.7, 81	86, 2.4, 89	98, 2.5, 99	85, 1.7, 88
	98, 3.8, 99	50, 1.7, 53	66, 3.4, 74	78, 3.3, 84	94, 2.6, 97	76, 3.7, 82
$(k_5, p_1)$	97, 2.3, 98	43, 2.6, 53	60, 3.1, 70	75, 2.4, 82	94, 2.3, 97	72, 2.2, 79
	95, 2.5, 97	39, 3, 47	52, 3.5, 64	70, 2.6, 77	92, 2.6, 95	68, 2.3, 74
	91, 4.4, 94	27, 2.6, 33	48, 3.4, 49	59, 4.2, 68	88, 4.1, 91	54, 4.3, 65
$(k_5, p_2)$	98, 2.5, 99	52, 2.2, 60	75, 2, 80	79, 2.3, 88	96, 2.4, 98	76, 2.5, 85
	97, 3.3, 99	45, 2.8, 54	69, 2.2, 76	75, 2.6, 83	93, 3, 96	70, 2.6, 78
	93, 4.4, 97	35, 3, 40	59, 3.5, 66	65, 1.6, 68	88, 5.2, 93	56, 2.9, 67
$(k_5, p_3)$	99, 1.8, 100	64, 1.6, 69	83, 1.7, 87	89, 2, 93	97, 2.5, 98	85, 1.9, 89
	99, 2, 99	56, 2.2, 64	80, 2, 83	86, 2, 90	95, 2.8, 98	81, 2.1, 85
	97, 2.7, 98	44, 3.2, 52	68, 2.2, 74	78, 2.7, 82	91, 3.6, 94	65, 2.6, 72

TAB. 8.3 – Influence des paramètres sur les résultats de cartographie. A une cellule du tableau, est associée la séquence d'image utilisée ( $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  ou  $\mathcal{F}_{5a}$ ) et les paramètres utilisés ( $k_1=1$ ,  $k_2=1.25$ ,  $k_3=1.5$ ,  $k_4=1.75$ ,  $k_5=2$ ,  $p_1=0.01$ ,  $p_2=0.005$  et  $p_3=0.001$ ). Pour chaque cellule, il est possible de déterminer une carte de p-valeur, cette dernière est résumée ensuite par neuf mesures consignées dans la dite-cellule. Les trois chiffres situés sur la même ligne sont respectivement de gauche à droite la valeur de la courbe ROC, de la courbe  $\alpha$  et de la courbe  $\beta$  pour une abscisse de 0.01 (en haut), de 0.005 (au milieu) et de 0.001 (en bas). A noter que les résultats pour les courbes  $\beta$  et les courbes COR ont été multipliés par 100 de manière à obtenir une mesure entre 0 et 100. On rappelle que la courbe  $\alpha$  correspond au nombre de fausses alarmes observées sur le nombre de fausses alarmes attendues en fonction de la p-valeur estimée et que la courbe  $\beta$  est le pourcentage de bonnes détections en fonction de la p-valeur estimée.

intrinsèquement d'obtenir de meilleurs pouvoirs discriminants. Néanmoins, en raison de la faiblesse du nombre d'événements composant la séquence, il est difficile pour le modèle statistique d'analyser de telles séquences. L'algorithme converge souvent de façon prématurée vers un modèle très contraint qui déclare actif un ensemble réduit de signaux. Ces signaux se révèlent être des signaux actifs pour lesquels l'étape de prétraitement s'est très bien déroulée. En conclusion, il est certain que le taux de détection a une influence importante sur les résultats de cartographie. Ceci est naturel dans la mesure où le pouvoir discriminant intrinsèque n'est pas identique pour tous les taux de détection. Choisir le taux de détection optimal peut donc être un problème difficile. Aussi, dans la prochaine partie, en réponse à ce problème, nous utilisons plusieurs taux de détection différents.

En ce qui concerne l'influence de la  $p$ -valeur, on remarque tout d'abord que pour une échelle de détection fixée, les résultats sont d'autant plus satisfaisants que la  $p$ -valeur est faible (mis à part, parfois, pour le taux de détection  $k_1$ , cf. couple  $(k_1, p_3)$  avec la séquence  $\mathcal{F}_{1,b}$ ). En effet, lorsque la variabilité des signaux actifs est peu importante, l'utilisation d'une  $p$ -valeur faible permet de ne pas introduire un nombre important de fausses alarmes dans la base d'apprentissage sans pour autant courir le risque de perdre certaines catégories de signaux actifs. Pour conclure, de manière analogue au choix du seuil de détection  $k$ , le choix de la  $p$ -valeur a une influence importante sur les résultats. L'étude de son influence que nous venons d'effectuer sur séquences synthétiques semble militer pour une  $p$ -valeur la plus petite possible. Cette conclusion doit cependant être pondérée eu égard à la relative faible variabilité des signaux actifs synthétiques considérés. En effet, en utilisant une  $p$ -valeur très petite, certains signaux actifs peuvent ne pas rentrer dans la base d'apprentissage. Le modèle converge ainsi vers un modèle trop contraint et non représentatif de l'ensemble des signaux actifs. C'est ce que l'on observe pour le couple  $(k_1, p_3)$  avec la séquence  $\mathcal{F}_{1,b}$ .

## 8.2 Influence des paramètres sur données réelles

Contrairement aux données synthétiques, la vérité terrain reste inconnue pour les données réelles. Aussi, nous avons uniquement évalué l'influence des différents paramètres sur les résultats de cartographie. Précisément, nous examinons tour à tour l'influence du seuil de détection  $k$  utilisé dans l'étape de détection de ruptures de dynamiques (RDs), l'influence de la  $p$ -valeur utilisée pour seuiller les cartes d'activation, et enfin l'influence du modèle initial  $\lambda_1$  sur les résultats de cartographie. Pour pouvoir évaluer l'influence des paramètres sur les résultats de cartographie, d'autres mesures ont dû être définies. Ces mesures seront utilisées plus loin dans ce chapitre pour évaluer l'influence de la  $p$ -valeur et du modèle initial  $\lambda_1$ . Nous les présentons dans le cadre du paramètre  $k$ .

Nous définissons tout d'abord le taux d'inclusion entre deux cartes d'activation  $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  de la façon suivante :  $I_{i/j}^k = \frac{\text{card}(\mathcal{A}_i^k \cap \mathcal{A}_j^k)}{\text{card}(\mathcal{A}_j^k)}$ . L'indice  $k$  est utilisé pour indiquer que l'étude porte sur ce paramètre. Les indices  $i$  et  $j$  désignent la  $i$ -ème et la  $j$ -ème valeur prise par le paramètre  $k$  utilisé pour obtenir respectivement les cartes  $\mathcal{A}_i^k$  and  $\mathcal{A}_j^k$ . Le numérateur de  $I_{i/j}^k$  représente le nombre de voxels actifs qu'ont en commun les deux cartes et le dénominateur représente le nombre de voxels actifs de  $\mathcal{A}_j^k$ . Le taux d'inclusion peut varier entre 0 ( $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  n'ont pas de voxels actifs en commun) et 1 (tous les voxels actifs de  $\mathcal{A}_j^k$  sont également des voxels actifs de  $\mathcal{A}_i^k$ ). A partir du taux d'inclusion, on définit la mesure de similarité entre les cartes  $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  de la façon suivante :  $S_{i,j}^k = 0.5 \times (I_{i/j}^k + I_{j/i}^k)$ . La mesure de similarité varie entre 0 ( $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  n'ont aucun voxel actif en commun) à 1 ( $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  sont identiques).

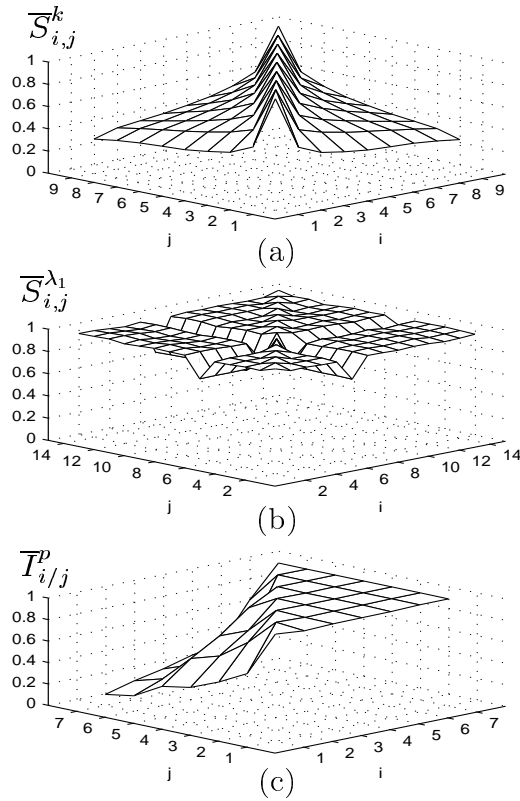


FIG. 8.2 – Influence des paramètres sur les résultats. (a) Influence du paramètre  $k$  de l'étape de prétraitement, (le taux de similarité moyen est représenté sur l'axe  $(O \vec{u}_z)$ ), (b) influence du modèle initial  $\lambda_1$ , (le taux de similarité moyen est représenté sur l'axe  $(O \vec{u}_z)$ ), (c) influence de la  $p$ -valeur (le taux d'inclusion moyen est représenté sur l'axe  $(O \vec{u}_z)$ ).

Afin d'examiner l'influence du seuil de détection  $k$  sur les résultats de cartographie, des cartes d'activation  $\mathcal{A}_i^k$  et  $\mathcal{A}_j^k$  ont donc été calculées pour  $i, j = 1, \dots, 9$ , et  $k = 1, 1.125, 1.25, 1.375, 1.5, 1.625, 1.75, 1.875$  et  $2$  sur l'ensemble des 119 séquences à disposition. Ensuite, les mesures d'inclusion  $I_{i/j}^k$  et  $I_{j/i}^k$  ont été déterminées et moyennées sur l'ensemble des données. Finalement, une mesure  $\overline{S}_{i,j}^k$  a été calculée pour chaque paire d'indice  $i, j$ . Les mesures obtenues sont représentées par la surface de la figure 8.2.a. La surface est par définition symétrique par rapport à la ligne  $\overline{S}_{i,i}^k = 1$ . En s'écartant de cette ligne, la mesure de similarité diminue. On illustre ainsi la grande influence de  $k$ , et par delà  $k$ , la grande influence du choix de l'échelle haute,  $s_h$ , de décomposition en ondelettes du signal IRMf sur les résultats.

Une valeur raisonnable pour le paramètre  $k$  semble se situer entre  $1.375$  et  $1.625$  ( $i, j \sim 4, 5, 6$ ). Ceci permet d'assurer une mesure de similarité supérieure à  $0.6$ . En pratique,  $k$  est fixé à  $1.5$  pour l'analyse des données réelles. Une brève expertise montre également que les résultats seraient plus satisfaisants avec  $k=1.5$  qu'avec  $k=1.25$ . Ceci reste cependant *a priori* une caractéristique propre aux données utilisées.

On s'intéresse maintenant à la  $p$ -valeur utilisée lors du seuillage. Changer la  $p$ -valeur change la carte d'activité : on s'attend à ce que la taille d'une région active augmente lorsque l'on augmente la  $p$ -valeur. Par conséquent, notre but est de comparer les taux d'inclusion plutôt que les taux de similarité. Dans ce but, les cartes d'activation  $\mathcal{A}_i^p$  et  $\mathcal{A}_j^p$  sont déterminées

pour  $i, j = 1, \dots, 7$  et  $p$  dans  $\{0.001, 0.003, 0.005, 0.008, 0.01, 0.03, 0.05\}$ , et pour l'ensemble des séquences à disposition. Ensuite, les taux d'inclusions  $I_{i/j}^p$  sont calculés et moyennés sur l'ensemble des données IRMf à disposition. Les taux d'inclusion  $\bar{T}_{i/j}^p$  sont représentés par la surface de la figure 8.2.c. Comme attendu, on observe que les valeurs diminuent rapidement pour  $i < j$ . De plus, si  $i \geq j$ , la surface est plate et proche de 1. Ceci signifie que les voxels actifs de  $\mathcal{A}_i^p$  incluent les voxels actifs de  $\mathcal{A}_j^p$ . Les cartes d'activation obtenues sont donc cohérentes les unes par rapport aux autres.

Pour évaluer l'influence du modèle initial  $\lambda_1$  sur les résultats, 14 modèles  $\lambda_{1,i}$  ont été créés à partir de  $\lambda_1$  par des variations successives des sous-ensembles  $\mathbf{A}_1$  (paramètres de la matrice de transition),  $\mathbf{B}_1$  (paramètres des lois d'observation des états *evp*),  $\mathbf{B}'_1$  (paramètres des lois d'observation des *efp*),  $\mathbf{D}_1$  (paramètres des lois de durée) ou  $\mathbf{E}_1$  (probabilités d'émission). Ensuite, comme précédemment, les mesures de similarité  $\bar{S}_{i,j}^{\lambda_1}$  ont été déterminées pour  $i, j = 1, \dots, 14$ . Les résultats sont représentés par la surface de la figure 8.2.b. Cette dernière est presque plate, toutes les valeurs étant au-dessus de 0,9, à l'exception des valeurs de  $i$  et  $j$  égales à 6. Le modèle initial  $\lambda_{1,6}$  a un écart type  $\sigma d_0$  très faible. Aussi, s'il existe peu de variabilité entre les signaux actifs synchrones, la valeur  $\sigma d_0$  restera toujours faible et les éventuels signaux en retard ne seront jamais détectés. En pratique, on conclut que le modèle  $\lambda_1$  a une influence très mineure sur les résultats dès lors que ce dernier est peu contraint. Il doit en effet être potentiellement représentatif de tous les signaux actifs pour pouvoir converger vers un modèle définissant au mieux les signaux actifs contenus dans la base.

## 8.3 Bilan de l'approche de cartographie par MSMCSE

### 8.3.1 Les points positifs

L'approche proposée possède tout d'abord l'avantage d'être non supervisée. Alors que les principales méthodes inférentielles nécessitent de fortes hypothèses sur la forme de la réponse et sur la chronologie d'activation, la méthode détermine à partir des données et sans aucune intervention de l'utilisateur, le modèle représentatif de l'activité grâce notamment aux capacités d'apprentissage du modèle markovien.

Il faut cependant noter que trois types de paramètres peuvent potentiellement avoir une influence sur les résultats. Le modèle initial  $\lambda_1$  n'a en pratique que très peu d'influence dès lors qu'il est peu contraint. Le choix de la p-valeur a une influence sur les résultats (ce qui est tout à fait normal). Cependant, nous avons vérifié la cohérence des résultats, les zones déterminées avec une petite p-valeur étant incluses dans les zones déterminées avec une p-valeur plus importante. Nous avons également mis en valeur sur données synthétiques que les p-valeurs étaient bien estimés. Le dernier paramètre pouvant avoir une influence sur les résultats est le paramètre  $k$ . Dans la prochaine partie, nous utilisons plusieurs détecteurs par signaux IRMf ce qui permet de ne pas restreindre son choix à une unique valeur de  $k$ .

L'approche proposée offre également de nouvelles fonctionnalités comme l'analyse des retards d'activation, des modes d'activation et de la forme de la réponse hémodynamique.

Finalement, cette première approche permet de valider le principe sur lequel elle se fonde : en cas d'activité, les RDs détectées dans le signal IRMf actif doivent s'aligner dans une certaine mesure sur les transitions « off-on » du paradigme. On dispose donc d'une approche de modélisation très intéressante dont on ne tire toutefois pas suffisamment profit. A cause d'une très grande sensibilité à la qualité de l'étape de prétraitement, nous n'avons par exemple pas clairement mis en évidence l'intérêt de considérer les non stationnarités du signal IRMf

pour prendre en compte la grande variabilité de la réponse à une stimulation. Cependant, la méthode proposée est largement améliorable de par la richesse du formalisme markovien caché, et plus largement des approches de modélisation par réseaux bayésiens dynamiques.

### 8.3.2 Les améliorations envisageables

Pour pouvoir améliorer notre approche, il faut tout d'abord prendre conscience de ses faiblesses. Notre approche possède principalement deux limitations. Tout d'abord, elle est très sensible au bruit et de manière plus générale à la performance de l'étape de prétraitement. En effet, les résultats obtenus se dégradent très vite dès lors que l'étape de prétraitement est moins performante. Enfin, elle ne prend pas en compte des signaux actifs sous-représentés statistiquement. Nous citons ci-après l'ensemble des éléments qui peuvent expliquer ces faiblesses et nous donnons une première ébauche de solution.

#### 8.3.2.1 Un unique modèle d'activité ?

La définition d'un modèle d'activité unique à l'ensemble des signaux actifs n'est pas une solution satisfaisante dès lors qu'il existe une grande variabilité de signaux actifs. Il se pose tout d'abord le problème de sous-représentativité de certaines catégories de signaux actifs. Puisque sous-représentés, ces derniers auront une vraisemblance faible et ne seront pas considérés comme des signaux actifs par la suite. Il se pose enfin le problème de la capacité du modèle à représenter plusieurs modes d'activité à la fois. On conçoit aisément qu'il est difficile pour un même modèle de modéliser convenablement différents types d'activité. Par exemple, si une moitié des signaux actifs répond à une transition sur deux alors que l'autre moitié répond à chaque transition, on obtient en définitive un modèle moyen qui ne caractérise pas véritablement les deux types d'activité.

Par la suite, ce problème est résolu en définissant un modèle d'activité local propre à un voxel  $v$  et à son voisinage.

#### 8.3.2.2 Modélisation des durées inter-événements

Il existe une inadéquation importante entre la modélisation des lois de durée et notre application. Pour illustrer cette dernière, on considère un paradigme avec 11 transitions « off-on » au temps 1,11,21,31,41,51,61,71,81,91,101 (transitions fictives incluses). Supposons que la

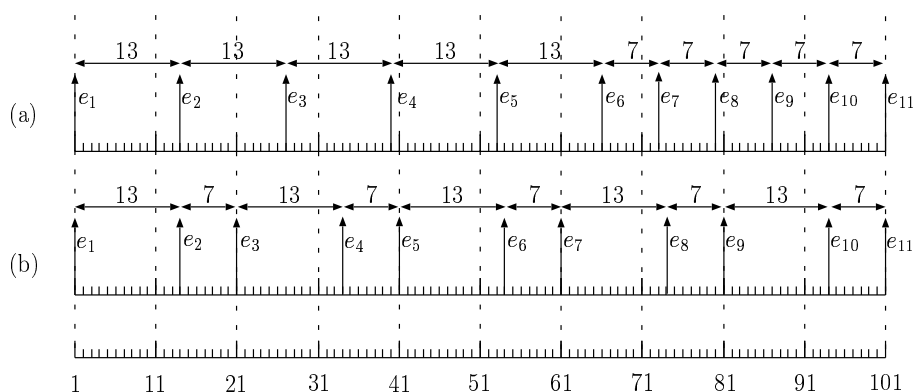


FIG. 8.3 – Illustration du problème de la modélisation des lois de durée (voir texte)

durée associée à chacun des états cachés suit une loi gaussienne de moyenne 10 et d'écart type  $\sigma$  ( $\sigma$  ne dépendant pas de l'état) et que les événements  $e_i$  soient associés à la  $i$ -ème transition du paradigme (voir figure 8.3), alors, les probabilités  $p_1$  et  $p_2$  associées aux séquences de la figure 8.3.a et b sont écrites ci-après (sous réserve de ne considérer que les lois de durée) :

$$p_1 = p_2 = (\mathcal{N}(7; 10, \sigma))^5 \times (\mathcal{N}(13; 10, \sigma))^5 \quad (8.1)$$

Alors que la séquence (b) est beaucoup mieux alignée sur les transitions du paradigme que la séquence (a), les deux séquences ont même vraisemblance ! Ainsi, la modélisation des durées de type inter-événement est mal adaptée à notre application. Notons également que les durées entre événements sont supposées indépendantes alors que ce n'est pas le cas, la longueur de la séquence étant fixée.

Pour illustrer d'une autre manière l'inadéquation entre la modélisation des lois de durée et notre application, nous avons calculé par simulation de Monte Carlo la probabilité d'occupation des états *evp* non fictifs en fonction du temps sous réserve que la matrice de transition n'autorise aucun saut d'états. Deux exemples sont illustrés sur la figure 8.4. Il apparaît clairement un étalement au cours du temps, étalement qui diminue ensuite en raison de la dernière loi d'observation. Cet étalement est d'autant plus important que le nombre de transitions dans le paradigme est important.

### 8.3.2.3 Caractérisation ou discrimination ?

La méthode proposée est basée sur la caractérisation des signaux actifs plutôt que sur la discrimination entre signaux actifs et signaux inactifs. A partir de la base d'apprentissage, on détermine un modèle d'activité  $\lambda$  au sens du maximum de vraisemblance. On espère ainsi que ce modèle soit représentatif de l'activité, c'est-à-dire que  $P[\mathbf{E}|\lambda, \mathcal{H}_1]$  soit grand et que  $P[\mathbf{E}|\lambda, \mathcal{H}_0]$  soit petit,  $\mathcal{H}_1$  et  $\mathcal{H}_0$  représentant respectivement l'hypothèse d'activité et d'inactivité cérébrale. Il serait sans doute plus adapté ici d'utiliser des critères discriminants.

Par la suite, nous conservons une approche basée sur la caractérisation des données d'un point de vue algorithmique mais les paramètres du modèle sont contraints à caractériser l'activité. Le modèle conserve ainsi une discrimination importante entre signaux actifs et signaux inactifs. Par exemple, des contraintes sur la matrice de transition tels que  $a_2 \leq a_1$  sont utilisées,  $a_1$  et  $a_2$  étant respectivement la probabilité de sauter 0 et 1 transition. Si le nombre de sauts d'états pour les signaux actifs est très important, on aboutit, grâce à la contrainte, à l'équation  $a_1 = a_2$ . Il est alors à espérer que d'autres paramètres du modèle puissent permettre une bonne discrimination. Sans utiliser de contraintes, on peut aboutir à un modèle pour lequel  $a_2 > a_1$ . Si la vraisemblance sous l'hypothèse  $\mathcal{H}_1$  est augmentée (maximum de vraisemblance), il est à craindre que la vraisemblance sous l'hypothèse  $\mathcal{H}_0$  augmente également de manière importante conduisant ainsi au risque de diminuer le pouvoir discriminant du modèle.

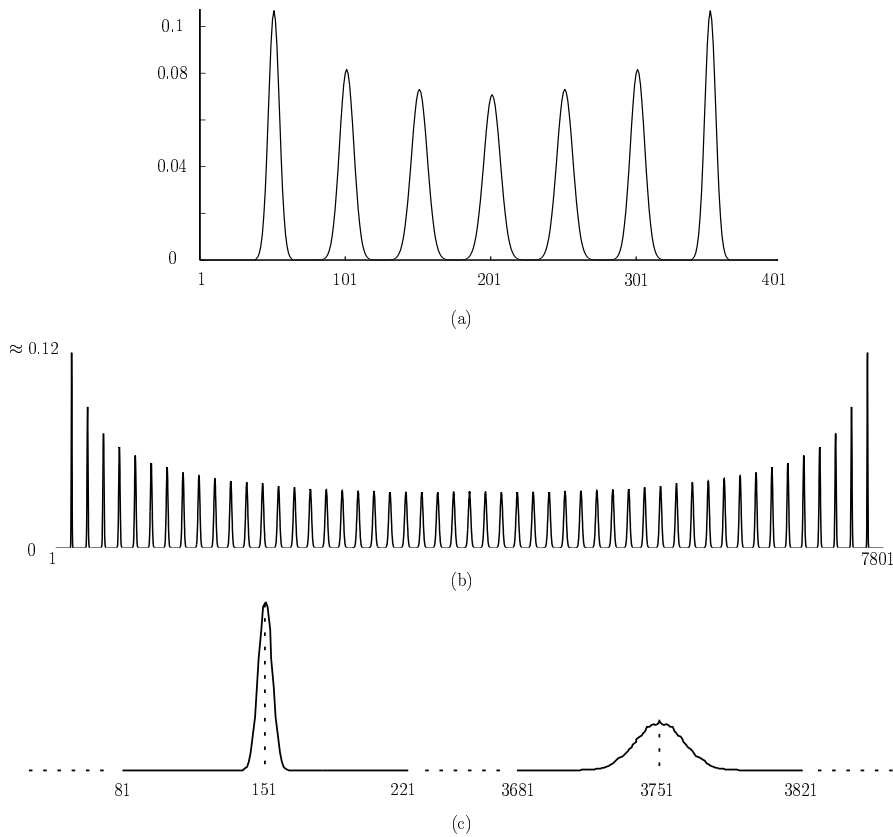


FIG. 8.4 – Densité de probabilité d’occupation des états *evp* en fonction du temps. (a) Le paradigme est constitué de 9 transitions (transitions fictives incluses), la  $i$ -ème transition ayant lieu à l’instant  $50 \times (i - 1) + 1$ . Chaque loi de durée est une loi normale de moyenne 50 et d’écart-type 4. (b) Le paradigme est constitué de 53 transitions, la  $i$ -ème transition ayant lieu à l’instant  $150 \times (i - 1) + 1$ . Chaque loi de durée est une loi normale de moyenne 150 et d’écart-type 3,3. La figure (c) est un zoom sur la figure (b). Il est à noter que les densités de probabilité des figure (a) et (b) ne somment pas à 1 mais respectivement à 7 et à 51.



# Conclusion partielle

Nous avons présenté une nouvelle méthode de cartographie en IRMf cérébrale. L'activation est formulée au niveau du voxel en termes de couplage entre la séquence de RDs détectées dans le signal IRMf et les transitions « off-on » du paradigme. Ce principe de couplage est largement validé par les résultats obtenus. La formulation de l'activité en de tels termes nous permet tout d'abord d'apporter des informations supplémentaires tels que la carte de retard d'activation, l'analyse des modes d'activité et l'analyse de la réponse hémodynamique. Cette formulation nous permet enfin de nous abstenir de faire un nombre important d'hypothèses concernant notamment la forme de la réponse hémodynamique et la chronologie d'activation. Cependant, la très grande sensibilité de la méthode proposée à la performance de l'étape de prétraitement nous a empêché de tirer entièrement profit de ses bonnes propriétés. Par exemple, quand l'activité n'est pas maintenue pendant tout le bloc, les RDs sont plus difficiles à détecter rendant l'étape de prétraitement moins efficace. Les résultats de cartographie en pâtissent alors immédiatement ! Nous avons tenté de cerner les causes de ces faiblesses. Nous remettons en cause l'utilisation d'un unique modèle d'activité pour caractériser l'ensemble des signaux actifs et nous remettons également en cause la modélisation des lois de durée inter-événements. Cependant, nous n'incriminons pas la modélisation markovienne cachée. La résolution des trois problèmes, à savoir les problèmes de l'apprentissage, du décodage et de l'apprentissage s'est avérée d'un grand intérêt dans le cadre de la cartographie en IRMf.

Par la suite, nous conservons une modélisation markovienne cachée mais des solutions sont apportées à chacun des problèmes mentionnés. De plus, nous proposons également de prendre en compte le contexte spatial et nous nous plaçons dans un cadre multi-détecteurs de manière à augmenter la robustesse de l'approche.



## Troisième partie

Le modèle markovien caché de  
multiples séquences d'événements :  
application en cartographie IRMf  
cérébrale



# Introduction

Nous avons présenté dans la partie II une méthode de cartographie en IRMf cérébrale basée sur les non stationnarités du signal IRMf. Plus précisément, l'activation est formulée au niveau du voxel en termes de couplage entre la séquence de RDs détectées dans le signal IRMf et les transitions « off-on » du paradigme. Ce principe permet en partie de limiter le nombre d'hypothèses *a priori* sur la forme de la réponse et sur la chronologie d'activation. Cependant, la modélisation utilisée dans la partie précédente a montré des faiblesses expliquées en partie (i) par l'utilisation d'un unique modèle d'activité pour caractériser l'ensemble des signaux actifs et (ii) par une « mauvaise » modélisation des lois de durée inter-événements.

Fort de cette expérience, nous proposons ici une autre approche basée sur le même principe d'alignement. Nous conservons une modélisation markovienne cachée mais des solutions sont apportées à chacun des problèmes mentionnés. De plus, de manière à augmenter la robustesse de l'algorithme, nous proposons également de prendre en compte le contexte spatial et de nous placer dans un cadre multi-détecteurs. Sous l'hypothèse d'activité fonctionnelle au voxel  $v$ , les séquences d'événements associées à  $v$  et au voisinage de  $v$  doivent désormais s'aligner dans une certaine mesure sur les transitions du paradigme (paradigme en blocs) ou sur les stimuli eux-mêmes (paradigme événementiel). Nous résolvons ce problème dans un cadre statistique de fusion de multiples séquences d'événements : le modèle markovien caché de multiples séquences d'événements (MMCMSE).

Nous abordons tout d'abord le principe de modélisation. Nous présentons ensuite de façon détaillée le modèle markovien développé : le MMCMSE. Nous voyons ensuite comment ce modèle peut s'appliquer à la cartographie en IRMf cérébrale. Nous présentons et nous discutons enfin les résultats obtenus.



# Chapitre 9

## Principes et méthodes

### 9.1 Principe d'alignement multi-séquences

La méthode de cartographie proposée dans cette partie repose sur le même principe énoncé en partie II, à savoir qu'en présence d'activité locale au voxel  $v$ , la séquence de RDs détectées dans le signal IRMf  $f^v(\cdot)$  correspondant doit s'aligner sur les transitions « off-on » du paradigme. Cependant, afin d'accroître les performances de détection d'activation, le principe d'alignement temporel est étendu aux signaux IRMf voisins du voxel  $v$  de manière à prendre en compte l'information spatiale de voisinage. Les zones corticales actives ayant un volume important par rapport à celui d'un voxel, il n'est en effet pas dénué de sens de caractériser l'activité au voxel  $v$  en utilisant également les voxels voisins de  $v$ . D'autre part, toujours pour gagner en robustesse, l'étape de détection des RDs pour un signal IRMf s'appuie dorénavant sur  $K$  détecteurs opérant en parallèle.

De façon plus formelle, si  $\mathbf{E}^{1,v}, \mathbf{E}^{2,v}, \dots, \mathbf{E}^{K,v}$  dénotent les  $K$  séquences de RDs détectées au voxel  $v$ , et si  $\mathcal{N}(v)$  désigne le voisinage de  $v$ , alors l'ensemble des séquences supposé caractériser l'activité au voxel  $v$  devient :

$$\mathbf{E}^v = \{\mathbf{E}^{k,v'}\}_{1 \leq k \leq K, v' \in \{v, \mathcal{N}(v)\}} \quad (9.1)$$

A titre d'exemple, considérant un voisinage de type 26-connexité et trois détecteurs opérant en parallèle ( $K=3$ ), l'ensemble  $\mathbf{E}^v$  caractérisant l'activité au voxel  $v$  est formé de  $(3 \times (26+1))=81$  séquences d'événements.

### 9.2 Problématique

La problématique sera ici définie dans le cadre des paradigmes en blocs. Elle reste toutefois la même pour les paradigmes événementiels. Comme dans la dernière partie, nous faisons l'hypothèse qu'une transition « off-on » du paradigme peut être associée au maximum à une RD (par canal d'observation) et qu'une RD peut être associée au maximum à une transition du paradigme. Etablir un alignement temporel entre la séquence de transitions du paradigme et les multiples séquences d'événements soulève différents problèmes illustrés à la figure 9.1 :

- Une transition « off-on » n'est pas nécessairement associée à des RDs, le patient n'ayant pas par exemple répondu à un bloc de stimulation.

Dans une moindre mesure, tous les détecteurs n'ont pas obligatoirement détecté l'événement d'intérêt. Ainsi, la transition « off-on » est associée à des RDs de certaines séquences uniquement.

- Chaque RD détectée ne correspond pas forcément à une transition « off-on » du paradigme, auquel cas il s’agit soit d’une fausse alarme, soit d’une RD associée à un DRH de non-intérêt.
- Des asynchronismes existent entre les RDs détectées dans les différentes séquences d’une part et entre la transition du paradigme et les RDs associées d’autre part.
- Doit-on considérer comme équivalentes toutes les séquences d’événements ? Autrement dit, ne peut-on pas avoir confiance en certaines séquences d’événements plus qu’en d’autres ?

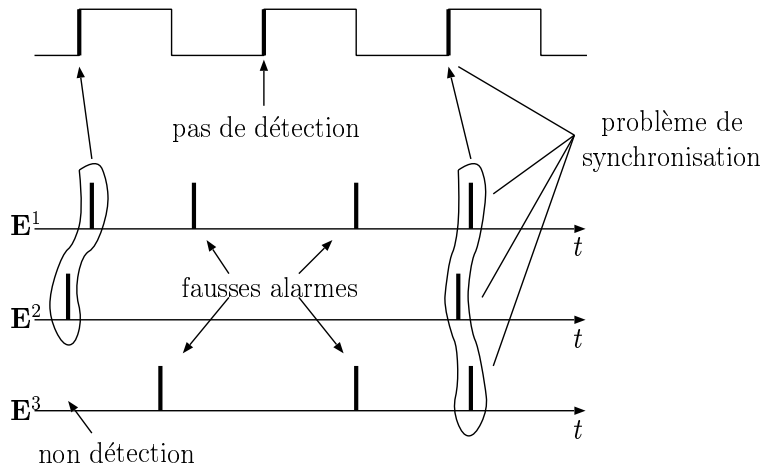


FIG. 9.1 – Problème d’alignement multi-séquences. Les RDs associées à la même transition sont regroupées.

Finalement, les observations associées aux séquences d’événements peuvent être informatives ou pas sur le processus analysé, redondantes et/ou complémentaires, incertaines, incomplètes et asynchrones. La répartition de ces observations sur l’axe des temps n’est pas, pour autant, désordonnée. Elle obéit à des modes propres à chaque séquence et communs à plusieurs séquences. Les modes communs suggèrent l’existence d’interrelations entre les séquences d’observations. L’analyse de ces interrelations relève du cadre de la fusion de données et plus précisément du cadre de la fusion multi-séquences.

### 9.3 Stratégie de résolution

Nous avons vu dans la seconde partie qu’un premier modèle devait idéalement modéliser l’étape de préparation des données alors qu’un second pouvait être lié à la caractérisation de l’activité cérébrale ou à la discrimination entre l’activité cérébrale recherchée et toute autre activité de fond. Cependant, une telle modélisation n’a pas été privilégiée pour deux raisons majeures. Il nous a avant tout semblé difficile de spécifier convenablement ces deux modèles. D’autre part, une telle approche fait craindre un risque certain de sur-représentation des données.

De par le nouveau cadre dans lequel nous nous plaçons, cette voie semble intéressante. Nous pouvons désormais estimer la « vraie » séquence de DRH (celle fournie par un observateur idéal) à travers un ensemble de séquences d’événements. On observe ainsi la même réalité à travers différentes séquences d’événements ou à travers différents canaux d’observation. Cette multiplicité des canaux nous offre un cadre plus riche pour estimer la vraie séquence à partir des observations (modèle du détecteur). Ensuite, pour simplifier le modèle biologique, nous



supposons qu'il n'existe que des DRHs d'intérêt (DRHI). Les DRHs de non intérêt seront considérés comme des fausses alarmes du système de détection. Le modèle biologique et le modèle des détecteurs sont détaillés ci-après.

### 9.3.1 Le modèle biologique

Le modèle biologique permet de caractériser l'activité cérébrale que l'on cherche à détecter ou permet de discriminer l'activité cérébrale recherchée de toute autre activité de fond. Dans notre cadre, nous faisons l'hypothèse que le modèle biologique « privilégie » les séquences de DRHI pour lesquelles le retard d'activation est constant tout au long du paradigme et pour lesquelles le nombre de DRHI est important. Cette hypothèse se justifie d'un point de vue biologique. Il est nécessaire alors de déterminer la vraisemblance de la séquence de DRHI, que nous notons  $\mathbf{E}_{DRHI}^v$ , conditionnellement au modèle d'activité biologique  $\lambda_B$  :

$$P[\mathbf{E}_{DRHI}^v | \lambda_B] \quad (9.2)$$

### 9.3.2 Le modèle des détecteurs

Le modèle des détecteurs  $\lambda_D$  doit lier statistiquement les observations effectuées  $\mathbf{E}_{obs}^v$  et une séquence d'événements DRHI  $\mathbf{E}_{DRHI}^v$ . Ce lien s'effectue à travers la loi suivante :

$$P[\mathbf{E}_{obs}^v | \lambda_D, \mathbf{E}_{DRHI}^v] \quad (9.3)$$

Les équations (9.2) et (9.3) permettent ainsi de déterminer la vraisemblance des observations sachant les deux modèles, cette vraisemblance pouvant être considérée comme une mesure de l'activité :

$$P[\mathbf{E}_{obs}^v | \lambda_D, \lambda_B] = \sum_{\mathbf{E}_{DRHI}^v} P[\mathbf{E}_{obs}^v | \lambda_D, \mathbf{E}_{DRHI}^v] P[\mathbf{E}_{DRHI}^v | \lambda_B] \quad (9.4)$$

Il est à remarquer que la séquence de DRHI n'est pas observable directement. Il est donc nécessaire d'intégrer sur cette dernière. En pratique, le modèle développé est un unique modèle permettant à la fois de prendre en compte le modèle biologique et le modèle des détecteurs. Il convient ici de donner son principe.

### 9.3.3 Le modèle développé

La stratégie de résolution que nous empruntons s'appuie sur la notion de scénarios explicatifs de l'ensemble  $\mathbf{E}^v$  de séquences d'événements observés au voxel  $v$ . Un scénario, noté  $\mathcal{S}$ , repose sur l'association (figure 9.2.b) :

- des  $N$  séquences de RDs détectés constituant l'ensemble  $\mathbf{E}^v$ . Les RDs détectées sont appelées *événements de surface* ;
- d'une  $N + 1$ -ème séquence d'événements, les DRHIs, induite par le paradigme d'activation. Les DRHIs sont appelés *événements profonds* ;
- de signatures définies comme l'association d'une ou plusieurs RDs détectées sur différents canaux d'observation.
- des contraintes de causalité entre les signatures conditionnellement à la séquence de DRHI.

La séquence d'événements profonds (les DRHIs) n'est pas observable directement. Elle est néanmoins observée indirectement au travers des séquences d'événements de surface (les RDs) qu'elle est supposée avoir produites, l'occurrence d'un événement profond se traduisant en surface par l'émission d'une signature. Notons que pour être observable et donc détectable, un événement profond doit pour le moins donner lieu à un événement au moins en surface. Autrement dit, une signature est nécessairement composée au minimum d'une RD détectée.

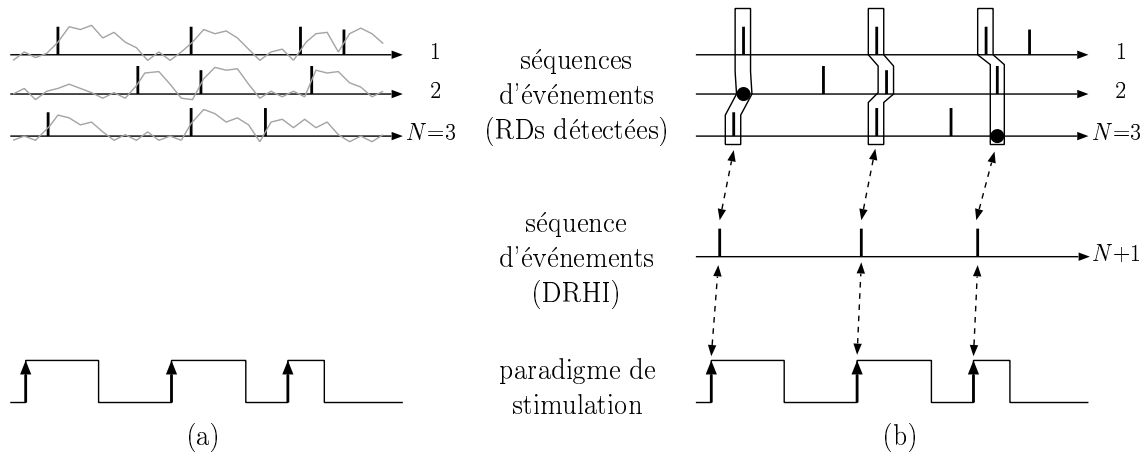


FIG. 9.2 – Position du problème. (a) ensemble de séquences  $\mathbf{E}^v$ . Chaque séquence constituant  $\mathbf{E}^v$  est observée au voxel  $v$  ou dans son voisinage avec éventuellement plusieurs détecteurs. (b) un scénario possible de l'ensemble  $\mathbf{E}^v$  représenté en (a) sous l'hypothèse d'activité du voxel  $v$ . Les événements encadrés correspondent à une signature (un point noir signifie la non détection de l'événement d'intérêt). Nous voyons le lien étroit qui existe entre une transition « off-on » du paradigme, un événement profond et une signature.

Grâce à l'introduction du principe des scénarios, le problème d'alignement des séquences de RDs sur la séquence de transitions du paradigme se ramène à un problème d'alignement simple entre une séquence de DRHI et la séquence de transitions du paradigme. Cette stratégie de fusion est formellement décrite à l'intérieur du modèle markovien caché de multiples séquences d'événements.

Pour effectuer le lien avec le paragraphe précédent, le modèle biologique permet de lier la séquence de DRHI avec les transitions du paradigme alors que le modèle des détecteurs permet de lier les séquences observées avec la séquence de DRHI. Le modèle markovien caché de multiples séquences d'événements permet ainsi de prendre en compte à la fois la mesure et les principes biologiques.

# Chapitre 10

## Le modèle markovien caché de multiples séquences d'événements

Un modèle markovien caché de multiples séquences d'événements (MMCMSE) est un modèle probabiliste à deux niveaux d'information,  $\{X_u, O_u\}$ , dédié à la modélisation et à l'analyse d'un processus aléatoire composé d'événements non directement observables, le processus *profond*. Ce dernier est observé au travers de  $N$  processus stochastiques de nature événementielle, les processus de *surface*. Les processus de surface sont corrélés : ils correspondent à des observations différentes du même processus profond. Ces  $N$  canaux d'observation différents apportent ainsi des informations complémentaires et/ou redondantes sur une même réalité.

La partie cachée d'un MMCMSE,  $\{X_u\}$ , modélise l'ordre des événements profonds le long de l'axe  $u$ . La partie observable,  $\{O_u\}$ , modélise (i) les instants d'occurrence des événements profonds, (ii) la traduction en surface du processus profond et (iii) les éventuelles fausses alarmes détectées en surface.

Dans le cadre de l'IRMf, le processus profond correspond au processus d'activation neuronale, une réalisation de ce dernier étant une séquence d'événements de DRHI. Les processus de surface correspondent quant à eux aux multiples séquences de RDs détectées au voxel  $v$  et dans son voisinage.

### 10.1 Séquences d'événements, signatures, scénario et séquence d'observations

L'ensemble des notations introduites dans cette section est illustré figure 10.1.

#### 10.1.1 Séquences d'événements

Une étape de prétraitement détecte indépendamment pour chaque processus de surface les événements d'intérêt. Soit  $I = [0, T]$  l'intervalle de temps pendant lequel les processus de surface sont observés. On note  $\mathbf{E}^{k,m}$ ,  $1 \leq k \leq K$ ,  $1 \leq m \leq M$  la séquence d'événements détectés sur  $I$  dans le  $k$ -ème processus de surface par le  $m$ -ème détecteur, et  $\boldsymbol{\tau}^{k,m}$  l'ensemble des instants de détection associé à  $\mathbf{E}^{k,m}$ . Aucune distinction n'étant faite entre les différentes séquences d'événements, on écrit sans perte de généralité  $\mathbf{E}^{(i)}$  et  $\boldsymbol{\tau}^{(i)}$ ,  $1 \leq i \leq N$  avec  $N = M.K$ . L'ensemble des séquences observées est finalement :  $\mathbf{E} = \{\mathbf{E}^{(i)}\}_{1 \leq i \leq N}$ . Un exemple de séquence est représenté à la figure 10.1.a.

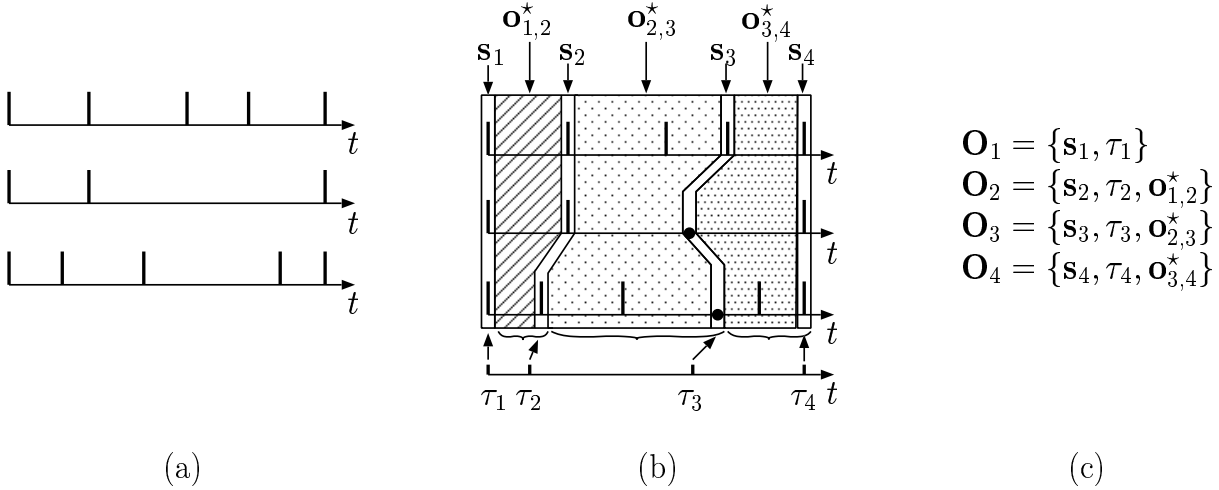


FIG. 10.1 – Séquences d'événements, scénario, séquence d'observations. (a) Ensemble  $\mathbf{E}$  des séquences d'événements ( $N=3$ ). (b) Scénario  $\mathcal{S}$  pouvant expliquer  $\mathbf{E}$ . Les ronds à l'intérieur des signatures correspondent à des événements non détectés. (c) Construction d'une séquence d'observations  $\mathbf{O} = \mathbf{O}_1\mathbf{O}_2\mathbf{O}_3\mathbf{O}_4$  à partir d'un scénario valide  $\mathcal{S}$  et d'un ensemble  $\mathbf{E}$  de séquences d'événements de surface.

### 10.1.2 Signatures

Une signature est la partie « observable » d'un événement profond. Elle est formée de  $N$  événements de surface (un événement par processus de surface). Les événements composant une signature peuvent être soit observés (*evp observés* : *evpo*) soit non observés (*evp manquants* : *evpø*). Un *evpø* a pour origine une non détection de l'événement d'intérêt sur un canal d'observation. Par exemple, la signature  $\mathbf{s}_3$  de la figure 10.1.b possède deux *evpø*s (ronds noirs) et un *evpo* (trait vertical). Un événement détecté qui n'est pas un *evpo* est *a fortiori* un *événement faux-positif (efp)*. Un *efp* n'est donc pas associé au processus profond. Il résulte d'une fausse alarme de l'étape de détection.

### 10.1.3 Scénario

Comme déjà vu dans le chapitre précédent, un scénario, noté  $\mathcal{S}$ , est explicatif de l'ensemble  $\mathbf{E}$  des séquences observées. Plus spécifiquement, un scénario  $\mathcal{S}$  est caractérisé par les temps d'occurrence des événements profonds et des événements de surface associés. Formellement, un scénario  $\mathcal{S}$  est défini comme une série de  $U$  vecteurs d'instant de dimension  $N + 1$  :

$$\mathcal{S} = \{\tau_u, \tau_u^{(1)}, \dots, \tau_u^{(N)}\}_{1 \leq u \leq U} \quad (10.1)$$

Le terme  $U$  représente le nombre total d'événements profonds. Les instants  $\tau_u$  et  $\tau_u^{(i)}$  représentent respectivement l'instant d'occurrence du  $u$ -ème événement profond et l'instant d'occurrence de l'événement de la  $u$ -ème signature détecté sur le  $i$ -ème canal d'observation. Nous rappelons que  $\boldsymbol{\tau}^{(i)}$  représente l'ensemble des instants de détection associés à  $\mathbf{E}^{(i)}$ . Si  $\tau_u^{(i)} \in \boldsymbol{\tau}^{(i)}$ , alors l'événement a été réellement détecté (*evpo*), sinon, cela signifie que l'événement n'a pas été détecté (*evpø*). Une mesure  $m_u^{(i)}$  est associée avec  $\tau_u^{(i)}$ . La mesure est vide si l'événement est un *evpø*. L'ensemble  $\mathbf{s}_u = \{\tau_u^{(i)}, m_u^{(i)}\}_{1 \leq i \leq N} = \{\boldsymbol{\tau}_u^*, \mathbf{m}_u^*\}$  représente la *signature* du  $u$ -ème événement profond.

Pour qu'un scénario  $\mathcal{S}$  soit valide et puisse expliquer les séquences  $\mathbf{E}$ , il doit vérifier les contraintes suivantes :

$$\begin{aligned}
 \tau_u, \tau_u^{(i)} &\in I && (i) \\
 \tau_1 = \tau_1^{(i)} = 0, \tau_U = \tau_U^{(i)} = T &&& (ii) \\
 (u' > u) \Rightarrow (\tau_{u'} > \tau_u \text{ et } \tau_{u'}^{(i)} > \tau_u^{(i)}) &&& (iii) \\
 \forall u, \exists i : \tau_u^{(i)} \in \boldsymbol{\tau}^{(i)} &&& (iv)
 \end{aligned} \tag{10.2}$$

(ii) provient du fait que deux signatures fictives sont introduites respectivement au début ( $u = 1$ ) et à la fin ( $u = U$ ) des séquences pour permettre de modéliser convenablement les lois de durée. (iii) signifie que les signatures préservent l'ordre des événements profonds. (iv) stipule qu'au moins un événement détecté doit être présent par signature, ceci afin de satisfaire la condition d'observabilité de l'événement profond associé.

Un exemple de scénario valide est donné à la figure 10.1.b. Il est composé de 4 signatures dont 2 sont fictives. Des exemples de scénarios non valides sont illustrés à la figure 10.2. Le premier, figure 10.2.b, ne vérifie pas la condition (iv) de l'équation (10.2) alors que le second, figure 10.2.c, ne vérifie pas la condition (iii) de l'équation (10.2).

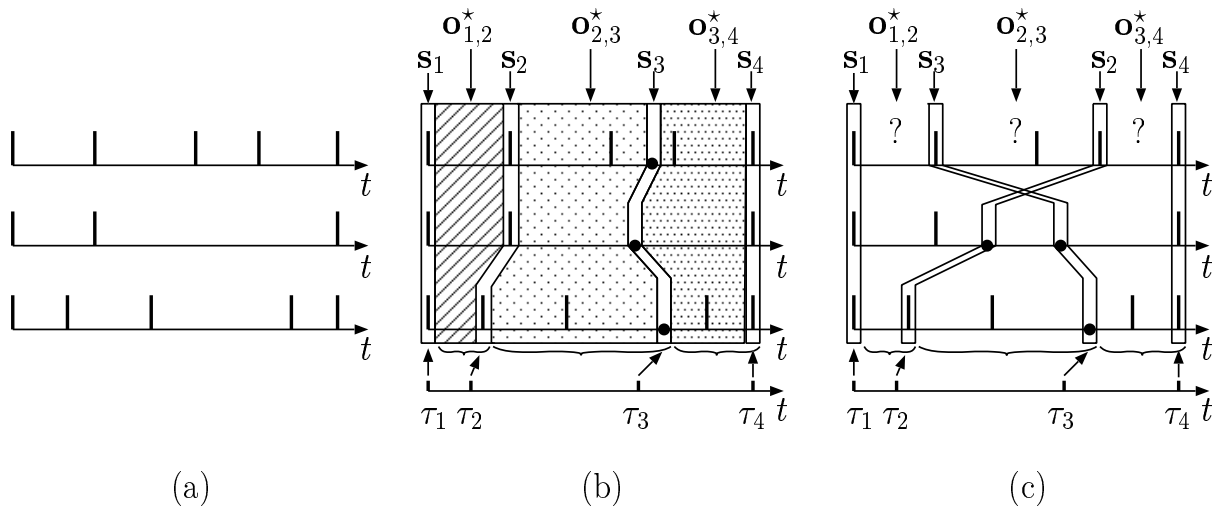


FIG. 10.2 – (a) ensemble  $\mathbf{E}$  de séquences d'événements de surface. (b,c) Deux scénarios ne remplissant pas les conditions de validité de l'équation (10.2).

#### 10.1.4 Séquence d'observations

Par définition, la séquence d'observation  $\mathbf{O} = \mathbf{O}_1 \mathbf{O}_2 \dots \mathbf{O}_U$  d'un MMCMSE résulte de la combinaison de  $\mathbf{E}$  avec un scénario valide  $\mathcal{S}$ . Comme le montre la figure 10.1.b, les signatures définissent un découpage de  $\mathbf{E}$  à travers les différentes séquences d'événements. Etant donné  $\mathbf{E}$  et  $\mathcal{S}$ ,  $\mathbf{O}$  est unique. En pratique, la  $u$ -ème observation de  $\mathbf{O}$ ,  $\mathbf{O}_u$ , est constituée d'un temps en profondeur  $\tau_u$ , d'une signature  $\mathbf{s}_u$  et de tous les événements détectés entre  $\mathbf{s}_{u-1}$  et  $\mathbf{s}_u$ , notés  $\mathbf{o}_{u-1,u}^*$ , si bien que (voir figure 10.1) :

$$\begin{cases} \mathbf{O}_1 = \{\tau_1, \mathbf{s}_1\} \\ \mathbf{O}_u = \{\tau_u, \mathbf{s}_u, \mathbf{o}_{u-1,u}^*\} \quad \text{si } u \neq 1 \end{cases} \tag{10.3}$$

## 10.2 Définition du MMCMSE

### 10.2.1 Éléments d'un MMCMSE

Un MMCMSE est complètement défini par :

- son espace d'état  $\mathcal{Q} = \{q_i\}_{1 \leq i \leq C}$  où  $C - 2$  représente le nombre de classes d'événements profonds. Les états  $q_1$  et  $q_C$  modélisent respectivement les signatures initiale et finale.
- sa matrice de transition  $\mathbf{A} = \{a_{ij}\}$  : en raison de la spécificité des états  $q_1$  et  $q_C$ , elle est définie pour  $1 \leq i < C$  et  $1 < j \leq C$  par  $a_{ij} = P[X_u = q_i | X_{u-1} = q_j]$ .
- les poids associés à chaque canal  $\mathbf{w} = (w_1 \dots w_N)^T$  avec la contrainte  $\sum_{k=1}^N w_k = 1$ .  $w_k$  reflète la confiance que l'on a dans le  $k$ -ème processus de surface.
- les lois d'observation  $\mathbf{B} = \{b_{ij\mathbf{w}}(\cdot|\cdot)\}_{1 \leq i < C, 1 < j \leq C}$  : en raison de l'équation (10.2.iii), l'hypothèse traditionnelle d'indépendance des observations conditionnellement aux états ne peut pas être faite. A la place, on pose :

$$P[\mathbf{O}_u | \mathbf{O}_1^{u-1}, \mathbf{X}_1^u, \mathbf{W} = \mathbf{w}] = P[\mathbf{O}_u | \mathbf{O}_{u-1}, X_{u-1} = q_i, X_u = q_j, \mathbf{w}] \quad (10.4)$$

$$= b_{ij\mathbf{w}}(\mathbf{O}_u | \mathbf{O}_{u-1}) \quad (10.5)$$

En respect avec l'équation (10.4), le graphe de dépendance d'un MMCMSE est montré figure 10.3.b. Un MMCMSE est entièrement caractérisé par la donnée de son vecteur de paramètres  $\lambda = \{\mathcal{Q}, \mathbf{A}, \mathbf{B}\}$  et du vecteur de poids  $\mathbf{w}$ . Le vecteur  $\mathbf{w}$  n'est pas inséré dans le vecteur de paramètres  $\lambda$  pour des raisons de simplicité. Par la suite, le vecteur  $\mathbf{w}$  est supposé connu. Ceci se justifie dans la mesure où il est toujours possible d'estimer au préalable la confiance que l'on accorde à chacun des canaux d'observation à partir du rapport signal sur bruit et des performances du détecteur par exemple. Le modèle  $\lambda$  est donc caractéristique d'un vecteur  $\mathbf{w}$ . Dans le cadre de notre application, le vecteur  $\mathbf{w}$  sera estimé itérativement à partir du nombre de bonnes détections associées à chaque canal.

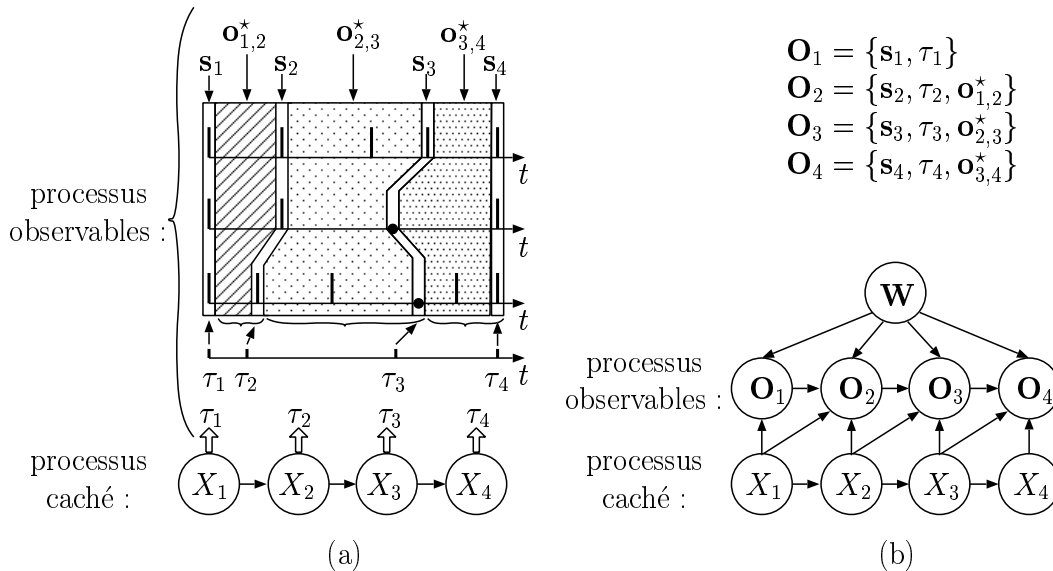


FIG. 10.3 – Graphe de dépendance d'un MMCMSE. (a) Graphe de dépendance de la partie cachée et observation associée. (b) graphe de dépendance d'un MMCMSE.

### 10.2.2 Lois d'observation et MMCMSE

La dimension des lois d'observation d'un MMCMSE peut être très importante au regard du nombre  $N$  de séquences d'événements de surface (canaux d'observation) considérées et des nombreuses relations de dépendance entre elles. A partir de la définition du contenu d'une observation et du contenu d'une signature, la loi d'observation  $b_{ij\mathbf{w}}(\cdot)$  (équation 10.5) peut s'écrire :

$$\begin{aligned} b_{ij\mathbf{w}}(\mathbf{O}_u|\mathbf{O}_{u-1}) &= P(\tau_u|\tau_{u-1}, \mathbf{m}_{u-1}^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{o}_{u-2,u-1}^*, X_{u-1} = q_i, X_u = q_j, \mathbf{w}) \times \\ &\quad P(\tau_u^*|\tau_u, \tau_{u-1}, \mathbf{m}_{u-1}^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{o}_{u-2,u-1}^*, X_{u-1} = q_i, X_u = q_j, \mathbf{w}) \times \\ &\quad P(\mathbf{m}_u^*|\tau_u, \boldsymbol{\tau}_u^*, \tau_{u-1}, \mathbf{m}_{u-1}^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{o}_{u-2,u-1}^*, X_{u-1} = q_i, X_u = q_j, \mathbf{w}) \times \\ &\quad P(\mathbf{o}_{u-1,u}^*|\tau_u, \boldsymbol{\tau}_u^*, \mathbf{m}_u^*, \tau_{u-1}, \mathbf{m}_{u-1}^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{o}_{u-2,u-1}^*, X_{u-1} = q_i, X_u = q_j, \mathbf{w}) \end{aligned} \quad (10.6)$$

Selon l'application considérée, différentes hypothèses simplificatrices peuvent être effectuées. Par exemple, en ne considérant pas les termes grisés de l'équation (10.6), la loi d'observation  $b_{ij\mathbf{w}}(\cdot)$  peut s'écrire :

$$b_{ij\mathbf{w}}(\mathbf{O}_u|\mathbf{O}_{u-1}) = b_{0ij\mathbf{w}}(\tau_u|\tau_{u-1})b_{1j\mathbf{w}}(\boldsymbol{\tau}_u^*|\tau_u, \boldsymbol{\tau}_{u-1}^*)b_{2j\mathbf{w}}(\mathbf{m}_u^*)b_{3j\mathbf{w}}(\mathbf{o}_{u-1,u}^*|\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*) \quad (10.7)$$

Cette simplification apparait naturelle :

- les lois  $b_{0ij\mathbf{w}}(\cdot)$  modélisent les instants d'occurrence des événements en profondeur : elles sont représentatives de la loi de durée entre deux événements profonds consécutifs associés aux états cachés  $q_i$  et  $q_j$ .
- Les lois  $b_{1j\mathbf{w}}(\cdot)$  modélisent les instants d'occurrence des événements de surface, c'est-à-dire la forme de la signature résultant de l'asynchronisme entre événements détectés ou non sur les différents canaux d'observation.
- Les lois  $b_{2j\mathbf{w}}(\cdot)$  sont représentatives des mesures associées aux signatures quand le processus caché est dans l'état  $q_j$ .
- Les lois  $b_{3j\mathbf{w}}(\cdot)$  modélisent les fausses alarmes survenant entre deux signatures ainsi que leur mesure.

Néanmoins, pour un nombre important de canaux, la dimension des lois  $b_{1j\mathbf{w}}(\cdot)$ ,  $b_{2j\mathbf{w}}(\cdot)$  et  $b_{3j\mathbf{w}}(\cdot)$  demeure toujours élevée. Pour la réduire encore, une étape de réduction des données peut être envisagée. Dans notre application, nous avons opté pour l'introduction de trois fonctions notées  $f_{\tau_u, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^1(\boldsymbol{\tau}_u^*)$ ,  $f_{\mathbf{w}}^2(\mathbf{m}_u^*)$  et  $f_{\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^3(\mathbf{o}_{u-1,u}^*)$  caractérisant respectivement la forme de la signature, les mesures des événements composant la signature et l'ensemble des événements détectés entre  $\mathbf{s}_{u-1}$  et  $\mathbf{s}_u$  :

$$\begin{cases} b_{1j\mathbf{w}}(\boldsymbol{\tau}_u^*|\tau_u, \boldsymbol{\tau}_{u-1}^*) &= b_{1j}(f_{\tau_u, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^1(\boldsymbol{\tau}_u^*)) \times b'_{1j}(\boldsymbol{\tau}_u^*|\tau_u, \boldsymbol{\tau}_{u-1}^*, f_{\tau_u, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^1(\boldsymbol{\tau}_u^*)) & (i) \\ b_{2j\mathbf{w}}(\mathbf{m}_u^*) &= b_{2j}(f_{\mathbf{w}}^2(\mathbf{m}_u^*)) \times b'_{2j}(\mathbf{m}_u^*|f_{\mathbf{w}}^2(\mathbf{m}_u^*)) & (ii) \\ b_{3j\mathbf{w}}(\mathbf{o}_{u-1,u}^*|\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*) &= b_{3j}(f_{\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^3(\mathbf{o}_{u-1,u}^*)) \times \\ &\quad b'_{3j}(\mathbf{o}_{u-1,u}^*|\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, f_{\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, \mathbf{w}}^3(\mathbf{o}_{u-1,u}^*)) & (iii) \end{cases} \quad (10.8)$$

## 10.3 Problèmes d'inférence

### 10.3.1 Présentation

Les trois problèmes fondamentaux que sont l'évaluation, l'apprentissage et le décodage classiquement rencontrés dans le cadre des MMCs [186] se posent également dans le cadre des

MMCMSE. Cependant, la formulation des problèmes diffère légèrement en raison des données manquantes des processus observables. Par exemple, la figure 10.4.a (simple copie de la figure 10.1) représente l'ensemble  $\mathbf{E}$  des séquences d'événements détectés. Il s'agit là de l'observation que l'on effectue. Cette observation n'est pas à confondre avec l'observation du MMCMSE. Une observation possible du MMCMSE est représentée à la figure 10.4.b et c. Elle résulte de la combinaison de la séquence  $\mathbf{E}$  avec un scénario valide  $\mathcal{S}$  explicatif de  $\mathbf{E}$ . Toutes les informations définissant un scénario n'y sont pas observées comme les temps d'occurrence des *evp* et des événements profonds, le nombre de signatures et leur composition, modifiant ainsi la formulation des problèmes d'inférence.

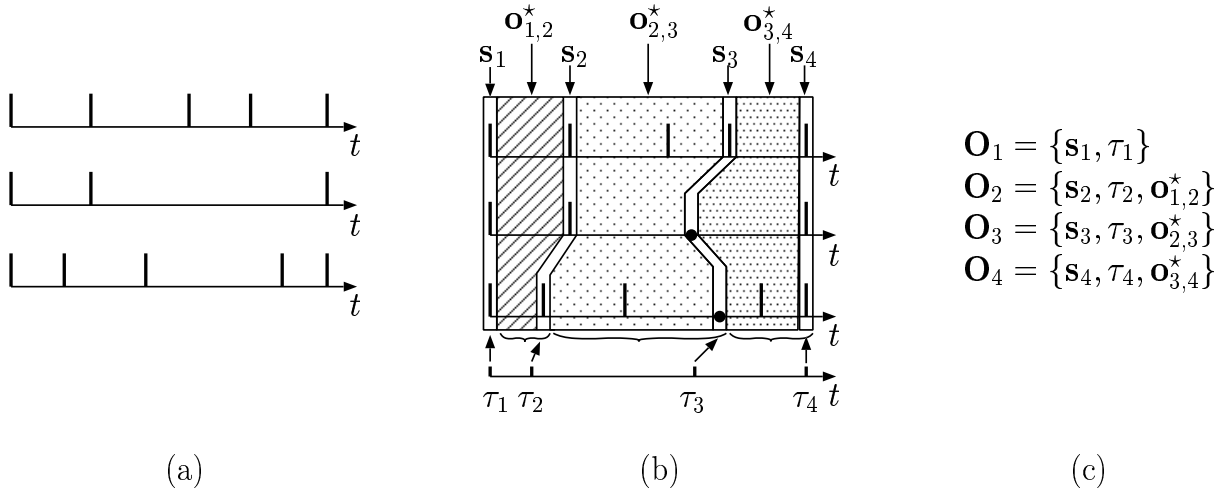


FIG. 10.4 – Observation effectuée et observation possible d'un MMCMSE. (a) Ensemble  $\mathbf{E}$  de séquences d'événements observés. (b) Scénario  $\mathcal{S}$  pouvant expliquer  $\mathbf{E}$  : ce scénario n'est pas observé. (c) La séquence d'observations du MMCMSE conditionnellement à  $\mathbf{E}$  et à  $\mathcal{S}$ . Elle contient les temps d'occurrence des *evp* et des événements profonds, le nombre de signatures et leur composition. Ces données ne sont pas observées mais manquantes.

### 10.3.1.1 Évaluation

L'évaluation consiste à déterminer la vraisemblance des séquences d'événements  $\mathbf{E}$  sachant le modèle  $\lambda$  et le poids des canaux  $\mathbf{w}$ . A partir de scénarios, cette dernière peut s'écrire :

$$\begin{aligned} P[\mathbf{E}|\lambda, \mathbf{w}] &= \sum_{\mathbf{O} \in \mathcal{S}(\mathbf{E})} P[\mathbf{O}|\lambda, \mathbf{w}] \\ &= \sum_{\mathbf{O} \in \mathcal{S}(\mathbf{E})} \sum_{\mathbf{X}} P[\mathbf{O}, \mathbf{X}|\lambda, \mathbf{w}] \end{aligned} \quad (10.9)$$

La séquence d'observation  $\mathbf{O}$  résulte de la combinaison de  $\mathbf{E}$  avec un scénario valide  $\mathcal{S}$  pris dans l'ensemble  $\mathcal{S}(\mathbf{E})$  de tous les scénarios valides explicatifs de  $\mathbf{E}$ .  $\mathbf{X}$  représente une séquence d'états cachés possible à l'origine de la séquence d'observation  $\mathbf{O}$ .

### 10.3.1.2 Décodage

Comme à la fois le scénario et les états cachés sont manquants, le problème du décodage peut être perçu de trois façons différentes. Etant donné  $\mathbf{E}$ ,  $\lambda$  et  $\mathbf{w}$ , il est possible de déterminer au sens du MAP :



- l'observation et la séquence d'états cachés qui expliquent au mieux  $\mathbf{E}$  (équation (10.10.i));
- l'observation qui explique au mieux  $\mathbf{E}$  (équation (10.10.ii));
- la séquence d'états cachés qui explique au mieux  $\mathbf{E}$  (équation (10.10.iii)).

$$\begin{cases} \{\mathbf{O}^{MAP}, \mathbf{X}^{MAP}\} = \operatorname{argmax}_{\mathbf{O} \in \mathcal{S}(\mathbf{E}), \mathbf{X}} P[\mathbf{O}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}] & (i) \\ \mathbf{O}^{MAP2} = \operatorname{argmax}_{\mathbf{O} \in \mathcal{S}(\mathbf{E})} P[\mathbf{O} | \boldsymbol{\lambda}, \mathbf{w}] & (ii) \\ \mathbf{X}^{MAP2} = \operatorname{argmax}_{\mathbf{X}} P[\mathbf{E}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}] & (iii) \end{cases} \quad (10.10)$$

Par la suite, le décodage est résolu suivant le critère de l'équation (10.10.i). Ce critère conduit à une information de décodage non contradictoire, à l'opposé de celle obtenue en résolvant les équations (10.10.ii) et (10.10.iii), lesquelles peuvent conduire à une probabilité jointe  $P[\mathbf{O}^{MAP2}, \mathbf{X}^{MAP2}]$  nulle.

### 10.3.1.3 Apprentissage

Soient  $\mathbf{w}$  et  $\mathbf{E}^1, \mathbf{E}^2, \dots, \mathbf{E}^R$ ,  $R$  ensembles de séquences d'événements supposés mutuellement indépendants, l'apprentissage consiste à ajuster l'ensemble de paramètres  $\boldsymbol{\lambda}$  pour maximiser la vraisemblance suivante :

$$\prod_{i=1}^R P[\mathbf{E}^i | \boldsymbol{\lambda}, \mathbf{w}] \quad (10.11)$$

Les trois problèmes d'inférence que nous venons de formuler explicitement ont été résolus à l'aide de versions adaptées des algorithmes standards des MMCs. Ces versions d'algorithmes ainsi que la stratégie empruntée pour les résoudre sont présentées ci-après.

## 10.3.2 Stratégie de résolution

La stratégie empruntée pour résoudre les problèmes d'évaluation, de décodage et d'apprentissage repose sur la définition d'un chemin caché permettant d'utiliser des versions adaptées des algorithmes standards des MMCs. Cette définition requiert d'introduire au préalable la notion de couple observation-état et les ensembles  $\mathcal{V}$  et  $\mathcal{W}$  associés.

### 10.3.2.1 Couple observation-état

Afin de sommer sur l'ensemble des observations (combinaison de  $\mathbf{E}$  avec un scénario de  $\mathcal{S}(\mathbf{E})$ ) et sur l'ensemble des états cachés pouvant être associés aux observations, nous introduisons la notion de couple observation-état. Un couple observation-état,  $\boldsymbol{\omega}_k$ , est défini comme l'agrégation d'un événement profond (défini par son temps d'occurrence,  $\tau_k$ , et sa classe  $n_k$ ) et de la signature associée à l'événement profond,  $\mathbf{s}_k = \{\boldsymbol{\tau}_k^*, \mathbf{m}_k^*\}$ . Ainsi :

$$\boldsymbol{\omega}_k = \{\tau_k, \boldsymbol{\tau}_k^*, \mathbf{m}_k^*, n_k\} \quad (10.12)$$

A noter que l'état caché  $n_k$  prend, par définition, ses valeurs dans l'ensemble  $\{q_i\}_{1 \leq i \leq C}$ . L'indice  $k$  dans les notations  $\tau_k$ ,  $\boldsymbol{\tau}_k^*$  et  $\mathbf{m}_k^*$  fait référence au  $k$ -ème couple observation-état. Au contraire, l'indice  $u$  dans les notations  $\tau_u$ ,  $\boldsymbol{\tau}_u^*$  et  $\mathbf{m}_u^*$  représentent les observations associées à la  $u$ -ème observation du MMCMSE. Ces notations ne doivent pas être confondues par la suite. Sommer à la fois sur l'ensemble des observations et sur l'ensemble des états cachés peut être réalisé grâce à l'ensemble des couples observation-état  $\mathcal{W} = \{\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_K\}$  qu'il est possible de construire à partir de  $\mathbf{E}$  à condition de faire l'hypothèse que  $\mathbf{o}_{u-2,u-1}^* \perp \mathbf{O}_u | \{\tau_{u-1}, \boldsymbol{\tau}_{u-1}^*, \mathbf{m}_{u-1}^*\}$  :

$$b_{ij\mathbf{w}}(\mathbf{O}_u | \mathbf{O}_{u-1}) \triangleq b_{ij\mathbf{w}}(\mathbf{O}_u | \tau_{u-1}, \boldsymbol{\tau}_{u-1}^*, \mathbf{m}_{u-1}^*, \mathbf{o}_{u-2,u-1}^*) = b_{ij\mathbf{w}}(\mathbf{O}_u | \tau_{u-1}, \boldsymbol{\tau}_{u-1}^*, \mathbf{m}_{u-1}^*) \quad (10.13)$$

Cependant, avant de construire l'ensemble  $\mathcal{W}$ , nous définissons l'ensemble  $\mathcal{V}$  des associations qu'il est possible d'effectuer entre les différents événements détectés de  $\mathbf{E}$ .

### 10.3.2.2 Association des événements détectés

Il s'agit simplement d'associer entre-eux des événements détectés en surface. Aucune contrainte particulière n'est à considérer si ce n'est que deux événements d'un même canal ne peuvent pas être associés ensembles (cf. définition de la signature, section 10.1.2, page 138). A titre d'exemple, en notant  $e_j^{(i)}$  le  $j$ -ième événement observé sur le canal  $i$ , et par  $e_\emptyset^i$  un événement non-observé sur le canal  $i$ , l'ensemble des associations qu'il est possible de créer à partir de l'ensemble  $\mathbf{E}$  de la figure 10.5 est :

- pour les signatures avec 1 *evpo* :  $\{e_2^{(1)}, e_\emptyset^{(2)}, e_\emptyset^{(3)}\}, \{e_3^{(1)}, e_\emptyset^{(2)}, e_\emptyset^{(3)}\}, \{e_4^{(1)}, e_\emptyset^{(2)}, e_\emptyset^{(3)}\}, \{e_\emptyset^{(1)}, e_2^{(2)}, e_\emptyset^{(3)}\}, \{e_\emptyset^{(1)}, e_\emptyset^{(2)}, e_2^{(3)}\}, \{e_\emptyset^{(1)}, e_\emptyset^{(2)}, e_3^{(3)}\}$ .
- pour les signatures avec 2 *evpos* :  $\{e_2^{(1)}, e_2^{(2)}, e_\emptyset^{(3)}\}, \{e_3^{(1)}, e_2^{(2)}, e_\emptyset^{(3)}\}, \{e_4^{(1)}, e_2^{(2)}, e_\emptyset^{(3)}\}, \{e_2^{(1)}, e_\emptyset^{(2)}, e_2^{(3)}\}, \{e_2^{(1)}, e_\emptyset^{(2)}, e_3^{(3)}\}, \{e_3^{(1)}, e_\emptyset^{(2)}, e_2^{(3)}\}, \{e_3^{(1)}, e_\emptyset^{(2)}, e_3^{(3)}\}, \{e_4^{(1)}, e_\emptyset^{(2)}, e_2^{(3)}\}, \{e_4^{(1)}, e_\emptyset^{(2)}, e_3^{(3)}\}, \{e_\emptyset^{(1)}, e_2^{(2)}, e_2^{(3)}\}, \{e_\emptyset^{(1)}, e_2^{(2)}, e_3^{(3)}\}$ .
- pour les signatures avec 3 *evpos* :  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}\}, \{e_2^{(1)}, e_2^{(2)}, e_3^{(3)}\}, \{e_3^{(1)}, e_2^{(2)}, e_2^{(3)}\}, \{e_3^{(1)}, e_2^{(2)}, e_2^{(3)}\}, \{e_3^{(1)}, e_2^{(2)}, e_3^{(3)}\}, \{e_4^{(1)}, e_2^{(2)}, e_3^{(3)}\}$ .
- pour les signatures « fictives » :  $\{e_1^{(1)}, e_1^{(2)}, e_1^{(3)}\}$  et  $\{e_5^{(1)}, e_3^{(2)}, e_4^{(3)}\}$

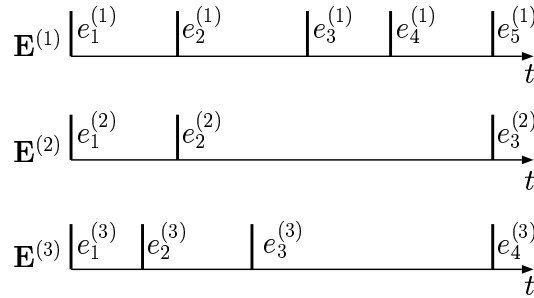


FIG. 10.5 – Exemple d'un ensemble de séquences d'événements  $\mathbf{E}$ .

### 10.3.2.3 Construction de l'ensemble des couples observation-état

L'ensemble des couples observation-état  $\mathcal{W} = \{\omega_1, \dots, \omega_K\}$  est construit à partir de l'ensemble  $\mathcal{V}$ . Pour chaque association de  $\mathcal{V}$ , un ensemble de couples observation-état peut être déterminé en définissant l'instant d'occurrence et la classe de l'événement en profondeur, ainsi que les temps d'occurrence des éventuels *evpøs*. A cette occasion, les variables aléatoires supposées continues jusqu'à présent (instants en profondeur et instants d'occurrence des *evpøs*) sont discrétisées.

L'ensemble  $\mathcal{W}$  est ensuite ordonné de manière à ce que la suite des valeurs de  $\{\tau_k\}_{k=1..K}$  soit croissante. En raison de la spécificité des états initiaux, on a  $n_1 = q_1$ ,  $n_K = q_C$  et  $\forall i \in [1, N], \tau_1^{(i)} = \tau_1 = 0$  et  $\tau_K^{(i)} = \tau_K = T$  ( $K$  est le cardinal de l'ensemble  $\mathcal{W}$ ).

### 10.3.2.4 Chemin caché $\mathbf{c}_\psi$

Un chemin caché  $\mathbf{c}_\psi$  est défini comme une suite ordonnée de couples observation-état  $\omega_k$ . Il est donc caractérisé par une fonction  $\psi_c$  injective de  $[1, l_c]$  dans  $[1, K]$ ,  $l_c$  désignant la longueur du chemin et  $K$  le cardinal de  $\mathcal{W}$  :

$$\mathbf{c}_\psi = \{\omega_{\psi_c(1)}, \omega_{\psi_c(2)}, \dots, \omega_{\psi_c(l_c)}\} \quad (10.14)$$

Pour être un chemin valide, la fonction  $\psi_c$  associée à  $\mathbf{c}_\psi$  doit vérifier les 4 conditions suivantes :

$$\begin{cases} \psi_c(1) = 1 \text{ et } \psi_c(l_c) = K & (i) \\ \forall j \in [1, l_c - 1], \tau_{\psi_c(j+1)} > \tau_{\psi_c(j)} & (ii) \\ \forall (i, j) \in [1, N] \times [1, l_c - 1], \tau_{\psi_c(j+1)}^{(i)} > \tau_{\psi_c(j)}^{(i)} & (iii) \\ \forall j \in [1, l_c - 1], a_{[n_{\psi_c(j)}][n_{\psi_c(j+1)}]} \neq 0 & (iv) \end{cases} \quad (10.15)$$

où  $a_{[q_i][q_j]} = a_{ij}$ . La contrainte (i) est liée à la spécificité de l'état initial et de l'état final alors que les contraintes (ii) et (iii) sont simplement une reformulation de l'équation (10.2.iii). Enfin, la contrainte (iv) est liée aux possibilités d'évolution du processus caché.

Nous avons ainsi atteint l'objectif dans la mesure où si l'on somme sur l'ensemble des chemins cachés  $\mathbf{c}_\psi$ , cela revient indirectement à sommer sur l'ensemble des observations (combinaison de  $\mathbf{E}$  avec les scénarios de  $\mathcal{S}(\mathbf{E})$ ) et sur l'ensemble des états cachés pouvant être associés à chacune des observations. Comme le processus est supposé markovien d'ordre 1, sommer sur l'ensemble des chemins cachés peut être réalisé en utilisant des algorithmes relativement proches des algorithmes standards. Dans ce but, les contraintes sur la fonction  $\psi_c$  sont reformulées localement au travers de deux ensembles : les ensembles  $\gamma_k$  et  $\gamma_k^{-1}$ ,  $k = 1 \dots K$ .

### 10.3.2.5 Ensemble $\gamma_k$ et $\gamma_k^{-1}$

On définit pour chaque couple  $\omega_k$  un ensemble d'indices  $\gamma_k$  représentatif des possibilités d'évolution du processus caché et observable. En pratique,  $\gamma_k$  est constitué de l'ensemble des indices des couples observation-état où peuvent aller les processus au « temps »  $u_o + 1$  sachant qu'ils sont au couple  $\omega_k$  au temps  $u_o$ .

De même, on définit  $\gamma_k^{-1}$  comme l'ensemble des indices des couples qu'ont pu emprunter les processus au temps  $u_o - 1$  pour arriver à  $\omega_k$  au temps  $u_o$ .

La figure 10.6 présente un ensemble  $\mathcal{W}$  de couples observation-état (les états ne sont pas explicitement mentionnés). Cet ensemble est ici constitué de 12 couples. Sans considérer la contrainte 10.15(iv), on obtient alors :

- $\gamma_2 = \{4, 5, 6, 7, 8, 9, 10, 11, 12\}$  et  $\gamma_2^{-1} = \{1\}$  ;
- $\gamma_6 = \{12\}$  et  $\gamma_6^{-1} = \{1, 2, 3\}$ .

Une fois les ensembles  $\gamma_k$  et  $\gamma_k^{-1}$  définies pour  $k = 1 \dots K$ , il est possible alors de résoudre les trois problèmes d'inférence.

## 10.4 Résolution du problème d'évaluation

Soit à déterminer la vraisemblance de  $\mathbf{E}$  sachant  $\lambda$  et  $\mathbf{w}$ . Afin de simplifier les écritures, nous notons par la suite  $b_{k,l}$  la vraisemblance que le processus soit au couple observation-état  $\omega_l$  au temps  $u + 1$  sachant qu'il était au couple observation-état  $\omega_k$  au temps  $u$ . En pratique,

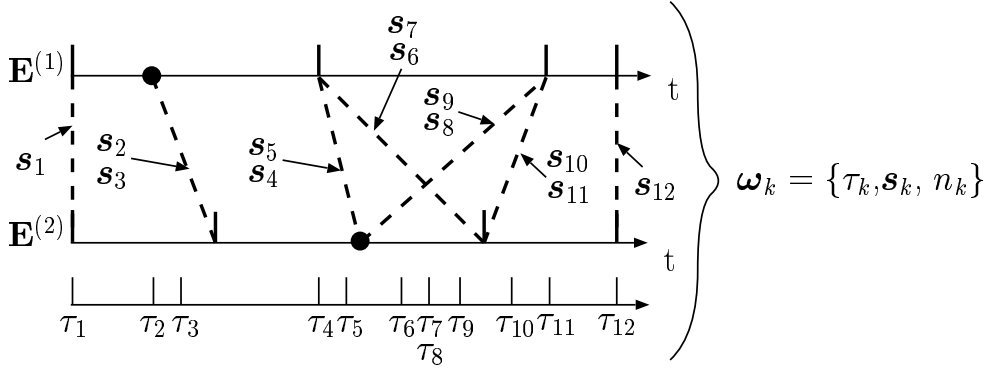


FIG. 10.6 – Ensemble  $\mathcal{W}$  de couples observation-état :  $\mathcal{W} = \{\omega_1, \dots, \omega_{12}\}$ . Les états ne sont pas explicitement mentionnés. Les ronds noirs représentent les non-détections, les traits verticaux les événements détectés et les traits en pointillés les signatures.

cette vraisemblance ne dépend pas de  $u$ . Conformément à l'équation 10.13, on obtient pour,  $k$  et  $l \in [1, K]^2$  :

$$b_{k,l} = b_{n_k n_l \mathbf{w}}(\tau_u = \tau_l, \tau_u^* = \tau_l^*, \mathbf{m}_u^* = \mathbf{m}_l^*, \mathbf{o}_{u,u-1}^* = \mathbf{E}_{k,l}^* | \tau_{u-1} = \tau_k, \tau_{u-1}^* = \tau_k^*, \mathbf{m}_{u-1}^* = \mathbf{m}_k^*) \quad (10.16)$$

où  $\mathbf{E}_{k,l}^*$  représente les événements de surface détectés entre les signatures  $\mathbf{s}_k$  et  $\mathbf{s}_l$ . Nous rappelons que le terme  $\tau_u$  de l'équation 10.16 représente l'instant d'occurrence du  $u$ -ème événement profond et en aucun cas le temps profond associé au  $u$ -ème couple signature-état. Il en est de même pour l'ensemble des variables ou vecteurs indicés par  $u$ . Cette notation est également conservée pour le reste du chapitre.

### 10.4.1 Evaluation directe

A partir de l'équation (10.16) et en notant  $\Gamma$  l'ensemble de tous les chemins valides  $\mathbf{c}_\psi$ , il vient :

$$\begin{aligned} P[\mathbf{E} | \boldsymbol{\lambda}, \mathbf{w}] &= \sum_{\mathbf{O} \in \mathcal{S}(\mathbf{E})} \sum_{\mathbf{X}} P[\mathbf{O}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}] \\ &= \sum_{\mathbf{c}_\psi \in \Gamma} \left( \prod_{i=1}^{l_c-1} a_{[n_{\psi_c(i)}][n_{\psi_c(i+1)}]} b_{\psi_c(i), \psi_c(i+1)} \right) \end{aligned} \quad (10.17)$$

où  $a_{[q_i][q_j]} = a_{ij}$ . Comme dans le cas standard, la propriété de Markov d'ordre 1 permet de mettre en place une procédure avant-arrière et donc de réduire de façon importante le coût calculatoire.

### 10.4.2 Evaluation récursive

La définition des probabilités avant et arrière,  $\boldsymbol{\alpha}$  et  $\boldsymbol{\beta}$ , s'écrit (pour  $k = 1 \dots K$ ) :

$$\begin{aligned} \alpha_{u,k} &= P[\mathbf{E}_{1,k}^* | \tau_u = \tau_k, \tau_u^* = \tau_k^*, \mathbf{m}_u^* = \mathbf{m}_k^*, X_u = n_k | \boldsymbol{\lambda}, \mathbf{w}] \\ \beta_{u,k} &= P[\mathbf{E}_{k,K}^* | X_u = n_k, \tau_u = \tau_k, \tau_u^* = \tau_k^*, \mathbf{m}_u^* = \mathbf{m}_k^*, \boldsymbol{\lambda}, \mathbf{w}] \end{aligned} \quad (10.18)$$

Le calcul de  $\boldsymbol{\alpha}$  se fait récursivement de la manière suivante :

$$\begin{array}{c}
\text{Initialisation} \\
\alpha_{1,1} = 1 \text{ et } \alpha_{1,l} = 0 \ (l = 2 \dots K) \\
\text{Récursion} \\
\text{pour } u = 1, 2 \dots, u_o - 1, \alpha_{u+1,l} = \sum_{k \in \gamma_l^{-1}} \alpha_{u,k} a_{[n_k][n_l]} b_{k,l} \ (l = 2 \dots K)
\end{array} \quad (10.19)$$

où  $u_o$  est déterminé lors du calcul des  $\alpha$  comme étant le plus petit entier naturel tel que  $\forall i \in [2, K - 1], \alpha_{u_o, i} = 0$ . Le calcul de  $\beta$  se fait également récursivement :

$$\begin{array}{c}
\text{Initialisation} \\
\forall u \in \mathbb{N}, \beta_{u,K} = 1 \text{ si } \alpha_{u,K} \neq 0, \text{ sinon } \beta_{u,K} = 0 \\
\text{Récursion} \\
\text{pour } u = u_o, u_o - 1, \dots, 1, \beta_{u,k} = \sum_{l \in \gamma_k} \beta_{u+1,l} a_{[n_k][n_l]} b_{k,l} \ (k = 1 \dots K - 1)
\end{array} \quad (10.20)$$

A partir de  $\alpha$  et de  $\beta$ , on obtient les relations suivantes :

$$\left\{ \begin{array}{ll}
\alpha_{u,K} = P[\mathbf{E}, \text{ le chemin caché a une longueur de } u | \boldsymbol{\lambda}, \mathbf{w}] & (i) \\
P[\mathbf{E} | \boldsymbol{\lambda}, \mathbf{w}] = \sum_u \alpha_{u,K} = \beta_{1,1} & (ii) \\
\alpha_{u,k} \beta_{u,k} = P[\mathbf{E}, X_u = n_k, \tau_u = \tau_k, \boldsymbol{\tau}_u^* = \boldsymbol{\tau}_k^*, \mathbf{m}_u^* = \mathbf{m}_k^* | \boldsymbol{\lambda}, \mathbf{w}] & (iii) \\
\sum_{k=1}^K \alpha_{u,k} \beta_{u,k} = P[\mathbf{E}, \text{ le chemin caché a au moins une longueur } \geq u | \boldsymbol{\lambda}, \mathbf{w}] & (iv)
\end{array} \right. \quad (10.21)$$

### 10.4.3 Exemple

Un exemple de calcul de  $\alpha$  et de  $\beta$  est donné ci-après. Il est également illustré figure 10.7. L'ensemble  $\mathcal{W}$  est formé de 6 couples :  $\mathcal{W} = \{\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_6\}$ . Les possibilités d'évolution du processus caché et observable s'écrivent enfin :  $\gamma_1 = \{2, 3, 4\}$ ,  $\gamma_2 = \{3\}$ ,  $\gamma_3 = \{5\}$ ,  $\gamma_4 = \{5\}$  et  $\gamma_5 = \{6\}$ . On pose enfin  $\epsilon_{ij} = a_{[n_i][n_j]} b_{ij}$ . Le calcul des  $\alpha$  et des  $\beta$  s'écrit finalement :

$$\begin{array}{ll}
\alpha_{1,1} = 1 & \beta_{5,6} = 1 \\
\alpha_{2,2} = \epsilon_{12}; \alpha_{2,3} = \epsilon_{13}; \alpha_{2,4} = \epsilon_{14} & \beta_{4,5} = \epsilon_{56}; \beta_{4,6} = 1 \\
\alpha_{3,3} = \alpha_{2,2}\epsilon_{23}; \alpha_{3,5} = \alpha_{2,4}\epsilon_{45} + \alpha_{2,3}\epsilon_{35} & \beta_{3,3} = \beta_{4,5}\epsilon_{35}; \beta_{3,5} = \beta_{4,6}\epsilon_{56} \\
\alpha_{4,5} = \alpha_{3,3}\epsilon_{35}; \alpha_{4,6} = \alpha_{3,5}\epsilon_{56} & \beta_{2,2} = \beta_{3,3}\epsilon_{23}; \beta_{2,3} = \beta_{3,5}\epsilon_{35}; \beta_{2,4} = \beta_{3,5}\epsilon_{45} \\
\alpha_{5,6} = \alpha_{4,5}\epsilon_{56} & \beta_{1,1} = \beta_{2,2}\epsilon_{12} + \beta_{2,3}\epsilon_{13} + \beta_{2,4}\epsilon_{14}
\end{array} \quad (10.22) \quad (10.23)$$

Toujours en considérant notre exemple, l'équation (10.21.i) s'écrit désormais :

$$\begin{array}{l}
P[\mathbf{E}, \text{ le chemin a une taille différente de 4 et de 5} | \boldsymbol{\lambda}, \mathbf{w}] = 0 \\
P[\mathbf{E}, \text{ le chemin a une taille de 4} | \boldsymbol{\lambda}, \mathbf{w}] = \alpha_{4,6} \\
P[\mathbf{E}, \text{ le chemin a une taille de 5} | \boldsymbol{\lambda}, \mathbf{w}] = \alpha_{5,6}
\end{array} \quad (10.24)$$

A partir de l'équation (10.21.ii), on obtient :

$$P[\mathbf{E} | \boldsymbol{\lambda}, \mathbf{w}] = \alpha_{4,6} + \alpha_{5,6} = \beta_{1,1} \quad (10.25)$$

Enfin, à partir de l'équation (10.21.iv) et comme les chemins de probabilité non nulle ont soit une longueur de 4 ou de 5, on a :

$$\begin{array}{l}
P[\mathbf{E} | \boldsymbol{\lambda}, \mathbf{w}] = \beta_{1,1} \\
= \alpha_{2,2}\beta_{2,2} + \alpha_{2,3}\beta_{2,3} + \alpha_{2,4}\beta_{2,4} = \alpha_{3,3}\beta_{3,3} + \alpha_{3,5}\beta_{3,5} = \alpha_{4,5}\beta_{4,5} + \alpha_{4,6}\beta_{4,6} \\
P[\mathbf{E}, \text{ le chemin a au moins une taille de 5} | \boldsymbol{\lambda}, \mathbf{w}] = \alpha_{5,6}
\end{array} \quad (10.26)$$

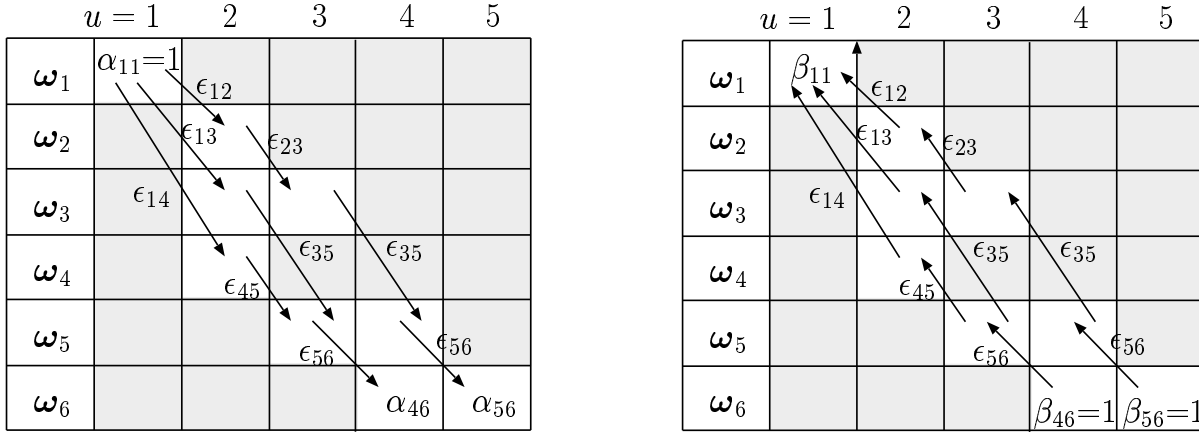


FIG. 10.7 – Représentation graphique du calcul des  $\alpha$  à gauche et des  $\beta$  à droite. A chaque case de coordonnée  $(i,j)$  est associée la valeur de  $\alpha_{i,j}$  ou de  $\beta_{i,j}$ . Les cases ombrées correspondent à des valeurs nulles. Sinon, la valeur associée à une case est obtenue en sommant sur l'ensemble des cases qui pointent sur cette dernière en pondérant par la valeur associée à la flèche.

## 10.5 Résolution du problème de décodage

Soit à déterminer, sachant  $\lambda$  et  $\mathbf{w}$ , la séquence d'observation et la séquence d'états qui expliquent au mieux  $\mathbf{E}$  au sens du MAP :

$$\{\mathbf{O}^{MAP}, \mathbf{X}^{MAP}\} = \operatorname{argmax}_{\mathbf{O} \in \mathcal{S}(\mathbf{E}), \mathbf{X}} P[\mathbf{O}, \mathbf{X} | \lambda, \mathbf{w}] \quad (10.27)$$

On note  $\Gamma_{u,k}$  l'ensemble des chemins cachés allant de  $\omega_1$  à  $\omega_k$  en émettant exactement  $u$  signatures. Chaque chemin  $\mathbf{c}_\psi \in \Gamma_{u,k}$  est caractérisé par une fonction  $\psi_c$  injective allant de  $[1,u]$  dans  $[1,k]$ .

$$\mathbf{c}_\psi = \{\omega_{\psi_c(1)}, \omega_{\psi_c(2)}, \dots, \omega_{\psi_c(u)}\} \quad (10.28)$$

Pour être valide, la fonction  $\psi$  doit vérifier les 4 conditions suivantes :

$$\begin{cases} \psi_c(1) = 1 \text{ et } \psi_c(u) = k & (i) \\ \forall j \in [1, u-1], \tau_{\psi_c(j+1)} > \tau_{\psi_c(j)} & (ii) \\ \forall (j, i) \in [1, u-1] \times [1, N], \tau_{\psi(j+1)}^{(i)} > \tau_{\psi(j)}^{(i)} & (iii) \\ \forall j \in [1, u-1], a_{[n_{\psi(j)}][n_{\psi(j+1)}]} \neq 0 & (iv) \end{cases} \quad (10.29)$$

On note  $\Psi_{u,k}(\mathbf{E})$  l'ensemble des fonctions injectives permettant de définir  $\Gamma_{k,u}$ . On définit finalement :

$$\delta_{u,k} = \max_{\psi \in \Psi_{u,k}(\mathbf{E})} P \left[ \{X_v = n_{\psi(v)}\}_{1 \leq v \leq u}, \{\mathbf{O}_v = \{\tau_{\psi(v)}, \boldsymbol{\tau}_{\psi(v)}^*, \mathbf{m}_{\psi(v)}^*, \mathbf{E}_{\psi(v-1)\psi(v)}^*\}\}_{1 \leq v \leq u} | \lambda, \mathbf{w} \right] \quad (10.30)$$

Le calcul de  $\delta$  s'effectue alors récursivement :

$$\begin{array}{c}
\text{Initialisation} \\
\delta_{1,1} = 1 \text{ et } \delta_{1,l} = 0 \ (l = 2 \dots K) \\
\text{Récursion} \\
\text{pour } u = 1, 2, \dots, u_o - 1, \delta_{u+1,l} = \max_{k \in \gamma_l^{-1}} \delta_{u,k} a_{[n_k][n_l]} b_{k,l} \ (l = 2 \dots K)
\end{array} \tag{10.31}$$

La détermination du chemin caché le plus vraisemblable est obtenue ensuite par des méthodes classiques de backtracking. A noter qu'on détermine les chemins les plus vraisemblables pour les différentes longueurs possibles. En effet, on a :

$$\delta_{u,K} = \max_{\mathbf{O} \in \mathcal{S}(\mathbf{E}), \mathbf{X}, L_X=u} P[\mathbf{O}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}] \tag{10.32}$$

où  $L_X$  représente la taille du chemin caché  $\mathbf{X}$ . Ainsi, dans l'exemple de la figure 10.7, on obtiendrait trois chemins : le chemin le plus probable de longueur 4 ( $C_1$ ), le chemin le plus probable de longueur 5 ( $C_2$ ) et le chemin le plus probable ( $C_1$  ou  $C_2$ ).

## 10.6 Résolution du problème d'apprentissage

Le corpus d'apprentissage est composé de  $R$  ensembles de séquences d'événements  $\mathbf{E}^1, \mathbf{E}^2, \dots, \mathbf{E}^R$  supposés mutuellement indépendants, chaque ensemble de séquences étant constitué de  $N$  séquences d'événements. L'apprentissage suivant le critère du MV conduit à résoudre l'équation suivante :

$$\boldsymbol{\lambda}^* = \operatorname{argmax}_{\boldsymbol{\lambda}} \prod_{r=1}^R P[\mathbf{E}^r | \boldsymbol{\lambda}, \mathbf{w}] \tag{10.33}$$

### 10.6.1 Définition et décomposition de la fonction auxiliaire

L'ensemble  $\mathbf{E}$  correspond aux données observées alors que la fonction  $\psi$  définit les états cachés et la partie observable manquante. La fonction auxiliaire peut alors s'écrire :

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') = \sum_{r=1}^R \sum_{\psi \in \Psi^r} P[\psi | \mathbf{E}^r, \boldsymbol{\lambda}, \mathbf{w}] \log P[\psi, \mathbf{E}^r | \boldsymbol{\lambda}', \mathbf{w}] \tag{10.34}$$

$\Psi^r$  représente l'ensemble des fonctions permettant de définir l'ensemble des chemins cachés explicatifs de  $\mathbf{E}^r$ . On obtient ainsi :

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}') = \sum_{r=1}^R \sum_{(k,l) \in [1,K^r]^2} \sum_{u \in \mathbb{N}} p^r[k, l, u] \log a'_{[n_k][n_l]} + \sum_{r=1}^R \sum_{(k,l) \in [1,K^r]^2} \sum_{u \in \mathbb{N}} p^r[k, l, u] \log b'_{k,l} \tag{10.35}$$

avec

$$p^r[k, l, u] = P[\psi(u) = k, \psi(u+1) = l, | \mathbf{E}^r, \boldsymbol{\lambda}, \mathbf{w}] \tag{10.36}$$

L'équation (10.35) est obtenue à partir des équations (10.37.i) et (10.37.ii) où  $\delta_\psi(k, l, u)$  vaut 1 si  $\psi(u) = k$  et  $\psi(u+1) = l$ , 0 sinon.

$$\begin{array}{ll}
\log P[\psi, \mathbf{E}^r | \boldsymbol{\lambda}'] = \sum_{(k,l) \in [1,K^r]^2} \sum_{u \in \mathbb{N}} \delta_\psi(k, l, u) (\log a'_{[n_k][n_l]} + \log b'_{k,l}) & (i) \\
\sum_{\psi \in \Psi^r} \delta_\psi(k, l, u) P[\psi | \mathbf{E}^r, \boldsymbol{\lambda}, \mathbf{w}] = p^r[k, l, u] & (ii)
\end{array} \tag{10.37}$$

### 10.6.2 Optimisation de la fonction auxiliaire

Conformément à l'équation (10.35), la fonction auxiliaire se présente comme une somme de deux termes optimisables séparément. L'optimisation du premier terme conduit à la formule de réestimation de la probabilité de transition :

$$a'_{ij} = \frac{\sum_{r=1}^R \sum_{k,l \in [1, K^r]^2; n_k^r = q_i, n_l^r = q_j} \sum_{u \in \mathbb{N}} p^r[k, l, u]}{\sum_{r=1}^R \sum_{k,l \in [1, K^r]^2; n_k^r = q_i} \sum_{u \in \mathbb{N}} p^r[k, l, u]} \text{ avec } p^r[k, l, u] = \frac{\alpha_{u,k} a_{[n_k][n_l]} b_{k,l} \beta_{u+1,l}}{P[\mathbf{E}^r | \boldsymbol{\lambda}, \mathbf{w}]} \quad (10.38)$$

L'optimisation du second terme relatif aux lois d'observation dépend intimement de la définition de ces lois. En l'état, il est difficile de résoudre le problème d'optimisation dans un cadre général. Toutefois, nous rappelons qu'il est possible d'utiliser des méthodes numériques pour optimiser la fonction auxiliaire.

## 10.7 Stratégie de diminution de la combinatoire

La construction de l'ensemble  $\mathcal{W}$  nécessite de déterminer au préalable l'ensemble  $\mathcal{V}$  des associations possibles entre des événements de surface. On rappelle qu'une association est constituée d'événements de surface pouvant éventuellement appartenir à une même signature. La construction de  $\mathcal{V}$  est problématique dans la mesure où le cardinal de cet ensemble devient vite très important quand le nombre de canaux  $N$  d'observation et le nombre de détections par canal augmentent. La valeur de ce cardinal est donnée équation (10.39) en prenant en compte la spécificité des événements fictifs :

$$|\mathcal{V}| = \sum_{k=1}^N \left[ \sum_{(c_1, \dots, c_k) \in \text{Comb}_k([1, N])} \prod_{i=1}^k (L_{c_i} - 2) \right] + 2 \quad (10.39)$$

Les termes  $c_1, \dots, c_k$  sont une combinaison de  $k$  valeurs prises dans l'ensemble  $[1, N]$  et  $L_{c_i}$  représente le nombre de détections dans le canal  $c_i$  sur l'intervalle de mesure considéré  $[0, T]$ . Pour avoir un ordre de grandeur du cardinal  $|\mathcal{V}|$ , on se place dans le cas où  $\forall i \in [1, N], L_i = L$ . L'équation (10.39) conduit alors à :

$$|\mathcal{V}| = \sum_{k=1}^N C_N^k (L - 2)^k + 2 \quad (10.40)$$

Le tableau 10.1 rassemble les valeurs de ce cardinal pour différentes valeurs de  $N$  et de  $L$ . Dans

$L \setminus N$	2	3	4	5	10	20	80
10	82	730	6562	59050	3.48E9	1.21E19	2.18E76
20	362	6860	130322	2.47E6	6.13E12	3.75E25	1.99E102
30	842	24390	707282	2.05E07	4.20E14	1.76E29	9.81E116
50	2402	117650	5.76E6	2.82E8	7.97E16	6.36E33	1.64E135
100	9802	970300	9.60E7	9.50E9	9.04E19	8.17E39	4.47EE159

TAB. 10.1 – Valeurs du cardinal de  $\mathcal{V}$  en fonction du nombre de canaux d'observation  $N$  et du nombre  $L$  d'événements détectés par canal.

une application de cartographie typique avec un paradigme événementiel,  $L$  est de l'ordre de



50 alors que  $N$  est égal à 81 (26-connexité et 3 détecteurs opérant en parallèle). Le cardinal de  $\mathcal{V}$  est alors de l'ordre de 1.64E135. Il est alors impossible de sommer sur l'ensemble  $\mathcal{V}$  tout entier. Différentes contraintes sont donc utilisées pour diminuer son cardinal.

### 10.7.1 Contrainte de nombre

Les associations possédant un nombre réduit d'événements peuvent être ignorés. Ainsi, cela revient à ne pas considérer les signatures possédant un nombre réduit d'*evpos*. Le nombre d'associations possédant au moins  $Min$  événements est donné équation (10.41.i), dans le cas général, et équation (10.41.ii) dans le cas où  $\forall i \in [1, N], L_i = L$  :

$$\begin{aligned} |\mathcal{V}| &= \sum_{k=Min}^N \left[ \sum_{(c_1, \dots, c_k) \in Comb_k([1, N])} \prod_{i=1}^k (L_{c_i} - 2) \right] + 2 & (i) \\ |\mathcal{V}| &= \sum_{k=Min}^N C_N^k (L - 2)^k + 2 & (ii) \end{aligned} \quad (10.41)$$

Le tableau 10.2 représente le nombre d'associations possédant au moins  $Min$  événements observés en fonction du nombre  $N$  de canaux et du nombre  $L$  de détections par canal. On remarque que le nombre est réduit de manière peu significative malgré les grandes valeurs de  $Min$  choisies. Le choix des contraintes de nombre peut être guidé par des connaissances *a*

$L \setminus N(Min)$	2(2)	3(2)	4(3)	5(4)	10(9)	20(19)	80(79)
10	66	706	6146	53250	2.41E9	4.03E18	1.94E73
20	326	6806	128306	2.41E6	5.55E12	2.69E25	1.43E101
30	786	24306	702466	2.02E7	4.01E14	1.50E29	2.28E116
50	2306	117506	5.75E6	2.81E8	7.84E16	5.97E33	8.41E134
100	9606	970006	9.60E7	9.50E9	9.00E19	8.03E39	3.60E159

TAB. 10.2 – Cardinal de  $\mathcal{V}$  avec une contrainte de nombre.

*priori* ou par les lois d'observation.

### 10.7.2 Contrainte de forme

La forme des associations au travers des canaux peut constituer un second critère de diminution de la combinatoire. Elle donne des informations sur l'asynchronisme éventuel entre détections d'une même signature. Par exemple, la figure 10.8 donne un exemple de 6 associations qu'il est possible de construire à partir d'un ensemble  $\mathbf{E}$  de séquences d'événements. On note que par exemple que la troisième association  $\mathcal{V}_3$  est « droite » alors que la cinquième ne l'est pas du tout.

Dans le cadre de la cartographie en IRMf cérébrale, aucun asynchronisme n'est attendu entre les différents canaux (la justification en étant donnée dans le prochain chapitre). Il est donc possible de considérer uniquement les associations pour lesquelles  $t_{max} - t_{min}$  est petit,  $t_{min}$  et  $t_{max}$  représentant les temps de surface des événements de l'association survenus le plus tôt et le plus tard respectivement (cf. association  $\mathcal{V}_5$  de la figure 10.8). On note *extension* la distance temporelle maximale autorisée entre  $t_{max}$  et  $t_{min}$ . Il est à noter que la valeur du seuil *extension* peut être guidée par des connaissances *a priori* ou par les lois d'observation sur le processus observable.

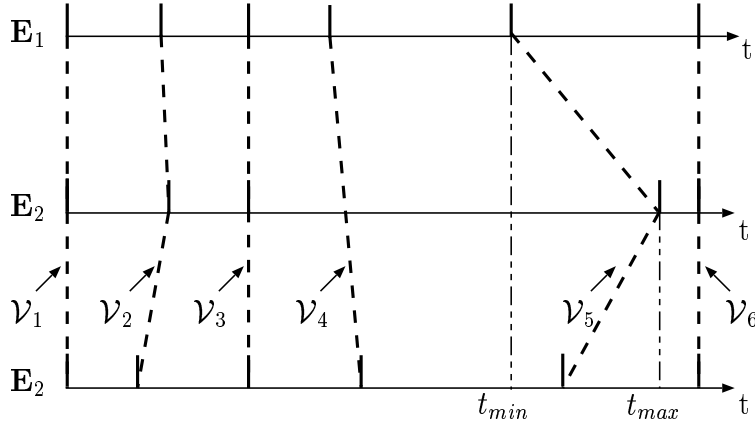


FIG. 10.8 – Exemple d'associations possibles. Les états ne sont pas explicitement mentionnés.

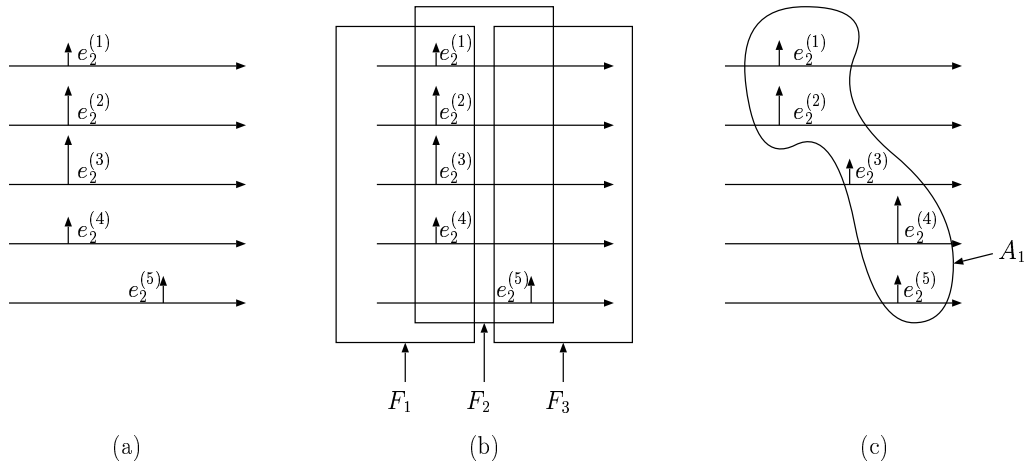


FIG. 10.9 – Contrainte de maximalité. (a) Exemple de cas où la contrainte de maximalité est trop forte (voir texte) (b) Exemple de contrainte de maximalité locale (voir texte) (c) Assouplissement de la contrainte de maximalité locale (voir texte)

### 10.7.3 Contrainte de maximalité

La contrainte de maximalité sous-tend un principe d'association effectué de manière à obtenir des associations de tailles les plus importantes possibles. Une association est dite *maximale* s'il n'est pas possible de rajouter un événement observé à l'association. Pour l'explicitier, supposons que  $N = 20$  canaux et que dans chaque canal, un unique événement de surface soit détecté au même instant  $t$ , alors, il existe encore  $\sum_{p=15}^{20} C_{20}^p = 21700$  associations non fictives avec une contrainte de taille ( $Min=15$ ) et une contrainte de forme ( $extension=1$ ). Ne considérer que les signatures maximales permet dans l'exemple précédent de passer de 21700 possibilités à une unique possibilité!

Il peut cependant apparaître des cas où la contrainte de maximalité est trop restrictive. Il existe par exemple dans le cas de la figure 10.9.a une unique association possible :  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}, e_2^{(4)}, e_2^{(5)}\}$  sous l'hypothèse que  $t_2^5 - t_2^1 < extension$ . Or, il serait cependant plus judicieux de laisser au modèle statistique la possibilité d'analyser plusieurs associations dont notamment l'association  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}, e_2^{(4)}, \phi\}$ .

Une méthode consiste à utiliser une contrainte de maximalité locale : on détermine l'ensemble des associations maximales lorsque l'on voit les différentes séquences au travers d'une fenêtre de taille  $l$ . Avec le même exemple (figure 10.9.b), selon la position de la fenêtre, on obtient 3 associations différentes :  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}, e_2^{(4)}, \phi\}$  pour la fenêtre  $F_1$ ,  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}, e_2^{(4)}, e_2^{(5)}\}$  pour  $F_2$  et l'association  $\{\phi, \phi, \phi, \phi, e_2^{(5)}\}$  pour  $F_3$ . Pour une fenêtre donnée, il est possible de déterminer le cardinal de l'ensemble des associations maximales. Si ce nombre est trop important, on peut choisir aléatoirement un nombre donné de signatures maximales ou de considérer uniquement les associations les plus vraisemblables.

La contrainte de maximalité locale peut encore être assouplie. Une fois une association maximale locale  $\mathbf{V}_i$  créée,  $d_i$  nouvelles associations non maximales  $\Delta_i = \{\mathbf{V}_i^1, \dots, \mathbf{V}_i^{d_i}\}$  peuvent être déterminées à partir de  $\mathbf{V}_i$  :

$$\begin{cases} \forall (m, n) \in [1, d_i]^2, m \neq n, \mathbf{V}_i^m \neq \mathbf{V}_i^n \neq \mathbf{V}_i \\ \forall m \in [1, d_i], \mathbf{V}_i^m \cap \mathbf{V}_i = \mathbf{V}_i^m \end{cases} \quad (10.42)$$

Des contraintes peuvent être utilisées lors de la création de l'ensemble  $\Delta_i$ . Par exemple, les événements observés au même instant peuvent être considérés comme indissociables pour la formation des signatures non maximales. Ainsi, dans l'exemple de la figure 10.9.c, l'association  $A_1$  permet de créer 6 associations non maximales :  $\{e_2^{(1)}, e_2^{(2)}, \phi, \phi, \phi\}$ ,  $\{\phi, \phi, \phi, e_2^{(4)}, e_2^{(5)}\}$ ,  $\{\phi, \phi, e_2^{(3)}, \phi, \phi\}$ ,  $\{e_2^{(1)}, e_2^{(2)}, \phi, e_2^{(4)}, e_2^{(5)}\}$ ,  $\{e_2^{(1)}, e_2^{(2)}, e_2^{(3)}, \phi, \phi\}$  et  $\{\phi, \phi, e_2^{(3)}, e_2^{(4)}, e_2^{(5)}\}$ . Il est possible également d'utiliser des contraintes probabilistes. A partir d'une association maximale, on peut déterminer un ensemble d'associations non maximales les plus vraisemblables.

### 10.7.4 Conclusion

En résumé, il semble difficile en pratique de sommer sur l'ensemble des associations  $\mathbf{V}$  eu égard à sa taille très importante. Aussi, des contraintes d'association, relevant de la fusion de données, peuvent être envisagées pour réduire cet ensemble de manière à conserver les associations les plus vraisemblables. On rappelle enfin que l'ensemble  $\mathbf{W}$  est construit à partir de chaque association de l'ensemble  $\mathbf{V}$ . A partir de chacune des associations, il est nécessaire de définir l'instant en profondeur, l'état caché et le temps de surface des éventuels événements non détectés. Plus généralement, la construction de l'ensemble  $\mathbf{W}$  dépend fortement de l'application et elle est directement liée à la construction de l'ensemble  $\mathbf{V}$  (les différentes contraintes peuvent être différentes selon l'état caché associé à la signature). L'objectif reste cependant le même : obtenir l'ensemble des couples observation-état les plus vraisemblables. Nous avons ici présenté trois contraintes : une contrainte de taille, une contrainte de forme et une contrainte de maximalité. Ces contraintes sont utilisées dans notre application.

On rappelle enfin que l'ensemble des couples observation-état est ensuite fourni au modèle statistique qui analyse ces dernières. Pour la résolution du problème de l'évaluation, il est important d'utiliser des contraintes peu restrictives de manière à approcher le plus convenablement possible la vraisemblance désirée (moins les contraintes sont restrictives, plus on somme sur un nombre important de configurations et donc plus on s'approche ainsi de la vraisemblance recherchée). Au contraire, le problème du décodage, pour obtenir des résultats satisfaisants, ne nécessite pas obligatoirement l'utilisation de contraintes trop souples dans la mesure où il n'est pas nécessaire de perdre son temps à analyser des couples signature-état peu vraisemblables lorsque l'on cherche la configuration finale la plus vraisemblable. Cette manière de faire reste tout de même heuristique dans la mesure où la « solution » peut être constituée d'un couple observation-état peu vraisemblable. En termes d'apprentissage, la très

grande taille de l'espace d'intégration nous incite enfin à utiliser des contraintes relativement fortes de manière à obtenir un nombre réduit de configurations possibles sur lesquelles l'apprentissage peut alors s'effectuer correctement.

# Chapitre 11

## Application en cartographie IRMf cérébrale

### 11.1 Présentation générale

Nous rappelons ici notre hypothèse de travail : en présence d'activité fonctionnelle au voxel  $v$ , les séquences constituant  $\mathbf{E}^v$  doivent s'aligner, dans une certaine mesure, sur la séquence de transitions « off-on » du paradigme. L'ensemble des séquences  $\mathbf{E}^v$  est obtenu par prétraitement du signal IRMf local à  $v$ ,  $f^v(\cdot)$ , et des signaux avoisinant le voxel,  $f^{v'}(\cdot)$ , où  $v' \in \mathcal{N}(v)$ , le tout dans un contexte multi-détecteurs. Le problème d'alignement est résolu au travers du modèle markovien caché de multiples séquences d'événements (MMCMSE). Le processus profond du MMCMSE est le processus neuronal prenant place au voxel  $v$ . Les processus de surface qui lui sont corrélés sont les signaux IRMf observés à  $v$  et dans le voisinage spatial  $\mathcal{N}(v)$  de  $v$ . En cas d'activité, les événements du processus profond doivent s'aligner dans une certaine mesure sur les transitions « off-on » du paradigme.

Deux grandes différences sont à noter par rapport à l'approche scalaire à une seule séquence développée dans la partie précédente :

(i) plusieurs séquences d'événements sont désormais considérées dans le calcul de l'activité locale au voxel  $v$ , prenant ainsi en compte l'information spatiale de voisinage dans un contexte multi-détecteurs. On espère ainsi gagner en robustesse par rapport à l'approche scalaire.

(ii) On ne cherche pas comme auparavant à définir un modèle d'activité représentatif de l'ensemble des signaux actifs. On définit plutôt un modèle d'activité propre à chaque voxel  $v$ . La démarche est la suivante. L'espace paramètre du modèle est tout d'abord contraint à représenter l'« activité cérébrale ». Par exemple, en ce qui concerne la matrice de transition, on contraint ces paramètres de manière à ce que la probabilité de ne pas sauter d'états soit supérieure ou égale à celle de sauter un état. On considère en effet que sauter des états est plus représentatif de l'inactivité que de l'activité. Un ensemble de contraintes est ainsi utilisé sur l'ensemble des paramètres. Les paramètres du modèle sont ensuite appris au sens du maximum de vraisemblance :

$$\boldsymbol{\lambda}^{v,\star} = \arg \max_{\boldsymbol{\lambda}} P[\mathbf{E}^v | \boldsymbol{\lambda}] \quad (11.1)$$

En cas d'inactivité, à cause de la contrainte sur l'espace des paramètres, la méthode d'apprentissage n'arrive pas à déterminer un modèle pouvant expliquer  $\mathbf{E}^v$ . Elle converge ainsi vers un modèle peu contraint, les probabilités se répartissant sur l'ensemble de l'espace d'observation. La vraisemblance  $P[\mathbf{E}^v | \boldsymbol{\lambda}^{v,\star}]$  est alors faible. A l'inverse, en cas d'activité, la méthode d'apprentissage fournit un modèle expliquant bien l'observation  $\mathbf{E}^v$  : la vraisemblance  $P[\mathbf{E}^v | \boldsymbol{\lambda}]$  est

alors élevée.

Le synoptique de traitement est représenté à la figure 11.1. Il est valable pour chaque voxel  $v$  de la séquence IRMf à analyser. Dans une première étape, l'ensemble des séquences  $\mathbf{E}^v$  associé au voxel  $v$  est obtenu par prétraitement du signal IRMf  $f^v(\cdot)$  et des signaux  $f^{v'}(\cdot)$  où  $v' \in \mathcal{N}(v)$  (section 11.2). Dans une seconde étape, les états cachés et la topologie du modèle sont définis à partir du paradigme de stimulation (section 11.3). Ensuite, on estime conjointement les paramètres du modèle associé au voxel  $v$  ainsi que l'activité cérébrale locale au voxel  $v$  dans le but d'obtenir la carte d'activation (section 11.4).

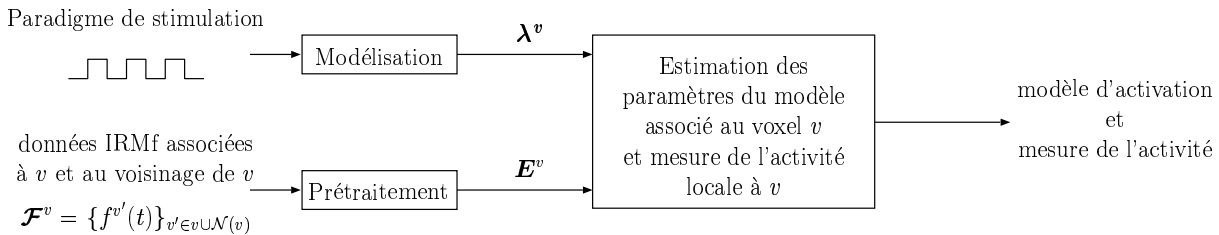


FIG. 11.1 – Synoptique de traitement au voxel  $v$

## 11.2 Prétraitement du signal IRMf

Le signal IRMf local à  $v$ ,  $f^v(\cdot)$ , et les signaux IRMf voisins,  $f^{v'}(\cdot)$ ,  $v' \in \mathcal{N}(v)$  sont analysés par un banc de détecteurs de RDs. Pour un voisinage de type 26-connexité et avec  $K=3$  détecteurs, l'ensemble  $\mathbf{E}^v$  est formé de  $3 \times (26 + 1) = 81$  séquences d'événements. Les RDs sont détectées avec la même méthode que précédemment. Pour chaque voxel, plusieurs séquences de RDs sont obtenues pour différentes échelles de décomposition en ondelettes du signal IRMf. De plus, pour simplifier le processus observable, les mesures  $m^{(\cdot)}$  associées aux événements sont ignorées.

A titre d'exemple, l'ensemble  $\mathbf{E}^v$  de séquences de RDs détectées pour un voisinage de type 6-connexité et trois détecteurs opérant en parallèle est illustré figure 11.2 pour un signal actif en  $v$ , et figure 11.3 pour un signal inactif en  $v$ . Le signal local à  $v$  est de couleur orange alors que les signaux dans son voisinage sont de couleur verte. A chaque signal IRMf correspondent trois séquences de RDs indiquées par des croix. Les trois séquences ont été obtenues, de haut en bas, pour un taux de détection  $k$  égal à 0.7, 1 et 1.4. Les bonnes détections sont encadrées et celles associées à une même transition du paradigme sont reliées entre elles (résultat fourni par l'algorithme que nous allons présenter). Les transitions du paradigme sont représentées par des lignes verticales marrons, ici au nombre de huit.

## 11.3 Modélisation

### 11.3.1 Processus caché

La modélisation du processus caché se fait à partir de la séquence de transition « off-on » du paradigme (pour les paradigmes en blocs) et à partir des différents stimuli pour les paradigmes événementiels (figure 11.4). Par la suite, on ne considère que le cas des paradigmes événementiels en sachant que ce qui est écrit est valide également pour un paradigme en bloc.

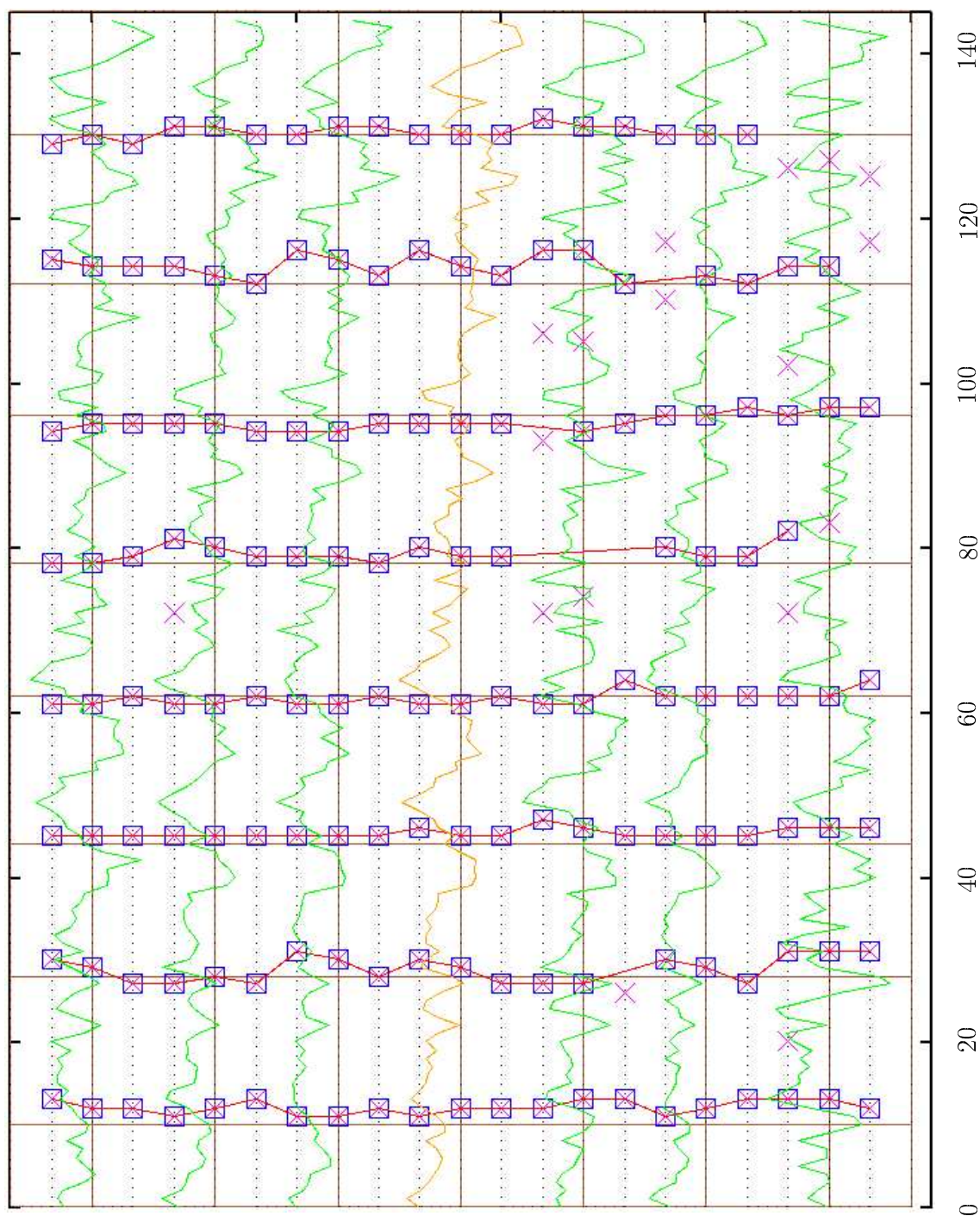


FIG. 11.2 – Ensemble  $E^v$  de séquences de RDs obtenu pour un signal actif au sortir de l'étape de prétraitement. Les associations entre les différentes RDs ont été quant à elles obtenues après lancement de l'algorithme présenté dans ce chapitre.

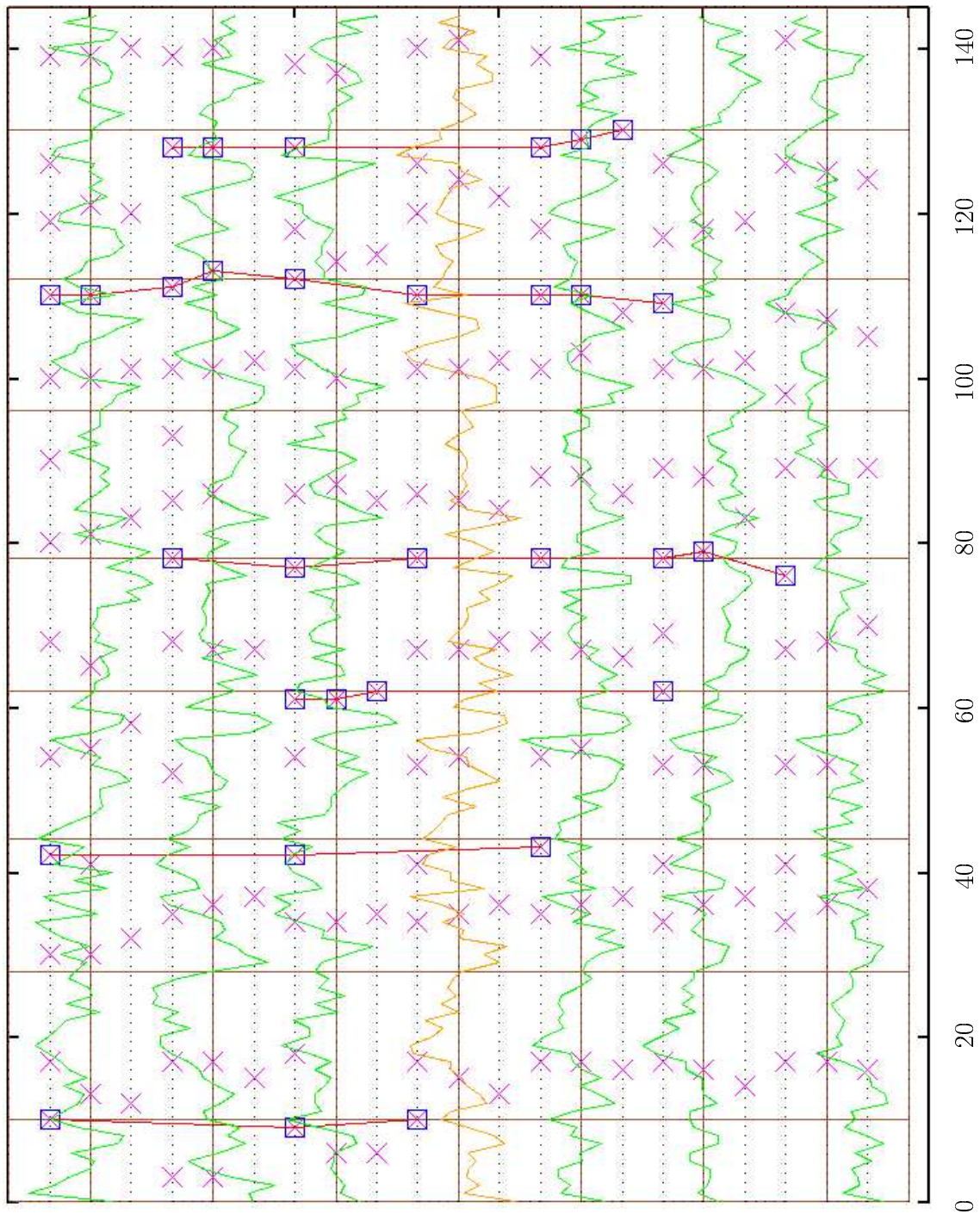


FIG. 11.3 – Ensemble  $\mathbf{E}^v$  de séquences de RDs obtenu pour un signal inactif au sortir de l'étape de prétraitement. Les associations entre les différentes RDs ont été quant à elles obtenues après lancement de l'algorithme présenté dans ce chapitre.



Nous rappelons qu'au maximum un événement profond peut être associé à une stimulation du paradigme et qu'au maximum une stimulation du paradigme peut être associée à un événement profond. Les états cachés et les possibilités d'évolution du processus caché sont ensuite spécifiés de la façon suivante :

- *Etats cachés* : un état caché  $q_i$  est associé à chaque stimulus du paradigme. L'indice  $i$  est représentatif de l'ordre d'apparition du stimulus dans le paradigme créant ainsi une bijection entre les états cachés et les stimuli. Plus précisément, l'état  $q_i$  modélise l'événement profond induit par le  $i$ -ème stimulus. Les états initiaux et finaux  $q_1$  et  $q_C$  permettent de modéliser le début et la fin de la séquence. Cela revient à rajouter deux stimuli fictifs au début et à la fin du paradigme de stimulation. Si  $P$  désigne le nombre de stimuli du paradigme, stimuli fictifs inclus, alors le cardinal  $C$  de l'espace d'états du modèle est égal à  $P$ .
- *Matrice de transition* : La topologie de la chaîne est de type gauche-droite, avec pour contrainte supplémentaire  $a_{ii} = 0, \forall i$ . Comme précédemment, on ajoute la contrainte :  $a_{ij} = 0$  si  $j > i + \Delta$  avec  $\Delta \sim 4$ . Un exemple de topologie de chaîne est représenté figure 11.4 pour  $\Delta = 2$  et  $C = 5$ .

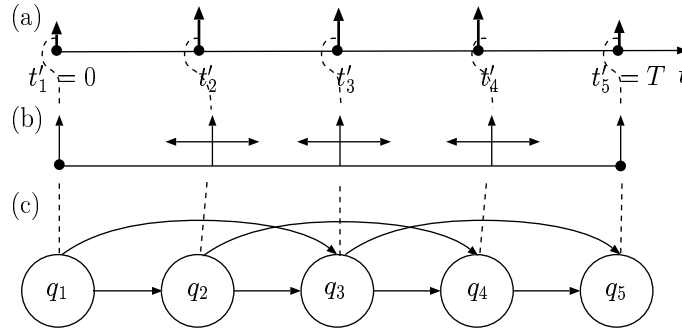


FIG. 11.4 – Modélisation de l'activité cérébrale. (a) Paradigme de stimulation. (b) Représentation de la séquence des événements profonds induite par le paradigme de stimulation (les flèches horizontales montrent l'incertitude des temps d'occurrence des événements). (c) Modèle d'activation.

### 11.3.2 Processus observable

#### 11.3.2.1 Contrainte de sommation des lois d'observation

Les contraintes liées à l'évolution du processus observable ( $\tau_u > \tau_{u-1}$  et  $\forall i, \tau_u^{(i)} > \tau_{u-1}^{(i)}$ ) sont codées au travers des lois d'observation. En effet, on a défini les lois  $b_{1j\mathbf{w}}(\boldsymbol{\tau}_u^* | \tau_u, \boldsymbol{\tau}_{u-1}^*)$  et  $b_{0ij\mathbf{w}}(\tau_u | \tau_{u-1})$  permettant d'imposer par exemple la contrainte  $b_{0ij\mathbf{w}}(\tau_u | \tau_{u-1}) = 0$  si  $\tau_u \leq \tau_{u-1}$ . On vérifie par conséquent la contrainte de sommation à 1 :  $\sum_{\mathbf{O}} P[\mathbf{O} | \boldsymbol{\lambda}] = 1$  où  $\mathbf{O}$  est une observation du MMCMSSE.

En pratique, vérifier la contrainte de sommation à 1 peut être effectuée de la manière suivante. Considérons toujours l'exemple de la loi  $b_{0ij\mathbf{w}}(\tau_u | \tau_{u-1})$ . Une loi  $b_{0ij\mathbf{w}}(\tau_u)$  ne permettant pas de vérifier les contraintes de sommation à 1 est définie. Il suffit alors d'écrire (cf. figure 11.5) :

$$b_{0ij\mathbf{w}}(\tau_u | \tau_{u-1}) = \begin{cases} \frac{1}{1 - \int_{t=-\infty}^{\tau_{u-1}} b_{0ij\mathbf{w}}(t) dt} b_{0ij\mathbf{w}}(\tau_u) & \text{si } \tau_u > \tau_{u-1} \\ 0 & \text{sinon} \end{cases} \quad (11.2)$$

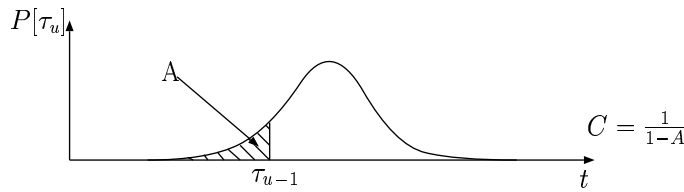


FIG. 11.5 – Lois d'observation et contrainte. Sous la contrainte  $\tau_u > \tau_{u-1}$ , la loi d'observation  $P[\tau_u]$  doit être multipliée par un coefficient multiplicateur  $C = \frac{1}{1-A}$  où  $A$  est l'aire hachurée sous la courbe (cf. équation (11.2)).

L'équation (11.2) met clairement en évidence que la loi sans contrainte est renormalisée de manière à intégrer à 1 sur l'intervalle  $[\tau_{u-1}, +\infty[$ . Une telle renormalisation n'est cependant pas nécessaire pour les paradigmes en blocs dans la mesure où la durée séparant deux événements profonds est très grande devant l'incertitude existant sur l'instant de ces derniers. Toutefois, une telle procédure de renormalisation devrait être utilisée pour les paradigmes événementiels car la durée séparant deux événements profonds peut être très faible. Néanmoins, la procédure de renormalisation n'est pas effectuée dans la mesure où elle attribue intrinsèquement une vraisemblance plus importante à certains stimuli qu'à d'autres en fonction de leur distance relative (cf. terme  $1/(1 - \int_{t=-\infty}^{\tau_u-1} b_{0ij\mathbf{w}}(t)dt)$ ). Par conséquent, ce terme est fixé à 1. Les lois utilisées n'explicitent donc plus les contraintes liées à l'évolution du processus observable. On n'utilise par exemple plus que la loi  $b_{0ij\mathbf{w}}(\tau_u)$ . Ainsi, des séquences d'observation  $\mathbf{O}$  ne respectant pas les contraintes pourront désormais avoir une vraisemblance non nulle. Il s'ensuit que la contrainte de sommation à 1 n'est plus respectée :  $\sum_{\mathbf{O}} P[\mathbf{O}|\boldsymbol{\lambda}] < 1$ . Cependant, lors du calcul de  $P[\mathbf{E}]$ , la sommation  $\sum_{\mathbf{O} \in \mathcal{S}(\mathbf{E})} P[\mathbf{O}|\boldsymbol{\lambda}]$  ne s'effectue que sur des observations valides car les contraintes sont toujours prises en compte lors de l'élaboration des chemins cachés.

### 11.3.2.2 Prise en compte du retard d'activité

En présence d'un retard d'activité locale  $l$ , on suppose que la séquence d'événements profonds doit s'aligner « idéalement » sur la séquence de stimuli du paradigme translatée de  $l$ . Conscient des limites d'une modélisation des durées du type inter-événements, une variable aléatoire  $L$  dont la valeur correspond au retard d'activité est introduite. Ainsi, le terme  $b_{0ij\mathbf{w}}(\tau_u|\tau_{u-1}) = P[\tau_u|\tau_{u-1}, X_{u-1} = q_i, X_u = q_j, \mathbf{w}]$  de l'équation (10.7) page 141 est remplacé par le terme  $b_{0jl\mathbf{w}}(\tau_u) = P[\tau_u|X_u = q_j, L = l, \mathbf{w}]$ . A noter qu'il aurait fallu définir le terme  $P[\tau_u|\tau_{u-1}, X_u = q_j, L = l, \mathbf{w}]$  pour pouvoir vérifier la contrainte de sommation à 1. Le graphe de dépendance du modèle est finalement défini à la figure 11.6.

### 11.3.2.3 Expression finale des lois d'observation

Les lois définies n'ont pas pour objectif de caractériser au mieux les données mais de permettre une bonne discrimination entre signaux actifs et signaux inactifs. La densité de probabilité s'exprime à partir de trois fonctions caractéristiques  $f^i(\cdot)_{1 \leq i \leq 3}$  conformément à l'équation (10.8), page 141. Cependant, on ne considère pas les termes grisés, ni les données utiles à la vérification des contraintes de sommation à 1. Les termes grisés ne sont pas considérés car ils sont « non informatifs », l'information désirée ayant déjà été obtenue à travers les fonctions caractéristiques. Aussi, ils sont considérés comme constants. Même constants, ils ont une

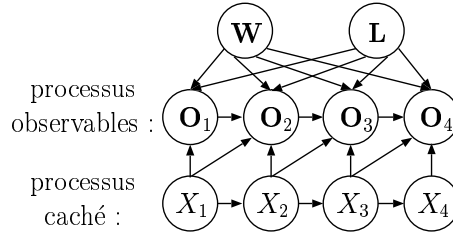


FIG. 11.6 – Graphe de dépendance utilisé dans le cadre de la cartographie en IRMf cérébrale. Pour modéliser convenablement le retard d'activité, une nouvelle variable aléatoire  $L$  est introduite.

influence sur la vraisemblance, la constante n'étant pas multipliée le même nombre de fois suivant la longueur du chemin caché. Cependant, une étape de normalisation de la vraisemblance suivant la longueur du chemin caché est présentée par la suite. La constante n'a alors plus aucune influence. C'est pourquoi, on fixe cette dernière à 1. On obtient finalement :

$$b_{ijklw}(\mathbf{O}_u | \mathbf{O}_{u-1}) = b_{0jlw}(\tau_u) \times b_{1j}(f_{\tau_u, w}^1(\boldsymbol{\tau}_u^*)) \times b_{2j}(f_w^2(\mathbf{m}_u^*)) \times b_{3j}(f_{\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, w}^3(\mathbf{o}_{u-1, u}^*)) \quad (11.3)$$

Les fonctions caractéristiques  $f^i(\cdot)_{1 \leq i \leq 3}$  prennent leur valeur dans  $\mathbb{R}^+$  et sont représentatives de l'activité dès lors que leurs valeurs sont petites. En pratique, on pose :

$$\begin{aligned} f_{\tau_u, w}^1(\boldsymbol{\tau}_u^*) &= \frac{\sqrt{\sum_{i=1}^N w_i \mathbf{1}_{\tau^{(i)}(\tau_u^{(i)}) \cdot (\tau_u^{(i)} - \tau_u)^2}}}{\sqrt{\sum_{i=1}^N w_i \mathbf{1}_{\tau^{(i)}(\tau_u^{(i)})}} & (i) \\ f_w^2(\mathbf{m}_u^*) &= 1 - \sum_{i=1}^N w_i \mathbf{1}_{m_u^{(i)} \neq \phi} & (ii) \\ f_{\boldsymbol{\tau}_u^*, \boldsymbol{\tau}_{u-1}^*, w}^3(\mathbf{o}_{u-1, u}^*) &= \frac{\sum_{i=1}^N w_i \cdot fp(\tau_{u-1}^{(i)}, \tau_u^{(i)})}{\sum_{i=1}^N w_i \cdot (\tau_u^{(i)} - \tau_{u-1}^{(i)})} & (iii) \end{aligned} \quad (11.4)$$

$\mathbf{1}_{\tau^{(i)}(\tau_u^{(i)})}$  et  $\mathbf{1}_{m_u^{(i)} \neq \phi}$  sont deux fonctions indicatrices qui valent 1 si  $\tau_u^{(i)} \in \tau^{(i)}$  ou si  $m_u^{(i)} \neq \phi$ , respectivement, 0 sinon. On rappelle que  $\tau^{(i)}$  représente l'ensemble des instants de détection associés à la  $i$ -ème séquence  $\mathbf{E}^{(i)}$ .  $fp(\tau_{u-1}^{(i)}, \tau_u^{(i)})$  représente enfin le nombre de fausses alarmes détectées dans le canal  $i$  sur l'intervalle ouvert  $]\tau_{u-1}^{(i)}, \tau_u^{(i)}[$ .

La fonction caractéristique  $f^1(\cdot)$  est une mesure du caractère rectiligne de l'alignement des événements composant la signature au travers des canaux et de l'adéquation entre le temps en profondeur et les différents temps de surface ( $f^1(\cdot) = 0$  correspond à une synchronisation parfaite des événements détectés en surface et de l'événement en profondeur).

La fonction caractéristique  $f^2(\cdot)$  correspond au taux de détection de la signature ( $f^2(\cdot) = 0$  signifie que tous les événements ont été détectés). Les deux caractéristiques  $f^1(\cdot)$  et  $f^2(\cdot)$  sont petites en cas d'activité. En effet, nous avons vu dans la partie II que des événements d'intérêt sont détectés pour plusieurs échelles avec des instants de détection ne variant que très peu (approche multi-détecteurs). De plus, on peut considérer que deux voxels voisins actifs possèdent le même retard d'activité dans la mesure où l'imprécision sur le temps de détection est beaucoup plus importante que la durée représentative d'une éventuelle désynchronisation entre deux voxels voisins (prise en compte de l'information spatiale).

La fonction caractéristique  $f^3(\cdot)$  représente enfin le taux moyen de fausses alarmes par unité de temps ( $f^3(\cdot) = 0$  signifie qu'il n'y a pas de fausses alarmes entre deux signatures). Nous concevons que ce choix puisse largement être discuté. Il est uniquement basé sur l'expérience : le

nombre de fausses alarmes des signaux actifs est souvent faible comparativement à celui enregistré pour les signaux inactifs. On peut légitimement penser qu'il se pose néanmoins un problème lorsque le paradigme est composé de plusieurs tâches d'intérêt (tâches A,B,C...) et qu'une zone corticale est activée pour toutes les tâches d'intérêt. Lorsque l'on désire effectuer la cartographie de la tâche  $A$  uniquement, un nombre important de fausses alarmes va alors apparaître pour les signaux de la dite zone. Un signal actif peut donc parfois être caractérisé par un nombre important de fausses alarmes. Malgré cela, nous avons observé des bons résultats de cartographie (dans le cas contraire, il aurait été possible de ne pas modéliser les fausses alarmes). D'après notre analyse, quand la caractéristique  $f^3(\cdot)$  ne permet pas de discriminer les signaux actifs des signaux inactifs, les autres caractéristiques peuvent quant à elles toujours permettre une bonne discrimination. En d'autres termes, il n'est pas gênant que tous les signaux actifs ne possèdent pas un nombre réduit de fausses alarmes. Par contre, il aurait sûrement été plus gênant d'observer des signaux inactifs avec peu de fausses alarmes!

En résumé, les caractéristiques  $f^i(\cdot)_{1 \leq i \leq 3}$  prennent leur valeur dans  $\mathbb{R}^+$ . En présence d'activité, les valeurs rendues par ces fonctions sont petites. Pour contraindre le modèle à être représentatif de l'activité et non de l'inactivité, on définit la loi monotone décroissante  $\mathcal{N}'(x; u)$ <sup>1</sup> suivant  $x$ ,  $u$  étant le paramètre de la loi :

$$\mathcal{N}'(x; u) = \begin{cases} \frac{2}{u\sqrt{2\pi}} \exp(-\frac{x^2}{2u^2}) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (11.5)$$

A partir de cette loi, sont définies les trois lois  $b_{1j}(\cdot)$ ,  $b_{2j}(\cdot)$ , et  $b_{3j}(\cdot)$  de l'équation (11.3) comme suit :

$$\begin{cases} b_{1j}(x) = \mathcal{N}'(x; u_j^1) & (i) \\ b_{2j}(x) = \mathcal{N}'(x; u_j^2) & (ii) \\ b_{3j}(x) = \mathcal{N}'(x; u_j^3) & (iii) \end{cases} \quad (11.6)$$

Ne reste alors à définir dans l'équation (11.3) que la loi  $b_{0j\mathbf{w}}(\tau_u)$ . Si  $X_u = q_j$  alors, en cas d'activité, on s'attend à ce que l'instant d'occurrence de l'événement profond soit  $t'_j + l$  où  $l$  est le retard d'activité et  $t'_j$  l'instant d'occurrence du stimulus associé à  $q_j$ . Le principe de causalité devrait également imposer que l'événement profond prenne place après la stimulation du paradigme avec laquelle il est associé. Cependant, nous n'avons pas considéré ce principe au regard d'éventuels phénomènes d'anticipation du patient. On définit finalement la loi de la manière suivante :

$$b_{0j\mathbf{w}}(x) = \mathcal{N}(x; l + t'_j, \sigma_{j\mathbf{w}}^0) \quad (11.7)$$

### 11.3.3 Liaison de paramètres

Pour réduire le nombre de paramètres du modèle à estimer, le concept de liaison de paramètres [186] est utilisé. Nous rappelons qu'un modèle d'activité propre à chaque voxel  $v$  est défini. Il est donc primordial, pour rendre l'étape d'apprentissage la plus fiable possible, de limiter le nombre de paramètres du modèle.

Nous notons également qu'aussi bien le retard d'activité que le poids des canaux n'ont pas été insérés dans le vecteur de paramètres  $\boldsymbol{\lambda}$ . Les paramètres du modèle définis ici sont donc valables pour une valeur particulière du retard d'activité et du poids des canaux.

<sup>1</sup>Le terme  $u$  ne correspond pas à l'écart-type de la loi. Cependant, si on observe les valeurs  $x_1, \dots, x_N$ , l'estimation du paramètre  $u$  au sens du MV donne :  $u^2 = \frac{1}{N} \sum_{i=1}^N x_i^2$ .

### 11.3.3.1 Matrice de transition

On utilise la même matrice de transition que dans la partie II (voir sous-section 6.3.2.1 page 94). Nous rappelons par exemple que si  $\Delta = 4$  et  $C = 7$ , la matrice de transition s'écrit :

$$\mathbf{A} = \begin{pmatrix} a_{12} & a_{13} & a_{14} & a_{15} & 0 & 0 \\ 0 & a_1 & a_2 & a_3 & a_4 & 0 \\ 0 & 0 & a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & 0 & a_1 & a_2 & a_3 + a_4 \\ 0 & 0 & 0 & 0 & a_1 & a_2 + a_3 + a_4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (11.8)$$

Toutefois, trois modifications sont apportées à cette matrice. (i) Les coefficients  $\{a_{12}, \dots, a_{1\Delta+1}\}$  permettant de modéliser spécifiquement la mise en place de l'activité induite par le protocole expérimental sont tous fixés à  $\frac{1}{\Delta}$ . En effet, ces derniers ne permettent pas de discriminer un signal actif d'un signal inactif. (ii) L'ensemble  $\{a_1, \dots, a_\Delta\}$  modélisant la poursuite de l'activité est contraint de manière à caractériser l'activité. En supposant que l'activité est d'autant plus importante que le nombre d'états sautés est faible, nous imposons ainsi la contrainte  $a_1 \geq a_2 \geq \dots \geq a_\Delta$ . Dans ce but, on définit une loi de probabilité discrète notée  $\mathcal{A}$  prenant ses valeurs dans l'ensemble  $\{1, 2, \dots, \Delta\}$ . Cette dernière possède un hyper-paramètre  $a$ ,  $a \in [\frac{1}{\Delta}, 1[$ . On pose par la suite  $\forall i \in [1, \Delta]$ ,  $a_i = \mathcal{A}(i; a)$ . Plus  $a$  est grand, plus la distribution est resserrée autour de la valeur 1, et plus  $a$  est faible, plus la distribution se rapproche de la distribution uniforme<sup>2</sup> :

$$\begin{cases} \mathcal{A}(1; a) \geq \mathcal{A}(2; a) \geq \dots \geq \mathcal{A}(\Delta; a) \\ \sum_{i=1}^{\Delta} \mathcal{A}(i; a) = 1 \\ \mathcal{A}(1; \frac{1}{\Delta}) = \mathcal{A}(2; \frac{1}{\Delta}) = \dots = \mathcal{A}(\Delta; \frac{1}{\Delta}) \\ \mathcal{A}(1; 1^-) = 1 \text{ et } \mathcal{A}(2; 1^-) = \mathcal{A}(3; 1^-) = \dots = \mathcal{A}(\Delta; 1^-) = 0 \end{cases} \quad (11.9)$$

(iii) La dernière modification concerne les probabilités  $a_{iC}$  ( $C - \Delta + 1 \leq i \leq C - 1$ ). Ces probabilités permettent de discriminer les signaux actifs uniquement en début du support des signaux actifs sur tout le support mais elles ne permettent pas de discriminer un signal actif d'un signal inactif. Par conséquent, les probabilités  $a_{iC}$  ( $C - \Delta + 1 \leq i \leq C - 1$ ) sont fixées à  $\frac{1}{\Delta}$ .

La matrice de transition est entièrement caractérisée par la donnée de  $\Delta$  et par le paramètre  $a$  définissant la forme de la loi discrète. En définitive, si  $\Delta = 4$  et  $C = 7$ , la matrice de transition s'écrit :

$$\mathbf{A} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0 & \mathcal{A}(1; a) & \mathcal{A}(2; a) & \mathcal{A}(3; a) & \mathcal{A}(4; a) & 0 \\ 0 & 0 & \mathcal{A}(1; a) & \mathcal{A}(2; a) & \mathcal{A}(3; a) & 0.25 \\ 0 & 0 & 0 & \mathcal{A}(1; a) & \mathcal{A}(2; a) & 0.25 \\ 0 & 0 & 0 & 0 & \mathcal{A}(1; a) & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 0.25 \end{pmatrix} \quad (11.10)$$

La matrice de transition ainsi obtenue n'est donc plus une matrice stochastique et n'est pas à proprement parler une matrice de transition d'une chaîne de Markov. La liaison de paramètres et les contraintes *a priori* ne sont cependant pas compatibles.

<sup>2</sup>Plus précisément, la loi discrète prend ses valeurs dans l'ensemble  $\{1, 2, \dots, \Delta\}$ . Elle est définie par un hyper-paramètre  $a$  prenant ses valeurs dans  $[\frac{1}{\Delta}, 1[$  par la relation : pour  $k = 1 \dots \Delta$ ,  $P[X = k] = \mathcal{A}(k; a) = a^{k \times i + 1}$ . La contrainte de sommation ( $\sum_{k=1}^{\Delta} \mathcal{A}(k; a) = 1$ ) fait qu'une unique valeur de  $i$  peut être associée à  $a$  ( $a \in [\frac{1}{\Delta}, 1[$ ). La fonction qui à une valeur de  $a$  fait correspondre la valeur de  $i$  est strictement croissante avec  $f(\frac{1}{\Delta}) = 0$  et  $f(1^-) = +\infty$ .

### 11.3.3.2 Lois d'observation

Toujours avec l'objectif de limiter au maximum le nombre de paramètres du modèle, aucune distinction statistique n'est effectuée entre les différents états (non fictifs). Ainsi, on pose pour  $1 < i < C$  :  $u_i^1 = u^1$ ,  $u_i^2 = u^2$  et  $u_i^3 = u^3$  (équation (11.6)) et on pose enfin  $\sigma_{j|w}^0 = \sigma^0$  (équation (11.7)). L'ensemble des lois d'observation est finalement caractérisé par 4 paramètres uniquement :  $\sigma^0$ ,  $u^1$ ,  $u^2$  et  $u^3$ .  $\sigma^0$  modélise la distribution temporelle des événements profonds,  $u^1$  le caractère rectiligne des signatures,  $u^2$  le taux de détection des *evpos* dans les signatures et  $u^3$  le taux de fausses alarmes.

En résumé, le modèle est défini par 6 paramètres  $\lambda = \{\sigma^0, u^1, u^2, u^3, u^4, \Delta, a\}$ . Gardons cependant à l'esprit que le modèle est défini pour une valeur particulière du retard  $l$  et des poids des canaux  $w$ .

## 11.4 Algorithme de cartographie

En raison de la richesse du modèle développé et des informations qu'il apporte, l'algorithme de cartographie peut prendre des formes variées. Par conséquent, il est nécessaire d'effectuer des choix. Ces choix seront justifiés mais il est certain qu'ils conservent toujours une part d'arbitraire et que d'autres algorithmes de cartographie auraient pu être envisagés.

Comme nous l'avons déjà mentionné dans la présentation générale de ce chapitre, le principe général de l'algorithme de cartographie consiste à définir pour chaque voxel  $v$  un modèle d'activité à partir de l'ensemble des séquences d'événements  $E^v$ . Si le voxel  $v$  est inactif, la définition du modèle telle qu'elle a été faite dans la partie précédente fait que la méthode d'apprentissage converge vers un modèle peu explicatif de  $E^v$ . Le modèle répartit donc les probabilités sur l'ensemble de l'espace d'observation. La vraisemblance  $P[E^v|\lambda]$  est donc faible. Au contraire, si le voxel  $v$  est actif, la méthode d'apprentissage fournit un modèle contraint expliquant très bien l'observation  $E^v$  : la vraisemblance  $P[E^v|\lambda]$  est donc grande.

Plus précisément, l'algorithme de cartographie est constitué de deux grandes étapes. Dans une première étape, pour chaque voxel  $v$ , les paramètres du modèle  $\lambda^v$  et un ensemble d'informations caractérisant l'activité locale au voxel  $v$  sont estimés. Ensuite, dans une seconde étape, il est nécessaire à partir de l'ensemble de ces informations de déterminer une carte d'activité.

### 11.4.1 Estimation des paramètres

#### 11.4.1.1 Critère utilisé

Le critère du maximum de vraisemblance n'est pas adapté dans la mesure où le modèle privilégie intrinsèquement les chemins cachés courts par rapport aux chemins longs (voir suite). En pratique, on estime conjointement, pour chaque voxel  $v$  et à partir de  $E^v$  :

- un scénario explicatif de  $E^v$  ou de manière équivalente une observation  $O^{*,v}$  ;
- la séquence d'état cachée  $X^{*,v}$  associée à  $O^{*,v}$  ;
- le retard d'activité  $l^{*,v}$  ;
- le modèle d'activité  $\lambda^{*,v}$  ;
- le poids des différents canaux  $w^{*,v}$ .

sous le critère :

$$\{O^{*,v}, X^{*,v}, l^{*,v}, \lambda^{*,v}\} = \underset{O \in \mathcal{S}(E^v), X, l, \lambda}{\operatorname{argmax}} L_X^2 \sqrt{P[O, X|l, \lambda, w(O)]} \quad (11.11)$$

où  $L_X$  représente la longueur du chemin caché  $\mathbf{X}$ , c'est-à-dire le nombre d'événements profonds. Le vecteur de poids  $\mathbf{w}(\mathbf{O})$  est calculé de manière déterministe à partir de l'observation  $\mathbf{O}$  par la relation :

$$\forall i \in [1, N], w_i(\mathbf{O}) = \frac{D_i}{\sum_{j=1}^N D_j} \quad (11.12)$$

où  $D_i$  désigne le nombre d'*evpos* dans le canal  $i$ . En particulier, on a  $\mathbf{w}^{*,v} = \mathbf{w}(\mathbf{O}^{*,v})$ . La confiance que l'on attribue à un canal dépend uniquement du nombre d'*evpos* et du terme de normalisation. En effet, pour ne pas diminuer l'influence des détecteurs plus sensibles, les fausses alarmes ne sont pas prises en compte dans la détermination du poids des canaux.

Le choix d'un tel critère se justifie comme suit. On prend tout d'abord la racine  $L_X - 1$  de la vraisemblance dans la mesure où le modèle privilégie les chemins courts par rapport aux chemins longs. En effet, on a :

$$P[\mathbf{O}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}, l] = \prod_{u=2}^{L_X} (a_{X_{u-1}X_u} b_{X_{u-1}X_u l \mathbf{w}}(\mathbf{O}_u | \mathbf{O}_{u-1})) \quad (11.13)$$

Le calcul de la vraisemblance  $P[\mathbf{O}, \mathbf{X} | \boldsymbol{\lambda}, \mathbf{w}, l]$  comporte la multiplication de  $L - 1$  termes de probabilité de transition et de vraisemblance d'une observation. Le calcul de la vraisemblance d'une observation s'exprime enfin comme la multiplication de 3 termes dans notre application à part pour la dernière observation (un seul terme). Ainsi, la racine  $L_X - 1$  de la vraisemblance représente approximativement la vraisemblance moyenne d'une observation multipliée par une probabilité de transition. Par conséquent, les chemins courts et les chemins longs sont « traités » de la même façon par le modèle. Cependant, comme on peut s'attendre à ce que les chemins cachés associés à des signaux actifs soient de taille importante, il serait intéressant que le modèle privilégie les chemins longs aux chemins courts. C'est pourquoi, la racine  $L_X - 1$  de la vraisemblance est multipliée par le terme  $L_X^2$ . Nous avons observé en pratique que le terme  $L_X^2$  était très peu discriminant, la discrimination se faisant principalement à partir de la racine  $L_X - 1$  de la vraisemblance.

#### 11.4.1.2 Optimisation du critère

Le critère de l'équation (11.11) est optimisé plusieurs fois pour un ensemble de retard d'activation. Nous retenons finalement le retard d'activation et les paramètres associés à ce retard qui maximisent le critère.

En pratique, le critère de l'équation (11.11) est résolu en fixant le retard d'activité. Le retard  $l$  étant fixé, le critère est optimisé de manière itérative. Connaissant le vecteur des poids  $\mathbf{w}_i$ , il est possible d'estimer l'observation  $\mathbf{O}_{i+1}$ , la séquence d'états cachés  $\mathbf{X}_{i+1}$  et le modèle d'activation  $\boldsymbol{\lambda}_{i+1}$  sous le critère de l'équation (11.14) en utilisant une version adaptée de l'algorithme de l'apprentissage de Viterbi. Connaissant l'observation, il est possible enfin d'estimer le vecteur des poids des canaux avec l'équation (11.12). L'algorithme est finalement illustré figure 11.7.

$$\{\mathbf{O}_{i+1}, \mathbf{X}_{i+1}, \boldsymbol{\lambda}_{i+1}\} = \underset{\mathbf{O} \in \mathcal{S}(\mathbf{E}^v), \mathbf{X}, \boldsymbol{\lambda}}{\operatorname{argmax}} L_X^2 L_X^{-1} \sqrt{P[\mathbf{O}, \mathbf{X} | l, \boldsymbol{\lambda}, \mathbf{w}_i]} \quad (11.14)$$

Le poids des différents canaux  $\mathbf{w}_0$  est initialement fixé à  $\frac{1}{N}$ . Le modèle initial  $\boldsymbol{\lambda}_0$  est choisi peu contraint : le paramètre  $a$  relatif à la loi discrète de la matrice de transition est fixé à 1 (loi uniforme) ; les paramètres des lois, c'est à dire  $\sigma^1, u^2, u^3, u^4$  sont déterminés au travers de simples statistiques effectuées sur l'ensemble  $\mathcal{W}$  calculé à partir du retard d'activité  $l$  et de  $\mathbf{w}_0$ .

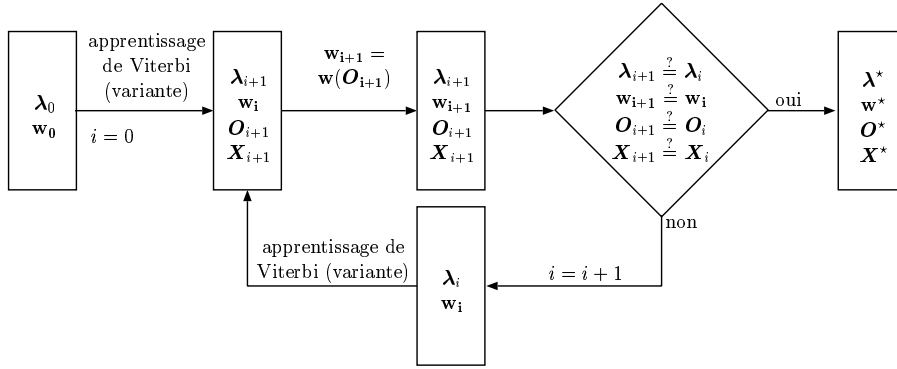


FIG. 11.7 – Optimisation du critère de l'équation (11.11), le retard d'activation étant fixé

### 11.4.1.3 Stratégie de diminution de la complexité algorithmique

L'ensemble des couples observation-état est déterminé dès lors que les poids des canaux et le retard d'activité sont fixés. On utilise tout d'abord une contrainte de maximalité locale : on détermine l'ensemble des associations maximales lorsque l'on voit les différentes séquences au travers d'une fenêtre de taille *extension*. *extension* correspond ainsi à l'écart temporel maximal accepté entre deux événements de surface associés à un même événement du processus profond. Par exemple, les trois événements de la figure 11.8.a peuvent former une éventuelle association si  $\tau_u^{(2)} - \tau_u^{(4)}$  est strictement inférieur au paramètre *extension*.

On utilise ensuite une contrainte de taille, la taille de l'association étant définie comme la somme des poids des canaux pour lesquels un événement a été détecté. On note *tailleMin* la taille minimale des associations considérées. Par exemple, l'association éventuelle des trois événements de la figure 11.8.a a une taille  $w_1 + w_2 + w_4$ . Il faut donc que  $w_1 + w_2 + w_4 > \text{tailleMin}$  pour que l'association soit envisagée.

A chaque association, le temps profond et les temps de surface des *evpos* sont fixés à la moyenne des temps de surface des *evpos*. Cette moyenne est pondérée par le poids des canaux associés aux *evpos*. En d'autres termes, le temps en profondeur est déterminé au sens du maximum de vraisemblance connaissant uniquement les temps de surface des événements détectés (équations (11.6.i) et (11.4.i)) Par exemple dans le cas de l'association de la figure 11.8.a, le temps en profondeur est fixé par :

$$\tau_u = \frac{w_1 \tau_u^{(1)} + w_2 \tau_u^{(2)} + w_4 \tau_u^{(4)}}{w_1 + w_2 + w_4} \quad (11.15)$$

Ne pas utiliser un ensemble de temps en profondeur pour chaque association permet de limiter de façon importante la complexité algorithmique. Cependant, ce critère est sous-optimal dans le sens où l'on ne prend pas en compte les autres temps en profondeur de la séquence de DRHI pour estimer le temps en profondeur de la signature considérée. En ce qui concerne les temps de surface des événements non détectés, ces derniers ont peu d'influence dans la mesure où ils interviennent uniquement pour définir un découpage temporel permettant d'associer les fausses alarmes aux observations du MMCMSE.

Pour former un couple observation-état, il reste ensuite à déterminer les états cachés. On associe une observation à l'état caché  $q_i$  si et seulement si le temps profond appartient à l'intervalle  $[t'_i + l - ecm, t'_i + l + ecm]$ , *ecm* correspondant donc à l'écart maximal autorisé entre ce qui est attendu et ce qui est observé et  $t'_i$  étant l'instant d'occurrence du  $i$ -ème stimulus du



paradigme. Par exemple, la figure 11.8.b représente le couple observation-état (l'état est  $q_3$ ) que l'on peut créer à partir des événements de la figure 11.8.a.

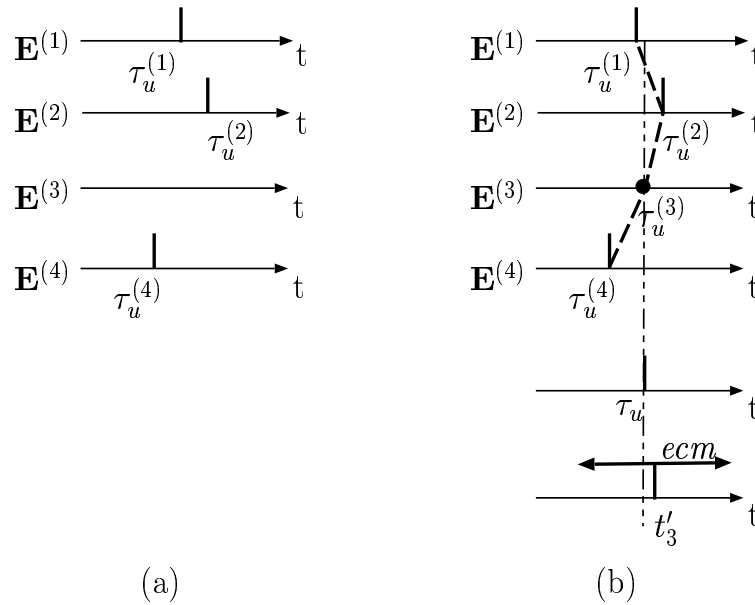


FIG. 11.8 – Stratégie de création des couples observation-état. (a) Séquences d'événements. (b) Couple observation état

En définitive, les contraintes utilisées permettent de définir l'ensemble des couples observation-états  $\mathcal{W}$  qui ont une vraisemblance importante. La construction de cet ensemble nécessite la donnée de trois paramètres : *extension*, *tailleMin* et *ecm*.

## 11.4.2 Cartographie

### 11.4.2.1 Problématique

A chaque voxel  $v$ , est associé l'ensemble  $\{\mathbf{O}^{*,v}, \mathbf{X}^{*,v}, l^{*,v}, \boldsymbol{\lambda}^{*,v}, \mathbf{w}^{*,v}\}$ <sup>3</sup> ainsi qu'une mesure  $p^v = L_{X^{*,v}}^2 L_{X^{*,v}}^{-1} \sqrt{P[\mathbf{O}^{*,v}, \mathbf{X}^{*,v} | l^{*,v}, \boldsymbol{\lambda}^{*,v}, \mathbf{w}^{*,v}]}$ . La mise à jour de la carte d'activité n'est pas aussi simple que dans l'approche précédente. L'approche consistant par exemple à associer à chaque voxel la valeur  $p^v$  n'est pas satisfaisante. En effet, les séquences formant  $\mathbf{E}^v$  sont soit observées en  $v$ , soit observées dans le voisinage de  $v$ . Ainsi, certains voxels  $v$ , à proximité de voxels actifs, peuvent avoir une vraisemblance  $p^v$  importante alors qu'ils sont inactifs. Ainsi, il nous semble intéressant de prendre en compte la notion de poids des canaux lors de la mise à jour de la carte d'activité. La stratégie est expliquée dans la sous-section 11.4.2.2.

Une fois la carte définie, il est nécessaire de seuiller cette dernière. Dans ce but, la carte définie est transformée en une carte de p-valeur. Pour cela, il faut déterminer la densité de probabilité de la valeur associée à un voxel  $v$ , sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale du voxel  $v$  et de ses voisins. Dans ce but, il est nécessaire de construire un ensemble de séquences  $\mathbf{E}$  représentatif de l'inactivité cérébrale, la séquence  $\mathbf{E}$  étant une observation du MMCMSE.  $\mathbf{E}$  est construit à partir de l'ensemble des séquences  $\mathbf{E}^v$  où  $v$  est un voxel intracrânien. Par exemple,

<sup>3</sup>Cet ensemble de données peut également être exploité de différentes façons de manière à caractériser l'activité. Comme dans l'approche scalaire, on peut déterminer une carte de retard, visualiser les modes d'activité et déterminer la réponse hémodynamique.

la figure 11.9 montre la construction de  $\mathbf{E}$  à partir de  $\mathbf{E}^{v_1}$ ,  $\mathbf{E}^{v_2}$  et  $\mathbf{E}^{v_3}$ . La construction de  $\mathbf{E}$  s'avère désormais plus complexe que dans la partie précédente. En effet, autant pour les signaux inactifs que pour les signaux actifs, les séquences constituant  $\mathbf{E}^v$  ne sont pas indépendantes car elles sont observées à un voxel voisin et/ou à travers un détecteur différent. Cette corrélation intrinsèque doit également exister à l'intérieur des séquences de l'ensemble  $\mathbf{E}$  créées. D'une part, il est donc nécessaire de déstructurer les séquences d'événements de manière à ce que  $\mathbf{E}$  soit représentatif de l'inactivité cérébrale, d'autre part, la déstructuration ne doit pas être trop importante pour conserver les corrélations intrinsèques qui existent entre des séquences d'événements différentes appartenant à une même observation. La seconde difficulté provient ensuite du fait que la valeur associée à un voxel est déterminée à partir d'un ensemble de calculs effectués à partir des poids et des valeurs de vraisemblance  $p^v$ . La procédure décrite section 11.4.2.3 essaie au mieux de prendre en considération les difficultés énoncées ci-dessus.

#### 11.4.2.2 Mise à jour de la carte d'activité

Nous notons  $g^v(c)$  le voxel qui a permis d'obtenir la  $c$ -ème séquence de  $\mathbf{E}^v$ . La carte d'activité est alors mise à jour suivant l'algorithme 11.1

---

#### Algorithme 11.1 Mise à jour de la carte d'activation

---

*\*\*\*Initialisation\*\*\**

**Pour**  $v \in \mathcal{I}$  **Faire**

-  $\mathcal{I}_v = 0$

**Fin Pour**

*\*\*\*Calcul de la carte d'activité\*\*\**

**Pour**  $v \in \mathcal{I}$  **Faire**

**Pour**  $c = 1 \dots N$  **Faire**

-  $\mathcal{I}_{g^v(c)} = \mathcal{I}_{g^v(c)} + p^v \times \min(\frac{1}{N}, w_c^{*,v})$

**Fin Pour**

**Fin Pour**

---

L'algorithme 11.1 se justifie car sans considérer les effets de bord, il associe à chaque voxel  $v$  actif la valeur  $p > 0$  et à chaque voxel inactif la valeur 0 sous les hypothèses présentées ci-après. Parmi les  $N$  canaux associés à  $\mathbf{E}^v$ , si  $N_1$  canaux sont représentatifs de l'activité et  $N_2$  sont représentatifs de l'inactivité ( $N_1 + N_2 = N$ ), alors :

- les poids des canaux exprimant l'activité sont de  $\frac{1}{N_1}$  et ceux exprimant l'inactivité sont nuls (sous l'hypothèse que  $N_1 \neq 0$ ).
- $p^v = p > 0$  si  $N_1 \neq 0$ , 0 sinon.
- les canaux associés à un même voxel expriment soit tous l'inactivité soit tous l'activité.

#### 11.4.2.3 Carte de p-valeur

La carte d'activité est enfin transformée en une carte de p-valeur. La densité de probabilité de la vraisemblance associée à un voxel  $v$  sous l'hypothèse  $\mathcal{H}_0$  d'inactivité est inférée expérimentalement en générant aléatoirement des séquences d'événements représentatives de  $\mathcal{H}_0$  à partir de l'ensemble des séquences d'événements intracrâniennes.

On commence tout d'abord par déterminer une séquence de RDs représentative de l'hypothèse  $\mathcal{H}_0$  à partir de l'ensemble des séquences de RDs détectées à l'échelle de résolution la plus basse. On utilise pour cela la méthode proposée dans l'approche scalaire. On note  $t_1 = 1$ ,

$t_2, \dots, t_k = T$  les instants de détection correspondant à la séquence créée et  $v_i$  le numéro du voxel qui a été utilisée pour générer l'observation ayant lieu à  $t = t_i$ . On découpe l'intervalle discret  $[1, T]$  en  $k$  intervalles disjoints :

$$[1, T] = \gamma_1 \cup \gamma_2 \cup \dots \cup \gamma_k \text{ avec } \begin{cases} \gamma_1 = [t_1, \frac{t_1+t_2}{2}[ \\ \gamma_i = [\frac{t_{i-1}+t_i}{2}, \frac{t_i+t_{i+1}}{2}[ \text{ si } i = 2 \dots k-1 \\ \gamma_k = [\frac{t_{k-1}+t_k}{2}, t_k] \end{cases} \quad (11.16)$$

Ensuite, pour ne pas déstructurer de manière trop importante les différentes séquences et conserver les corrélations qui existent entre les séquences, on crée, pour chaque intervalle  $\gamma_i$ , les différentes séquences par simple recopie des séquences  $\mathbf{E}^{k,v'}$  où  $1 \leq k \leq K$  et où  $v'_i = v_i \cup \mathcal{N}(v_i)$ ,  $K$  étant le nombre de séquences de RDs détectées par signal IRMf. Un exemple est donné figure 11.9.

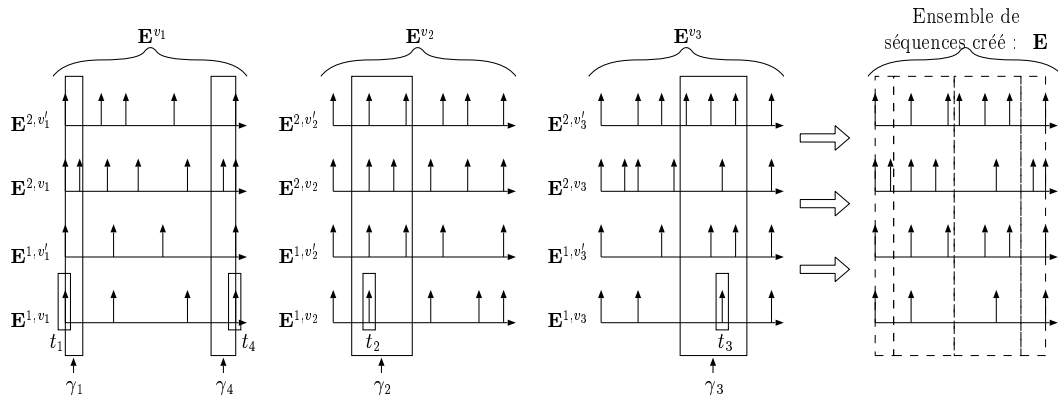


FIG. 11.9 – Illustration de l'algorithme de création d'un ensemble de séquence d'événements  $\mathbf{E}$  représentative de  $\mathcal{H}_0$  à partir de  $\mathbf{E}^{v_1}$ ,  $\mathbf{E}^{v_2}$  et  $\mathbf{E}^{v_3}$ .

A partir d'un ensemble de séquences  $\mathbf{E}$ , il est possible d'estimer comme précédemment l'ensemble  $\{\mathbf{O}^*, \mathbf{X}^*, l^*, \boldsymbol{\lambda}^*, \mathbf{w}^*\}$  ainsi que  $p = L_{\mathbf{X}^*}^2 \sqrt{L_{\mathbf{X}^*}^{-1} P[\mathbf{O}^*, \mathbf{X}^* | l^*, \boldsymbol{\lambda}^*, \mathbf{w}^*]}$ . On détermine ensuite la mesure  $\sum_{i=1}^N p \times \min(w_i^*, \frac{1}{N})$  à partir de laquelle on construit la densité de probabilité désirée sous l'hypothèse  $\mathcal{H}_0$  d'inactivité cérébrale.

La validation expérimentale a montré des résultats décevants en ce qui concerne le calcul de la p-valeur. Cependant, ceci ne remet pas en cause la méthode mais uniquement la procédure de calcul de la p-valeur. Il conviendra d'utiliser des procédures décrites dans la littérature [168] pour estimer plus fidèlement les p-valeurs (ceci n'a pas encore été réalisé à ce jour). Il s'agit, en utilisant des méthodes de permutation, de construire de nouvelles séquences d'images IRMf représentatives de l'inactivité cérébrale à chaque voxel  $v$  et qui possèdent les mêmes caractéristiques que la séquence originelle en termes de corrélation temporelle et/ou spatiale. Ces séquences ainsi constituées peuvent ensuite être utilisées pour une estimation plus fidèle des p-valeurs.



# Chapitre 12

## Résultats expérimentaux : paradigmes en blocs

Nous traitons aussi bien dans le cas réel que synthétique, les mêmes données que précédemment. Nous rappelons que les données synthétiques sont présentées dans la section 7.1.1, page 101 alors que les données réelles sont présentées dans la section 7.2.1, page 110.

Nous commençons par présenter les résultats obtenus sur données synthétiques et sur données réelles dans le cas de paradigmes en blocs. Les résultats présentés ont été obtenus avec un voisinage de type 26-connexité et avec un détecteur détectant trois séquences de RDs par signal IRMf avec  $k = 0.7, 1, 1.3$ . Nous évaluons ensuite l'influence des paramètres sur les résultats de détection de RDs et/ou de cartographie.

### 12.1 Résultats

#### 12.1.1 Résultats sur données synthétiques

##### 12.1.1.1 Rappel

Nous rappelons que six séquences synthétiques d'images IRMf 3-D, notées  $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$  ont été construites en additionnant, dans des zones définies comme actives, un signal d'activation synthétique au bruit. Dans chaque séquence, quatre cubes, de taille  $6 \times 6 \times 6$  voxels, ont été définis comme zones actives (voir figure 7.1, page 101). Les séquences  $\mathcal{F}_{1a}$  et  $\mathcal{F}_{1b}$  illustrent un cas d'application idéal où ce qui observé comme signal d'activation est conforme en tous points à ce qui est attendu. Pour cela, les fonctions d'excitation sont toutes égales à la fonction de base  $x_0(\cdot)$  associée au paradigme. De même, les réponses hémodynamiques sont toutes égales au modèle de réponse utilisé dans SPM. Seul le rapport signal à bruit  $m$  varie, avec  $m = 2$  pour  $\mathcal{F}_{1a}$  et  $m = 1$  pour  $\mathcal{F}_{1b}$ . Nous rappelons ensuite que  $\mathcal{F}_{2a}$  illustre différents cas de non maintien de la réponse hémodynamique dans un bloc de stimulation, que  $\mathcal{F}_{3a}$  est représentative de différents cas de variabilité de forme de la réponse hémodynamique à un bloc de stimulation, que  $\mathcal{F}_{4a}$  a été conçue pour illustrer différents cas de retard d'activation sur un bloc de stimulation et que  $\mathcal{F}_{5a}$  illustre enfin différents cas de phénomènes d'habituation et d'apprentissage.

Afin de tester plus avant la robustesse de la méthode en contexte de faible rapport signal sur bruit, nous avons complété le jeu des six séquences précédentes de sept nouvelles séquences. Ces séquences sont notées :  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$ ,  $\mathcal{F}_{1d}$ ,  $\mathcal{F}_{2b}$ ,  $\mathcal{F}_{3b}$ ,  $\mathcal{F}_{4b}$  et  $\mathcal{F}_{5b}$ . Les indices  $a$ ,  $b$ ,  $c$  et  $d$  sont représentatifs du rapport signal sur bruit  $m$  utilisé dans les zones définies comme actives, avec

$m = 2$  pour l'indice  $a$ ,  $m = 1$  pour l'indice  $b$ ,  $m = 0.5$  pour l'indice  $c$  et  $m = 0.25$  pour l'indice  $d$ . Par exemple, les séries  $\mathcal{F}_{i,a}$  et  $\mathcal{F}_{i,b}$  ( $i = 2 \dots 5$ ) ne diffèrent que par le rapport signal sur bruit ( $m = 2$  pour  $\mathcal{F}_{i,a}$  et  $m = 1$  pour  $\mathcal{F}_{i,b}$ ). Ces valeurs de  $m$  ont été sélectionnées de manière à couvrir les différents cas de figure rencontrés en pratique.

Comme précédemment, les résultats de cartographie de la méthode MMCMSE sont comparés à ceux obtenus par la méthode SPM, version SPM2. Pour une évaluation plus juste des résultats, nous rappelons que deux implémentations de SPM ont été mises en œuvre. L'implémentation SPMa correspond à une mise en œuvre « standard » du logiciel par les médecins en routine clinique. L'implémentation SPMb correspond, quant à elle, au cas d'application idéal où les signaux synthétiques d'activation utilisés pour créer les données sont réinjectés dans la matrice de dessin de SPM.

### 12.1.1.2 Influence du bruit

Les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{1a}$ ,  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$  et  $\mathcal{F}_{1d}$  permettent d'évaluer la robustesse des différentes méthodes au bruit. Rappelons que pour ce jeu de séquences, les implémentations SPMa et SPMb sont rigoureusement identiques pour les raisons invoquées section 7.1.2.2, page 105.

Les résultats obtenus sur les séquences d'images  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$  et  $\mathcal{F}_{1d}$  sont représentés sur les figures 12.1 et 12.2. A noter d'emblée que les résultats obtenus sur la séquence  $\mathcal{F}_{1a}$  ne sont pas présentés, les deux méthodes conduisant à des résultats optimaux de détection. Les cartes

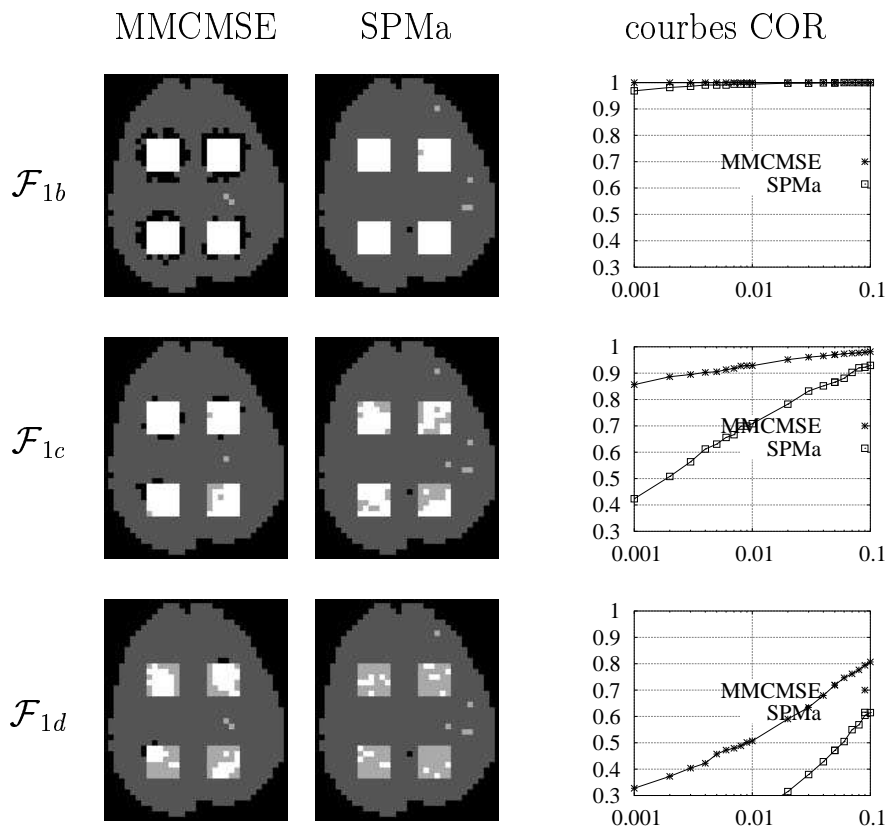


FIG. 12.1 – Résultats sur données synthétiques (voir texte)

obtenues avec la méthode proposée sont présentées sur la première colonne et celles obtenues avec SPMa sont présentées sur la seconde colonne. Les cartes de la figure 12.1 sont seuillées pour une PFA de 0.005 alors que les cartes de la figure 12.2 sont seuillées pour une p-valeur estimée de 0.005. Pour toutes les cartes, le blanc représente une bonne détection, le gris clair une fausse alarme ou une non détection. Le gris foncé représente les signaux non actifs légitiment non détectés. Enfin, le noir représente soit l'extérieur de la coupe, soit des détections dans la zone qui entoure les zones actives. La troisième colonne de la figure 12.1 correspond aux courbes COR alors que les troisièmes et quatrièmes colonnes de la figure 12.2 correspondent respectivement aux courbes  $\beta$  et  $\alpha$ . On peut se référer à la section 7.1.2.1 de la page 103 pour un rappel sur la définition de ces courbes.

L'analyse des courbes COR (figure 12.1) montre que la méthode de cartographie par MMCMSE surpasse, quelque soit le rapport signal sur bruit  $m$ , la méthode SPMa pourtant placée en contexte idéal d'analyse. La prise en compte de l'information spatiale dans une approche multi-détecteurs explique la grande robustesse de la méthode au bruit. Ces premiers résultats valident la méthode de cartographie développée.

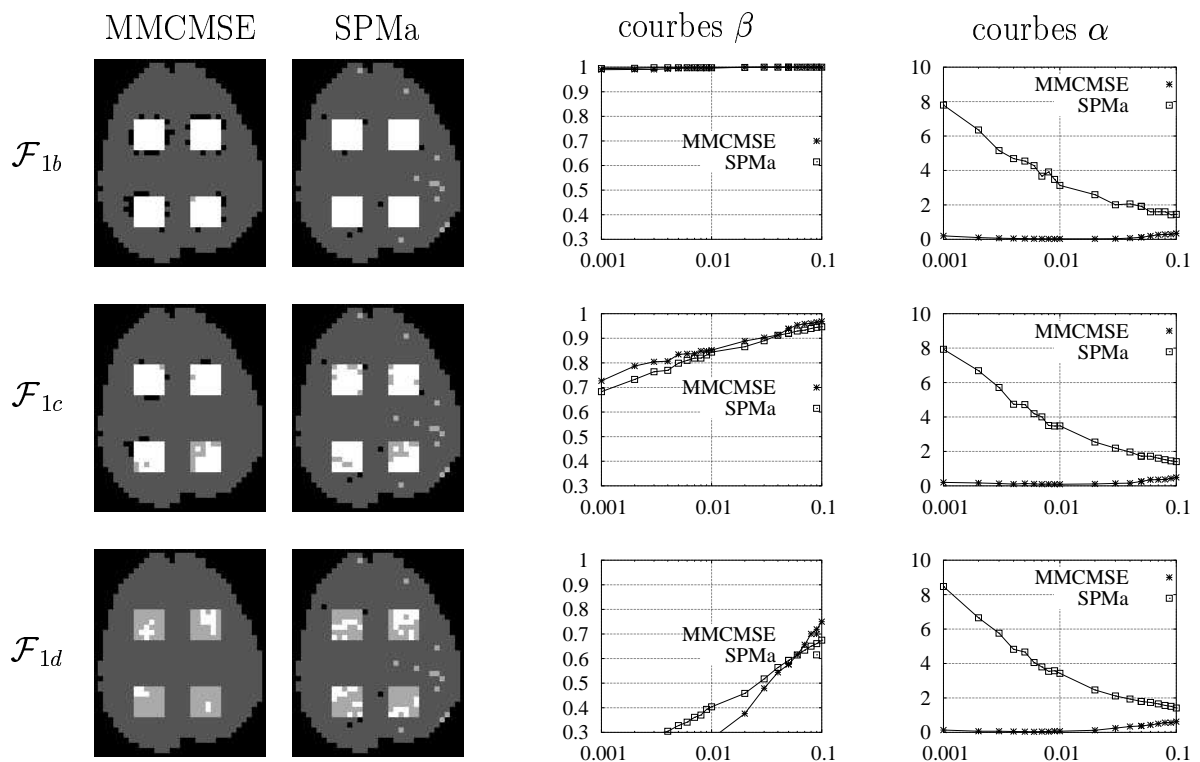


FIG. 12.2 – Résultats sur données synthétiques (voir texte)

L'analyse des courbes  $\alpha$  de la figure 12.2 montrent le peu crédit que l'on peut accorder à la p-valeur estimée, aussi bien pour SPMa que pour la méthode proposée. Les courbes  $\alpha$  obtenues pour SPMa montrent que cette méthode détecte en pratique un nombre de fausses alarmes cinq fois supérieur, en moyenne, à celui attendu compte tenu de la p-valeur sélectionnée ( $p$  étant dans l'intervalle  $[0.01, 0.001]$ ). L'estimation de la p-valeur dans SPMa apparaît donc optimiste. A l'inverse, les courbes  $\alpha$  obtenues pour notre méthode montrent que celle-ci détecte un nombre de fausses alarmes cinq à dix fois inférieur, en moyenne, à celui attendu dans le même intervalle de p-valeur. L'estimation de la p-valeur dans notre méthode apparaît donc

largement pessimiste. Cet écart entre p-valeurs estimées par SPMA et par notre méthode est clairement illustré sur les imagerie de la figure 12.2. Bien qu'obtenues pour la même p-valeur de 0.005, les imagerie SPMA montrent un nombre de fausses alarmes plus important que celui observé dans les imagerie de notre méthode. Le biais dans l'estimation de la p-valeur, mis en avant par les courbes  $\alpha$  n'est pas non plus sans conséquence sur l'analyse des courbes  $\beta$ . Ces courbes montrent des performances de détection d'activation équivalentes pour les deux méthodes SPMA et MMCMSE et ce quelque soit le rapport signal à bruit considéré. Ce fait va à l'encontre des courbes COR obtenues figure 12.1 lesquelles montrent la supériorité de notre méthode par rapport à SPMA en contexte de fort ou faible rapport signal à bruit. L'analyse conjointe des courbes  $\alpha$  et  $\beta$  permet de lever cette ambiguïté et de rétablir la vérité des courbes COR. La prise en compte des courbes  $\alpha$  doit conduire à corriger les courbes  $\beta$  en translatant à droite celles obtenues par SPMA et à gauche celles obtenues par notre méthode.

Nous verrons par la suite que l'estimation pessimiste de la p-valeur constitue la principale limitation de notre méthode. Cependant, cette limitation n'est liée qu'à la méthode d'estimation de la p-valeur décrite 11.4.2.3, page 168. De plus, elle pourrait être levée par l'utilisation de méthodes plus évoluées d'estimation de la p-valeur présentées dans la littérature [168].

### 12.1.1.3 Variabilité du signal actif

Nous présentons maintenant les résultats obtenus lorsque l'on introduit de la variabilité au niveau de la composante d'activation. Nous rappelons que  $\mathcal{F}_{2b}$  illustre différents cas de non maintien de la réponse hémodynamique dans un bloc de stimulation, que  $\mathcal{F}_{3b}$  est représentative de différents cas de variabilité de forme de la réponse hémodynamique à un bloc de stimulation, que  $\mathcal{F}_{4b}$  a été conçue pour illustrer différents cas de retard d'activation sur un bloc de stimulation et que  $\mathcal{F}_{5b}$  illustre enfin différents cas de phénomènes d'habituation et d'apprentissage (voir figure 7.2, page 103). Nous rappelons enfin que l'implémentation SPMA correspond à l'utilisation courante du logiciel alors que SPMb correspond à un cas idéal où l'on injecte dans la matrice de dessin les véritables composantes d'activation des différents signaux actifs.

Les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{2b}$ ,  $\mathcal{F}_{3b}$ ,  $\mathcal{F}_{4b}$  et  $\mathcal{F}_{5b}$  sont représentés sur les figures 12.3 et 12.4 avec la même méthode que précédemment à l'exception d'une nouvelle colonne, utilisée pour insérer les imagerie obtenues avec SPMb. Il est à noter que les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{2a}$ ,  $\mathcal{F}_{3a}$ ,  $\mathcal{F}_{4a}$  et  $\mathcal{F}_{5a}$  ne sont pas présentés dans la mesure où toutes les méthodes donnent de très bons résultats avec ces séquences. Aussi est-il important de diminuer le rapport signal sur bruit. Nous rappelons que le cas  $m = 1$  (rapport signal sur bruit considéré ci-après) est un rapport signal sur bruit tout à fait réaliste.

Les résultats obtenus avec la séquence  $\mathcal{F}_{2b}$  (première ligne de l'image 12.3, non maintien de la réponse) montrent clairement que notre méthode surpasse très facilement toutes les implémentations de SPM, à la fois SPMA (courante) et SPMb (idéale). Dans la partie précédente, nous n'avions pas clairement mis en évidence le manque de robustesse de SPMA au non maintien de la réponse à un bloc de stimulation. En effet, de très bons résultats étaient obtenus avec la séquence  $\mathcal{F}_{2a}$  (cf. figure 7.9 de la page 109). Pour observer ce manque de robustesse, un niveau de bruit plus important est nécessaire. Ce niveau de bruit reste néanmoins réaliste. Nous rappelons que les séries  $\mathcal{F}_{2a}$  et  $\mathcal{F}_{2b}$  ne diffèrent que par leur rapport signal sur bruit, la série  $\mathcal{F}_{2b}$  étant la plus bruitée ( $m = 2$  pour  $\mathcal{F}_{2a}$  et  $m = 1$  pour  $\mathcal{F}_{2b}$ ). SPMb permet certes d'obtenir des résultats plus satisfaisants que SPMA, mais ils restent moins bons que ceux obtenus avec notre méthode pour  $m = 1$ .

Les résultats obtenus avec la séquence  $\mathcal{F}_{3b}$  (seconde ligne de l'image 12.3, variabilité de



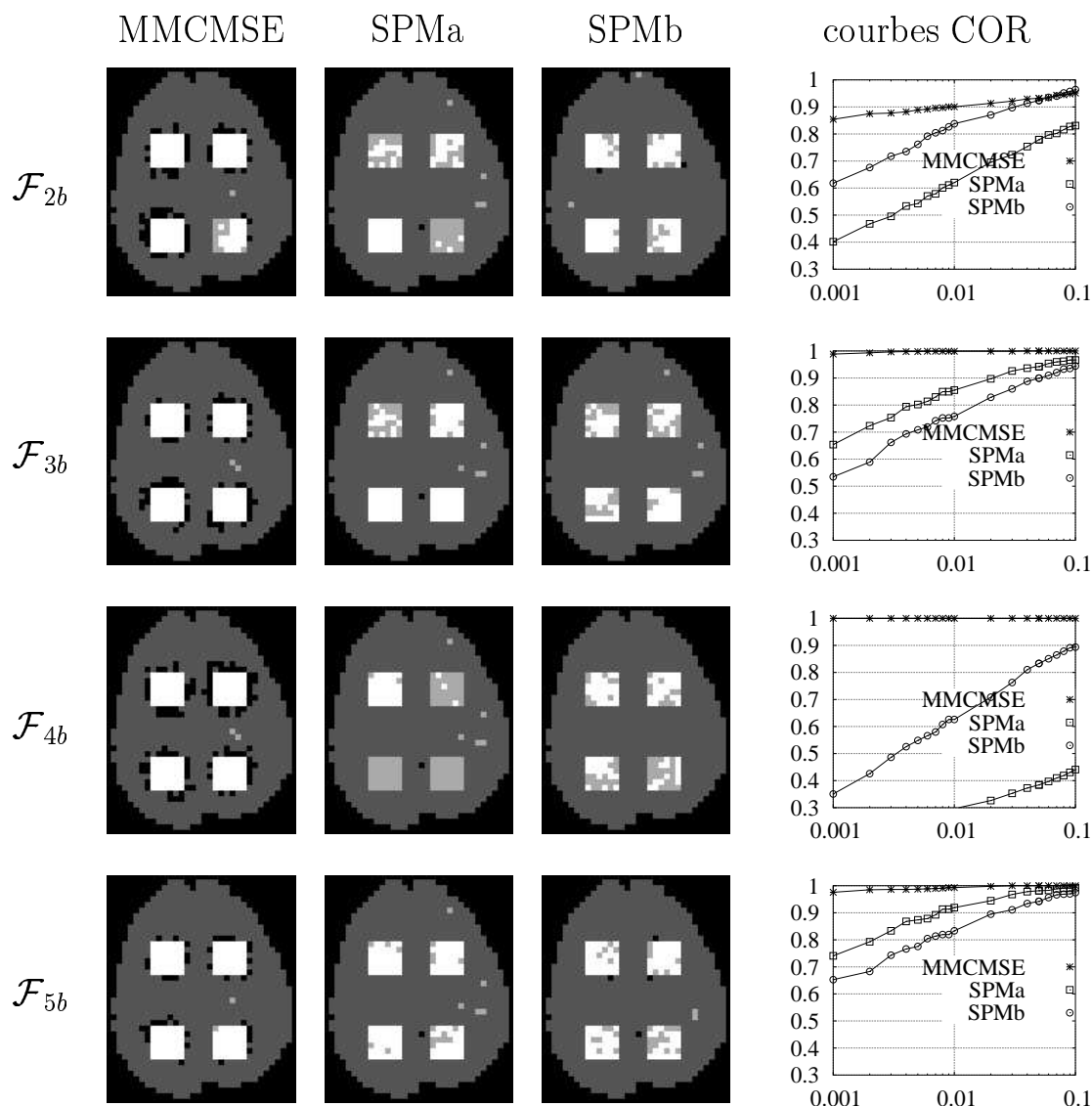


FIG. 12.3 – Résultats sur données synthétiques (voir texte)

la réponse hémodynamique) montrent encore une fois la supériorité de la méthode proposée. Cette dernière est très robuste à la variabilité de la réponse hémodynamique contrairement à SPMa et SPMb. Le manque de robustesse de SPMa et de SPMb à la variabilité de la réponse hémodynamique est ici mis en évidence grâce à la présence d'un niveau de bruit relativement important mais toujours réaliste. Nous rappelons en effet que de très bons résultats étaient observés avec la séquence  $\mathcal{F}_{3a}$  (cf. figure 7.9 de la page 109). Il convient également de noter que les résultats obtenus avec SPMb ne sont pas meilleurs que ceux obtenus avec SPMa. Ainsi, utiliser plusieurs signaux explicatifs à l'intérieur de la matrice de dessin ne conduit pas à une meilleure immunité au bruit.

Les résultats obtenus avec la séquence  $\mathcal{F}_{4b}$  (troisième ligne de l'image 12.3, introduction de retard d'activité) montrent la capacité de la méthode à prendre en compte les retards d'activation. Au contraire, SPMa ne détecte pas les zones avec un fort retard d'activation et SPMb, bien que détectant toutes les zones d'activité, possède un taux de détection relativement

faible.

Les résultats obtenus avec la séquence  $\mathcal{F}_{5b}$  (troisième ligne de l'image 12.3, introduction de différents phénomènes d'apprentissage) montrent finalement encore une fois l'intérêt de notre approche. La « modification » de la matrice de transition nous permet désormais de modéliser de façon convenable à la fois les signaux dont l'activité augmente au cours du temps et les signaux dont l'activité diminue au cours du temps.

De manière générale, les résultats obtenus mettent clairement en évidence la supériorité de notre méthode par rapport aux deux implémentations de SPM. La méthode s'avère très efficace pour prendre en compte toutes les variabilités du signal IRMf actif. L'implémentation « idéale » SPMb a clairement montré ses limites. L'introduction dans la matrice de dessin d'un ensemble de composantes d'activation attendues semble très peu robuste au bruit.

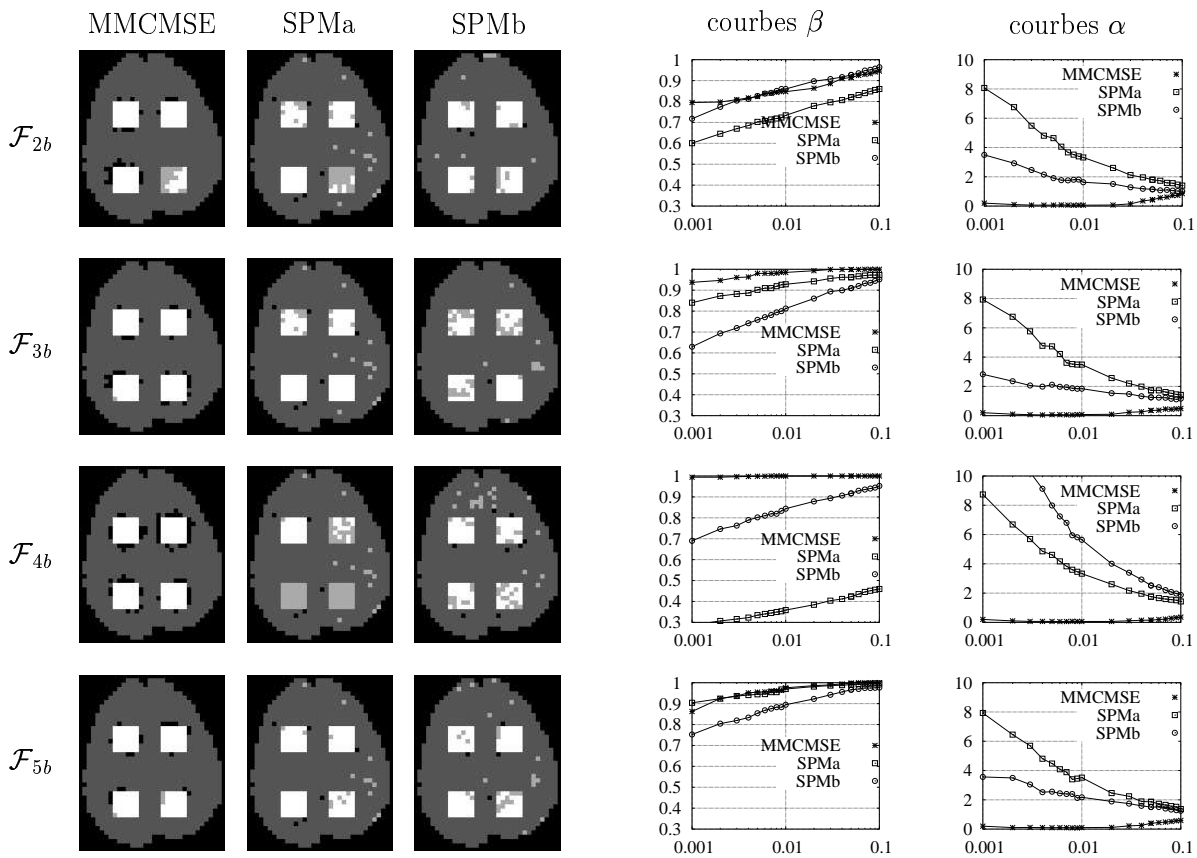


FIG. 12.4 – Résultats sur données synthétiques (voir texte)

Les courbes  $\alpha$  de la figure 12.4 montrent encore une fois le biais important observé dans l'estimation de la p-valeur, aussi bien pour SPMa, pour SPMb que pour la méthode proposée. Les méthodes SPMa et SPMb ont tendance à détecter un nombre de fausses alarmes beaucoup plus important que prévu alors que c'est l'inverse pour notre méthode.

### 12.1.2 Résultats sur données réelles

Dans le cadre de notre collaboration avec l'Institut de Physique Biologique, nous rappelons que nous disposons d'une base de 119 séquences d'images IRMf acquises sur 17 sujets sains

âgés entre 18 et 39 ans, chacune formée de 145 images. Les sujets ont accompli des tâches de traitement linguistique selon trois protocoles différents de stimulation. Les trois protocoles de stimulation, à savoir, VERBE, AUDIO, et VISUEL, ont été conçus pour déterminer les zones corticales impliquées dans la recherche de mots et dans le traitement lexicologique auditif et visuel. Des informations détaillées sur les trois paradigmes expérimentaux peuvent être obtenues à la section 7.2.1, page 110 ou dans [160].

Des exemples de cartes d'activation sont représentés figure 12.5.a,b et c pour les protocoles VERBE, AUDIO, et VISUEL, respectivement. Des cartes d'activation obtenues avec la méthode SPM sont également insérées pour comparaison.

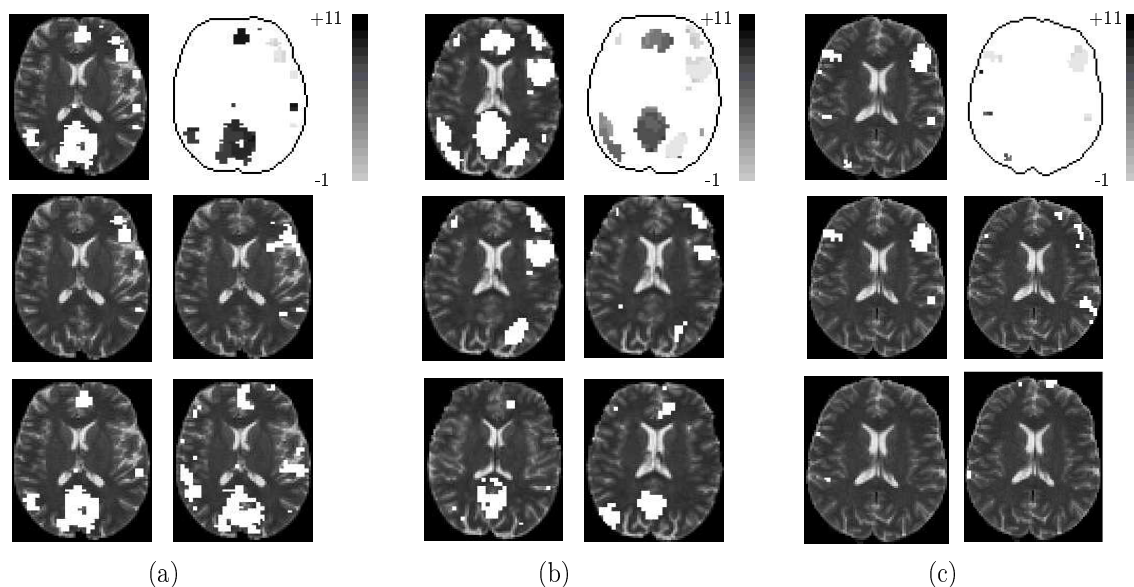


FIG. 12.5 – Cartes d'activation obtenues avec la méthode proposée (26-connexité et trois séquences de RDs détectées par signal IRMf avec  $k = 0.7, 1, 1.3$ ) et SPM pour les trois paradigmes (de gauche à droite, VERBE, AUDIO, VISUEL). Chaque groupe de six cartes se compose tout d'abord d'une carte d'activation basée sur la méthode MMCMSE (en haut à gauche), de la carte de retard d'activation associée (en haut à droite). La carte au milieu à gauche représente uniquement les zones actives synchrones détectées par notre méthode alors que la carte en bas à gauche représente les zones déphasées de  $\pi$ . Les cartes de SPM sont représentées au milieu à droite et en bas à droite. Pour la carte au milieu, on cherche à détecter les signaux en phase alors que pour la carte en bas, on a cherché à détecter les signaux déphasés de  $\pi$ .

Chaque groupe de six cartes se compose tout d'abord d'une carte d'activation basée sur la méthode MMCMSE (en haut à gauche), de la carte de retard d'activation associée (en haut à droite). La carte au milieu à gauche représente uniquement les zones actives synchrones (retard d'activation estimé compris entre -2 et 2 échantillons) alors que la carte en bas à gauche représente les zones déphasées de  $\pi$  (les paradigmes sont périodiques avec une durée de l'état de repos égale à celle de l'état d'activation). Les cartes obtenues avec SPM sont représentées au milieu à droite et en bas à droite. Pour la carte au milieu, on a cherché à détecter les signaux en phase alors que pour la carte en bas, on a cherché à détecter les signaux déphasés de  $\pi$ . Pour ce faire, le traitement a été lancé deux fois en utilisant une exécution standard du logiciel SPM2 (correspondant à SPMa) pour deux paradigmes différents, le paradigme originel et le

paradigme originel décalé d'une demi-période. Les cartes SPM ont été seuillées en utilisant les champs aléatoires gaussiens et une p-valeur corrigée de 0.05 (champ gaussien) alors que les cartes obtenues avec la méthode proposée ont été obtenues pour une p-valeur de 0.005. Il est à noter qu'il est difficile de comparer les résultats en termes d'extension spatiale des zones d'activité dans la mesure où des p-valeurs différentes sont utilisées pour le seuillage. De plus, nous avons vu dans la partie précédente qu'il pouvait y avoir des biais très importants dans l'estimation des p-valeurs.

Les résultats obtenus avec la méthode proposée et le logiciel SPM présentent des activations dans les mêmes secteurs corticaux. De plus, la méthode proposée présente l'avantage d'être non supervisée, les signaux déphasés ayant été détectés sans intervention de l'utilisateur. À noter que notre méthode détecte pour le paradigme AUDIO (figure 12.5.b) une zone à fort retard. Cette dernière n'apparaît cependant plus quand on considère uniquement les signaux actifs déphasés de  $\pi$ . Cela vient du fait que le paradigme AUDIO, contrairement aux autres paradigmes, n'est pas exactement périodique. Une activité sur les fronts descendants du paradigme ne correspond donc pas exactement à une activité retardée sur les fronts montants du paradigme.

Plus généralement, 21 séquences d'événements ont été traitées avec notre méthode dans le but d'effectuer un traitement de groupe. Nous n'avons cependant encore eu aucun véritable retour médical.

## 12.2 Influence des paramètres

### 12.2.1 Etude sur données synthétiques

#### 12.2.1.1 Etape de détection

Nous n'avons pas utilisé les mêmes taux de détection que dans la première approche par MSMCSE. En effet, l'approche par MSMCSE ne donnait pas des résultats satisfaisants pour de très faibles taux de détection ( $k$  petit) : l'algorithme d'apprentissage convergait souvent de façon prématurée vers un modèle très contraint déclarant actif un ensemble réduit de signaux. Ces signaux se révélaient être des signaux actifs pour lesquels l'étape de prétraitement s'était très bien déroulée (ce problème avait été mentionné mais aucun résultat n'a été montré pour l'illustrer). Nous n'avons plus ce type de problèmes avec l'approche par MMCMSE.

Nous proposons de quantifier la qualité de l'étape de prétraitement pour chacun des nouveaux détecteurs. Nous avons utilisé cinq taux de détection différents ( $k = 0.7, 1, 1.3, 1.6$  et  $2$ ). On rappelle que  $k$  correspond au nombre moyen de détections par signal IRMf divisé par le nombre de transitions « off-on » dans le paradigme, le paradigme étant ici constitué de sept transitions « off-on ». Pour mesurer la qualité de l'étape de prétraitement, nous utilisons les mêmes mesures que précédemment.

Pour rappel, afin de quantifier les performances de l'étape de détection, cinq mesures ont été utilisées : le nombre moyen de RDs détectées par signal actif,  $\mu_{nA}$ , le nombre moyen de RDs détectées par signal inactif,  $\mu_{nI}$ , les écarts-types associés,  $\sigma_{nA}$  et  $\sigma_{nI}$ , ainsi que la mesure discriminante  $|\mu_{nA} - \mu_{nI}| / (\sigma_{nA} + \sigma_{nI})$ . Les résultats sont consignés dans le tableau 12.1 pour les 5 détecteurs et les séries  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$  et  $\mathcal{F}_{1d}$ . On rappelle que les séries diffèrent uniquement par le rapport signal sur bruit, la séquence  $\mathcal{F}_{1b}$  étant la séquence la moins bruitée et  $\mathcal{F}_{1d}$  la séquence la plus bruitée.

Il apparaît tout d'abord que le pouvoir de discrimination basé sur le nombre de détections varie de façon très importante avec le choix du détecteur, le choix  $k = 0.7$  est le détecteur

	0.7	1	1.3	1.6	2
$\mathcal{F}_{1b}$	(6.44, 0.7) (4.81, 0.98) 0.97	(6.97, 0.49) (6.98, 1.06) 0.	(7.84, 0.93) (9.14, 1.16) 0.62	(9.53, 1.39) (11.26, 1.27) 0.65	(12.87, 1.7) (14.03, 1.45) 0.36
$\mathcal{F}_{1c}$	(5.67, 0.9) (4.85, 0.98) 0.43	(7.04, 0.77) (6.98, 1.06) 0.03	(8.7, 1.11) (9.1, 1.16) 0.17	(10.77, 1.29) (11.19, 1.27) 0.16	(13.82, 1.48) (13.98, 1.44) 0.05
$\mathcal{F}_{1d}$	(5.21, 0.89) (4.87, 0.98) 0.18	(7.12, 0.91) (6.97, 1.06) 0.07	(9.15, 1.14) (9.07, 1.16) 0.03	(11.27, 1.25) (11.17, 1.27) 0.03	(14.18, 1.42) (13.97, 1.44) 0.07

TAB. 12.1 – Nombre moyen de RDs détectées par signal IRMf actif et inactif pour les séquences synthétiques  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$ ,  $\mathcal{F}_{1d}$  et les cinq détecteurs ( $k=0.7, 1, 1.3, 1.6$  et  $2$ ). A chaque cellule du tableau sont associées 5 mesures, avec dans l'ordre de lecture,  $\mu_{nA}$ ,  $\sigma_{nA}$ ,  $\mu_{nI}$ ,  $\sigma_{nI}$  et la mesure discriminante.

permettant une meilleure discrimination alors que  $k = 1$  est le détecteur permettant la plus mauvaise discrimination. De plus, le pouvoir de discrimination chute très rapidement quand le rapport signal sur bruit diminue.

Cinq autres mesures ont été utilisées pour prendre en compte l'information de localisation temporelle : le nombre moyen de bonnes détections par signal IRMf actif,  $\mu_{dA}$ , le nombre moyen de fausses alarmes (pouvant être considérées comme des bonnes détections) par signal IRMf inactif,  $\mu_{dI}$ , les écarts-types associés,  $\sigma_{dA}$  et  $\sigma_{dI}$ , enfin la mesure discriminante  $|\mu_{dA}-\mu_{dI}|/(\sigma_{dA}+\sigma_{dI})$ . On pourra se référer à la figure 8.1 de la page 118 pour un rappel des définitions. Les mesures ont été effectuées pour deux tailles de fenêtre différentes, à savoir  $d = 1$  et  $d = 2$ , et pour les cinq seuils de détection considérés plus haut. Les valeurs obtenues sont consignées dans le tableau 12.2.

	0.7	1	1.3	1.6	2
$\mathcal{F}_{1b}$	(3.7, 1.37) (0.63, 0.76) 1.44 (5.08, 1.29) (1.05, 0.94) 1.8	(4.08, 1.44) (0.93, 0.88) 1.35 (5.65, 1.14) (1.56, 1.08) 1.84	(3.52, 1.41) (1.25, 0.99) 0.94 (5.14, 1.29) (2.1, 1.15) 1.24	(3.38, 1.4) (1.57, 1.08) 0.72 (5.01, 1.28) (2.61, 1.24) 0.95	(3.59, 1.35) (2.01, 1.16) 0.62 (5.26, 1.19) (3.31, 1.29) 0.78
$\mathcal{F}_{1c}$	(2.26, 1.29) (0.63, 0.77) 0.79 (3.38, 1.41) (1.06, 0.95) 0.98	(2.63, 1.4) (0.93, 0.88) 0.74 (4.03, 1.44) (1.56, 1.08) 0.98	(2.44, 1.3) (1.24, 0.99) 0.52 (3.86, 1.38) (2.09, 1.14) 0.7	(2.49, 1.3) (1.56, 1.08) 0.39 (3.91, 1.33) (2.59, 1.24) 0.51	(2.84, 1.34) (2, 1.15) 0.33 (4.44, 1.31) (3.3, 1.29) 0.43
$\mathcal{F}_{1d}$	(1.42, 1.1) (0.64, 0.77) 0.41 (2.22, 1.26) (1.06, 0.95) 0.52	(1.75, 1.17) (0.93, 0.88) 0.4 (2.82, 1.35) (1.56, 1.08) 0.51	(1.87, 1.16) (1.24, 0.99) 0.29 (3.04, 1.27) (2.08, 1.14) 0.39	(2.01, 1.22) (1.56, 1.07) 0.19 (3.33, 1.33) (2.59, 1.24) 0.28	(2.38, 1.27) (2, 1.15) 0.15 (3.9, 1.35) (3.29, 1.29) 0.23

TAB. 12.2 – Caractéristiques de l'étape de prétraitement en termes de bonnes détections (pour les signaux actifs) et de fausses alarmes pouvant être considérées comme des bonnes détections par le modèle (pour les signaux inactifs). Les différentes mesures sont calculées pour les séquences synthétiques  $\mathcal{F}_{1b}$ ,  $\mathcal{F}_{1c}$ ,  $\mathcal{F}_{1d}$  et les cinq détecteurs ( $k=0.7, 1, 1.3, 1.6$  et  $2$ ). A chaque cellule du tableau sont associées dix mesures, avec dans l'ordre de lecture,  $\mu_{1I}$ ,  $\sigma_{1I}$ ,  $\mu_{1A}$ ,  $\sigma_{1A}$ , la mesure discriminante associée ( $d=1$ ) et  $\mu_{2I}$ ,  $\sigma_{2I}$ ,  $\mu_{2A}$ ,  $\sigma_{2A}$  et la mesure discriminante associée ( $d=2$ ).

Les conclusions tirées dans la seconde partie restent vraies : le pouvoir de discrimination est plus important pour les détecteurs opérant à basse échelle. Il est à noter enfin que l'information de localisation temporelle (tableau 12.2) est beaucoup moins sensible au bruit que l'information de nombre de RDs (tableau 12.1).

### 12.2.1.2 Cartographie

**12.2.1.2.1 Les paramètres utilisés** Nous avons évalué l'influence de trois groupes de paramètres. Un premier groupe concerne l'étape de prétraitement (nombre de détecteurs et taux de détection), un second concerne la connexité utilisée (définition du voisinage) et un dernier est en rapport avec la stratégie utilisée pour former les couples observation-état.

On note  $C_{26}$  la 26-connexité,  $C_6$  la 6-connexité et  $C_0$  la 0-connexité (non prise en compte du voisinage spatial).

On désigne par  $\mathbf{a}_i$ ,  $\mathbf{b}_i$  et  $\mathbf{c}_i$  trois détecteurs permettant de détecter respectivement une, deux ou trois séquences de RDs par signal IRMf. Chaque détecteur est donc caractérisé par la donnée d'un, de deux ou de trois taux de détection  $k$  de sorte que  $\mathbf{b}_3 = \{0.7, 1.6\}$  désigne un détecteur de séquence de RDs opérant à  $k = 0, 7$  et  $k = 1, 6$ . Les détecteurs considérés sont les suivants :

- $\mathbf{a}_1 = \{0.7\}$ ,  $\mathbf{a}_2 = \{1\}$ ,  $\mathbf{a}_3 = \{1.3\}$ ,  $\mathbf{a}_4 = \{1.6\}$  et  $\mathbf{a}_5 = \{2\}$ .
- $\mathbf{b}_1 = \{0.7, 1\}$ ,  $\mathbf{b}_2 = \{0.7, 1.3\}$ ,  $\mathbf{b}_3 = \{0.7, 1.6\}$ ,  $\mathbf{b}_4 = \{0.7, 2\}$ ,  $\mathbf{b}_5 = \{1, 1.3\}$ ,  $\mathbf{b}_6 = \{1, 1.6\}$ ,  $\mathbf{b}_7 = \{1, 2\}$ ,  $\mathbf{b}_8 = \{1.3, 1.6\}$ ,  $\mathbf{b}_9 = \{1.3, 2\}$  et  $\mathbf{b}_{10} = \{1.6, 2\}$
- $\mathbf{c}_1 = \{0.7, 1, 1.3\}$ ,  $\mathbf{c}_2 = \{0.7, 1, 1.6\}$ ,  $\mathbf{c}_3 = \{0.7, 1, 2\}$ ,  $\mathbf{c}_4 = \{0.7, 1.3, 1.6\}$ ,  $\mathbf{c}_5 = \{0.7, 1.3, 2\}$ ,  $\mathbf{c}_6 = \{0.7, 1.6, 2\}$ ,  $\mathbf{c}_7 = \{1, 1.3, 1.6\}$ ,  $\mathbf{c}_8 = \{1, 1.3, 2\}$ ,  $\mathbf{c}_9 = \{1, 1.6, 2\}$  et  $\mathbf{c}_{10} = \{1, 1.6, 2\}$ .

Ainsi, pour un voisinage de type  $C_{26}$  et le détecteur  $\mathbf{c}_1$ , le nombre de séquences de RDs détectés sera de  $(26+1) \times 3 = 81$  séquences alors qu'avec le couple  $C_0$  et  $\mathbf{a}_1$ , l'observation sera composée d'une seule et unique séquence de RDs.

Enfin, le dernier groupe de paramètres dont l'influence est étudiée concerne la stratégie utilisée pour former l'ensemble des couples observation-état. Il s'agit des paramètres *extension* (contrainte de forme), *tailleMin* (contrainte de nombre) et *ecm* (règle pour associer les états cachés) (voir section 11.4.1.3, page 166). Pour l'étude de l'influence des paramètres, nous avons utilisé 3 jeux de paramètres différents allant du moins contraint au plus contraint :

- $\mathbf{A}_1 = \{ \text{extension} = 5, \text{ecm} = 5, \text{tailleMin} = 0.3 \}$ ,
- $\mathbf{A}_2 = \{ \text{extension} = 4, \text{ecm} = 4, \text{tailleMin} = 0.4 \}$ ,
- $\mathbf{A}_3 = \{ \text{extension} = 3, \text{ecm} = 3, \text{tailleMin} = 0.5 \}$ .

A noter que les résultats présentés précédemment ont été obtenus avec le jeu  $\mathbf{A}_1$ .

L'influence des paramètres est étudiée sur la séquence  $\mathcal{F}_{1c}$  uniquement. De plus, le retard d'activité n'est pas estimé mais fixé à 0. Sachant que le retard d'activité est nul, le fixer à 0 permet de gagner un temps important de traitement.

**12.2.1.2.2 Résultats** Le tableau 12.3 présente les résultats obtenus avec les détecteurs  $\mathbf{a}_i$  (avec les 3 stratégies de recherche des couples observation-état et les trois types de connexité) alors que le tableau 12.4 présente les résultats obtenus avec les détecteurs  $\mathbf{b}_i$  et  $\mathbf{c}_i$ .

Nous regardons tout d'abord l'influence du voisinage spatial. Les 3 premières lignes du tableau 12.3 correspondent au cas où l'on ne prend pas en compte le voisinage spatial (connexité  $C_0$ ). Il est alors à souligner que les résultats sont beaucoup plus satisfaisants que ceux obtenus dans la première approche par MSMCSE. Avec cette méthode, on obtient en effet 0 % de bonne détection pour un taux de fausse alarme de 0.01 (avec la série  $\mathcal{F}_{1c}$ , résultat non présenté dans la thèse). Ceci valide notre stratégie de modélisation des lois de durée et notre algorithme de cartographie. Les lignes 4 à 9 montrent les résultats lorsque l'on prend seulement en compte le voisinage spatial (6-connexité pour les lignes 4 à 6 et 26-connexité pour les lignes 7 à 9). Les résultats sont très dégradés lorsque l'on prend en compte les 6 premiers voisins alors qu'ils sont largement meilleurs en 26-connexité. Ceci peut paraître *a priori* surprenant. Il convient donc d'interpréter ce fait.

	$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{a}_3$	$\mathbf{a}_4$	$\mathbf{a}_5$
$\mathbf{A}_1, C_0$	46, 1.3, 50	48, 1.3, 50	39, 1.2, 41	12, 3.4, 25	12, 2.2, 20
	36, 2, 47	36, 1.7, 45	28, 1.3, 31	8, 3.9, 19	8, 3, 16
	33, 6.6, 40	21, 0.5, 18	9, 1.7, 17	8, 9.7, 12	4, 2.8, 7
$\mathbf{A}_2, C_0$	44, 1.6, 50	47, 1.2, 50	37, 1.3, 40	12, 2.6, 22	12, 1, 12
	36, 2.1, 44	38, 0.9, 37	29, 1.8, 35	8, 3.1, 16	8, 1.2, 9
	20, 1.9, 26	22, 1.2, 23	11, 1.1, 12	3, 2.6, 5	3, 0.1, 1
$\mathbf{A}_3, C_0$	39, 1.4, 45	46, 1.4, 50	33, 1.5, 40	11, 2.2, 17	11, 2, 16
	30, 2.4, 41	37, 1.4, 41	24, 1.5, 29	8, 2, 11	8, 1.5, 9
	17, 1.9, 21	18, 2.1, 26	8, 1.4, 10	3, 2.8, 6	3, 0, 0
$\mathbf{A}_1, C_6$	2, 1.43, 3	1, 2.08, 2	0, 12.3, 47	0, 6.95, 22	0, 6.43, 15
	1, 1.79, 2	1, 1.71, 1	0, 8.63, 1	0, 11.51, 7	0, 11.31, 5
	0, 5.98, 1	0, 1.82, 0	0, 3.36, 0	0, 37.18, 0	0, 8.61, 0
$\mathbf{A}_2, C_6$	2, 1.12, 3	2, 2.04, 7	0, 5.05, 21	0, 6.27, 28	0, 6.64, 31
	1, 0.9, 1	1, 1.6, 2	0, 2.92, 1	0, 11.15, 21	0, 10.11, 18
	0, 0.94, 0	0, 4.5, 1	0, 4.64, 0	0, 27.7, 0	0, 37.71, 3
$\mathbf{A}_3, C_6$	56, 6.08, 83	21, 2.32, 68	21, 3.37, 44	0, 4.44, 9	0, 4.64, 11
	56, 12.15, 83	5, 3.97, 61	21, 5.94, 37	0, 6.28, 2	0, 6.66, 4
	56, 60.77, 83	2, 4.84, 4	21, 22.92, 24	0, 20.91, 1	0, 6.39, 0
$\mathbf{A}_1, C_{26}$	98, 0.91, 98	97, 0.46, 93	80, 0.63, 74	42, 0.11, 20	42, 0.05, 10
	96, 1.43, 97	94, 0.3, 85	69, 0.95, 68	36, 0.17, 15	36, 0.07, 4
	92, 3.16, 95	69, 1.28, 76	54, 2.02, 55	17, 0.34, 6	16, 0.34, 0
$\mathbf{A}_2, C_{26}$	92, 2.57, 96	96, 1.14, 96	88, 0.36, 83	48, 0.36, 36	48, 0.3, 34
	88, 3.87, 95	95, 0.65, 94	84, 0.48, 79	40, 0.3, 28	40, 0.04, 19
	76, 5.65, 88	86, 1.48, 91	56, 1.28, 64	24, 0.13, 4	25, 0.13, 0
$\mathbf{A}_3, C_{26}$	35, 1.26, 35	63, 4.71, 80	23, 16.43, 72	9, 1.9, 14	10, 2.66, 17
	35, 2.51, 35	58, 9.41, 80	16, 32.86, 72	5, 1.52, 7	6, 4.07, 14
	35, 12.57, 35	58, 47.06, 80	11, 2.96, 13	2, 1.68, 3	3, 2.69, 4

TAB. 12.3 – Influence des paramètres sur les résultats de cartographie dans le cas où l'on a 1 détecteur. A une cellule du tableau, est associée le détecteur ( $\mathbf{a}_1$ ,  $\mathbf{a}_2$ ,  $\mathbf{a}_3$ ,  $\mathbf{a}_4$  ou  $\mathbf{a}_5$ ), l'algorithme ( $\mathbf{A}_1$ ,  $\mathbf{A}_2$  ou  $\mathbf{A}_3$ ) et la connexité ( $C_0$ ,  $C_6$  ou  $C_{26}$ ). Pour chaque cellule, il est possible de déterminer une carte de p-valeurs, cette dernière est résumée ensuite par 9 mesures consignées dans la dite-cellule. Les trois chiffres situés sur la même ligne sont respectivement de gauche à droite la valeur de la courbe ROC, de la courbe  $\alpha$  et de la courbe  $\beta$  pour une abscisse de 0.01 (en haut), de 0.005 (au milieu) et de 0.001 (en bas). A noter que les résultats pour les courbes  $\beta$  et les courbes COR ont été multipliés par 100 de manière à obtenir une mesure entre 0 et 100. On rappelle que la courbe  $\alpha$  correspond au nombre de fausses alarmes observées sur le nombre de fausses alarmes attendues en fonction de la p-valeur estimée et que la courbe  $\beta$  est le pourcentage de bonnes détections en fonction de la p-valeur estimée.

Notons  $p$  la probabilité de détecter un événement susceptible d'être associé au  $i$ -ème stimulus du paradigme pour un canal fixé, le signal étant supposé inactif (hypothèse  $\mathcal{H}_0$ ). On suppose que  $p$  ne dépend pas du canal et que les différents canaux sont mutuellement indépendants. La probabilité  $q(l)$  d'avoir au moins  $l$  événements associés à la  $i$ -ème transition du paradigme s'écrit donc :

$$q(l) = \sum_{i=l}^N C_i^N p^i (1-p)^{N-i} \quad (12.1)$$

On rappelle que  $N$  représente le nombre de canaux. La figure 12.6.a représente cette probabilité en fonction de  $p$  pour 3 valeurs de  $l$  ( $l = 2, 3, 4$ ) dans le cas  $N = 7$  alors que la figure 12.6.b représente cette même probabilité dans le cas où  $N = 27$  et avec  $l = 9, 12, 14$ . La droite  $y = x$  représente enfin le cas où l'on a un unique canal. Ainsi, lorsque la courbe est au-dessus de la droite  $y = x$ , cela signifie que la probabilité de pouvoir associer une association d'événements à un bloc de stimulation est plus importante que dans le cas où l'on a un unique canal.

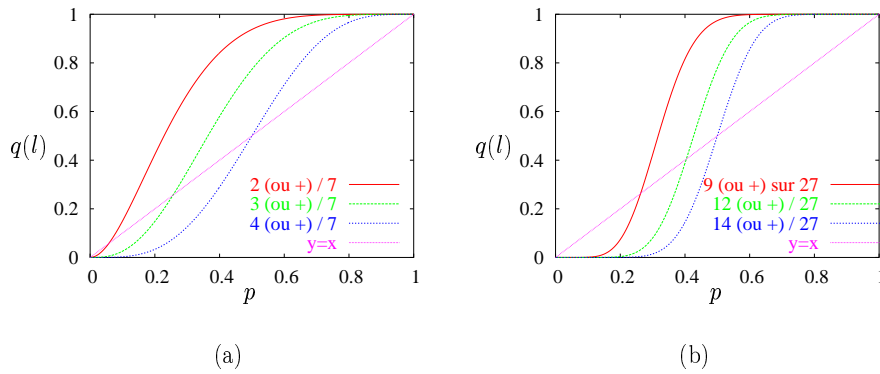


FIG. 12.6 – Probabilité  $q(l)$  pour 3 valeurs distinctes de  $l$  dans le cas (a)  $N = 7$  et (b)  $N = 27$ .

Le cas  $l=2$  de la figure 12.6.a correspond à une taille de signature proche de 0.3 si les canaux sont de poids égaux. Cette taille correspond approximativement à la contrainte de taille de l'algorithme le moins contraint  $\mathbf{A}_1$ . Ainsi, dans le cas où l'on a 7 canaux (figure 12.6.a) et qu'on utilise l'algorithme  $\mathbf{A}_1$ , la probabilité d'associer des événements à un état caché de la chaîne est relativement importante, même pour des faibles valeurs de  $p$ . Par conséquent, même pour des signaux inactifs, il est fort probable que chaque transition du paradigme soit associée à une signature. La forme de la signature, le nombre d'événements détectés dans la signature, le nombre de fausses alarmes et les instants des événements profonds deviennent principalement les mesures sur lesquelles la discrimination peut s'effectuer. Ces mesures sont discriminantes dès lors que le rapport signal sur bruit est élevé. On obtient en effet de très bons résultats avec la séquence  $\mathcal{F}_{1a}$  en utilisant le jeu de paramètres ( $\mathbf{a}_1$ ,  $C_6$  et  $\mathbf{A}_1$ ) (la courbe COR passe par exemple par le point (0.005,0.92)). Cependant, dès lors qu'il est difficile de différencier la signature d'un signal actif de celle provenant d'un signal inactif, les résultats se dégradent très vite. C'est pourquoi les résultats obtenus en prenant en compte un faible voisinage peuvent conduire à des résultats moins bons que ceux obtenus sans prendre en compte le voisinage. Pour résoudre ce problème, il est possible d'utiliser des contraintes de nombre plus fortes lors de la création des couples observation-état. La courbe reste alors beaucoup plus longtemps sous la droite  $y = x$  (figure 12.6.a). Les résultats obtenus avec l'algorithme  $A_3$  en 6-connexité sont donc plus satisfaisants que ceux obtenus avec l'algorithme  $A_1$ . Cependant, il peut alors se poser le problème de l'observabilité des événements d'intérêt. Dans le cas où l'on considère



27 canaux, les différentes courbes restent sous la droite  $y = x$  sur un intervalle relativement important (figure 12.6.b). Les résultats en 26-connexité sont donc plus satisfaisants.

L'influence du voisinage spatial sur la qualité des résultats est très importante. Il faut noter également que les résultats se dégradent lorsque le taux de détection devient trop important (pour  $\alpha_3$ ,  $\alpha_4$  et  $\alpha_5$ ). Augmenter le taux de détection diminue en fait la distance qui peut exister entre un signal actif et un signal inactif (cf. tableau 13.2). On note également que les résultats sont d'autant meilleurs que la stratégie de création des couples observation-état est moins contrainte. Cela se comprend facilement au travers du tableau 13.2 où l'on voit qu'augmenter «  $d$  » augmente le pouvoir discriminant des détecteurs.

Les résultats en utilisant 2 ou 3 détecteurs sont consignés dans le tableau 12.4. L'analyse du tableau 12.4 se fonde en partie sur les conclusions que l'on a déjà tirées.

Dès lors que le nombre de canaux est faible (couples  $\mathbf{A}_i, c_i, C_0$  (3 canaux),  $\mathbf{A}_i, b_i, C_0$  (2 canaux)), les résultats sont mauvais. En augmentant le nombre de canaux (couples  $\mathbf{A}_i, c_i, C_6$  (21 canaux),  $\mathbf{A}_i, b_i, C_6$  (14 canaux)), les résultats s'améliorent petit à petit. Enfin, avec un nombre important de canaux, (couples  $\mathbf{A}_i, c_i, C_{26}$  (81 canaux),  $\mathbf{A}_i, b_i, C_{26}$  (54 canaux)), les résultats sont très améliorés. Dès lors que le nombre de canaux est supérieur ou égal à 27 environ (utilisation de la 26-connexité avec un, deux ou trois détecteurs), les résultats sont satisfaisants et sont quasiment identiques. Le nombre de canaux à partir duquel les résultats ne sont pas améliorés dépend fortement du bruit. En utilisant la série  $\mathcal{F}_{1b}$ , les résultats obtenus en 6-connexité avec 3 détecteurs, c'est-à-dire 18 canaux, sont aussi bons que ceux obtenus en 26-connexité avec 3 détecteurs.

Conformément au tableau 13.2, on remarque encore que les résultats se dégradent petit à petit lorsque les taux de détection des différents détecteurs augmentent.

Enfin, de manière générale, le calcul de la p-valeur est relativement instable et peu de crédit peut être attribué à son estimation. Dès lors que le nombre de canaux est important et qu'on utilise un algorithme peu contraint ( $\mathbf{A}_1$  ou  $\mathbf{A}_2$ ), il semble que le nombre de fausses alarmes soit beaucoup plus petit que le nombre attendu. Il est enfin difficile de trouver un comportement « prévisible » avec l'algorithme le plus contraint.

## 12.2.2 Etude sur données réelles

### 12.2.2.1 Protocole

Nous évaluons ici l'influence, de l'étape de prétraitement (nombre de détecteurs et taux de détection), et de la stratégie utilisée pour former les couples observation-état, sur les résultats de cartographie avec des données réelles. Nous nous plaçons dans le cadre de la 26-connexité. De plus, avec les mêmes notations que précédemment, nous avons formé 12 couples détecteur-algorithme pour lesquelles les traitements ont été lancés. A chaque couple est associé un entier permettant de les numéroter. Les couples choisis ont permis d'obtenir des résultats relativement satisfaisants sur données synthétiques. Nous n'avons pas utilisé des couples dont on savait qu'ils allaient donner de mauvais résultats. Ainsi, nous n'utilisons pas par exemple l'algorithme le plus contraint  $\mathbf{A}_3$ . Nous avons finalement :

- $\{1, c_1, \mathbf{A}_1\}$ ,
- $\{2, c_1, \mathbf{A}_2\}$ ,
- $\{3, c_2, \mathbf{A}_1\}$ ,
- $\{4, c_2, \mathbf{A}_2\}$ ,
- $\{5, c_3, \mathbf{A}_1\}$ ,
- $\{6, c_3, \mathbf{A}_2\}$ ,

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_{10}$
$A_1, C_0$	5, 2.16, 15	2, 3.35, 4	0, 3.52, 1	1, 2.72, 1	10, 2.16, 22	2, 2.63, 6	0, 1.82, 1	3, 1.12, 3	1, 1.53, 3	2, 1.04, 3
	3, 2.96, 9	1, 2.78, 2	0, 4.81, 1	0, 4.07, 1	3, 3.11, 15	2, 3.91, 4	0, 2.42, 1	3, 1.25, 3	0, 1.64, 1	2, 1.09, 2
	1, 8.87, 5	0, 2.29, 1	0, 9.34, 0	0, 8.67, 1	1, 4.71, 3	2, 10.76, 2	0, 6.99, 0	2, 2.15, 2	0, 0.74, 0	1, 0.47, 1
$A_2, C_0$	7, 0.92, 7	2, 3.05, 5	1, 3.42, 3	1, 1.7, 1	10, 1.74, 19	3, 1.91, 5	1, 2.08, 2	4, 1.56, 5	1, 1.65, 2	3, 1.11, 3
	4, 1.05, 4	1, 3.56, 3	1, 2.72, 1	0, 1.8, 1	3, 1.22, 4	2, 2.96, 4	0, 2.65, 1	3, 0.86, 3	0, 1.05, 0	2, 1.69, 3
	1, 2.15, 1	0, 3.9, 1	0, 6.05, 1	0, 1.75, 0	1, 2.42, 1	1, 6.99, 2	0, 1.68, 0	2, 0.54, 2	0, 3.16, 0	1, 2.29, 1
$A_3, C_0$	9, 1.51, 15	3, 1.21, 4	2, 1.59, 3	2, 1.79, 2	12, 1.05, 13	4, 2.45, 9	2, 2.9, 4	5, 1.26, 5	2, 2.66, 5	3, 1.12, 4
	4, 0.91, 4	2, 1.4, 3	2, 1.05, 2	2, 1.77, 2	6, 0.67, 4	2, 3.11, 6	1, 3.55, 3	3, 1.1, 3	1, 2.65, 2	2, 1.57, 3
	1, 0.54, 1	0, 2.22, 1	1, 0.47, 1	1, 0.13, 1	1, 2.69, 3	2, 8.2, 3	0, 3.83, 1	2, 0.61, 2	0, 0.81, 0	2, 4.5, 2
$A_1, C_6$	75, 1.86, 91	31, 2.16, 83	12, 2.45, 75	39, 2.58, 70	40, 2.05, 85	53, 1.67, 75	51, 0.85, 39	2, 3.3, 59	34, 1.61, 46	1, 2.56, 29
	38, 3, 87	8, 3.54, 73	5, 3.59, 57	39, 4.19, 56	40, 3.6, 77	16, 2.08, 57	4, 1.64, 34	2, 4.92, 42	15, 1.33, 26	1, 3.74, 21
	0, 7.6, 59	5, 12.1, 41	5, 9.34, 8	39, 17.61, 41	40, 11.03, 43	16, 7.46, 24	0, 6.79, 14	2, 13.92, 10	8, 4.77, 11	1, 13.04, 10
$A_2, C_6$	93, 1.21, 94	86, 0.85, 84	73, 1.08, 74	51, 1.53, 67	83, 1.12, 85	78, 0.83, 76	71, 1.24, 75	53, 2.43, 69	45, 2.01, 60	18, 1.78, 30
	87, 0.89, 86	75, 1.36, 79	44, 1.65, 70	15, 2.3, 57	69, 1.71, 81	69, 0.99, 69	59, 1.14, 62	21, 3.6, 61	33, 2.42, 48	7, 1.53, 13
	64, 1.95, 68	32, 2.49, 39	0, 3.97, 30	11, 5.18, 15	69, 6.59, 74	48, 3.36, 59	36, 3.56, 37	0, 8.94, 39	21, 5.38, 33	0, 4.44, 4
$A_3, C_6$	72, 7.41, 93	60, 8.5, 90	46, 1.6, 54	30, 2.25, 53	64, 1.04, 64	58, 0.71, 54	52, 1.2, 56	26, 2.05, 39	28, 1.41, 32	12, 1.57, 16
	62, 14.79, 93	47, 17.01, 90	27, 1.72, 41	13, 2.26, 36	52, 1.38, 58	47, 1.2, 49	38, 1.45, 46	26, 2.73, 29	18, 2.03, 29	5, 2.19, 12
	50, 4.24, 59	47, 7.39, 53	21, 5.58, 31	10, 5.98, 16	40, 4.77, 50	17, 1.14, 17	13, 4.03, 34	26, 11.09, 26	5, 2.96, 13	0, 3.03, 0
$A_1, C_{26}$	99, 0.49, 98	96, 0.28, 90	89, 0.15, 77	84, 0.22, 72	93, 0.31, 86	86, 0.21, 77	82, 0.03, 56	68, 0.47, 61	63, 0.11, 39	46, 0.07, 21
	98, 0.4, 95	94, 0.07, 78	84, 0.16, 67	81, 0.2, 62	88, 0.05, 77	83, 0.2, 69	77, 0.04, 49	62, 0.16, 41	52, 0.16, 31	39, 0.05, 10
	93, 0.13, 76	84, 0.13, 72	71, 0.2, 56	61, 0.2, 36	82, 0.13, 55	68, 0.27, 51	65, 0.13, 9	43, 0.13, 17	33, 0.4, 24	24, 0.27, 5
$A_2, C_{26}$	96, 26.58, 100	93, 4.18, 98	91, 1.09, 92	88, 0.58, 84	96, 0.21, 91	93, 0.18, 86	89, 0.34, 84	81, 0.26, 69	73, 0.44, 63	57, 0.3, 43
	94, 6.82, 98	91, 5.55, 98	89, 1.08, 89	83, 0.51, 78	90, 0.05, 85	90, 0.03, 75	86, 0.32, 81	73, 0.13, 57	65, 0.38, 56	48, 0.09, 27
	87, 20.1, 97	89, 5.04, 91	84, 0.07, 77	68, 1.61, 72	88, 0, 69	83, 0.07, 55	75, 0.34, 66	59, 0, 36	52, 0, 35	35, 0, 18
$A_3, C_{26}$	53, 1.18, 53	100, 0.96, 29	33, 1.1, 33	51, 2.49, 51	75, 5.36, 75	68, 5.07, 68	53, 9.57, 84	14, 23.37, 72	17, 36.78, 84	8, 1.98, 13
	53, 2.35, 53	29, 1.92, 29	33, 2.21, 33	51, 4.97, 51	75, 10.72, 75	68, 10.14, 68	47, 19.15, 84	11, 38.9, 69	13, 6.09, 30	6, 0.43, 4
	53, 11.76, 53	29, 9.61, 29	33, 11.03, 33	51, 24.87, 51	75, 53.58, 75	68, 50.69, 68	47, 6.72, 49	11, 7.6, 12	7, 6.59, 14	3, 0.87, 3
	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$
$A_1, C_0$	14, 1.08, 16	2, 1.09, 2	0, 1.21, 0	12, 1.41, 15	3, 0.89, 3	6, 1.57, 10	22, 1.63, 27	5, 2.07, 15	9, 1.68, 13	13, 0.81, 11
	8, 1.01, 8	1, 1.36, 1	0, 1.6, 0	6, 1.33, 8	1, 1.02, 1	1, 1.72, 3	15, 1.33, 19	2, 2.5, 10	3, 2.06, 9	7, 0.74, 6
	1, 1.21, 1	0, 1.01, 0	0, 2.42, 0	1, 1.08, 1	0, 1.88, 0	0, 2.15, 0	4, 1.28, 5	2, 7.19, 2	2, 4.97, 3	2, 1.61, 3
$A_2, C_0$	14, 2.53, 20	9, 1.4, 11	4, 1.71, 6	6, 2.64, 11	2, 2.46, 3	1, 0.65, 1	19, 1.52, 23	12, 1.73, 15	5, 2.06, 9	9, 1.2, 11
	12, 1.57, 14	7, 1.99, 9	2, 1.76, 4	3, 2.72, 6	2, 3.07, 2	0, 0.67, 0	14, 1.51, 16	9, 1.91, 11	3, 2.38, 7	3, 0.78, 2
	4, 1.41, 5	2, 2.49, 5	1, 4.1, 2	0, 4.97, 2	0, 1.75, 0	0, 2.15, 0	3, 2.49, 6	1, 1.34, 2	0, 1.08, 0	0, 0.94, 0
$A_3, C_0$	12, 1.78, 16	8, 2.7, 14	6, 1.59, 8	4, 2.4, 8	2, 3.33, 4	2, 1.48, 2	14, 1.56, 17	11, 1.36, 13	3, 1.79, 6	9, 1.29, 10
	9, 2.37, 14	5, 1.99, 8	3, 2.02, 6	2, 1.3, 3	2, 4.15, 3	1, 1.76, 2	9, 1.47, 12	8, 1.52, 10	1, 2.12, 4	4, 1.36, 5
	9, 4.77, 9	1, 3.9, 4	1, 2.62, 2	1, 1.68, 1	1, 10.15, 2	0, 2.29, 0	4, 3.36, 5	2, 0.74, 2	0, 2.42, 0	1, 0.4, 0
$A_1, C_6$	90, 1.63, 94	85, 1.15, 87	80, 0.86, 78	73, 1.53, 78	69, 1.09, 71	59, 0.78, 53	72, 1.39, 77	71, 0.58, 62	62, 0.42, 43	34, 0.81, 31
	80, 1.83, 89	78, 0.91, 75	68, 0.79, 61	46, 1.43, 65	56, 1.06, 58	44, 0.99, 44	41, 1.51, 63	60, 0.65, 51	47, 0.65, 39	23, 0.82, 18
	31, 3.76, 75	44, 2.62, 61	33, 2.49, 33	8, 4.24, 35	29, 2.08, 31	1, 3.56, 31	34, 4.37, 34	27, 2.08, 36	25, 2.29, 28	0, 2.49, 5
$A_2, C_6$	90, 0.59, 85	89, 0.84, 89	86, 1.38, 89	77, 0.73, 73	76, 0.86, 74	61, 1.5, 68	77, 0.71, 73	75, 0.83, 74	67, 0.42, 55	45, 0.67, 37
	83, 0.83, 80	85, 0.81, 82	83, 1.24, 84	64, 1.12, 65	68, 0.89, 65	49, 1.55, 55	68, 0.81, 65	68, 0.75, 64	57, 0.51, 47	33, 0.51, 22
	53, 1.14, 53	59, 1.08, 61	65, 1.75, 73	29, 1.95, 44	38, 0.54, 17	17, 2.62, 35	32, 2.29, 55	29, 1.21, 39	34, 0.47, 11	8, 1.34, 11
$A_3, C_6$	61, 5.65, 83	58, 5.03, 83	61, 6.31, 85	40, 1.75, 47	41, 3.42, 59	25, 2.07, 35	53, 0.99, 53	52, 0.99, 52	37, 1.15, 39	24, 1.87, 31
	51, 11.28, 83	50, 0.95, 48	58, 12.61, 85	32, 1.04, 32	31, 2.41, 43	18, 2.73, 29	41, 1.16, 44	41, 1.28, 45	26, 1.71, 34	17, 2.08, 24
	43, 3.9, 49	30, 1.41, 33	58, 15.13, 67	11, 1.68, 14	23, 5.51, 32	9, 3.09, 13	21, 2.22, 29	19, 1.68, 27	22, 3.23, 23	16, 5.04, 17
$A_1, C_{26}$	97, 0.49, 95	93, 0.46, 90	89, 0.03, 69	88, 0.23, 78	83, 0.26, 69	79, 0.17, 59	86, 0.52, 81	82, 0.29, 70	75, 0.13, 53	59, 0.12, 32
	95, 0.55, 92	90, 0.67, 89	85, 0.04, 59	84, 0.13, 69	78, 0.3, 66	73, 0.09, 50	81, 0.22, 72	75, 0.2, 61	69, 0.09, 45	51, 0.08, 22
	88, 0.94, 87	81, 0.4, 76	76, 0.07, 23	70, 0.27, 59	62, 0.13, 29	56, 0, 12	70, 0.67, 61	60, 0.34, 49	52, 0.13, 14	30, 0.13, 7
$A_2, C_{26}$	95, 2.18, 97	95, 1.32, 95	96, 1.08, 96	90, 0.13, 81	92, 0.17, 83	84, 0.34, 73	92, 0.25, 86	93, 0.04, 77	86, 0.3, 78	73, 0.07, 51
	93, 1.37, 94	93, 0.69, 91	94, 0.81, 94	88, 0.08, 76	89, 0.07, 71	77, 0.19, 58	90, 0.31, 85	89, 0.01, 68	82, 0.12, 65	64, 0.04, 39
	85, 0.74, 84	88, 1.28, 89	89, 0.2, 83	79, 0.13, 68	80, 0.07, 52	60, 0, 28	83, 0.13, 77	84, 0, 33	71, 0, 43	52, 0, 31
$A_3, C_{26}$	39, 1.2, 39	100, 0.95, 37	43, 1.16, 43	32, 1.67, 32	33, 2.23, 33	46, 6.18, 46	31, 5.4, 60	66, 6.85, 66	24, 12.91, 66	13, 33.07, 75
	39, 2.39, 39	37, 1.91, 37	43, 2.31, 43	32, 3.33, 32	33, 4.45, 33	46, 12.37, 46	25, 10.8, 60	66, 13.7, 66	20, 25.83, 66	12, 2.62, 15
	39, 11.97, 39	37, 9.55, 37	43, 11.56, 43	32, 16.67, 32	33, 22.25, 33	46, 61.85, 46	18, 3.29, 21	66, 68.5, 66	13, 2.42, 17	6, 2.42, 9

TAB. 12.4 – Influence des paramètres sur données synthétiques (même légende que le tableau 12.3)

- $\{7, \mathbf{b}_1, \mathbf{A}_1\}$ ,
- $\{8, \mathbf{b}_1, \mathbf{A}_2\}$ ,
- $\{9, \mathbf{b}_2, \mathbf{A}_1\}$ ,
- $\{10, \mathbf{b}_2, \mathbf{A}_2\}$ ,
- $\{11, \mathbf{b}_3, \mathbf{A}_1\}$ ,
- $\{12, \mathbf{b}_3, \mathbf{A}_2\}$ .

La vérité n'étant pas connue, on utilise comme précédemment la mesure de similarité entre deux cartes d'activation pour évaluer l'influence des paramètres (voir section 8.2, page 121 pour la définition). Nous rappelons que la mesure de similarité varie entre 0 (les deux cartes d'activation n'ont aucun voxel actif en commun) à 1 (les deux cartes d'activation sont identiques).

Pour chaque couple, il est possible d'obtenir une carte d'activation  $\mathcal{A}_i$  à partir d'une séquence IRMf ( $i$  correspond au numéro du couple détecteur-algorithme). En pratique, nous avons lancé les traitements sur trois séquences d'images différentes. Les mesures de similarité sont ensuite déterminées entre les différentes cartes  $\mathcal{A}_i$  et  $\mathcal{A}_j$ ,  $i \neq j$  pour les 3 séquences. Pour chaque indice  $i$ , on détermine la moyenne  $\mu_i$  de la similarité obtenue avec les autres jeux de paramètres. On note  $\sigma_i$  l'écart-type correspondant. Les résultats sont consignés sur la figure 12.7 pour 3 p-valeurs différentes, la p-valeur étant utilisée pour le seuillage de la carte.

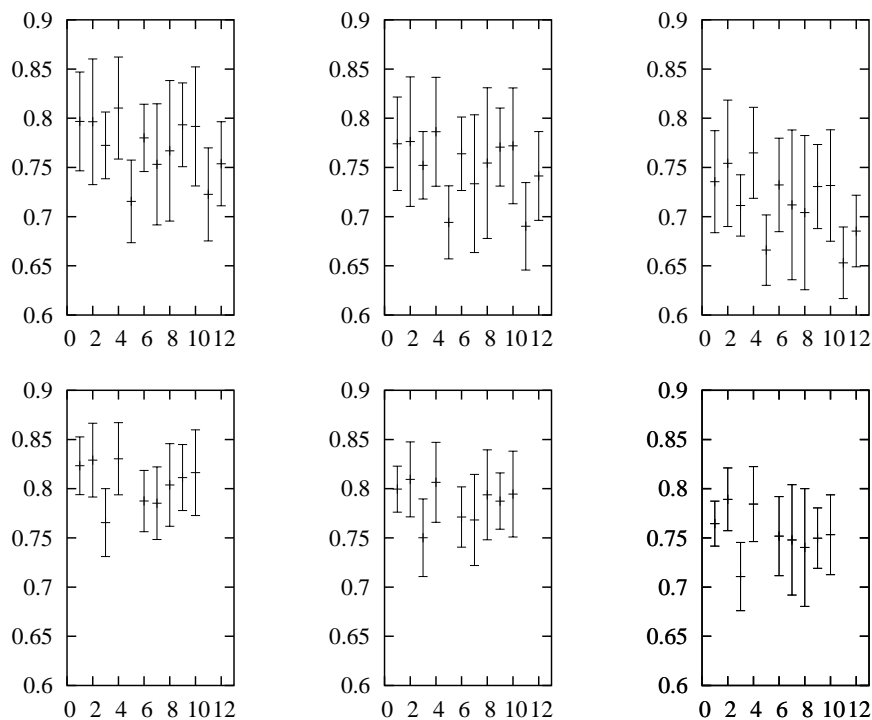


FIG. 12.7 – Influence des paramètres sur données réelles. Le numéro du couple est reporté en abscisse alors que la similarité moyenne est reportée sur l'axe des ordonnées. La longueur des traits correspond à 2 écarts-type. Un seuil de 0.01 est utilisé pour la colonne gauche, de 0.005 pour la colonne du milieu et de 0.001 pour la colonne de droite. La ligne du dessus correspond au cas où l'on considère tous les jeux de paramètres. La ligne du dessous correspond au cas où l'on s'intéresse uniquement aux jeux de paramètres  $i = 1, 2, 3, 4, 6, 7, 8, 9$  et 10.

### 12.2.2.2 Analyse

Il apparaît tout d'abord que les valeurs moyennes de similarité sont relativement importantes (supérieures à 0,65). Les jeux de paramètres 5, 11, 12 et dans une moindre mesure 3 sont ceux qui fournissent les cartes d'activations les plus différentes : ces paramètres correspondent effectivement à des taux de détection relativement importants. Il semble de plus que les résultats obtenus avec ces jeux de paramètres soient moins satisfaisants. Ainsi, nous obtenons le même constat pour les données réelles que pour les données synthétiques. Pour discriminer au mieux les signaux actifs des signaux inactifs, il est nécessaire d'utiliser des taux de détection relativement faibles. Sans considérer les jeux de paramètres 5, 11 et 12, les valeurs de similarité augmentent (elles dépassent toutes 0,7) alors que les écarts-types diminuent (seconde ligne de la figure 12.7).

En conclusion, le choix du détecteur a un rôle important car selon le taux de détection utilisé, son pouvoir de discrimination varie. On remarque cependant que pour les données de type « paradigme en blocs », les résultats ne dépendent que très peu du choix des paramètres. Aussi, la principale limitation provient du calcul de la p-valeur.

# Chapitre 13

## Résultats expérimentaux : paradigmes événementiels

Nous utilisons la même démarche que dans le chapitre précédent : nous commençons par présenter les résultats obtenus sur données synthétiques et sur données réelles. Nous finissons ensuite par étudier l'influence des paramètres. Les données traitées (de nature événementielle) diffèrent bien entendu de celles utilisées précédemment : nous les décrivons maintenant.

### 13.1 Données

#### 13.1.1 Séquence IRMf synthétiques

Pour construire les données synthétiques dans le cadre des paradigmes événementiels, nous utilisons la même démarche que dans le cadre des paradigmes en blocs (section 7.1.1, page 101). Nous rappelons ici son principe.

Trois séquences synthétiques d'images IRMf 3-D, notées  $\mathcal{F}_{Eb}$ ,  $\mathcal{F}_{Ec}$  et  $\mathcal{F}_{Ed}$ , ont été construites en additionnant, dans des zones définies comme actives, un signal d'activation synthétique au bruit. Dans chaque séquence, quatre cubes, de taille  $6 \times 6 \times 6$  voxels, ont été définies comme zones actives. Ces zones sont notées suivant l'ordre lexicographique  $Z_1$ ,  $Z_2$ ,  $Z_3$  et  $Z_4$ , respectivement (figure 13.1).

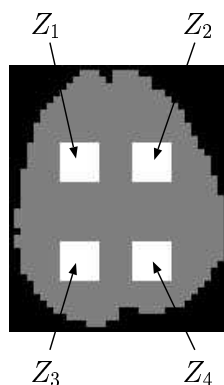


FIG. 13.1 – Masque d'activité définissant les zones actives ( $Z_1$ ,  $Z_2$ ,  $Z_3$  et  $Z_4$ ) et les zones inactives.

De façon plus formelle, pour chacune des trois séquences considérées, si  $\mathbf{X}_n$  désigne le signal IRMf au voxel  $n$ , alors :

$$\mathbf{X}_n = \mathbf{X}_{1n} + \alpha_n \mathbf{X}_{2n} \quad (13.1)$$

Les signaux  $\mathbf{X}_n$ ,  $\mathbf{X}_{1n}$  et  $\mathbf{X}_{2n}$  sont de longueur égale, à savoir  $T = 270$  :

- $\mathbf{X}_{1n}$  représente le bruit observé au voxel  $n$  dans une séquence  $\mathcal{F}'_0$ . Cette séquence a été acquise avec un imageur IRM Bruker S200 2 Teslas (antenne de volume, EPI, orientation axiale, 32 coupes,  $64 \times 64$  pixels, volume voxel =  $4 \times 4 \times 4mm$ , TE=43ms, TR=5s). Pendant l'acquisition, le sujet n'a été soumis à aucun paradigme de stimulation. Il lui a été demandé au préalable de ne pas bouger et de rester le plus calme possible. Parce qu'acquise sous l'hypothèse d'inactivité cérébrale, la séquence  $\mathcal{F}'_0$  peut être considérée comme constituée de bruit seul. Les images ont ensuite été recalées sur la première image de la séquence [169]. Les séquences événementielles étant souvent acquises avec une fréquence d'échantillonnage relativement importante, nous avons fait par la suite comme si la séquence  $\mathcal{F}'_0$  avait été obtenue avec un TR de 3s.
- $\mathbf{X}_{2n}$  représente le signal synthétique d'activation ajouté au bruit dans les zones définies comme actives (cas  $\alpha_n \neq 0$ ). Hors de ces zones, aucun signal d'activation n'est ajouté au bruit (cas  $\alpha_n = 0$ ). Le signal d'activation résulte de la convolution d'une fonction déterministe  $x_0(\cdot)$  représentative du paradigme d'activation avec le modèle de réponse hémodynamique proposé par le logiciel SPM (équation (2.12) de la page 36), notée  $h_{SPM}(\cdot)$  :

$$\mathbf{X}_{2n}(t) = (x_0 * h_{SPM})(t) \quad (13.2)$$

La fonction  $x_0(\cdot)$  est nulle à tout instant à part pour les instants d'occurrence d'un stimulus auquel cas elle vaut 1. Le paradigme, de type événementiel, est composé de 34 stimuli (figure 13.2). La séquence d'instants d'occurrence des stimuli est obtenue à partir de celle d'un paradigme « réel » présenté dans la prochaine section.

- $\alpha_n$  représente un coefficient d'activité égal à 0 en dehors des zones actives. Pour chaque signal actif, le coefficient  $\alpha_n$  est calculé à partir de la spécification du rapport signal sur bruit  $m$  de l'expérience, défini comme le rapport de l'amplitude de la réponse sur l'écart-type du bruit [170]. Ce dernier est estimé sur le bruit auquel on a au préalable soustrait la dérive de ligne de base. Les séquences  $\mathcal{F}_{Eb}$ ,  $\mathcal{F}_{Ec}$  et  $\mathcal{F}_{Ed}$  ont respectivement un rapport signal sur bruit  $m$  égal à 1, 0.5 et 0.25.

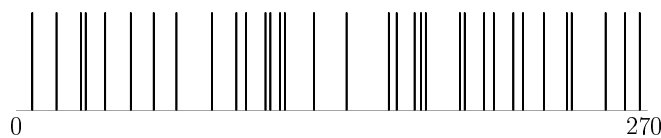


FIG. 13.2 – Séquence de stimuli du paradigme

Une fois les différentes séquences créées, un filtrage spatial gaussien isotrope (largeur à mi-hauteur de 4 mm) et un filtrage spatial temporel ( $\sigma = \sqrt{8}$  secondes) sont appliqués. Des exemples de signaux actifs et inactifs sont illustrés figure 13.3. Il faut noter que les différentes séquences ne diffèrent que par le rapport signal sur bruit.

### 13.1.2 Séquences IRMf réelles

Trois séquences IRMf événementielles ont été acquises à partir d'un même sujet sain sans problème neurologique antérieur. Chaque séquence se compose de 270 images 3-D. Les images

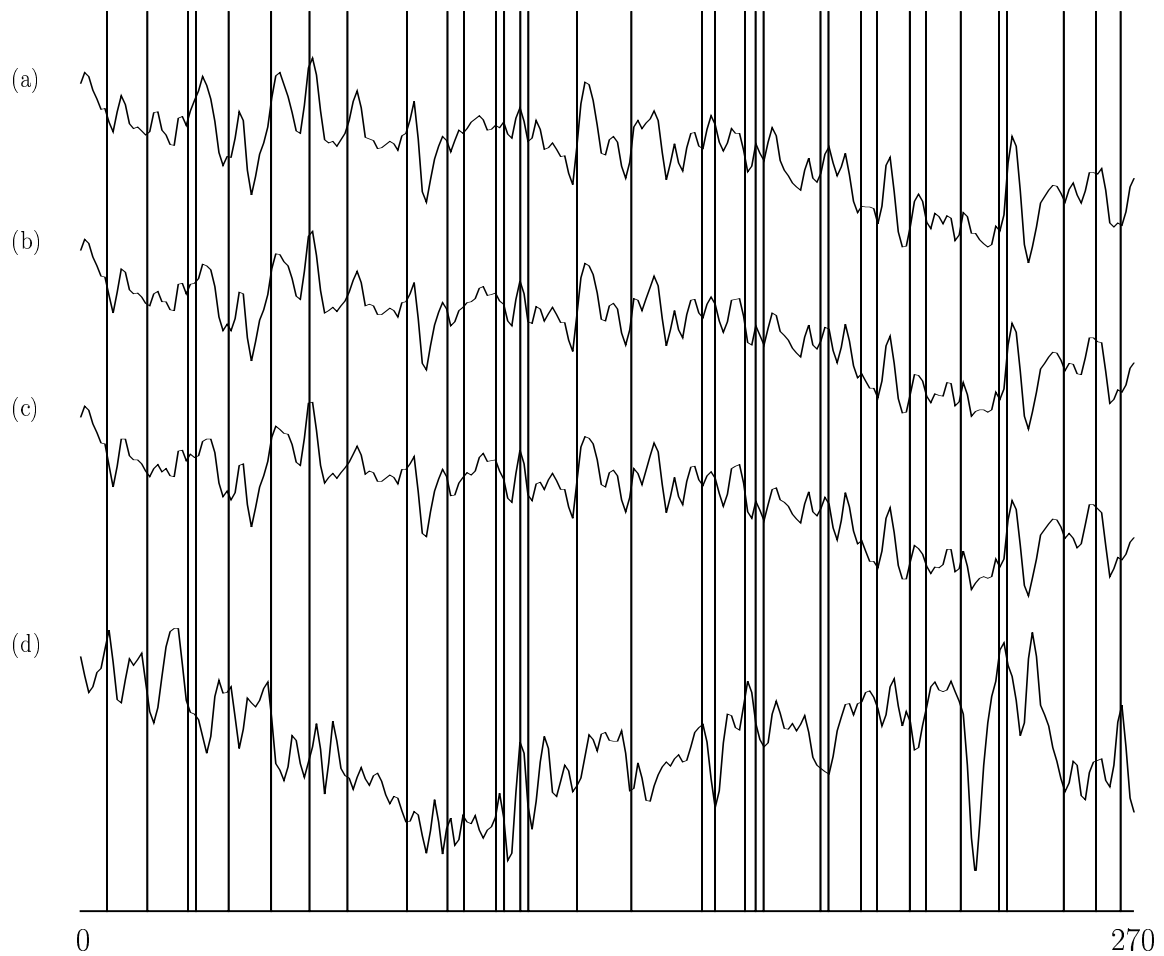


FIG. 13.3 – Signaux IRMf actifs extraits des séquences synthétiques (a)  $\mathcal{F}_{Eb}$ , (b)  $\mathcal{F}_{Ec}$  (b) et (c)  $\mathcal{F}_{Ed}$  pour le même voxel. (d) Exemple de signal inactif. Les traits verticaux représentent les instants de stimulation du paradigme.

ont été acquises avec un imageur S200 Bruker 2 Teslas (antenne de volume, EPI, 24 coupes, 64 x 64 pixels, volume voxel = 4 x 4 x 4 mm, TE=43ms, TR=4s). Le signal EEG a été acquis également au cours du paradigme. Les différents volumes cérébraux sont séparés de 1,6 secondes, temps pendant lequel le signal EEG est enregistré. Le protocole de stimulation utilisé est de type « oddball ». Il est demandé au patient de détecter l'apparition d'images spécifiques dans le flot d'images noir et blanc qui lui est présenté durant la session. Ces images spécifiques, appelées cibles, consistent en la lettre « X » affichée sur fond noir. Les autres images sont soit des images facilement nommables par le sujet, chacune présentée une seule fois, appelées nouveautés, soit des images représentant la lettre « O » sur fond noir, appelées distracteurs. Les cibles sont au nombre de 34. Leurs instants d'occurrence sont indiqués figure 13.4.a. Les nouveautés sont au nombre de 39. Leurs instants d'occurrence sont indiqués figure 13.4.b. Les distracteurs sont présentés quant à eux entre les cibles et les nouveautés. Pendant la session, il est demandé au sujet de signaler l'apparition des cibles en appuyant sur un bouton à l'aide de l'index droit. Le lecteur est renvoyé à [75] pour de plus amples informations sur le paradigme événementiel utilisé.

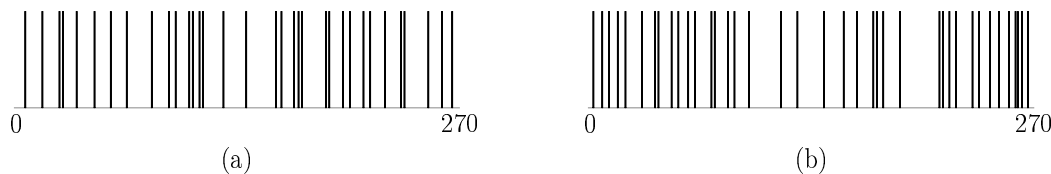


FIG. 13.4 – Paradigme événementiel. Instants de présentation (a) des cibles (lettre « X » sur fond noir) (b) des nouveautés.

## 13.2 Résultats

### 13.2.1 Résultats sur données synthétiques

Les résultats de cartographie de la méthode MMCMSE sont comparés à ceux obtenus par la méthode SPM, version SPM2. Il n'est pas utile dans cette partie d'utiliser deux implémentations différentes du logiciel. Nous rappelons que l'implémentation SPMA correspond à l'utilisation courante du logiciel. On utilise alors un unique régresseur d'intérêt, à savoir le modèle de réponse hémodynamique proposé par ce logiciel convolué avec le paradigme d'activation où plus précisément le signal  $x_0(\cdot)$ . Ainsi, dans le cas des séries  $\mathcal{F}_E$ , le cas SPMA correspond également au cas idéal dans la mesure où la composante d'activation injectée dans les données correspond exactement à la composante d'activation donnée par défaut par le logiciel SPM.

Les résultats obtenus sur les séquences d'événements  $\mathcal{F}_{Eb}$ ,  $\mathcal{F}_{Ec}$  et  $\mathcal{F}_{Ed}$  sont présentés sur les figures 13.5 et 13.6. Les cartes obtenues avec notre méthode (26-connexité, algorithme peu contraint  $\mathbf{A}_1$  (voir section 12.2.1.2.1, page 180) pour l'obtention des couples observation-état et un détecteur déterminant trois séquences de RDs par signal IRMf avec  $k = 0.5, 0.6, 0.7$ ) sont présentées sur la première colonne alors que celles obtenues avec SPMA sont présentées sur la seconde colonne. Les cartes de la figure 13.5 sont seuillées pour une PFA de 0.005 alors que les cartes de la figure 13.6 sont seuillées pour une p-valeur estimée de 0.005. Les couleurs ont les mêmes significations que précédemment : le blanc représente une bonne détection, le



gris clair une fausse alarme ou une non détection. Le gris foncé représente les signaux non actifs légitimement non détectés. Enfin, le noir représente soit l'extérieur de la coupe, soit des détections dans la zone qui entoure les zones actives. La troisième colonne de la figure 13.5 correspond aux courbes COR alors que les troisièmes et quatrièmes colonnes de la figure 13.6 correspondent respectivement aux courbes  $\beta$  et  $\alpha$ . On peut se référer à la section 7.1.2.1, page 103 pour un rappel des définitions des courbes  $\beta$  et  $\alpha$ .

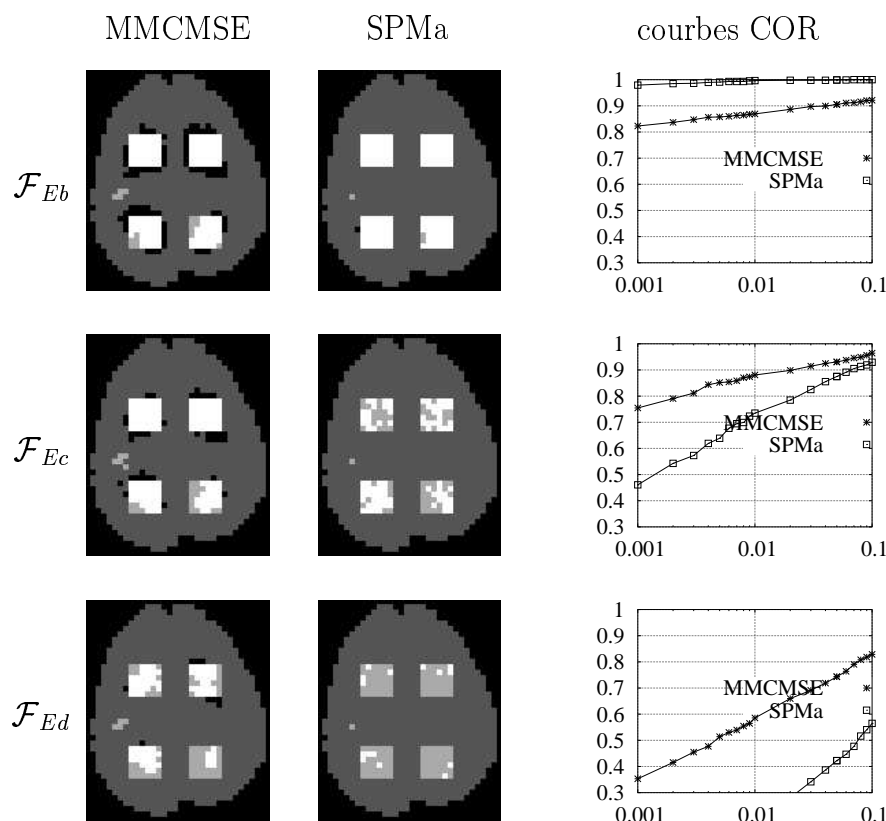


FIG. 13.5 – Résultats sur données synthétiques (voir texte)

En termes de courbes COR (figure 13.5), les résultats obtenus avec la séquence  $\mathcal{F}_{Eb}$  (première ligne de la figure 13.5) montrent des résultats satisfaisants de SPMa et de notre méthode. Les résultats obtenus avec SPMa sont clairement meilleurs. Ceci s'explique par le fait que l'étape de prétraitement donne des résultats peu satisfaisants (de nombreux événements d'intérêt ne sont pas détectés) malgré un fort rapport signal sur bruit ( $m = 1$ ).

Dès lors que l'on diminue le rapport signal sur bruit (seconde et troisième ligne de la figure 13.5, séquences  $\mathcal{F}_{Ec}$  et  $\mathcal{F}_{Ed}$ ), nos résultats se dégradent légèrement alors que les résultats obtenus avec SPM se dégradent très fortement. Notre méthode s'avère encore une fois beaucoup plus robuste au bruit que SPMa (on se place pourtant dans un cas idéal pour SPM : les composantes d'activation des signaux actifs correspondent à celles injectées dans la matrice de dessin). Les niveaux de bruit considérés restent toutefois réalistes.

Les courbes  $\alpha$  de la figure 13.6 illustrent encore une fois le biais important observé sur l'estimation des p-valeurs, aussi bien pour SPMa que pour notre méthode. Nous arrivons ainsi au même constat que pour les paradigmes en blocs : la méthode SPMa a tendance à détecter un nombre de fausses alarmes beaucoup plus important que prévu alors que c'est l'inverse pour

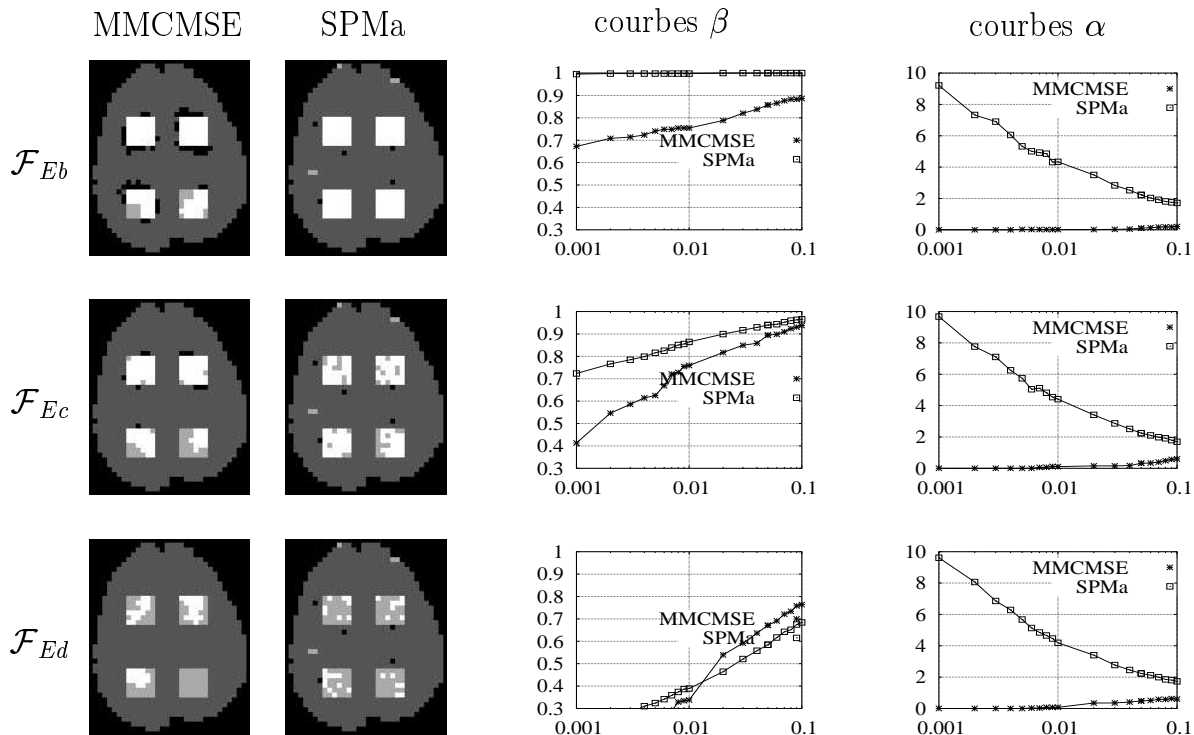


FIG. 13.6 – Résultats sur données synthétiques (voir texte)

notre méthode. Ceci est vrai pour les trois séquences (figure 13.6). Les courbes  $\beta$  obtenues avec SPM apparaissent plus satisfaisantes que les courbes COR du fait d'une estimation optimiste de la p-valeur (à une p-valeur de 0.01 est associée par exemple une PFA réelle de 0.05 environ). Au contraire, les courbes  $\beta$  obtenues avec notre méthode apparaissent moins satisfaisantes que les courbe COR, l'estimation des p-valeurs par notre méthode étant pessimiste (à une p-valeur de 0.01 est associée par exemple une PFA beaucoup plus petite). Ceci explique pourquoi les courbes  $\beta$  de notre méthode se trouvent en dessous de celles de SPMa.

### 13.2.2 Résultats sur données réelles

Les résultats obtenus par la méthode MMCMSE et par la méthode SPM ont été analysés et comparés par un neurologue. De façon générale, toutes les zones attendues actives en réponse aux images cibles et aux images nouveautés ont été détectées par la méthode MMCMSE dans les trois séquences d'images IRMf. La figure 13.7 correspond à des résultats obtenus avec le paradigme relatif aux images cibles.

De façon plus détaillée, le premier modèle d'activation construit sur le paradigme relatif aux images cibles a détecté le gyrus frontal inférieur (gauche et droit), le gyrus supramarginal postérieur droit, la région fronto-basale droite (BA 10), le gyrus cingulate antérieur, l'aire moteur correspondant à l'appui sur le bouton par l'index, le sulcus intrapariétal et central gauche, le thalamus (noyaux pulvinar et dorsomédian). De façon surprenante, notre approche a également détecté une zone de faible activité, précisément l'hippocampe antérieur gauche, laquelle a déjà été citée dans la littérature pour des paradigmes d'activation similaires. Pour comparaison, cette zone d'activité n'a pu être détectée par SPM qu'après avoir suffisamment baissé le seuil de détection, avec bien entendu pour corollaire d'augmenter la fausse alarme.

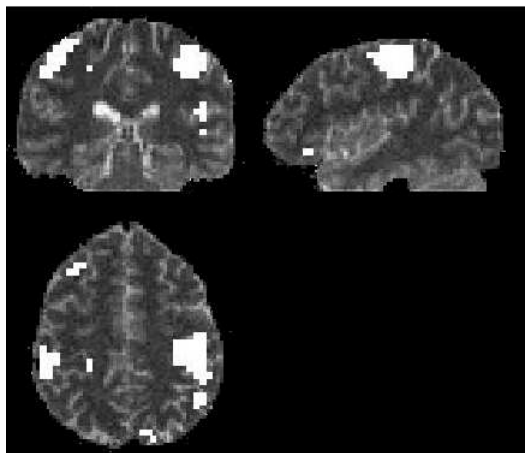


FIG. 13.7 – Cartes d’activation obtenues avec le paradigme relatif aux images cibles avec une p-valeur de 0.01.

Le second modèle d’activation construit sur le paradigme relatif aux images nouveautés a détecté le sulcus frontal inférieur gauche, les régions occipito-temporales inférieures (gauches et droites) et les aires pariétales postérieures gauches. Très peu de fausses alarmes ont été observées, chaque zone déclarée faussement active par la méthode MMCMSE n’excédant pas la taille du voxel.

Ces résultats, bien qu’obtenus sur un échantillon limité de séquences IRMf et de paradigmes événementiels, sont de nature à valider la méthode de cartographie proposée et en particulier le principe sur lequel elle se base : le couplage entre les RDs détectées dans le signal IRMf et les stimuli du paradigme.

## 13.3 Influence des paramètres

L’influence des paramètres est tout d’abord étudiée sur données synthétiques et ensuite sur données réelles.

### 13.3.1 Influence des paramètres sur données synthétiques

Nous proposons, comme précédemment, d’analyser tout d’abord l’influence du choix des détecteurs sur les performances de l’étape de prétraitement. Nous voyons ensuite l’influence du choix des paramètres sur les résultats de cartographie.

#### 13.3.1.1 Etape de détection

Nous avons utilisé 6 taux de détection différents :  $k=0.5, 0.6, 0.7, 0.8, 0.9$  et 1. Nous notons que les taux de détection sont plus petits que les taux de détection utilisés pour des paradigmes en blocs. En effet, le nombre de stimuli est très important comparativement aux nombres d’images dans le cadre des paradigmes événementiels. Des très hautes fréquences d’analyse sont par exemple nécessaires pour obtenir un taux de détection  $k=1$ . Nous sommes alors limités par le support temporel de l’ondelette analysante à échelle haute de décomposition.

Pour rappel, afin de quantifier les performances de l’étape de détection, cinq mesures ont été utilisées : le nombre moyen de RDs détectées par signal actif,  $\mu_{n_A}$ , le nombre moyen de

RDs détectées par signal inactif,  $\mu_{nI}$ , les écarts-types associés,  $\sigma_{nA}$  et  $\sigma_{nI}$ , ainsi que la mesure discriminante  $|\mu_{nA}-\mu_{nI}|/(\sigma_{nA}+\sigma_{nI})$ . Les résultats sont consignés dans le tableau 13.1 pour les six détecteurs et les séries  $\mathcal{F}_{Eb}$ ,  $\mathcal{F}_{Ec}$  et  $\mathcal{F}_{Ed}$ . Les séries diffèrent uniquement par le rapport signal sur bruit, la séquence  $\mathcal{F}_{Eb}$  étant la séquence la moins bruitée et  $\mathcal{F}_{Ed}$  la séquence la plus bruitée.

	0.5	0.6	0.7	0.8	0.9	1
$\mathcal{F}_{Eb}$	(18.85, 1.48) (17.37, 1.67) 0.46	(22.03, 1.55) (20.89, 1.82) 0.33	(24.98, 1.76) (24.41, 2) 0.15	(28.02, 1.81) (27.99, 2.19) 0	(30.86, 1.97) (31.58, 2.56) 0.15	(34.11, 2.14) (35.15, 2.98) 0.2
$\mathcal{F}_{Ec}$	(18.32, 1.5) (17.39, 1.67) 0.29	(21.76, 1.64) (20.9, 1.83) 0.24	(25.16, 1.8) (24.4, 2) 0.2	(28.19, 1.9) (27.99, 2.19) 0.04	(31.2, 2.16) (31.57, 2.56) 0.07	(34.38, 2.35) (35.14, 2.98) 0.14
$\mathcal{F}_{Ed}$	(18.12, 1.57) (17.4, 1.68) 0.22	(21.67, 1.62) (20.9, 1.83) 0.22	(25.2, 1.77) (24.4, 2) 0.21	(28.32, 1.92) (27.98, 2.19) 0.08	(31.35, 2.13) (31.56, 2.56) 0.04	(34.43, 2.4) (35.13, 2.99) 0.12

TAB. 13.1 – Nombre moyen de RDs détectées par signal IRMf actif et inactif pour les séquences synthétiques  $\mathcal{F}_{Eb}$ ,  $\mathcal{F}_{Ec}$ ,  $\mathcal{F}_{Ed}$  et six détecteurs  $k = 0.5, 0.6, 0.7, 0.8, 0.9$  et  $1$ . A chaque cellule du tableau sont associées 5 mesures, avec dans l'ordre de lecture,  $\mu_{nA}$ ,  $\sigma_{nA}$ ,  $\mu_{nI}$ ,  $\sigma_{nI}$  et la mesure discriminante.

Le pouvoir de discrimination en termes de nombre moyen de RDs détectées est moins grand que dans le cadre des paradigmes en blocs (cf. tableau 12.1 de la page 179). Si les détecteurs avec les taux de détection les plus faibles semblent être les plus discriminants, il faut noter que le pouvoir discriminant des différents détecteurs reste très faible.

Cinq autres mesures ont été utilisées pour prendre en compte l'information de localisation temporelle : le nombre moyen de bonnes détections par signal IRMf actif,  $\mu_{dA}$ , le nombre moyen de fausses alarmes (pouvant être considérées comme des bonnes détections) par signal IRMf inactif,  $\mu_{dI}$ , les écarts-types associés,  $\sigma_{dA}$  et  $\sigma_{dI}$ , enfin la mesure discriminante  $|\mu_{dA}-\mu_{dI}|/(\sigma_{dA}+\sigma_{dI})$ . On pourra se référer à la figure 8.1 de la page 118 pour un rappel des définitions. Les mesures ont été effectuées pour deux tailles de fenêtre différentes, à savoir  $d = 1$  et  $d = 2$ , et pour les cinq seuils de détection considérés plus haut. Les résultats sont regroupés dans le tableau 13.2.

	0.5	0.6	0.7	0.8	0.9	1
$\mathcal{F}_{Eb}$	(8.28, 2.21) (4.08, 1.82) 1.04 (14.05, 2.31) (8.06, 2.2) 1.32	(10.21, 2.54) (4.96, 2) 1.15 (16.32, 2.34) (9.72, 2.34) 1.41	(11.56, 2.59) (5.85, 2.11) 1.21 (18.3, 2.15) (11.49, 2.45) 1.48	(12.17, 2.65) (6.75, 2.21) 1.11 (19.09, 2.13) (13.26, 2.51) 1.25	(12.44, 2.66) (7.62, 2.35) 0.96 (20.08, 2.08) (15.06, 2.58) 1.07	(12.43, 2.68) (8.44, 2.58) 0.75 (21.18, 2.21) (16.79, 2.69) 0.89
$\mathcal{F}_{Ec}$	(6.37, 2.08) (4.08, 1.82) 0.58 (11.41, 2.22) (8.07, 2.2) 0.75	(7.75, 2.28) (4.96, 2) 0.65 (13.6, 2.39) (9.72, 2.34) 0.82	(8.96, 2.48) (5.85, 2.11) 0.67 (15.61, 2.41) (11.48, 2.45) 0.84	(9.6, 2.49) (6.75, 2.21) 0.6 (16.67, 2.34) (13.26, 2.51) 0.7	(10.09, 2.55) (7.61, 2.35) 0.5 (17.91, 2.36) (15.05, 2.58) 0.57	(10.43, 2.6) (8.44, 2.58) 0.38 (19.12, 2.37) (16.78, 2.69) 0.46
$\mathcal{F}_{Ed}$	(5.34, 1.98) (4.08, 1.82) 0.33 (9.92, 2.15) (8.07, 2.2) 0.42	(6.38, 2.12) (4.96, 2) 0.34 (11.78, 2.31) (9.72, 2.34) 0.44	(7.48, 2.21) (5.85, 2.11) 0.37 (13.77, 2.32) (11.48, 2.45) 0.48	(8.12, 2.36) (6.75, 2.21) 0.29 (15.08, 2.38) (13.26, 2.51) 0.37	(8.73, 2.35) (7.61, 2.35) 0.23 (16.42, 2.42) (15.05, 2.58) 0.27	(9.26, 2.41) (8.44, 2.58) 0.16 (17.69, 2.47) (16.77, 2.69) 0.17

TAB. 13.2 – Caractéristique de l'étape de prétraitement en termes de bonnes détections (pour les signaux actifs) et de fausses alarmes pouvant être considérées comme des bonnes détections par le modèle (pour les signaux inactifs). Les différentes mesures sont calculées pour les trois séries et les six détecteurs. A chaque cellule du tableau sont associées 10 mesures, avec dans l'ordre de lecture,  $\mu_{1I}$ ,  $\sigma_{1I}$ ,  $\mu_{1A}$ ,  $\sigma_{1A}$ , la mesure discriminante associée ( $d=1$ ) et  $\mu_{2I}$ ,  $\sigma_{2I}$ ,  $\mu_{2A}$ ,  $\sigma_{2A}$  et la mesure discriminante associée ( $d=2$ ).

Les valeurs de discrimination sont désormais du même ordre que celles obtenues avec des paradigmes en blocs (cf. tableau 12.2 de la page 179). Par conséquent, l'information de localisation temporelle dans le cadre des paradigmes événementiels est primordiale.

Nous notons également que le pouvoir de discrimination ne varie que très peu en fonction des taux de détection même si les détecteurs opérant à haute échelle ( $k = 0,9$  et  $k = 1$ ) semblent moins discriminants que les autres.

Il convient enfin de relever que, dans tous les cas, le nombre de bonnes détections est petit par rapport au nombre de stimuli, c'est-à-dire 34. Par conséquent, il peut être envisagé par la suite d'améliorer l'étape de prétraitement dans le cadre des paradigmes événementiels en proposant d'autres méthodes de détection des RDs.

### 13.3.1.2 Cartographie

**13.3.1.2.1 Les paramètres utilisés** Nous nous plaçons directement dans le cadre de la connexité  $C_{26}$ . En effet, l'étape de prétraitement étant de qualité très moyenne, il est important de prendre en compte le contexte spatial pour que les résultats soient satisfaisants. Nous étudions alors, comme dans le cas des paradigmes en blocs (section 12.2.1.2, page 180), l'influence de l'étape de prétraitement et de la stratégie utilisée pour la création des couples observation-état sur les résultats de cartographie.

On désigne par  $\mathbf{a}_i$ ,  $\mathbf{b}_i$  et  $\mathbf{c}_i$  trois détecteurs permettant de détecter respectivement une, deux ou trois séquences de RDs par signal IRMf. Chaque détecteur est donc caractérisé par la donnée d'un, de deux ou de trois taux de détection  $k$ . Les détecteurs considérés sont les suivants :

- $\mathbf{a}_1 = \{0.5\}$ ,  $\mathbf{a}_2 = \{0.6\}$ ,  $\mathbf{a}_3 = \{0.7\}$ ,  $\mathbf{a}_4 = \{0.8\}$ ,  $\mathbf{a}_5 = \{0.9\}$  et  $\mathbf{a}_6 = \{1\}$ .
- $\mathbf{b}_1 = \{0.5, 0.6\}$ ,  $\mathbf{b}_2 = \{0.6, 0.7\}$ ,  $\mathbf{b}_3 = \{0.7, 0.8\}$ ,  $\mathbf{b}_4 = \{0.8, 0.9\}$  et  $\mathbf{b}_5 = \{0.9, 1\}$
- $\mathbf{c}_1 = \{0.5, 0.6, 0.7\}$ ,  $\mathbf{c}_2 = \{0.6, 0.7, 0.8\}$ ,  $\mathbf{c}_3 = \{0.7, 0.8, 0.9\}$  et  $\mathbf{c}_4 = \{0.8, 0.9, 1\}$ .

Le dernier groupe de paramètres dont l'influence est étudiée concerne la stratégie utilisée pour former l'ensemble des couples observation-état. Il s'agit des paramètres *extension* (contrainte de forme), *tailleMin* (contrainte de nombre) et *ecm* (règle pour associer les états cachés) (voir section 11.4.1.3, page 166). Nous avons utilisé les mêmes jeux de paramètres que dans le cadre des paradigmes en blocs. Nous les rappelons ci-dessous du moins contraint au plus contraint :

- $\mathbf{A}_1 = \{ \textit{extension} = 5, \textit{ecm} = 5, \textit{tailleMin} = 0.3 \}$ ,
- $\mathbf{A}_2 = \{ \textit{extension} = 4, \textit{ecm} = 4, \textit{tailleMin} = 0.4 \}$ ,
- $\mathbf{A}_3 = \{ \textit{extension} = 3, \textit{ecm} = 3, \textit{tailleMin} = 0.5 \}$ .

L'étape de préparation des données fournissant des résultats de qualités très moyennes, il est important de ne pas utiliser des contraintes trop fortes pour la création des couples observation-état. Les résultats obtenus avec le jeu de paramètres le plus contraint  $\mathbf{A}_3$  ne sont en effet pas satisfaisants. Par la suite, seuls les résultats obtenus avec les jeux de paramètres  $\mathbf{A}_1$  et  $\mathbf{A}_2$  sont présentés.

**13.3.1.2.2 Résultats** Les résultats sont consignés dans le tableau 13.3. Nous commençons par analyser les résultats en termes de courbes COR. Contrairement au cas des paradigmes en blocs, il est important d'avoir un nombre très important de canaux pour obtenir des résultats satisfaisants (26-connexité avec 2 ou 3 séquences de RDs détectées par signal IRMf). Par exemple, les résultats obtenus avec les détecteurs  $\mathbf{a}_i$  sont moins bons que ceux obtenus avec les détecteurs  $\mathbf{b}_i$  et  $\mathbf{c}_i$ .

Conformément aux résultats du tableau 13.2 où l'on avait vu que le pouvoir discriminant des différents détecteurs était relativement constant, on note également que le choix des taux de détection n'influence que très peu les résultats de cartographie. Les résultats sont donc cohérents.

Nous notons enfin que les résultats sont améliorés avec l'algorithme peu contraint  $\mathbf{A}_1$ , ce qui est tout à fait compréhensible eu égard à l'étape de préparation des données. On voit en effet d'après le tableau 13.2 que le pouvoir de discrimination augmente légèrement quand la largeur de la fenêtre  $d$  augmente. De plus, avec l'algorithme  $\mathbf{A}_2$ , il se pose un problème important d'observabilité des événements d'intérêt à cause de la grande valeur utilisée pour le paramètre *tailleMin*. Par conséquent, beaucoup de signaux ont une vraisemblance finale nulle!

Nous poursuivons notre analyse en regardant les courbes  $\alpha$ . Comme on l'a déjà mentionné plusieurs fois, le calcul de la p-valeur est relativement instable et peu de crédit peut être accordé à cette estimation. Cependant, on peut noter que l'utilisation de l'algorithme peu contraint  $\mathbf{A}_1$  conduit à une estimation pessimiste des p-valeurs (à une p-valeur  $p$  est associée une PFA beaucoup plus petite que  $p$ ). Le comportement de la courbe  $\alpha$  avec l'algorithme plus contraint  $\mathbf{A}_2$  est difficile à prévoir.

Il est enfin à noter que dans notre étude, nous nous sommes limités à l'utilisation de détecteurs déterminant trois séquences de RDs par signal IRMf. Pour justifier ce choix, nous avons utilisé un détecteur déterminant cinq séquences de RDs par signal IRMf ( $k = 0.5, 0.6$  et  $0.7, 0.8$  et  $0.9$ ). Nous obtenons alors des résultats similaires (très légèrement moins bons) que ceux obtenus avec les 3 détecteurs ( $k = 0.5, 0.6$  et  $0.7$ ). Cela montre que la nouvelle information introduite par les 2 nouveaux détecteurs ( $k=0.8$  et  $0.9$ ) est redondante de celle apportée par les 3 autres détecteurs. Néanmoins, même si seulement 3 détecteurs sont utiles, l'utilisation de 5 détecteurs reste plus pratique pour l'utilisateur car son choix est plus simple à effectuer. Il suffit donc de considérer un nombre important de détecteurs. Certains seront utiles, d'autres non. Le choix particulier des détecteurs n'est cependant plus à la charge de l'utilisateur.

### 13.3.2 Influence des paramètres sur données réelles

Nous évaluons ici l'influence, de l'étape de prétraitement et de la stratégie utilisée pour former les couples observation-état, sur les résultats de cartographie. Nous nous plaçons dans le cadre de la 26-connexité. De plus, avec les mêmes notations que précédemment (nous rajoutons tout de même les détecteurs  $\mathbf{c}_5 = \{0.9, 1, 1.1\}$  et  $\mathbf{b}_6 = \{1, 1.1\}$ ), nous avons formé 22 couples détecteur-algorithme pour lesquelles le traitement a été lancé. Nous les donnons ci-après. A chaque couple est associé un entier permettant de numéroter les couples :

- $\{1, \mathbf{b}_1, \mathbf{A}_1\}$ ,
- $\{2, \mathbf{b}_1, \mathbf{A}_2\}$ ,
- $\{3, \mathbf{b}_2, \mathbf{A}_1\}$ ,
- $\{4, \mathbf{b}_2, \mathbf{A}_2\}$ ,
- $\{5, \mathbf{b}_3, \mathbf{A}_1\}$ ,
- $\{6, \mathbf{b}_3, \mathbf{A}_2\}$ ,
- $\{7, \mathbf{b}_4, \mathbf{A}_1\}$ ,
- $\{8, \mathbf{b}_4, \mathbf{A}_2\}$ ,
- $\{9, \mathbf{b}_5, \mathbf{A}_1\}$ ,
- $\{10, \mathbf{b}_5, \mathbf{A}_2\}$ ,
- $\{11, \mathbf{b}_6, \mathbf{A}_1\}$ ,
- $\{12, \mathbf{b}_6, \mathbf{A}_2\}$ ,

	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
$A_1, C_{26}, \mathcal{F}_{Ec}$	65, 0.03, 31	78, 0.03, 58	89, 0, 60	79, 0, 26	76, 0, 14	71, 0, 6
	58, 0.04, 25	75, 0.01, 51	87, 0, 50	73, 0, 14	71, 0, 11	64, 0, 4
	49, 0.13, 18	63, 0, 41	80, 0, 27	61, 0, 3	64, 0, 3	54, 0, 1
$A_2, C_{26}, \mathcal{F}_{Ec}$	58, 0.2, 36	81, 0.33, 64	83, 0.26, 71	74, 0.99, 73	66, 0.97, 66	58, 0.17, 48
	47, 0.22, 30	74, 0.27, 50	77, 0.19, 63	67, 1.01, 67	58, 1.21, 60	54, 0.04, 36
	28, 0.67, 19	37, 0.87, 27	63, 0.4, 47	48, 0.74, 43	44, 0.47, 36	45, 0, 22
$A_1, C_{26}, \mathcal{F}_{Ed}$	20, 0.02, 1	37, 0.05, 13	44, 0, 8	38, 0, 0	40, 0, 0	22, 0, 0
	14, 0.04, 0	32, 0.03, 10	37, 0, 3	30, 0, 0	34, 0, 0	16, 0, 0
	6, 0.07, 0	17, 0, 2	26, 0, 1	11, 0, 0	22, 0, 0	8, 0, 0
$A_2, C_{26}, \mathcal{F}_{Ed}$	2, 0.1, 0	19, 0.46, 9	24, 0.44, 16	16, 1.56, 20	19, 0.6, 14	14, 0.06, 4
	1, 0.15, 0	10, 0.43, 3	17, 0.27, 8	11, 2.23, 17	11, 0.85, 10	10, 0.01, 1
	0, 0.47, 0	0, 0.94, 0	7, 1.08, 7	3, 1.88, 5	4, 1.08, 4	6, 0, 1

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$a_1$	$a_2$	$a_3$	$a_4$
$A_1, C_{26}, \mathcal{F}_{Ec}$	82, 0.37, 77	88, 0.39, 82	85, 0.05, 70	82, 0.01, 51	80, 0, 44	88, 0.26, 80	85, 0.11, 76	85, 0.05, 68	84, 0, 47
	78, 0.22, 72	85, 0.26, 76	82, 0, 57	77, 0, 45	74, 0, 27	85, 0.04, 71	81, 0.03, 64	81, 0, 59	80, 0, 34
	72, 0.07, 61	74, 0.13, 65	75, 0, 47	68, 0, 24	67, 0, 6	75, 0, 61	75, 0, 50	74, 0, 32	69, 0, 18
$A_2, C_{26}, \mathcal{F}_{Ec}$	78, 0.65, 71	86, 0.92, 85	84, 2.53, 92	77, 3.34, 87	75, 1.63, 79	84, 0.83, 84	87, 2.2, 91	84, 4.94, 94	81, 3.5, 89
	69, 0.85, 68	82, 1.44, 84	80, 2.58, 87	71, 3.52, 81	69, 0.28, 59	80, 0.87, 78	82, 2.45, 87	78, 4.45, 90	73, 2.47, 82
	46, 1.34, 53	72, 2.35, 77	66, 1.55, 66	67, 9.82, 77	58, 0.13, 47	54, 2.35, 70	82, 6.66, 83	72, 9.48, 84	66, 2.49, 69
$A_1, C_{26}, \mathcal{F}_{Ed}$	51, 0.73, 49	55, 0.18, 35	51, 0.05, 23	47, 0.03, 7	39, 0, 2	59, 0.16, 41	49, 0.24, 33	49, 0.01, 11	46, 0.01, 8
	46, 0.85, 44	45, 0.09, 23	44, 0.03, 13	37, 0.01, 3	32, 0, 1	51, 0.08, 30	41, 0.04, 18	42, 0, 3	40, 0, 3
	31, 0, 0	29, 0, 8	30, 0, 1	19, 0, 0	19, 0, 0	35, 0, 16	27, 0, 4	28, 0, 1	20, 0, 0
$A_2, C_{26}, \mathcal{F}_{Ed}$	11, 0.63, 8	33, 0.7, 29	28, 3.76, 55	23, 4.19, 47	27, 1.94, 37	22, 0.85, 21	32, 2.8, 50	32, 4.14, 52	30, 2.66, 45
	7, 0.47, 4	26, 0.67, 21	20, 4.8, 47	16, 6.45, 42	19, 2.41, 30	16, 0.48, 7	22, 3.09, 39	32, 5.62, 45	21, 1.92, 29
	0, 1.28, 0	6, 1.95, 15	18, 8.34, 26	5, 6.99, 18	10, 2.76, 14	3, 1.55, 3	9, 3.63, 18	32, 21.31, 41	10, 3.5, 19

TAB. 13.3 – Influence des paramètres sur les résultats de cartographie. A une cellule du tableau, est associée le détecteur ( $a_i, b_i, c_i$ ), l’algorithme ( $A_1$  ou  $A_2$ ), la connexité ( $C_{26}$ ) et la séquence traitée ( $\mathcal{F}_{Ec}$  ou  $\mathcal{F}_{Ed}$ ). Pour chaque cellule, il est possible de déterminer une carte de p-valeur, cette dernière est résumée ensuite par 9 mesures consignées dans la dite-cellule. Les trois chiffres situés sur la même ligne sont respectivement de gauche à droite la valeur de la courbe ROC, de la courbe  $\alpha$  et de la courbe  $\beta$  pour une abscisse de 0.01 (en haut), de 0.005 (au milieu) et de 0.001 (en bas). A noter que les résultats pour les courbes  $\beta$  et les courbes COR ont été multipliés par 100 de manière à obtenir une mesure entre 0 et 100. On rappelle que la courbe  $\alpha$  correspond au nombre de fausses alarmes observées sur le nombre de fausses alarmes attendues en fonction de la p-valeur estimée et que la courbe  $\beta$  est le pourcentage de bonnes détections en fonction de la p-valeur estimée.

- {13,  $\mathbf{c}_1$ ,  $\mathbf{A}_1$  },
- {14,  $\mathbf{c}_1$ ,  $\mathbf{A}_2$  },
- {15,  $\mathbf{c}_2$ ,  $\mathbf{A}_1$  },
- {16,  $\mathbf{c}_2$ ,  $\mathbf{A}_2$  },
- {17,  $\mathbf{c}_3$ ,  $\mathbf{A}_1$  },
- {18,  $\mathbf{c}_3$ ,  $\mathbf{A}_2$  },
- {19,  $\mathbf{c}_4$ ,  $\mathbf{A}_1$  },
- {20,  $\mathbf{c}_4$ ,  $\mathbf{A}_2$  },
- {21,  $\mathbf{c}_5$ ,  $\mathbf{A}_1$  },
- {22,  $\mathbf{c}_5$ ,  $\mathbf{A}_2$  }.

La vérité terrain n'étant pas connue, on utilise comme précédemment la mesure de similarité entre deux cartes d'activation pour évaluer l'influence des paramètres (voir section 8.2, page 121 pour la définition). Nous rappelons que la mesure de similarité varie entre 0 (les deux cartes d'activation n'ont aucun voxel actif en commun) à 1 (les deux cartes d'activation sont identiques).

Pour chaque couple, il est possible d'obtenir une carte d'activation  $\mathcal{A}_i$  à partir d'une séquence IRMf ( $i$  correspond au numéro du couple détecteur-algorithme). En pratique, nous avons lancé les traitements sur trois séquences d'image différentes. Les mesures de similarité sont ensuite déterminées entre les différentes cartes  $\mathcal{A}_i$  et  $\mathcal{A}_j$ ,  $i \neq j$  pour les 3 séquences. Pour chaque indice  $i$ , on détermine la moyenne  $\mu_i$  de la similarité obtenue avec les autres jeux de paramètres. On note enfin  $\sigma_i$  l'écart-type correspondant. Les résultats sont consignés sur la figure 13.8 pour 3 p-valeurs différentes, la p-valeur étant utilisée pour le seuillage de la carte.

Il apparaît tout d'abord que les valeurs moyennes de similarité sont relativement faibles montrant la grande influence des paramètres sur les résultats de cartographie. Contrairement à ce que l'on avait pu noter dans le cas synthétique, les résultats obtenus avec l'algorithme  $\mathbf{A}_2$  fournissent de bons résultats sur données réelles. De plus, il semble que les résultats se dégradent dès lors que le taux de détection est faible. Ceci ne contredit pas les résultats précédents. Selon la nature de la série, les détecteurs peuvent en effet avoir des pouvoirs discriminants différents.

En considérant uniquement les jeux de paramètres 9, 10, 11, 12, 19, 20, 21 et 22 correspondant à de hauts taux de détection, les valeurs de similarité augmentent et sont de l'ordre de 0.5 et les écarts-type diminuent (seconde ligne de la figure 13.8). En pratique, la valeur de 0.5 est relativement élevée car elle est obtenue après seuillage de la carte de p-valeur, la p-valeur pouvant être mal estimée. Le tableau 13.3 nous laisse par exemple à penser que la courbe  $\alpha$  est toujours au dessous de 1 en utilisant l'algorithme  $\mathbf{A}_1$  alors qu'elle se trouve souvent au dessus de 1 avec l'algorithme  $\mathbf{A}_2$ .



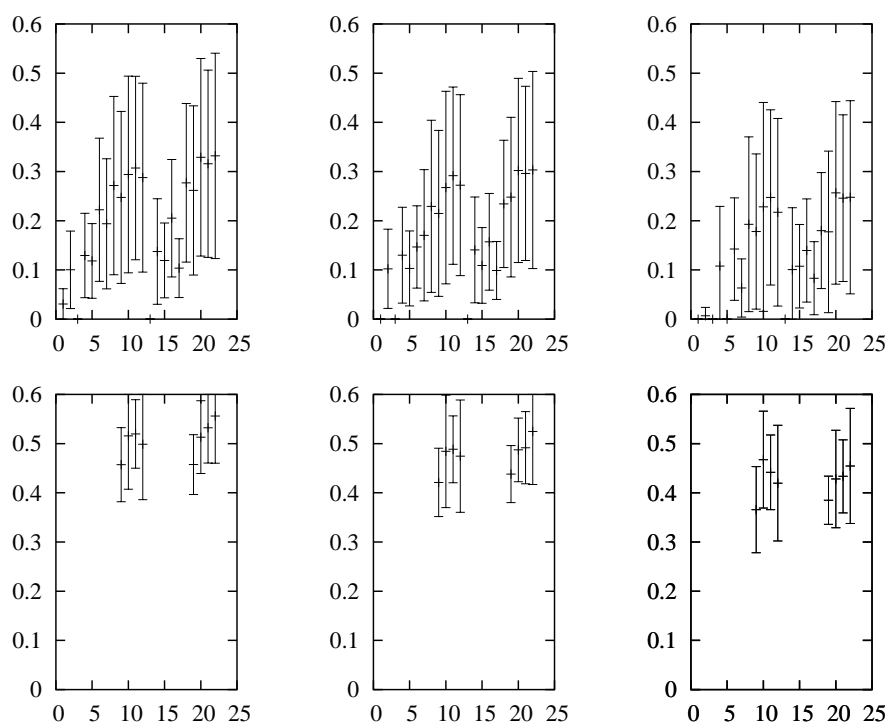


FIG. 13.8 – Influence des paramètres sur données réelles. Le numéro du couple est reporté en abscisse alors que la similarité moyenne est reportée sur l’axe des ordonnées. La longueur des traits correspond à 2 écarts-type. Un seuil de 0.01 est utilisé pour la colonne gauche, de 0.005 pour la colonne du milieu et de 0.001 pour la colonne de droite. La ligne du dessus correspond au cas où l’on considère tous les jeux de paramètres. La ligne du dessous correspond au cas où l’on s’intéresse uniquement aux jeux de paramètres  $i = 9, 10, 11, 12, 19, 20, 21$  et 22.



# Conclusion partielle

La méthode de cartographie proposée dans cette partie repose sur le même principe énoncé en partie II, à savoir qu'en présence d'activité locale au voxel  $v$ , la séquence de RDs détectées dans le signal IRMf  $f^v(\cdot)$  correspondant doit s'aligner sur les transitions « off-on » du paradigme (paradigme en blocs) ou sur les stimuli (paradigme événementiel). Cependant, afin d'accroître les performances de détection d'activation, le principe d'alignement temporel est étendu aux signaux IRMf voisins du voxel  $v$  de manière à prendre en compte l'information spatiale de voisinage. D'autre part, toujours pour gagner en robustesse, l'étape de détection des RDs pour un signal IRMf s'appuie dorénavant sur  $K$  détecteurs opérant en parallèle.

L'évaluation sur paradigmes en blocs et paradigmes événementiels a montré que la méthode donne de très bons résultats comparativement à ceux de SPM et à ceux présentés dans la partie précédente. La formulation de l'activité en termes de couplage entre les RDs et le « paradigme » nous permet de ne pas faire de fortes hypothèses sur la forme de la réponse à un bloc de stimulation et sur la chronologie d'activation. Nous avons montré que la méthode, contrairement à SPM, était très robuste au non-maintien de la réponse hémodynamique dans un bloc de stimulation, à la variabilité de la réponse hémodynamique, au retard d'activation et à divers phénomènes d'habituation et d'apprentissage. Enfin, de par la prise en compte de l'information spatiale et de par le cadre multi-détecteur utilisé, la méthode est beaucoup plus robuste au bruit que les autres méthodes considérées.

De manière générale, les résultats valident largement l'approche de cartographie proposée en termes à la fois de principe de modélisation (traiter le signal IRMf uniquement à travers ses non-stationnarités), de choix de modélisation (utilisation du MMCMSE) et d'algorithme de cartographie.

Il convient cependant de s'interroger sur les principales limitations de la méthode proposée. Le choix des paramètres de l'algorithme ne semble pas être une véritable limitation même s'ils ont une influence sur les résultats. L'utilisateur peut adopter un détecteur déterminant un nombre important de RDs (5 par exemple) pour limiter l'influence de son choix. Nous avons aussi montré que l'algorithme de création des couples observation-état doit être choisi peu contraint pour obtenir des résultats satisfaisants. La principale limitation de l'approche provient essentiellement de la méthode utilisée pour estimer la p-valeur. Cette étape d'estimation peut cependant se faire à partir des méthodes de permutation présentées dans la littérature, méthodes que nous n'avons pas eu le temps d'implanter.

La poursuite logique des travaux consiste tout d'abord à doter l'approche d'un module fiable d'estimation de la p-valeur. Ensuite, différents aspects de l'algorithme peuvent être étudiés plus en détail. Il est par exemple possible de perfectionner l'étape de prétraitement en ajoutant d'autres types de détecteurs des événements d'intérêt. Il est envisageable également d'étudier des approches plus complexes pour estimer le poids des canaux. Certaines pourraient tirer profit d'informations anatomiques, aucune activité n'étant par exemple attendue dans la matière blanche.



# Conclusion générale et perspectives

## Conclusion générale

Nous avons présenté dans cette thèse deux approches de modélisation et d'analyse statistique de séquences d'événements par modèles de Markov cachés. Ces deux approches sont particulièrement adaptées à la modélisation et à l'analyse de processus aléatoires de nature événementielle.

Elles procèdent en deux étapes. La première étape, générique, détecte et caractérise les événements d'intérêt composant le processus événementiel à analyser. La seconde étape analyse, sur la base d'un modèle markovien caché adapté, la distribution temporelle des événements détectés.

Les deux approches de modélisation se distinguent par le nombre de séquences d'événements, ou canaux d'observation, composant le processus observable du modèle. La première approche, par MSMCSE, considère une seule et unique séquence d'événements. La seconde approche, par MMCMSE, est capable de gérer de multiples canaux d'observation dans un cadre formel riche de fusion–association d'événements asynchrones entre canaux. Elle se distingue des approches par fusion markovienne existantes par les points suivants :

- La stratégie de fusion–association d'événements utilisée permet l'analyse simultanée d'un nombre important de canaux d'observation. Nous avons utilisé jusqu'à 81 canaux d'observation, les approches similaires n'en considérant généralement que 2 ou 3.
- L'approche proposée autorise, sous certaines contraintes de causalité, des asynchronismes importants entre événements associés.
- La partie cachée de notre modèle reste simple. Contrairement aux approches par fusion markovienne existantes, basées sur plusieurs chaînes de Markov (une chaîne par canal d'observation) aux liens de dépendance plus ou moins complexes, notre approche repose sur une unique chaîne de Markov, auto-régressive et représentative de l'ensemble des événements observés sur les différents canaux<sup>1</sup>.
- Sous certaines hypothèses d'indépendance conditionnelle, l'expression de la loi d'observation attachée à un état caché reste simple et générale à la fois, offrant ainsi un large éventail d'applications possibles en fusion d'observations et de séquences.

En relation avec ces approches de modélisation, nous avons développé deux méthodes originales et non supervisées de cartographie des zones d'activation en IRM fonctionnelle cérébrale. Les deux méthodes sont fondées sur le même principe d'alignement temporel entre séquences d'événements. En présence d'activation neuronale, les ruptures de dynamiques d'intérêt induites dans le signal IRMf s'alignent, au retard d'activation près, sur les transitions « off-on » du paradigme d'activation, pour les paradigmes en blocs, et sur les stimuli du paradigme, pour

---

<sup>1</sup>D'un point de vue conceptuel, il apparaît que le MMCMSE se rapproche des modèles markoviens segmentaux. Dans les deux cas, la partie cachée est simplifiée de manière importante alors que la loi d'observation est largement complexifiée.

les paradigmes événementiels. Ce principe, novateur en cartographie IRMf cérébrale, a pour triple avantage :

- de limiter sensiblement le nombre d’hypothèses a priori quant à la forme et à la chronologie du signal IRMf actif, par rapport aux méthodes fondées sur le modèle linéaire général,
- d’élargir, dans un contexte de forte variabilité du signal IRMf actif, l’espace des signaux d’activation à détecter,
- de pouvoir introduire, dans un contexte de faible rapport signal à bruit, des outils puissants de détection–caractérisation de transitoires que sont les ruptures de dynamique du signal IRMf actif.

En prenant en compte l’information spatiale de voisinage de chaque voxel, dans un cadre multi-détecteurs, la méthode de cartographie par MMCMSE développée est de loin plus riche que celle à base de MSMCSE. Les études que nous avons menées sur séquences IRMf synthétiques et réelles, pour des paradigmes en blocs et pour des paradigmes événementiels, ont clairement démontré la grande robustesse de cette méthode face au bruit d’une part, et en présence de retards d’activation, de variations de forme de la réponse hémodynamique, de non maintien de cette dernière en réponse à un bloc de stimulation, ou de phénomènes d’habituation et d’apprentissage d’autre part. Dans tous les cas d’étude considérés, la méthode de cartographie par MMCMSE surpasse une implémentation standard de SPM, méthode de cartographie fonctionnelle cérébrale faisant référence dans le domaine. Dans la plupart de ces cas, les performances de détection d’activation par MMCMSE égalent voire dépassent celles d’une implémentation idéale de SPM. Ces résultats, obtenus en contexte de cartographie non supervisée, valident le principe général consistant à analyser une séquence IRMf à partir du seul champ spatio-temporel de ruptures de dynamique des signaux IRMf. Ils valident également le principe novateur d’alignement temporel entre séquences d’événements utilisé dans cette thèse pour mesurer le taux d’activité cérébrale locale. Ces résultats valident enfin le principe et la stratégie de fusion markovienne de séquences d’événements IRMf empruntés pour cartographier l’activité fonctionnelle cérébrale.

## Perspectives en IRMf

Parce que plus riche et plus général que le MSMCSE, nous ne discutons dans la suite que des perspectives du MMCMSE. Pour débiter, notons qu’un certain nombre de points de la méthode de cartographie par MMCMSE sont perfectibles :

- En ce qui concerne l’étape de prétraitement. Nous avons vu que la méthode de détection utilisée donnait des résultats peu satisfaisants dans le cadre des paradigmes événementiels. D’autres méthodes de détection pourraient être envisagés. Il serait peut-être intéressant d’étudier l’apport de la déconvolution piquée [158] dans le cadre de l’événementiel.
- Différents choix, bien que justifiés, ont été effectués. Il serait bon d’étudier plus en détail leur impact sur les résultats. Nous pensons plus particulièrement au critère utilisé pour estimer les paramètres, à la stratégie d’estimation des poids des canaux et aux différentes règles permettant de calculer la carte d’activité. Par exemple, les informations portées par l’IRM anatomique pourrait être utilisées pour estimer le poids des canaux.
- Pour être encore moins sensible au bruit, il peut être envisagé d’effectuer différents traitements à partir des résultats obtenus. Nous rappelons que nous avons défini un modèle d’activité par voxel. A partir de l’ensemble des modèles d’activité déterminés, des méthodes de clustering, basées sur le calcul de distance entre modèles markoviens

[186], pourraient être utilisées pour regrouper les voxels en régions actives, un modèle d'activité étant associé par région. La prise en compte de l'information spatiale pourrait être alors d'envergure, augmentant ainsi la robustesse au bruit. Par exemple, si la zone active est constituée de 200 voxels et que l'on détecte 3 séquences de RDs par signal IRMf, on aurait ainsi une observation composée de 600 séquences d'événements!

Nous avons validé dans cette thèse le principe suivant lequel l'analyse du signal IRMf pouvait s'effectuer à travers ses non stationnarités. A plus long terme, l'analyse des séquences d'événements vise à inférer sur la connectivité fonctionnelle du cerveau. La répartition des événements sur l'axe des temps n'est pas désordonnée. Elle obéit à des modes propres à chaque séquence et/ou communs à plusieurs séquences. Les modes propres peuvent se justifier à la lumière de connaissances physiologiques acquises sur les processus à analyser. Les modes communs suggèrent quant à eux l'existence d'interrelations entre observations et/ou entre séquences d'observations. D'un point de vue strictement médical, ces interrelations sont connues pour certaines, mal connues pour d'autres ou bien encore à découvrir. D'un point de vue méthodologique, leur exploitation algorithmique reste délicate compte tenu de la multiplicité des séquences et des observations à traiter, de leurs caractéristiques (les événements peuvent être informatifs ou pas sur les processus analysés, redondants et/ou complémentaires, incertains et incomplets). L'analyse des interrelations n'est explorée que depuis peu par la communauté du génie biomédical et relève de la fusion de données. Pour mettre à jour de telles relations, le MMCMSE peut être un outil puissant. Par exemple, pour étudier si deux régions sont reliées fonctionnellement, les RDs de la première région pourraient être comparées à celles de la seconde région au travers d'un MMCMSE. Le processus profond correspondrait au processus d'activation neuronale relatif aux deux régions alors que la signature, c'est-à-dire la partie observable d'un événement profond, serait en tous points identique à la notion définie dans cette thèse. Toujours dans la perspective d'étudier de nouvelles relations entre processus, il serait intéressant d'utiliser une démarche similaire pour mettre en évidence des relations entre le signal EEG (électroencéphalographie) et le signal IRMf, les différentes relations reliant ces signaux étant encore très largement méconnues.

## Perspectives en modélisation

Nous présentons dans un premier temps de nouvelles applications possibles du MMCMSE dans le domaine biomédical. Nous proposons ensuite deux évolutions théoriques possibles du modèle.

## Perspectives en génie biomédical

Nous avons vu dans cette thèse que l'analyse de l'activité cérébrale pouvait, grâce au MMCMSE développé, se ramener à celle du champ de ruptures de dynamique des signaux composant la séquence d'images IRMf d'entrée à analyser. Cette approche n'est pas spécifique de l'IRM fonctionnelle du cerveau. L'analyse de nombre de processus physiologiques rencontrés en génie biomédical peut être rapportée à celle d'un champ discret d'observations. Deux exemples :

- En neurologie, la recherche et la caractérisation de circuits neuronaux impliqués dans la genèse et la propagation des crises épileptiques peuvent être envisagées au travers de l'analyse conjointe du champ de ruptures de dynamique IRMf précité et du champ de grapho-éléments (pointes) portés par les signaux électroencéphalographiques (EEG)

enregistrés simultanément.

- En cardiologie, le dépistage précoce des troubles du rythme cardiaque repose sur l'observation et l'analyse conjointe de séquences de grapho-éléments (ondes P, QRS, T, ESV, ESSV) électrocardiographiques (ECG) et/ou hémodynamiques (signaux de pression sanguine internes/externes) et/ou phono-mécanographiques (bruits B1, B2).

Ces deux exemples soulignent l'existence de nouveaux champs d'application possibles du MMCMSE en génie biomédical.

## MMCMSE et segmentation d'images

Nous rappelons que le MMCMSE permet de déterminer selon différents critères une séquence d'événements profonds qui aurait pu donner naissance aux différentes séquences observées. En remplaçant la chaîne de Markov par un champ de Markov, il peut être envisagé de déterminer une carte de segmentation profonde qui aurait pu donner naissance aux différentes cartes de segmentation observées. En pratique, plusieurs méthodes de segmentation pourraient être utilisées sur une même image ou sur des images différentes représentatives de la même réalité (images multispectrales en imagerie astronomique, images multimodales en imagerie médicale), le modèle statistique analyserait ensuite les différents résultats de manière à obtenir une nouvelle carte de segmentation optimale selon différents critères. Une signature dans cette application correspondrait à un regroupement de régions (ou de point de contours) au travers des différentes cartes. La signature serait également la partie observable d'une région associée au processus profond. La partie cachée modéliserait les dépendances entre les différentes régions du processus profond.

Cette piste semble très intéressante dans la mesure où le nombre de méthodes de segmentations d'images est très important. De plus, des connaissances *a priori* sur certaines méthodes de détection de contours pourraient être incorporées dans le modèle statistique ou être apprises (précision du positionnement des contours de la région).

## MMCMSE et espace multi-échelles

D'un point de vue conceptuel, un lien étroit peut être effectué entre le MMCMSE et l'espace multi-échelles linéaire, ou *scale-space* linéaire<sup>2</sup>. Dans ce type d'approche, un ensemble de signaux (ou d'images) est créé par filtrage gaussien d'un signal ou d'une image originale. L'ensemble ainsi formé est paramétré par une valeur d'échelle associée à la fréquence du filtrage gaussien. Cet ensemble vérifie un ensemble de propriétés [128]. Par exemple, un extremum d'un signal monodimensionnel créé est obligatoirement associé à un extremum du signal original. Des théories permettent également de prévoir le comportement de certains points particuliers tels que le passage par zéro des dérivées, les extrema, au travers des échelles. Les structures d'intérêt dans chacun des signaux sont recherchées indépendamment et des liens sont ensuite établis entre les échelles. Cette étape est facilitée grâce aux résultats théoriques des espaces multi-échelles linéaire [142]. Dans le but de ne pas filtrer les zones d'intérêt, le *scale space* non linéaire a été introduit.

Notre modèle peut être une alternative intéressante pour l'analyse des différentes structures. Une signature correspondrait à un ensemble de structures d'intérêt observé à différentes échelles. La signature, comme précédemment, serait la partie observable du processus profond.

<sup>2</sup>Il est à noter que l'espace multi-échelle linéaire a été utilisé dans le cadre de l'analyse d'un ensemble de cartes d'activation IRMf (étude de groupe) [43].



L'ensemble des signatures envisageables pourrait être analysé par le modèle statistique dans le but par exemple de remonter à l'information cachée (cartes de structures d'intérêt). Ainsi, contrairement aux méthodes classiques, le modèle statistique pourrait se baser sur l'ensemble des informations au travers des échelles (observations en surface) pour obtenir une information sur le processus profond alors que les méthodes courantes utilisent des heuristiques pour obtenir une valeur particulière de la fréquence [143]. La résolution des problèmes de l'apprentissage et du décodage pourrait enfin être d'un grand intérêt. Il est à noter cependant que des travaux importants sont tout de même à mener pour aboutir à un tel modèle. La notion de signature doit tout d'abord être généralisée de manière à prendre en compte qu'à travers les échelles, une structure d'intérêt peut apparaître, peut disparaître, se diviser en deux nouvelles structures et fusionner avec une autre structure. Enfin, les liens de dépendance entre les différentes structures associées au processus profond ne peuvent pas toujours (dans le cas des images notamment) être une simple chaîne de Markov.



# Annexe A

## Liste des publications

### Revue internationale

S. Faisan, L. Thoraval, J.-P. Armspach, M.-N. Metz-Lutz et F. Heitz. Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models. *IEEE Transactions on Medical Imaging* (accepté).

S. Faisan, L. Thoraval, J.-P. Armspach, J. Foucher, M.-N. Metz-Lutz et F. Heitz. Hidden Markov event sequence models : towards unsupervised functional MRI brain mapping. *Academic Radiology*. Invited paper (accepté).

### Conférences internationales avec actes et comité de lecture

S. Faisan, L. Thoraval, J.-P. Armspach et F. Heitz. Hidden semi-Markov event sequence models : Application to brain functional MRI sequence analysis. *IEEE Int. Conf. Image Processing, ICIP 2002*, 880–883, Rochester, USA, Sep. 2002.

S. Faisan, L. Thoraval, J.-P. Armspach et F. Heitz. Unsupervised learning and mapping of brain fMRI signals based on hidden semi-Markov event sequence models. *Medical Image Computing and Computer Assisted Intervention, MICCAI 2003*, long paper, 75–82, Montreal, Canada, Nov. 2003.

S. Faisan, L. Thoraval, J.-P. Armspach, F. Heitz et J. Foucher. Unsupervised spatio-temporal detection of brain functional activation based on hidden Markov multiple event sequence models. *Medical Imaging : Image Processing, SPIE 2005* (accepté).

### Conférences nationales avec actes et comité de lecture

S. Faisan, L. Thoraval et J.-P. Armspach. Modèles semi-markoviens cachés de séquences d'événements : Application en cartographie IRM fonctionnelle cérébrale. *19ème Colloque GRETSI*, Paris, France, Sep. 2003.



# Bibliographie

- [1] G.K. Aguirre, E. Zarahn et M. D'Esposito. Empirical analyses of BOLD fMRI statistics ii. Spatially smoothed data collected under null-hypothesis and experimental conditions. *NeuroImage*, 5 :199–212, 1997.
- [2] G.K. Aguirre, E. Zarahn et M. D'Esposito. A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. *Magnetic Resonance Medicine*, 39 :500–505, 1998.
- [3] S.M. Aji et R.J. McEliece. The generalized distributive law. *IEEE Trans. on Information Theory*, 46(2) :325–343, 2000.
- [4] A. Aldroubi et M. Unser. *Wavelets in Medicine and Biology*. CRC press, 1996.
- [5] R. Almeida et A. Ledberg. Exact multivariate tests for brain imaging data. *Human Brain Mapping*, 16 :24–35, 2002.
- [6] A.H. Andersen, D.M. Gash et M.J. Avison. Principal component analysis of the dynamic response measured by fMRI : a generalized linear systems framework. *Magnetic Resonance Imaging*, 17(6) :795–815, 1999.
- [7] R. André-Obrecht, B. Jacob et N. Parlangeau. Audio-visual speech recognition and segmental master-slave HMM. *AVSP'97*, 1997.
- [8] J. Ashburner et K. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4) :254–266, 1999.
- [9] M. Askar et H. Derin. A recursive algorithm for the Bayes solution of the smoothing problem. *IEEE Transactions on Automatic Control*, AC-26 :558–561, 1981.
- [10] W.F. Auffermann, S.C. Ngan et X. Hu. Cluster significance testing using the bootstrap. *NeuroImage*, 17 :583–591, 2002.
- [11] L.R. Bahl, P.F. Brown, P.V. de Souza et R.L. Mercer. A new algorithm for the estimation of hidden Markov model parameters. *In Proc. ICASSP*, 493–496, 1988.
- [12] L.R. Bahl, P.F. Brown, P.V. de Souza et R.L. Mercer. Maximum mutual information estimation of hidden Markov model parameters for speech recognition. *In Proc. ICASSP*, 49–52, 1986.
- [13] J.K. Baker. Stochastic modeling for automatic speech understanding. *Speech Recognition*, 521–542, 1975.
- [14] D. Balslev, F.A. Nielsen et al. Cluster analysis of activity-time series in motor learning. *Human Brain Mapping*, 15 :135–145, 2002.
- [15] L.E. Baum. An inequality and associated maximisation technique in statistical estimation for probabilistic functions. *Inequalities*, 3 :1–8, 1972.
- [16] L.E. Baum et J.A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc.*, 73 :360–363, 1967.

- [17] L.E. Baum et T. Petrie. Statistical inference for probabilistic functions of Markov chains. *Annals Math. Stat.*, 37 :1554–1563, 1966.
- [18] L.E. Baum, T. Petrie, G. Soules et N. Weiss. A maximisation technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals Math. Stat.*, 41(1) :164–171, 1970.
- [19] L.E. Baum et G.R. Sell. Growth transformations for functions on manifolds. *Pac. J. Math.*, 27 :211–227, 1968.
- [20] R. Baumgartner, C. Windischberger et E. Moser. Quantification in functional magnetic resonance imaging : Fuzzy clustering vs correlation analysis correlation analysis. *Magnetic Resonance Imaging*, 16(2) :115–125, 1998.
- [21] Y. Benjamini et Y. Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *J.R. Stat. Soc.*, 289–300, 1995.
- [22] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc.*, 48(3) :259–302, 1986.
- [23] J. Bilmes. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, 1997.
- [24] J. Bilmes. Buried Markov models : A graphical-modeling approach to automatic speech recognition. *Computer, Speech and Language*, 17(2-3) :213–231, 2003.
- [25] R.M. Birn, R.W. Cox et P.A. Bandettini. Detection versus estimation in event-related fMRI : Choosing the optimal stimulus timing. *NeuroImage*, 15 :252–264, 2001.
- [26] R.M. Birn, Z.S. Saad et P.A. Bandettini. Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. *Human Brain Mapping*, 14 :817–826, 2001.
- [27] M. Borga et H. Knutsson. A canonical correlation approach to blind source separation. Technical Report LiU-IMT-EX-0062, Department of Biomedical Engineering, Linköping University, 2001.
- [28] M. Brand, N. Olivier et A. Pentland. Coupled hidden Markov models for complex action recognition. *CVPR*, 994, 1997.
- [29] M. Breakspear, M.J. Brammer, E.T. Bullmore, P. Das et L.M. Williams. Spatiotemporal wavelet resampling for functional neuroimaging data. *Human Brain Mapping*, 23 :1–25, 2004.
- [30] B.T. Logan et P.J. Moreno. Factorial HMMs for acoustic modeling. *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 813–816, 1998.
- [31] E.T. Bullmore, C. Long, J. Suckling, M.J. Fadili, G. Calvert, F. Zelaya, T.A. Carpenter et M. Brammer. Colored noise and computational inference in neurophysiological fMRI time series analysis : Resampling methods in time and wavelet domains. *Human Brain Mapping*, 12(2) :61–78, 2001.
- [32] M.A. Burock et A.M. Dale. Estimation and detection of event-related fMRI signals with temporally correlated noise : A statistical efficient and unbiased approach. *Human Brain Mapping*, 11 :249–260, 2001.
- [33] R.B. Buxton, E.C. Wong et L.R. Frank. Dynamics of blood flow and oxygen metabolism during brain activation : The balloon model. *Magnetic Resonance Medicine*, 39 :855–864, 1998.
- [34] D. Caplan et L. Moo. Cognitive conjunction and cognitive functions. *Neuroimage*, 21 :751–756, 2004.

- [35] J.D. Carew, G. Wahba, X. Xie, E.V. Nordheim et M.E. Meyerand. Optimal spline smoothing of fMRI time series by generalized cross-validation. *Neuroimage*, 18 :950–961, 2003.
- [36] G. Celeux et J. Diebolt. L’algorithme SEM : un algorithme d’apprentissage probabiliste pour la reconnaissance de mélanges de densités. *Revue de Statistiques Appliquées*, 34(2) :35–51, 1986.
- [37] G. Celeux et J. Diebolt. A stochastic approximation type EM algorithm for the mixture problem. *Stochastics and Stochastics Reports*, 41 :119–134, 1992.
- [38] Z. Chair et P. Parshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerosp. Electron. Syst.*, AES-22 :98–101, 1986.
- [39] K.H. Chuang, M.J. Chiu, C.C. Lin, K.M. Huang, P.J. Chiang et J.H. Chen. Model free functional MRI analysis using Kohonen clustering neural network. *IEEE Trans. Medical Imaging*, 18 :1117–1128, 1999.
- [40] P. Ciuciu, J.B. Poline, G. Marrelec, J. Idier, C. Pallier et H. Benali. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Trans. Medical Imaging*, 22 :1235–1251, 2003.
- [41] A.E. Cook et M.J. Russel. Improved duration modelling techniques in hidden Markov models using series-parallel configuration of states. *Proceedings of the Institute of Acoustics Autumn Conference on Speech and Hearing*, 299–306, 1986.
- [42] J. Coull, C. Frith, R. Dolan et P. Grasby. The neural correlates of the noradrenergic modulation of human attention, arousal and learning. *Eur. Journal of Neurosciences*, 9 :589–598, 1997.
- [43] O. Coulon, J.-F. Mangin, J.-B. Poline, V. Frouin et I. Bloch. Group analysis of individual activation maps using 3D scale-space primal sketches and a Markovian random field. *Spatial Statistics : Methodological Aspects and Applications*, M. Moore (Eds.), *Lecture Notes in Statistics*, (Springer-Verlag), 159 :201–211, 2001.
- [44] F. Crivello, T. Schormann, N. Tzourio-Mazoyer, P.E. Roland, K. Zilles et B.M. Mazoyer. Comparison of spatial normalization procedures and their impact on functional maps. *Human Brain Mapping*, 16(4) :228–250, 2002.
- [45] A.M. Dale. Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8 :109–114, 1999.
- [46] A.P. Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, 2 :25–36, 1992.
- [47] D.B. de Araujo, W. Tedeschi, A.C. Santos, J. Elias, U.P.C Neves et O. Baffa. Shannon entropy applied to the analysis of event-related fMRI time series. *NeuroImage*, 20 :311–317, 2003.
- [48] J.P. Delmas. Relations entre les algorithmes d’estimation itératives EM et ICE avec exemples d’application. *Quinzième colloque GRETSI*, 185–188, 1995.
- [49] J.P. Delmas. An equivalence of the EM and ICE algorithm for exponential family. *IEEE Trans. Signal Processing*, 45(10) :2613–2615, 1997.
- [50] B. Delyon, M. Lavielle et E. Moulines. - Stochastic computation - Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 27(1) :94–128, 1999.

- [51] A.P. Dempster, N.M. Laird et D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39(1) :1–38, 1977.
- [52] H. Derin et P.A. Kelly. Discrete-index Markov-type random processes. *Proceedings of the IEEE*, 77(10) :268–278, 1989.
- [53] X. Descombes, F. Kruggel et D.Y. von Cramon. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Trans. Medical Imaging*, 17 :1028–1039, 1998.
- [54] M. D’esposito, E. Zarahn et G.K. Aguirre. Event-related functional MRI : implications for cognitive psychology. *Psychol. Bull.*, 125(1) :155–164, 1999.
- [55] P.A. Devijver. Baum’s forward backward algorithm revisited. *Pattern Recognition Letter*, 39 :369–373, April 1985.
- [56] J.R. Duann, T.P. Jung, W.J. Kuo, T.C. Yeh, S. Makeig, H. Jen-Chuen et T.J. Sejnowski. Single trial variability in event-related BOLD signals. *NeuroImage*, 15 :823–835, 2002.
- [57] R.O. Duda, P.E. Hart et D.G. Stork. *Pattern Classification*. John Wiley and Sons, 2000.
- [58] Y. Ephraim, A. Dembo et L.R. Rabiner. A minimum discrimination approach for hidden Markov modeling. *In Proc. ICASSP*, 25–28, 1987.
- [59] Y. Ephraim, A. Dembo et L.R. Rabiner. A minimum discrimination approach for hidden Markov modeling. *IEEE Trans. on Information Theory*, 35(5) :1001–1013, 1989.
- [60] Y. Ephraim, A. Dembo et L.R. Rabiner. Review of several stochastic speech unit models. *Computer Speech and Language*, 6 :377–402, 1992.
- [61] F. Esposito, E. Formisano, E. Seifritz, R. Goebel, R. Morrone, G. Tedeschi et F. Di Salle. Spatial independent component analysis of functional MRI time-series : To what extent do results depend on the algorithm user? *Human Brain Mapping*, 16 :146–157, 2002.
- [62] B.S. Everitt et E.T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 7 :1–14, 1999.
- [63] M.J. Fadili et E.T. Bullmore. Wavelet-generalized least squares : a new BLU estimator of linear regression models with  $1/f$  errors. *NeuroImage*, 15 :217–232, 2001.
- [64] M.J. Fadili, S. Ruan, D. Bloyet et B. Mazoyer. A multistep unsupervised fuzzy clustering analysis of fMRI time series. *Human Brain Mapping*, 10 :160–178, 2000.
- [65] M.J. Fadili, S. Ruan, D. Bloyet et B. Mazoyer. On the number of clusters and the fuzziness index for unsupervised FCA applications to BOLD fMRI time series. *Medical Image Analysis*, 41 :124–131, 2001.
- [66] S. Faisan, L. Thoraval et J.-P. Armpach. Modèles semi-markoviens cachés de séquences d’événements : Application en cartographie IRM fonctionnelle cérébrale. *19ème Colloque GRETSI*, 2003.
- [67] S. Faisan, L. Thoraval, J.-P. Armpach et F. Heitz. Hidden semi-Markov event sequence models : Application to brain functional MRI sequence analysis. *IEEE Int. Conf. Image Processing ICIP*, 880–883, 2002.
- [68] S. Faisan, L. Thoraval, J.-P. Armpach et F. Heitz. Unsupervised learning and mapping of brain fMRI signals based on hidden semi-Markov event sequence models. *In Proc. MICCAI Conf.*, 75–82, 2003.
- [69] J.D. Ferguson. Recognition of isolated digits using hidden Markov models with continuous mixture densities. *In Proc. Symp. on the Application of Hidden Markov Models to Text and Speech*, 79–143, April 1980.



- [70] R. Fernandez et R.W. Picard. Modeling driver's speech under stress. *ISCA Workshop on Speech and Emotion*, 219–224, 2000.
- [71] S. Fine, Y. Singer et N. Tishby. The hierarchical hidden Markov model : analysis and applications. *Machine Learning*, 1998.
- [72] J.W. Fisher, E.R. Cosman, C. Wible et W.M. Wells. Adaptive entropy rates for fMRI time-series analysis. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 905–912, september 2001.
- [73] E. Formisano, F. Esposito, N. Kriegeskorte, G. Tedeschi, F. Di Salle et R. Goebel. Spatial independent component analysis of functional magnetic resonance imaging time-series : characterization of the cortical components. *Neurocomputing*, 49 :241–254, 2002.
- [74] G.D. Forney. The Viterbi algorithm. *Proc. IEEE*, 268–278, 1973.
- [75] J.R. Foucher, H. Otzenberger et D. Gounot. Where arousal meets attention : a simultaneous fMRI and EEG recording study. *NeuroImage*, 22 :688–697, 2004.
- [76] L. Freire, A. Roche et J.F. Mangin. What is the best similarity measure for motion correction in fMRI time series ? *IEEE Trans. Medical Imaging*, 21 :470–484, 2002.
- [77] O. Friman, M. Borga, P. Lundberg et H. Knutsson. Detection of neural activity in fMRI using maximum correlation modeling. *NeuroImage*, 15 :386–395, 2002.
- [78] O. Friman, M. Borga, P. Lundberg et H. Knutsson. Exploratory fMRI analysis by autocorrelation maximization. *NeuroImage*, 16 :454–464, 2002.
- [79] O. Friman, M. Borga, P. Lundberg et H. Knutsson. Adaptative analysis of fMRI data. *NeuroImage*, 19 :837–845, 2003.
- [80] K. Friston, C. Frith, P. Liddle, R. Dolan, A. Lammertsma et R. Frackowiak. The relationship between global and local changes in PET scans. *J. of Cerebral Blood Flow and Metabolism*, 10 :458–466, 1990.
- [81] K.J. Friston. Bayesian estimation of dynamical systems : an application to fMRI. *NeuroImage*, 16 :513–530, 2002.
- [82] K.J. Friston, J. Ashburner et al. SPM 97 course notes. *Wellcome Department of Cognitive Neurology, University College London*, 1997.
- [83] K.J. Friston, P. Flechter, O. Josephs, A. Holmes, M.D. Rugg et R. Turner. Event-related fMRI : Characterizing differential responses. *NeuroImage*, 7(1) :30–40, 1998.
- [84] K.J. Friston, C.D. Frith, R.S.J. Frackowiak et R. Turner. Characterizing dynamic brain responses with fMRI : a multivariate approach. *NeuroImage*, 2 :166–172, 1995.
- [85] K.J. Friston, P. Jezzard et R. Turner. The analysis of functional MRI time-series. *Human Brain Mapping*, 1 :153–171, 1994.
- [86] K.J. Friston, O. Josephs, G. Rees et R. Turner. Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine*, 39 :41–52, 1998.
- [87] K.J. Friston, O. Josephs, E. Zarahn, A. Holmes, S. Rouquette et J.-P. Poline. To smooth or not to smooth ? Bias and efficiency in fMRI time-series analysis. *NeuroImage*, 12 :196–208, 2000.
- [88] K.J. Friston, A. Mechelli, R. Turner et C.-J. Price. Nonlinear responses in fMRI : The ballon model, Volterra kernels and other hemodynamics. *NeuroImage*, 12 :466–477, 2000.
- [89] K.J. Friston et W. Penny. Posterior probability maps and SPMs. *NeuroImage*, 19 :1240–1249, 2003.

- [90] K.J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton et J. Ashburner. Classical and Bayesian inference in neuroimaging : theory. *NeuroImage*, 16 :465–483, 2002.
- [91] K.J. Friston, J.P. Poline, S. Strother, A.P. Holmes, C.D. Frith et R.S.J. Frackowiak. A multivariate analysis of PET activation studies. *Human Brain Mapping*, 4 :140–151, 1996.
- [92] K.J. Friston, E. Zarahn, O. Josephs, R.N.A. Henson et A.M. Dale. Stochastic designs in event-related fMRI. *NeuroImage*, 10 :607–619, 1999.
- [93] M. Gales et S. Young. Segmental HMMs for speech recognition. *Proc. European Conf. on Speech Commun. and Technology*, ASSP-33 :1579–1582, 1993.
- [94] S. Geman et D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.
- [95] C.R. Genovese. A Bayesian time-course model for functional magnetic resonance imaging data. *J. Am. Stat. Assoc.*, 95 :691–719, 2000.
- [96] G. Gerig, G. Szekely, G. Israel et M. Berger. Detection and characterization of unsharp. *Information Processing in Medical Imaging 95*, 165–176, 1995.
- [97] Z. Ghahramani. Keyword spotting in poorly printed documents using pseudo 2-d hidden Markov models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16 :842–848, 1994.
- [98] O. Ghitza et M.M. Sondhi. Hidden Markov models with templates as non-stationary states : an application to speech recognition. *Computer Speech and Language*, 2 :101–119, 1993.
- [99] G.H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9 :416–429, 1999.
- [100] P.S. Gopalakrishnan, D. Kanevsky, A. Nadas et D. Nahamoo. An inequality for rational functions with applications to some statistical estimation problems. *IEEE Trans. on Information Theory*, 37(1) :107–113, 1991.
- [101] C. Goutte, P. Toft, E. Rostrup, F.A. Nielsen et L.K. Hansen. On clustering fMRI time series. *NeuroImage*, 9 :298–310, 1998.
- [102] Y. Guédon et C. Coccozza-Thivent. Explicit state occupancy modelling by hidden semi-Markov models : application of Derin’s scheme. *Computer Speech and Language*, 77(2) :167–192, February 1990.
- [103] H. Guo et W. Hsu. A survey of algorithms for real-time Bayesian network inference. *AAAI/KDD/UAI-2002 Joint Workshop on Real-Time Decision Support and Diagnosis Systems*. Edmonton, July 2002.
- [104] E.M. Haacke, W. Lin, X. Hu et K. Thulborn. A current perspective of the status of understanding BOLD imaging and its use in studying brain function : a summary of the workshop at the University of North Carolina in Chapel Hill, 26-28 October. *NMR Biomed*, 14 :384–388, 2001.
- [105] G.E. Hagberg, G. Zito, F. Patria et J.N. Sanes. Improved detection of event-related functional MRI signals using probability functions. *NeuroImage*, 14 :293–309, 2001.
- [106] B. Hajek. Cooling schedules for optimal annealing. *Math. Oper. Res.*, 13 :311–329, 1988.
- [107] J. Hamilton. Analysis of time series subject to changes in regime. *J. Econometrics*, 45 :39–70, 1990.

- [108] D.A. Handwerker, J.M. Ollinger et M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21 :1639–1651, 2004.
- [109] N.V. Hartvig et J.L. Jensen. Spatial mixture modeling of fMRI data. *Human Brain Mapping*, 11 :233–248, 2000.
- [110] R.N.A. Henson, C. Buechel, O. Josephs et K. Friston. The slice-timing problem in event-related fMRI. *Neuroimage*, 99 :S125, 1999.
- [111] P. Højen-Sørensen, C.E. Rasmussen et L.K. Hansen. Bayesian modelling of fMRI time series. (Eds.) Sara A. Solla, Todd K. Leen and Klaus-Robert Müller, MIT Press, 754–760, 2000.
- [112] J. Hoey. Hierarchical unsupervised learning of facial expression categories. *ICCV Workshop on Detection and Recognition of Events in Video*, 2001.
- [113] J.B. Hopfinger, C. Buchel, A.P. Holmes et K.J. Friston. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. *NeuroImage*, 1111 :326–333, 2000.
- [114] X.D. Huang et M.A. Jack. Semi-continuous hidden Markov models for speech signals. *Computer Speech and Language*, 3 :239–251, 1989.
- [115] C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner et R. Turner. Image distortion correction in fMRI : a quantitative evaluation. *NeuroImage*, 16 :217–240, 2002.
- [116] Y.A. Ivanov et A.F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8) :852–872, 2000.
- [117] F. Jelinek. Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 64 :532–556, 1976.
- [118] M.I. Jordan. Learning in graphical models. *Kluwer Academic Publishers*, 113, 1998.
- [119] O. Josephs et N.A. Henson. Event-related functional magnetic resonance imaging : modelling, inference and optimization. *Phil. Trans. R. Soc. Lond. B*, 354 :1215–1228, 1999.
- [120] B.H. Juang, W. Chou et C.H. Lee. Minimum classification error rate methods for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 5 :266–277, 1997.
- [121] B.H. Juang et L.R. Rabiner. Mixture autoregressive hidden Markov models for speech signals. *IEEE Trans. on Acoust., Speech, and Signal Processing*, ASSP-33(6) :1404–1413, 1985.
- [122] B.H. Juang et L.R. Rabiner. A probabilistic distance measure for hidden Markov models. *AT&T Tech. J.*, 64(2) :391–408, February 1985.
- [123] B.H. Juang et S. Katagiri. Discriminative learning for minimum error classification. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 40 :3043–3054, 1992.
- [124] D. Jurafsky et J. H. Martin. Speech and language processing : An introduction to natural language processing. *Computational Linguistics, and Speech Recognition*. Prentice-Hall, 2000.
- [125] K. Katanoda, Y. Matsuda et M. Sugishita. A spatio-temporal regression model for the analysis of functional MRI data. *NeuroImage*, 17 :1415–1428, 2002.

- [126] P. Kenny, M. Lennig et P. Mermelstein. A linear predictive HMM for vector-valued observations with applications to speech recognition. *IEEE Trans. on Acoust., Speech, and Signal Processing*, ASSP-38(2) :220–225, 1990.
- [127] F. Kherif, J.P. Poline, G. Flandin, H. Benali, O. Simon, S. Dehaene et K.J. Worsley. Multivariate model specification for fMRI data. *NeuroImage*, 16 :1068–1083, 2002.
- [128] J. Koenderink. The structure of images. *Biological Cybernetics*, 50 :363–370, 1984.
- [129] F. Kruggel et D.Y. von Cramon. Temporal properties of the hemodynamic response in functional MRI. *Human Brain Mapping*, 8(4) :259–271, 1999.
- [130] F. Kruggel, D.Y. von Cramon et X. Descombes. Comparison of filtering methods for fMRI datasets. *NeuroImage*, 10 :530–543, 1999.
- [131] S. Kuo et O. Agazzi. Keyword spotting in poorly printed documents using pseudo 2-d hidden Markov models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16 :842–848, 1994.
- [132] J. Kybic, P. Thevanaz, A. Nirkko et M. Unser. Unwarping of unidirectionally distorted EPI images. *IEEE Trans. Medical Imaging*, 19(2) :80–93, 2000.
- [133] N. Lange et S.L. Zeger. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Appl. Stat., J. Roy. Stat. Soc., Series C*, 46(1) :1–29, 1997.
- [134] S. L. Lauritzen et D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems. *Proceedings of the Royal Statistical Society*, 50(B) :154–227, 1988.
- [135] S.L. Lauritzen. Graphical models. *Oxford U.K. : Clarendon*, 1996.
- [136] C.L. Lawson et R.J. Hanson. Solving least squares problems. *Prentice Hall, New York*, 1974.
- [137] H.G. Leimer. Optimal decomposition by clique separators. *Discrete Math.*, 113 :99–123, 1993.
- [138] S.E. Levinson. Continuously Variable Duration Hidden Markov Models for Speech Analysis. *Computer Speech and Language*, 1(1) :29–45, Mars 1986.
- [139] S.E. Levinson, L.R. Rabiner et M.M. Sondhi. An introduction to the application of the theory of probabilistic functions of a Markov process in automatic speech recognition. *Bell Society Technical Journal*, 62 :1035–1074, April 1983.
- [140] X. Li, M. Parizeau et R. Plamondon. Training hidden Markov models with multiple observations – a combinatory method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4) :371–377, 2000.
- [141] C.H. Liao, K.J. Worsley, J.P. Poline, J.A.D. Aston, G.H. Duncan et A.C. Evans. Estimating the delay of the fMRI response. *NeuroImage*, 16 :593–606, 2002.
- [142] T. Lindeberg. Scale-space theory : A basic tool for analysing structures at different scales. *J. of Applied Statistics (Supplement on Advances in Applied Statistics : Statistics and Images : 2)*, 21(2) :224–270, 1994.
- [143] T. Lindeberg et L. Bretzner. Real-time scale selection in hybrid multi-scale representations. *Proc. Scale-Space'03 (Springer Lecture Notes in Computer Science)*, 2695 :148–163, 2003.
- [144] T.-T. Liu et L.-R. Frank. Efficiency, power, and entropy in event-related fMRI with multiple trial types part I : theory. *NeuroImage*, 21 :387–400, 2004.

- [145] T.-T. Liu et L.-R. Frank. Efficiency, power, and entropy in event-related fMRI with multiple trial types part II : design of experiments. *NeuroImage*, 21 :401–413, 2004.
- [146] T.-T. Liu, L.-R. Frank, E.-C. Wong et R.-B. Buxton. Detection power, estimation efficiency and predictability in event-related fMRI. *NeuroImage*, 13 :759–773, 2001.
- [147] A. Ljolje, Y. Ephraim et L. Rabiner. Estimation of hidden Markov model parameters by minimizing empirical error rate. *In Proc. ICASSP*, 709–712, 1990.
- [148] B.R. Logan et D.B. Lowe. An evaluation of thresholding techniques in fMRI analysis. *NeuroImage*, 22 :95–108, 2004.
- [149] P.M. Macey, K.E. Macey, R. Kumar et R.M. Harper. A method for removal of global effects from fMRI time series. *Neuroimage*, 22 :360–366, 2004.
- [150] D.J.C. MacKay. Information theory, inference, and learning algorithms. *Cambridge University Press*, April 2003.
- [151] J. Marchini et A. Presanis. Comparing methods of analyzing fMRI statistical parametric maps. *NeuroImage*, 22 :1203–1213, 2004.
- [152] J.L. Marchini et B.D. Ripley. A new statistical approach to detecting significant activation in functional MRI. *Neuroimage*, 12 :366–380, 2000.
- [153] E. N. Marieb. *Anatomie et Physiologie humaines*. De Boeck Université, 1993.
- [154] G. Marrelec, H. Benali, P. Ciuciu, M. Péligrini-Issac et J.-P. Poline. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Human Brain Mapping*, 19 :1–17, 2003.
- [155] P. Masson et W. Pieczynski. SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE Trans. on Geoscience and Remote Sensing*, 31(3) :618–633, 1993.
- [156] A.R. McIntosh, F.L. Bookstein, J.V. Haxby et C.L. Grady. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, 3 :143–157, 1996.
- [157] M.J. McKeown, S. Makeig, G.G. Brown, T.P. Jung, S.S. Kindermann, A.J. Bell et T.J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6 :168–188, 1998.
- [158] J.M. Mendel. Optimal seismic deconvolution. *Academic Press, New York, NY, USA*, 1983.
- [159] N. Metropolis, A.M. Rosenbluth et A.E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087–1092, 1953.
- [160] M.N. Metz-lutz, I.J. Namer, D. Gounot, C. Kleitz, J.P. Armspach et P. Kehrli. Language functional neuro-imaging changes following focal left thalamic infarction. *NeuroReport*, 11(13) :2907–2912, 2000.
- [161] F.G. Meyer. Wavelet based estimation of a semi parametric generalized linear model of fMRI time-series. *IEEE Trans. Medical Imaging*, 22(3) :315–322, 2003.
- [162] F.G. Meyer et J. Chinrungrueng. Analysis of event-related fMRI data using local clustering bases. *IEEE Trans. Medical Imaging*, 22(8) :933–939, 2003.
- [163] A. Meziane, B. Jacob et R. André-Obrecht. Modélisation de la durée des sons dans un système de reconnaissance automatique de la parole. *SFA '97*, 1997.
- [164] K.L. Miller, W.M. Luh, T.T. Liu et al. Nonlinear temporal dynamics of the cerebral blood flow response. *Human Brain Mapping*, 13 :1–12, 2001.

- [165] U. Möller, M. Ligges, P. Gerorgiewa, C. Grünling, W.A. Kaiser, H. Witte et B. Blanz. How to avoid spurious cluster validation? A methodological investigation on simulated and fMRI data. *NeuroImage*, 17 :431–446, 2002.
- [166] A. Nefian et M. Hayes. Maximum likelihood training of the embedded HMM for face detection and recognition. *IEEE Intl. Conf. on Image Processing*, 2000.
- [167] F.H. Netter. *The Ciba collection of medical illustrations, nervous system (Vol. 1) part I and II. Summiit*. NJ : Hoechstetter Printing, 1995.
- [168] T.E. Nichols et A.P. Holmes. Nonparametric permutation tests for functional neuroimaging : a primer with examples. *Human Brain Mapping*, 15 :1–25, 2001.
- [169] C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer et D. Grucker. Registration of MR/MR and MR/SPECT brain images by fast stochastic optimization of robust voxel similarity measures. *NeuroImage*, 8 :30–43, 1998.
- [170] N. Lange, S.C. Strother, J.R. Anderson, F.A. Nielsen, A.P. Holmes, T. Kolenda, R. Savoy et L.K. Hansen. Plurality and resemblance in fMRI data analysis. *NeuroImage*, 10 :282–303, 1999.
- [171] Y. Normandin. An improved MMIE training algorithm for speaker independent, small vocabulary, continuous speech recognition. *In Proc. ICASSP*, 537–540, 1991.
- [172] J. Orchard, C. Greif, G.H. Golub, B. Bjornson et M.S. Atkins. Simultaneous registration and activation detection for fMRI. *IEEE Trans. Medical Imaging*, 22(11) :458–466, 2003.
- [173] M. Ostendorf, V. Digalakis et O. Kimball. From HMMs to segment models : a unified view of stochastic modeling for speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 4 :360–378, 1996.
- [174] E. Périlleux, B. Anselme et D. Richard. *Biologie Humaine. Anatomie, physiologie, santé*. Nathan, 1995.
- [175] J. Pfeuffer, J.C. McCullough, P.F. Van De Moortele, K. Ugurbil et X. Hu. Spatial dependence of the nonlinear BOLD response at short stimulus duration. *NeuroImage*, 18 :990–1000, 2003.
- [176] W. Pieczynski. Statistical image segmentation. *Machine Graphics and Vision*, 1(1/2) :261–268, 1992.
- [177] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Traitement du Signal*, 11(2) :141–153, 1994.
- [178] J.B. Poline et B. Mazoyer. Analysis of individual PET activation maps by detecting of high signal-to-noise ration pixel clusters. *J. of Cerebral Blood Flow and Metabolism*, 13 :425–437, 1993.
- [179] J.B. Poline, K.J. Worsley, A.C. Evans et K.J. Friston. Combining spatial extent and peak intensity to test for activations infunctional imaging. *NeuroImage*, 5 :83–96, 1997.
- [180] A.B. Poritz. Linear predictive hidden Markov models and the speech signal. *Proc. ICASSP, ASSP-33* :1291–1294, 1982.
- [181] P. Pérez et F. Heitz. Restriction of a Markov random field on a graph and multiresolution statistical image modeling. *IEEE Trans. on Information Theory*, 42(1) :180–190, 1996.
- [182] C. Price et K. Friston. Cognitive conjunction : a new approach to brain activation experiments. *NeuroImage*, 5 :261–270, 1997.
- [183] C. Price, R. Wise et R. Frackowiak. Demonstration of the implicit processing of visually presented words and pseudo-words. *Cer. Cortex*, 6 :62–70, 1996.

- [184] C. Price, R. Wise, S. Ramsay, K. Friston, D. Howard, K. Patterson et R. Frackowiak. Regional response differences within the human auditory cortex when listening to words. *Neurosciences Letters*, 146 :179–182, 1992.
- [185] P.-L. Purdon, V. Solo, R.-M. Weisskoff et E.-N. Brown. Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage*, 14 :912–923, 2001.
- [186] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2) :257–286, February 1989.
- [187] L.R. Rabiner, B.H. Juang, S.E. Levinson et M.M. Sondhi. Recognition of isolated digits using hidden Markov models with continuous mixture densities. *AT&T Technical Journal*, 64 :1211–1222, 1986.
- [188] N. Ramsey, V. Matthey, P.V. Gelderen, C. Moonen et D. Weinberger. Test-retest reliability of 3D functional MRI with PRESTO. *Human Brain Mapping [Suppl]*, 1 :316, 1995.
- [189] S. Rauch, P. Whalen et al. Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping*, 5 :124–132, 1997.
- [190] J.R. Raz, H. Zheng, H. Ombao et B. Turetsky. Statistical tests for fMRI based on experimental randomization. *Neuroimage*, 19 :226–232, 2003.
- [191] I. Rezek et S.J. Roberts. Estimation of coupled hidden Markov models with application to biosignal interaction modelling. *IEEE International Workshop on Neural Networks for Signal Processing*, page 994, 2000.
- [192] W. Richter et M. Richter. The shape of the fMRI BOLD response in children and adults changes systematically with age. *NeuroImage*, 20 :1122–1131, 2003.
- [193] B.R. Rosen, R.L. Buckner et A.M. Dale. Event-related functional MRI : past, present, future. *Proc. Natl. Acad. Sci. USA*, 95(3) :773–780, 1998.
- [194] M. Russel. A segmental HMM for speech pattern matching. *Proc. Int'l. Conf. on Acoust., Speech and Signal Processing*, ASSP-33 :499–502, 1993.
- [195] M. Russell et R. Moore. Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition. *In ICASSP*, 5–8, 1985.
- [196] Z.S. Saad, E.A. De Yoe et K.M. Ropella. Estimation of fMRI response delays. *NeuroImage*, 18 :494–504, 2003.
- [197] L. Saul et M. Rahim. Maximum likelihood and minimum classification error factor analysis for automatic speech recognition. *IEEE Trans. on Speech and Audio Processing*, 8(2) :115–125, 2000.
- [198] M. Smith, B. Pütz, D. Auer et L. Fahrmeir. Assessing brain activity through spatial Bayesian variable selection. *NeuroImage*, 20 :802–815, 2003.
- [199] P. Smyth. Clustering sequences with hidden Markov models. *NIPS*, 1996.
- [200] L. Stanberry, R. Nandy et D. Cordes. Cluster analysis of fMRI data using dendrogram sharpening. *Human Brain Mapping*, 20 :201–219, 2003.
- [201] M. Svensén, F. Kruggel et D.Y. von Cramon. Probabilistic modeling of single-trial fMRI data. *IEEE Trans. Medical Imaging*, 19(1) :25–35, 2000.
- [202] T.M. Talavage et W.B. Edmister. Nonlinearity of fMRI responses in human auditory cortex. *Human Brain Mapping*, 22 :216–228, 2004.

- [203] J. Tanabe, D. Miller, J. Tregellas, R. Freedman et F.G. Meyer. Comparison of detrending methods for optimal fMRI preprocessing. *NeuroImage*, 15 :902–907, 2002.
- [204] R.E. Tarjan. Decomposition by clique separators. *Discret Math*, 55 :221–232, 1985.
- [205] B. Thirion et O. Faugeras. Revisiting non-parametric activation detection on fMRI time series. In *MMBIA*, December 2001.
- [206] C.G. Thomas, R.A. Harshman et R.S. Menon. Noise reduction in BOLD-based fMRI using component analysis. *NeuroImage*, 17 :1521–1537, 2002.
- [207] L. Thoraval, J.-P. Armspach et I. Namer. Analysis of brain functional MRI time series based on continuous wavelet transform and stimulation-response coupling distance. In *Proc. MICCAI Conf.*, 881–888, 2001.
- [208] L. Thoraval, G. Carrault et J.J. Bellanger. Heart signal recognition by hidden Markov models : the ECG case. *Meth. Inform. Med.*, 33 :10–14, 1994.
- [209] A. Tsai, J.W. Fisher, C. Wible, W.M. Wells, J. Kim et A.S. Willsky. Analysis of functional MRI data using mutual information. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 473–480, september 1998.
- [210] A.L. Vazquez et D.C. Noll. Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage*, 7 :108–118, 1998.
- [211] V.D. Calhoun, J.J. Pekar, V.B. mcGinty, T.Adali, T.D. Watson et G.D. Pearlson. Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping*, 16 :158–167, 2002.
- [212] V.D. Calhoun, T.Adali, G.D. Pearlson et J.J. Pekar. A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14 :140–151, 2001.
- [213] V.D. Calhoun, T.Adali, G.D. Pearlson et J.J. Pekar. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human Brain Mapping*, 13 :43–53, 2001.
- [214] V.D. Calhoun, T.Adali, V.B. mcGinty et J.J. Pekar. fMRI activation in a visual-perception task : Network of areas detected using the general linear model and independent component analysis. *NeuroImage*, 14 :1080–1088, 2001.
- [215] A.J. Viterbi. Error bounds for convolutionnal codes and asymptotically optimum decoding algorithm. *IEEE Trans. on Information Theory*, 13 :260–269, 1967.
- [216] D. Wager et T.E. Nichols. Optimization of experimental design in fMRI : a general framework using a genetic algorithm. *NeuroImage*, 18 :293–309, 2001.
- [217] G.C. Wei et M.A. Tanner. A Monte-carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85 :699–704, 1990.
- [218] C.J. Wellekens. Explicit time correlation in hidden Markov models for speech recognition. *Proc. Int’l. Conf. on Acoust., Speech and Signal Processing*, 384–386, 1987.
- [219] A.S. Willsky. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8) :1396–1458, 2002.
- [220] A.M. Wink et J.B.T.M. Roerdink. Denoising functional MR images : a comparison of wavelet denoising and Gaussian smoothing. *IEEE Trans. Medical Imaging*, 23(3) :374–387, 2004.



- [221] A. Wismüller, O. Lange, D.R. Dersch, G.L. Leinsinger, K. Hahn, B. Pütz et D. Auer. Cluster analysis of biomedical image time-series. *Int. J. Comput. Vision*, 46(2) :103–128, 2002.
- [222] A. Witkin. Scale-space filtering. *International Joint Conference on Artificial Intelligence*, 1019–1023, 1983.
- [223] P. C. Woodland et D. Povey. Large scale MMIE training for conversational telephone speech recognition. *Proc. Speech Transcription Workshop*, 266–277, 2000.
- [224] M.W. Woolrich, M. Jenkinson, J.M. Brady et S.M. Smith. Fully Bayesian spatio-temporal modeling of fMRI data. *NeuroImage*, 23(2) :213–231, 2004.
- [225] M.W. Woolrich, B.D. Ripley, M. Brady et S.M. Smith. Temporal autocorrelation in univariate modeling of fMRI data. *NeuroImage*, 14 :1370–1386, 2001.
- [226] K. Worsley. Local maxima and the expected euler characteristic of excursion sets of  $\chi^2$ ,  $f$ , and  $t$  fields. *Adv. Appl. Prob.*, 26 :13–42, 1994.
- [227] K. Worsley et K. Friston. Analysis of fMRI time series revisited - again. *Neuroimage*, 2 :173–181, 1995.
- [228] K.J. Worsley, C.H. Lia, J. Aston, V. Petre, G.H. Duncan, F. Morales et A.C. Evans. A general statistic analysis for fMRI data. *NeuroImage*, 15 :1–15, 2002.
- [229] K.J. Worsley, J.P. Poline, K.J. Friston et A.C. Evans. Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage*, 6 :305–319, 1997.
- [230] G. Zweig et S. Russel. Probabilistic modeling with Bayesian networks for automatic speech recognition. *ICSLP'98 Proceedings*, 7 :3011–3014, 1998.