

**ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC**

**MÉMOIRE PRÉSENTÉ À  
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE**

**COMME EXIGENCE PARTIELLE  
À L'OBTENTION DE LA  
MAÎTRISE EN GÉNIE ÉLECTRIQUE  
M.Ing.**

**PAR  
Samuel DESLAURIERS-GAUTHIER**

**CLASSIFICATION DES POINTES ÉPILEPTIQUES EN  
ÉLECTRO-MAGNÉTO-ENCÉPHALOGRAPHIE**

**MONTRÉAL, LE 04 JANVIER 2010**

**©Samuel Deslauriers-Gauthier, 2010**

**PRÉSENTATION DU JURY**

**CE MÉMOIRE À ÉTÉ  
ÉVALUÉ**

**PAR UN JURY COMPOSÉ DE :**

M. Jean-Marc Lina, directeur de mémoire  
Département de génie électrique à l'École de technologie supérieure

M. Christian Gargour, président du jury  
Département de génie électrique à l'École de technologie supérieure

Mme Nicola Hagemester, membre du jury  
Département de génie de la production automatisée à l'École de technologie supérieure

**IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC**

**LE 18 DÉCEMBRE 2009**

**À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE**

## REMERCIEMENTS

Je souhaite tout d'abord remercier mon directeur de mémoire, Jean-Marc Lina qui m'a proposé ce sujet de recherche et qui m'a guidé à travers sa réalisation.

Je veux aussi remercier toute l'équipe du laboratoire LATIS pour leurs contributions à ce projet de maîtrise. Je souhaite remercier tout particulièrement Étienne Lemay qui a fourni le code nécessaire à la résolution des problèmes inverses présentés dans ce mémoire.

Je remercie aussi Christophe Grova et Eliane Kobayashi, de l'Institut de Neurologie de Montréal, qui ont fourni les données nécessaires à la réalisation de ce projet.

Finalement, je souhaite remercier mes parents et ma famille pour leur inébranlable confiance en moi.

# CLASSIFICATION DES POINTES ÉPILEPTIQUES EN ÉLECTRO-MAGNÉTO-ENCÉPHALOGRAPHIE

Samuel DESLAURIERS-GAUTHIER

## RÉSUMÉ

L'électroencéphalographie (EEG) et la magnétoencéphalographie (MEG) sont des outils essentiels de diagnostic et de traitement de l'épilepsie. Elles permettent d'observer des événements fortement associés à l'épilepsie, les pointes épileptiques. Ces pointes permettent entre autres de localiser la région du cerveau qui cause les symptômes de l'épilepsie. Toutefois, pour obtenir une localisation précise, les signaux EEG et MEG qui contiennent les pointes doivent avoir un rapport signal sur bruit (SNR) élevé. Une technique qui permet d'augmenter le SNR est de faire la moyenne de plusieurs signaux similaires. Cependant, comment peut-on s'assurer que les signaux sont suffisamment similaires pour en faire la moyenne? La solution consiste à effectuer la classification des pointes épileptiques.

Ce mémoire présente la méthodologie et son évaluation, de la conception d'une nouvelle technique de classification de pointes épileptiques mesurées en EEG et en MEG. Parce que cette nouvelle technique utilise la représentation des pointes dans l'espace des sources, elle permet de classifier des pointes morphologiquement similaires, mais provenant de sources distinctes. La performance de cet algorithme a été évaluée sur des signaux EEG et MEG simulés. Les résultats indiquent que la technique proposée permet de grouper les pointes qui possèdent une représentation dans l'espace des sources similaires. L'utilisation de l'algorithme sur des signaux épileptiques réels a permis de trouver des régions actives du cerveau qui n'apparaissaient pas lors de l'analyse traditionnelle.

**Mots-clés :** pointes épileptiques, classification, épilepsie, électroencéphalographie, magnétoencéphalographie.

## CLASSIFICATION DES POINTES ÉPILEPTIQUES EN ÉLECTRO-MAGNÉTO-ENCÉPHALOGRAPHIE

Samuel DESLAURIERS-GAUTHIER

### ABSTRACT

Electroencephalography (EEG) and magnetoencephalography (MEG) are indispensable tools used in the diagnostic and treatment of epilepsy. They measure signals that display events heavily linked to epilepsy: the epileptic spikes. These spikes are used by neurologists to confirm their diagnosis but also to localize the region of the brain that causes the pathology. Since current source localization techniques require a high signal to noise ratio (SNR), it is a common practice to average recordings which are assumed to contain similar events. However, how can we be sure the signals are similar enough to be averaged? The answer is to classify the spikes prior to signal averaging.

In the present work, we present the conception, methodology, and evaluation of a new classification technique based on the source representation of epileptic spikes. Because the source space is used to classify the spikes, the method is able to separate spikes with similar morphologies but generated by different sources. The performance of this algorithm was evaluated using simulated EEG and MEG signals. The results indicate that the method is able to group spikes with similar source representation even if their morphologies are similar. When applied to real data, the method allowed us to identify new active regions of the brain when compared to traditional analysis.

**Keywords :** epileptic spikes, classification, clustering, epilepsy, electroencephalography, magnetoencephalography.

## TABLE DES MATIÈRES

	Page
INTRODUCTION .....	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	7
1.1 Modélisation de l'activité cérébrale.....	7
1.2 Résolution du problème inverse.....	10
1.3 État de l'art sur la classification des pointes épileptiques .....	11
1.3.1 Les algorithmes de classification de type I.....	11
1.3.2 Les algorithmes de classification de type II.....	16
1.4 Problématique et objectifs.....	17
CHAPITRE 2 MÉTHODOLOGIE : CLASSIFICATION DES POINTES ÉPILEPTIQUE .....	20
2.1 Extraction des caractéristiques.....	20
2.2 Classification des pointes.....	24
CHAPITRE 3 ÉVALUATION DE LA PERFORMANCE ET SIMULATION DE SIGNAUX.....	27
3.1 Simulation de signaux cérébraux simple .....	27
3.2 Simulation de signaux cérébraux avec de l'activité physiologique.....	30
3.3 Simulation de signaux EEG et MEG mesurés simultanément.....	34
3.4 Résultats sur des mesures simulées.....	35
3.4.1 Les résultats EEG.....	38
3.4.2 Les résultats MEG.....	40
3.5 Classification combinée EEG et MEG.....	42
CHAPITRE 4 CLASSIFICATION DE POINTES ÉPILEPTIQUES RÉELLES .....	44
4.1 Classification de pointes épileptiques réelles en EEG .....	44
4.2 Classification de pointes épileptiques réelles en MEG .....	55
4.3 Classification multimodales de pointes épileptiques réelles.....	61
4.4 Délai entre les pointes EEG et MEG.....	64
CONCLUSION... ..	67
ANNEXE I BRAINENTROPY .....	70
BIBLIOGRAPHIE.....	76

## LISTE DES TABLEAUX

	Page
Tableau 3.1	Exemple d'une classification parfaite..... 35
Tableau 3.2	Exemple d'une classification avec quelques erreurs ..... 36
Tableau 3.3	Un exemple de résultats pour une classification EEG et MEG indépendante ..... 43
Tableau 3.4	Classification des pointes EEG et MEG simultanée..... 43
Tableau 4.1	Résultat de la classification EEG à 2 groupes ..... 52
Tableau 4.2	Résultat de la classification MEG à 2 groupes ..... 58
Tableau 4.3	Le résultat de la combinaison des classifications EEG et MEG ..... 62
Tableau 4.4	Moyenne et écart-type des délais pour chaque groupe ..... 66

## LISTE DES FIGURES

	Page
Figure 1.1	Modèle sphérique de l'activité cérébrale. .... 8
Figure 1.2	Exemple de groupes formés à l'aide d'un algorithme GTC. Adaptée de Wahlberg (2000, p. 861) ..... 13
Figure 1.3	Exemple de dendrogramme coupé pour classifier six éléments en trois groupes. .... 16
Figure 1.4	Deux pointes EEG simulées ayant des morphologies similaires, mais des générateurs différents. .... 18
Figure 2.1	Histogramme des scores MSP. .... 24
Figure 3.1	Générateur de pointe épileptique simulée à l'aide de trois fonctions gaussiennes. .... 28
Figure 3.2	Un exemple d'une pointe EEG simulé à l'aide d'une source unique. .... 29
Figure 3.3	Spectre bruitée de l'activité physiologique. .... 31
Figure 3.4	Le signal obtenu lorsqu'une transformée de Fourier inverse est appliquée au spectre de la Figure 3.3. .... 31
Figure 3.5	Deux exemples de signaux EEG simulés contenant de l'activité physiologique en plus d'une pointe. La pointe du graphique du haut a une amplitude de 2 et celle du bas a une amplitude de 6. .... 32
Figure 3.6	Deux exemples de signaux EEG simulés contenant de l'activité physiologique en plus d'une pointe d'amplitude 6. Le signal du haut a un SNR de 10 dB. Le signal du bas a un SNR de 15 dB. .... 34
Figure 3.7	Performance de l'algorithme de comparaison sur les mesures EEG possédant un SNR de 30 dB. .... 37
Figure 3.8	Performance de la classification EEG pour un SNR de 10 dB. .... 39
Figure 3.9	Performance de la classification EEG pour un SNR de 20 dB. .... 39
Figure 3.10	Performance de la classification EEG pour un SNR de 30 dB. .... 40
Figure 3.11	Performance de la classification MEG pour un SNR de 10 dB. .... 41
Figure 3.12	Performance de la classification MEG pour un SNR de 20 dB. .... 41



Figure 3.13	Performance de la classification MEG pour un SNR de 30 dB.....	42
Figure 4.1	L'image du haut montre la position du maximum de la fonction de puissance globale de la première pointe du jeu de données. L'image du bas présente la même position sur les mesures EEG.....	45
Figure 4.2	La moyenne de toutes les pointes EEG du jeu de données. La ligne verticale présente la position du maximum de la FPG lorsqu'elle est calculée sur la moyenne des mesures.....	46
Figure 4.3	La solution du problème inverse effectué sur la moyenne de toutes les pointes EEG du jeu de données.....	47
Figure 4.4	Les scores MSP sur la surface du cortex pour les pointes 3, 23, 44 et 70 du jeu de données.....	48
Figure 4.5	Les sources choisies pour représenter les pointes 2, 23, 44 et 70 du jeu de donnée.....	49
Figure 4.6	Le nombre de sources communes entre les pointes.....	50
Figure 4.7	L'organisation des groupes EEG. Les pointes sont organisées en fonction de la classification à 2 groupes.....	51
Figure 4.8	La moyenne des signaux EEG du groupe 1 (haut) et du groupe 2 (bas) d'une classification EEG à 2 groupes.....	53
Figure 4.9	La solution du problème inverse obtenue pour la moyenne des pointes EEG du groupe 1 de la classification EEG à 2 groupes.....	54
Figure 4.10	La solution du problème inverse obtenue pour la moyenne des pointes EEG du groupe 2 d'une classification EEG à 2 groupes.....	55
Figure 4.11	La moyenne de toutes les pointes MEG du jeu de données. La ligne verticale présente la position du maximum de la FPG lorsqu'elle est calculée sur la moyenne des mesures.....	56
Figure 4.12	La solution du problème inverse effectué sur la moyenne de toutes les pointes MEG du jeu de données.....	57
Figure 4.13	L'organisation des groupes MEG. Les pointes sont organisées en fonction de la classification à 2 groupes.....	58
Figure 4.14	La moyenne des signaux MEG du groupe 1 (haut) et du groupe 2 (bas) d'une classification à 2 groupes.....	59

Figure 4.15	La solution du problème inverse obtenue pour la moyenne des pointes MEG du groupe 1 de la classification à 2 groupes.....	60
Figure 4.16	La solution du problème inverse obtenue pour la moyenne des pointes MEG du groupe 2 de la classification à 2 groupes.....	61
Figure 4.17	La solution du problème inverse pour le premier groupe de la classification combinée EEG et MEG. ....	63
Figure 4.18	L'activation des sources pour les 18 pointes du groupe 2 de la classification EEG et MEG combinée. ....	64
Figure 4.19	Exemple de la mesure du délai entre la position du maximum de la FPG en EEG et en MEG. ....	65

## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

BEM	Boundary element method
EEG	Électroencéphalogramme/Électroencéphalographie
EMG	Électromyogramme/Électromyographie
FCM	Fuzzy C-means
FPG	Fonction de puissance global
GTC	Graph-theoretic categorization
ICA	Analyse en composantes indépendantes
IRM	Imagerie par résonance magnétique
IRMf	Imagerie par résonance magnétique fonctionnelle
LCR	Liquide céphalorachidien
MEG	Magnétoencéphalogramme/magnétoencéphalographie
MEM	Maximum d'Entropie sur la Moyenne
MSP	Multivariate source pre-localization
MUSIC	Multiple Signal Classification
RAP-MUSIC	Recursively applied and projected MUSIC
SNR	Ratio signal sur bruit
SVD	Décomposition en valeurs singulières

## INTRODUCTION

L'épilepsie est une maladie du système nerveux qui affecte entre 0.5 et 1.0 % de la population mondiale (Centre Hospitalier Universitaire Vaudois Lausanne: Épilepsie, 2009, 26 juin; Marieb, 2005). Le dictionnaire Larousse (*Le petit Larousse illustré*, 1998) en donne la définition suivante : « Maladie caractérisée par des crises d'activité excessive des neurones dans le cortex cérébral ». Le symptôme le plus grave chez les patients épileptiques est la crise d'épilepsie. Les crises les plus violentes, nommées *grand-mal*, sont caractérisées par une perte de conscience et des convulsions. Les crises de moindre importance, fréquente chez les jeunes patients, sont généralement accompagnées d'une perte de conscience de courte durée et de spasmes.

L'épilepsie est une maladie qu'on ne sait pas guérir. Toutefois, les symptômes de l'épilepsie peuvent être limités ou même éliminés à l'aide de médicaments pour 75 à 80 % des patients (Centre Hospitalier Universitaire Vaudois Lausanne: Épilepsie, 2009, 26 juin). Pour les patients où les symptômes de la maladie sont dangereux et pour lesquels les médicaments sont sans effet, la dernière option est souvent la chirurgie. Lors de la planification de cette chirurgie, les neurochirurgiens doivent déterminer la région du cerveau qui cause l'épilepsie, s'il y en a une. Pour trouver cette zone, les experts disposent de trois outils : l'imagerie par résonance magnétique (IRM), l'électroencéphalographie (EEG) et la magnéto-encéphalographie (MEG).

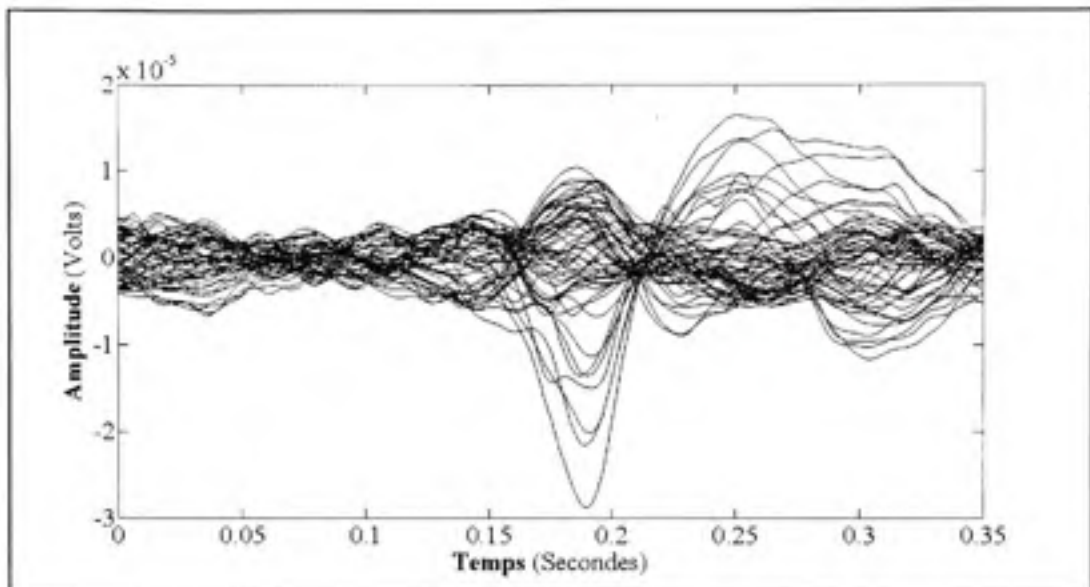
L'IRM peut être utilisée de deux manières distinctes dans le diagnostic de l'épilepsie. La première technique consiste à imager les différentes structures du cerveau à la recherche d'anomalie. La seconde est d'utiliser l'IRM pour mesurer l'activité fonctionnelle du cerveau. On parle alors d'imagerie par résonance magnétique fonctionnelle (IRMf). Dans ce cas, c'est le métabolisme ou la demande en oxygène du cerveau qui est quantifié. On suppose alors que les régions actives du cerveau ont une plus grande demande en oxygène que les régions inactives.

Contrairement à l'IRMf, l'EEG mesure une conséquence directe de l'activité cérébrale : la différence de potentiel sur la surface du scalp. Ces mesures sont obtenues en plaçant plusieurs électrodes sur la tête du sujet comme illustré par la figure suivante.



**Un sujet portant le casque EEG.**

En observant cette différence de potentiel varier au cours du temps, les neurologues distinguent certains événements qu'ils associent à l'épilepsie. Ces événements, qui prennent généralement la forme d'oscillations rapides de forte amplitude dans le signal, sont nommés *pointes épileptiques* ou *pointes intérieures*. En fait, les pointes sont si fortement liées à l'épilepsie qu'elles font de l'EEG un outil indispensable dans le diagnostic de la pathologie. La figure suivante présente une pointe épileptique mesurée en EEG. Chaque courbe du graphique correspond aux mesures recueillies par l'un des 56 capteurs du système d'acquisition.



**Une pointe épileptique mesurée en EEG.**

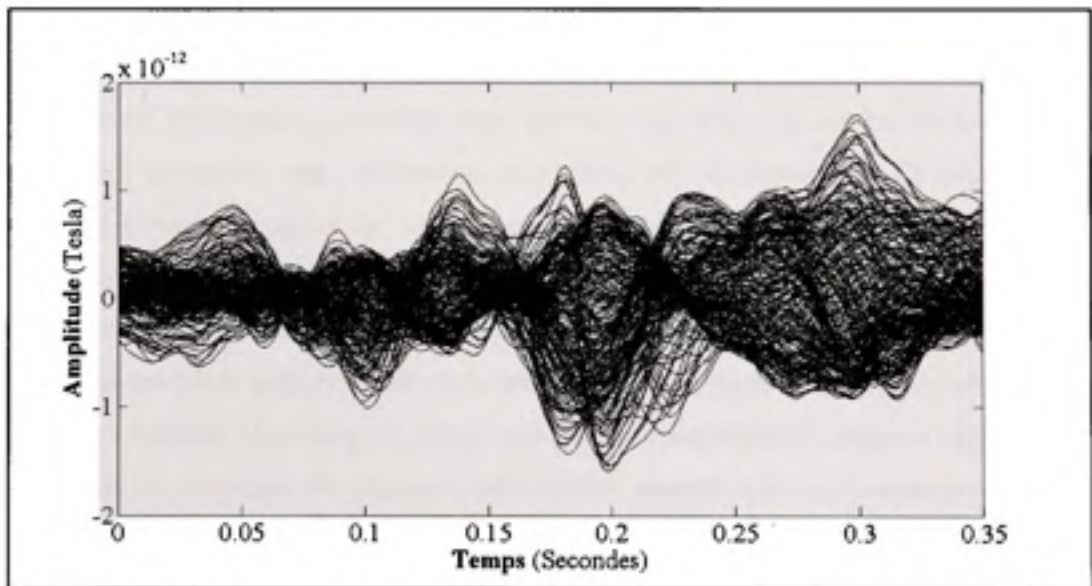
La MEG, comme l'EEG, permet d'obtenir une mesure directe de l'activité cérébrale. Dans ce cas, c'est l'amplitude ou le gradient du champ magnétique engendré par l'activité cérébrale qui est mesuré. L'équipement nécessaire pour les enregistrements MEG est beaucoup moins flexible que celui de l'EEG. La taille de l'équipement MEG est principalement due à la subtilité de l'information mesurée : le champ magnétique généré par l'activité cérébrale est de l'ordre de  $10^{-15}$  Tesla alors que le champ magnétique terrestre est de  $10^{-5}$  Tesla. Tout l'équipement MEG doit donc être placé dans une pièce isolée contre les champs magnétiques externes. De plus, l'amplitude des mesures est si faible qu'il faut avoir recours à des capteurs de type SQUID (de l'anglais *superconducting quantum interference devices*) qui fonctionnent à une température d'environ  $-150$  °C. Les appareils de réfrigération contribuent donc à la taille de l'équipement MEG. La figure suivante illustre l'équipement MEG.





**Présentation de l'équipement MEG.**

La figure suivante présente une pointe épileptique observée en MEG à l'aide de 272 capteurs.



**Une pointe épileptique mesurée en MEG.**

Bien que l'EEG et la MEG mesurent des effets directement liés à l'activité cérébrale, on observe une grande différence dans les enregistrements. Cette variation est expliquée par le

type d'information mesuré par chaque modalité. Les capteurs MEG, composés d'une ou de deux bobines, ne peuvent mesurer un champ magnétique que si celui-ci est perpendiculaire aux bobines. Inversement, l'activité cérébrale qui engendre un champ magnétique parallèle aux bobines ne sera pas vue en MEG. Par contre, le champ magnétique est très peu influencé par le milieu qu'il traverse ce qui permet d'obtenir un très bon signal de l'activité cérébrale qui génère un champ perpendiculaire aux bobines. Contrairement à la MEG, le signal mesuré en EEG dépend peu de l'orientation de la source de la différence de potentiel. Les mesures EEG contiennent donc une information sur l'activité cérébrale globale. Cependant, la propagation du potentiel est grandement influencée par le milieu traversé. L'EEG donnera donc un meilleur signal de l'activité cérébrale de surface, mais est plus difficile à modéliser de façon réaliste que la MEG. Puisque l'information mesurée est différente en EEG et en MEG, beaucoup de recherches en cours tentent de combiner les deux modalités afin d'exploiter les avantages de chacune d'elle.

Les pointes épileptiques ne sont pas seulement utilisées pour confirmer un diagnostic. On suppose aussi qu'elles sont générées par la région du cerveau qui est responsable des symptômes de l'épilepsie. La détermination de leur origine dans le cerveau permet donc aux neurologues d'acquérir une information importante sur la pathologie. Ce processus, qui permet de trouver la région du cerveau qui explique les mesures, est nommé problème inverse. Sans entrer dans les détails, mentionnons que la région identifiée lors de la résolution du problème inverse sera d'autant plus fiable que le rapport signal sur bruit des mesures sera élevé. Avec un faible rapport signal sur bruit, la solution trouvée risque d'être très diffuse ou simplement erronée. Une stratégie simple qui permet d'augmenter le rapport signal sur bruit est de faire la moyenne de plusieurs événements marqués par un neurologue. La partie commune des signaux, autrement dit la pointe, devrait être rehaussée. Un problème survient alors : **comment s'assurer que les événements sont similaires ou suffisamment similaires pour en faire la moyenne?** Répondre à cette question consiste à classer les pointes épileptiques sur un critère de similarité ou de ressemblance. Tel est le sujet du travail présenté ici. Il faut noter que cette question est d'autant plus importante que des détecteurs de pointes sont disponibles dans les outils logiciels utilisés par les neurologues (BESA: Brain



Electrical Source Analysis, 2009; Stellate, 2009). Fondés sur des critères assez arbitraires, ces détecteurs de pointes ne garantissent pas un inventaire cohérent des événements qu'ils détectent. Ce travail a fait l'objet d'une affiche (Deslauriers-Gauthier et al., 2009) présenté à l'atelier « Inverse problems in brain imagery and multimodal fusion » au Centre de Recherche Mathématiques de Montréal.

La contribution principale de ce mémoire est la conception d'un nouvel algorithme de classification pour les pointes intérieures. Ce nouvel algorithme se distingue des techniques habituelles en se basant sur la représentation des sources d'un modèle distribué. De plus, ce mémoire évalue une technique qui permet de combiner plusieurs classifications, ouvrant ainsi la porte à la classification multimodale. En marge de ces contributions, ce travail a fait l'objet de la conception d'un paquet logiciel nommé *BrainEntropy* en collaboration avec Étienne Lemay. Ce paquet a pour objectif le développement de méthodes numériques dédiées à l'analyse multimodale de l'activité cérébrale.

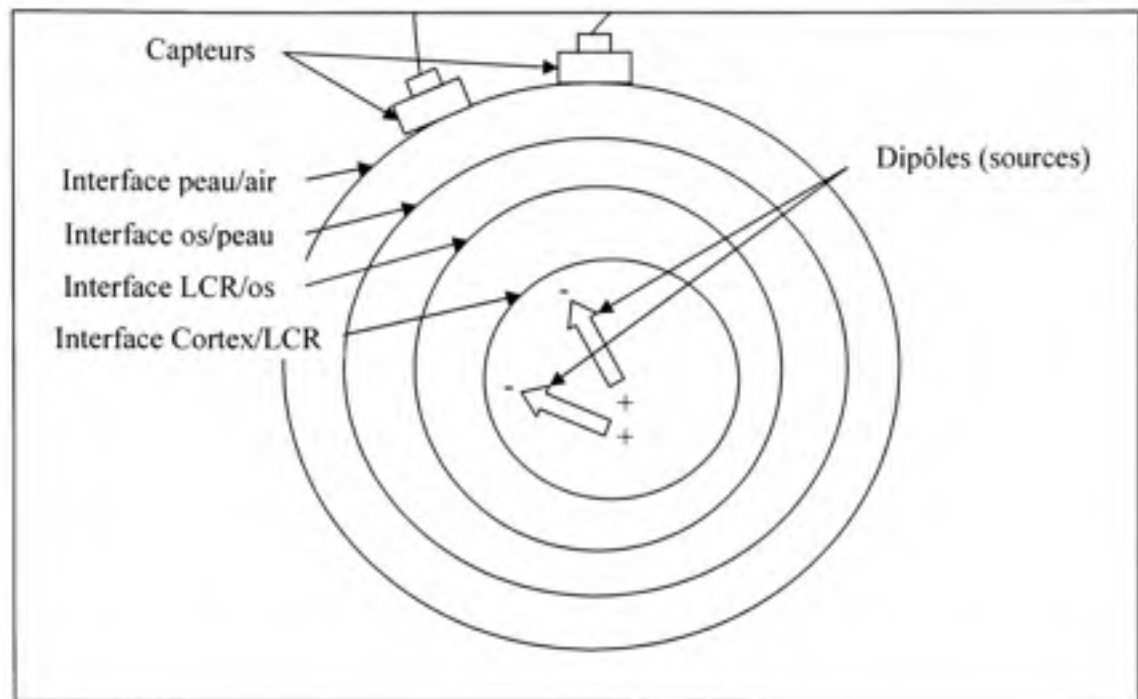
## CHAPITRE 1

### REVUE DE LITTÉRATURE

Plusieurs algorithmes de classification des pointes épileptiques ont déjà été proposés par d'autres auteurs. Cependant, avant de les présenter, il est nécessaire de décrire le modèle de l'activité cérébrale utilisé en neuroscience ainsi que la technique de résolution du problème inverse utilisée dans ce mémoire. La première section de ce chapitre présente donc la modélisation de l'activité cérébrale. La deuxième section expose brièvement la résolution du problème inverse. La troisième section présente les techniques de classification qui sont proposées par d'autres auteurs. La quatrième section présente les faiblesses des méthodes existantes et énonce les caractéristiques d'un algorithme qui pallie ces défauts.

#### 1.1 Modélisation de l'activité cérébrale

Le modèle de l'activité cérébrale présenté dans cette section permet d'obtenir les mesures si l'on connaît l'activité du cerveau, ce que l'on nomme le problème direct. Il permet aussi d'obtenir l'activité cérébrale qui explique les mesures, donc de résoudre le problème inverse, bien que cette opération soit beaucoup plus difficile. Le modèle le plus simple de l'activité cérébrale consiste à représenter la tête à l'aide de sphères concentriques, chaque sphère étant une des interfaces entre les différents tissus de la tête. Par exemple : le cortex, le liquide céphalorachidien (LCR), l'os et la peau. Au niveau du cortex, les colonnes de neurones responsables de l'activité électrique sont modélisées sous la forme de dipôles électriques (des générateurs similaires à de petites batteries) à l'interface de la matière blanche et de la matière grise. Ce sont les sources de l'activité électrique. La Figure 1.1 présente cette modélisation simple de l'activité cérébrale.



**Figure 1.1** Modèle sphérique de l'activité cérébrale.

Le problème qui consiste à obtenir les mesures à partir de l'activité des sources est un problème classique en électromagnétisme. Les mesures obtenues à l'aide de ce modèle dépendent de l'intensité, de la position et de l'orientation des sources. Elles seront aussi affectées par les propriétés physiques des différents tissus (les sphères du modèle). Si l'on suppose que les propriétés du modèle ne changent pas au cours du temps, les mesures dépendent alors uniquement des caractéristiques des sources. Donc, pour une intensité, une position et une orientation spécifique des sources, les mesures sont complètement définies. Cette relation source/mesure peut être modélisée à l'aide d'un opérateur nommé *matrice de gain*, généralement identifié par  $G$ . Parce que l'opérateur  $G$  contient un grand nombre de degrés de liberté, certaines hypothèses sont posées afin d'en réduire la complexité. Si l'on suppose que les mesures peuvent être expliquées à l'aide d'un seul dipôle, la matrice de gain fait le lien entre la position et l'orientation de ce dipôle et les mesures. On parle alors d'un

modèle dipolaire. Par contre, si l'on distribue plusieurs dipôles ayant une position et une orientation fixe à l'intérieur du modèle, il s'agit d'un modèle distribué. La matrice de gain représente alors la contribution de chaque dipôle sur les capteurs. Le calcul numérique d'une matrice de gain est loin d'être une opération triviale. Si l'on souhaite que les surfaces du modèle soient fidèles à l'anatomie du patient (et non des sphères), on obtient leur topologie en segmentant une image IRM du cerveau. Ensuite, il faut recalculer les capteurs sur les surfaces et établir les propriétés électromagnétiques du modèle (Geddes et Baker, 1967). Heureusement, beaucoup de recherches ont été effectuées sur ce sujet et des logiciels qui permettent de calculer une matrice de gain sont maintenant disponibles (Ermer et al., 2001). Si les différentes composantes de la tête sont modélisées à l'aide de surfaces (pas nécessairement des sphères), la méthode pour le calcul de la matrice de gain est alors nommée BEM (de l'anglais *Boundary element method*). La méthode BEM n'est pas la seule disponible, mais elle est la plus couramment utilisée (Baillet, Mosher et Leahy, 2001a).

La propagation des ondes électriques et magnétiques à travers le cerveau est modélisée à l'aide d'une équation linéaire. Dans le cadre de l'EEG, le problème direct consiste à calculer la différence de potentiel qui sera mesurée sur le scalp si l'activité cérébrale est connue. Sur un seul échantillon de temps, l'expression mathématique du problème direct est :

$$M = GJ + \varepsilon \quad (1.1)$$

Ici,  $M$  dénote le vecteur colonne qui contient les mesures sur  $N_c$  capteurs.  $J$  est le vecteur colonne qui contient l'intensité des  $N_s$  sources distribuées sur la surface corticale.  $G$  représente la matrice de gain, de dimension  $N_c \times N_s$ . Finalement,  $\varepsilon$  est le bruit de mesure. C'est un vecteur colonne de dimension  $N_c$  qui contient un bruit Gaussien. Cette modélisation devrait permettre d'obtenir l'activité des sources lorsqu'on connaît les mesures. Cette tâche, nommée problème inverse, est toutefois difficile à réaliser parce que le problème est mal posé.

## 1.2 Résolution du problème inverse

Comme il a été énoncé précédemment, la résolution du problème inverse permet d'obtenir l'activité des sources qui expliquent les mesures. Le problème inverse est dit mal posé parce que le nombre de degrés liberté du modèle est beaucoup plus grand que le nombre de capteurs ( $N_c \ll N_s$ ). Il admet donc une infinité de solutions. En plus, l'obtention de la « vraie » solution est compliquée par la présence du bruit de mesure. La résolution de ce problème est un sujet de recherche très actuel, mais plusieurs algorithmes permettent maintenant d'obtenir des résultats intéressants.

Afin de valider les résultats obtenus, plusieurs problèmes inverses seront résolus dans le cadre de ce mémoire. Pour obtenir des résultats comparables, seule la technique nommée maximum d'entropie sur la moyenne (MEM) sera utilisée. Le MEM est une technique probabiliste qui s'apparente aux techniques bayésiennes de problème inverse. Elle a été appliquée aux signaux biomédicaux par Clark (Clarke, 1989) et a récemment été améliorée par Amblard (Amblard, Lapalme et Lina, 2004). D'une façon générale, la méthode MEM tente de corriger une loi de référence établie sans observer les mesures. La loi de référence contient l'information connue a priori sur l'intensité des sources. Par exemple, que la distribution de l'intensité des sources suit une loi normale ou qu'il y a une variation continue de l'intensité des sources dans l'espace. Il est important de souligner que la loi de référence est établie sans observer les mesures. Cette loi est ensuite modifiée de telle manière qu'elle puisse expliquer les données. Le MEM trouve donc la solution qui offre le meilleur compromis entre les mesures et la loi de référence (l'information connue sur l'activité des sources). L'implémentation du MEM utilisée dans ce mémoire est celle proposée par Lemay (Lemay, 2009).

Une brève description de certaines techniques développées pour résoudre le problème inverse a été publiée par Baillet (Baillet, Mosher et Leahy, 2001b). Le lecteur est invité à consulter cet ouvrage pour obtenir plus d'information.

### 1.3 État de l'art sur la classification des pointes épileptiques

Tous les algorithmes de classification de pointes épileptiques possèdent deux étapes principales : l'extraction des caractéristiques et la classification. L'extraction des caractéristiques permet d'obtenir une petite quantité d'information qui résume chaque pointe. Dans la plupart des cas, la stratégie consiste à résumer l'essentiel d'une pointe à l'aide de son amplitude, de sa largeur temporelle, de son contenu spectral ou de sa décomposition en ondelettes (Hulata, Segev et Ben-Jacob, 2002). L'étape de la classification permet d'associer un numéro de groupe à chaque pointe en utilisant uniquement les caractéristiques retenues. Les pointes qui possèdent le même numéro de groupes sont alors jugées comme étant similaires.

Dans ce mémoire, les algorithmes de classification sont séparés en deux types distincts. Le type I englobe tous les algorithmes qui utilisent uniquement l'information des mesures. Les algorithmes de classification de type II utilisent aussi l'anatomie du patient pour poser des hypothèses sur la provenance des mesures. Elles font donc usage de la matrice de gain.

#### 1.3.1 Les algorithmes de classification de type I

Quian Quiroga (Quiroga, Nadasdy et Ben-Shaul, 2004) propose un algorithme de classification des pointes mesurées sur des électrodes intracrâniennes. Bien que cette technique n'ait pas été développée pour la classification des pointes épileptiques mesurées à la surface du scalp en EEG ou en MEG, elle peut être adaptée à ce contexte sans modifications majeures. L'algorithme de classification utilise la représentation en ondelette des mesures plutôt que les mesures elles-mêmes. Le choix du type d'ondelette dépend des caractéristiques du signal que l'on souhaite analyser. Parce que l'auteur n'a pas de critère spécifique sur ce choix, il propose l'ondelette la plus simple, soit celle de Haar.

La première étape de l'algorithme consiste à obtenir quelques coefficients en ondelettes qui représentent bien les pointes. Pour sélectionner les valeurs les plus pertinentes, l'auteur



propose de se baser sur la distribution empirique de chaque coefficient. Plus la distribution d'un coefficient s'éloigne d'une loi normale, plus il est jugé intéressant. Ce raisonnement est justifié par l'idée qu'une loi normale est difficilement séparable en deux groupes.

Les coefficients ainsi sélectionnés sont classifiés. L'algorithme de classification proposé par les auteurs traite les vecteurs de coefficients en ondelettes comme les éléments d'un système dynamique. Lorsque le système est mis en marche (suivant des règles qui définissent sa dynamique), les éléments similaires ont tendance à se regrouper. La classification est terminée lorsque le système est stable.

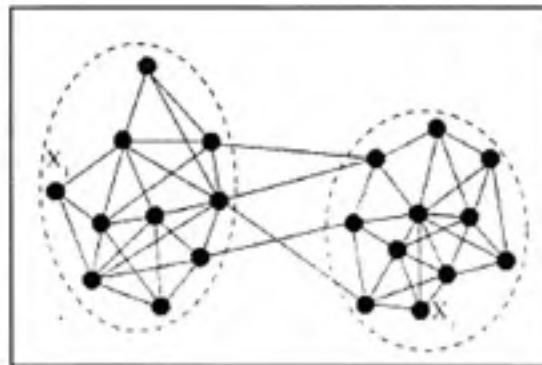
Certaines méthodes bayésiennes (Bar-Hillel, Spiro et Stark, 2006) et d'autres méthodes inspirées de la théorie des filtres et de la reconnaissance de formes ont aussi été utilisées (Wheeler et Heetderks, 1982) pour classifier les pointes mesurées avec des électrodes intracrâniennes. Toutefois, parce que ces techniques ne sont pas directement liées à la problématique de ce mémoire, elles ne seront pas présentées en détail. Toutes les autres méthodes présentées dans cette section ont été développées pour la classification des pointes épileptiques.

Les travaux de Wahlberg (Wahlberg et Lantz, 2000) comparent deux techniques de classification. La première, nommée *Fuzzy C-means* (FCM), utilise une transformée de Karhunen-Loève pour extraire quelques coefficients associé aux directions principales des mesures. Les vecteurs obtenus sont ensuite classifiés à l'aide d'un algorithme *fuzzy c-means*. Le *fuzzy c-means* est une version floue de l'algorithme *k-means* qui est présenté en détail dans la section 2.2 de ce mémoire. La deuxième technique, nommée *graph-theoretic categorization* (GTC), résume les mesures à l'aide de la distance entre deux pointes épileptiques  $i$  et  $j$ . En définissant par  $X_n$  la matrice de dimension  $N_c \times N_t$  où  $N_c$  correspond au nombre de capteurs et  $N_t$  au nombre d'échantillons, la distance est obtenue par :

$$\bar{X}_n = \frac{X_n}{\sqrt{\text{tr}(X_n^T X_n)}} \quad (1.2)$$

$$d_{i,j} = \sqrt{\text{tr}\left((\bar{X}_i - \bar{X}_j)^T (\bar{X}_i - \bar{X}_j)\right)} \quad (1.3)$$

où  $^T$  dénote la transposition. Ces distances sont utilisées pour regrouper les points à l'aide d'un algorithme *graph-theoretic clustering*. Ce type d'algorithme crée des groupes en utilisant les liens entre les points à classer. À la première itération, tous les points sont reliés entre eux. À chaque itération suivante, le lien entre les points les plus distants est éliminé. Les classes sont formées en regroupant les points qui sont liés à au moins  $k$  points du même groupe. La Figure 1.2 présente un exemple de groupes formés à l'aide d'un algorithme *graph-theoretic clustering*. Les paramètres de cet algorithme sont le nombre d'itérations à effectuer et la valeur de  $k$ .



**Figure 1.2 Exemple de groupes formés à l'aide d'un algorithme GTC.**

Adaptée de Wahlberg (2000, p. 861)

Les algorithmes *fuzzy C-means* et *graph-theoretic clustering* utilisés par Wahlberg permettent d'identifier les points qui ne font pas partie d'un groupe et qui sont donc des valeurs aberrantes. Selon l'auteur, cette caractéristique permet une classification plus robuste. La performance des deux algorithmes a été évaluée par les auteurs à l'aide de mesures EEG simulées et réelles. Afin de quantifier la performance, trois types d'erreurs ont été définies. Une erreur de type 1 (E1) est une pointe classée dans le mauvais groupe. Une erreur de type



2 (E2) est une pointe identifiée comme une valeur aberrante. Une erreur de type 3 (E3) est une valeur aberrante classée comme une pointe.

Les résultats obtenus en simulation étaient acceptables pour la technique GTC (E1 0% ; E2 0-9% ; E3 0-4%), mais était inacceptable pour la technique FCM (E1 0-28% ; E2 0-39% ; E3 1-24%). Avec des mesures réelles, les performances étaient comparées à la classification visuelle faite par un spécialiste. Encore une fois, les erreurs de classification du FCM étaient beaucoup trop élevées (E1 0-34% ; E2 3-48% ; E3 0-34%) alors que la performance de la technique GTC variait grandement en fonction des jeux de données. Pour certains patients, l'erreur était acceptable (E1 0% ; E2 0-5% ; E3 0%) alors que pour d'autres mesures, les erreurs de type 2 pouvaient atteindre 63%.

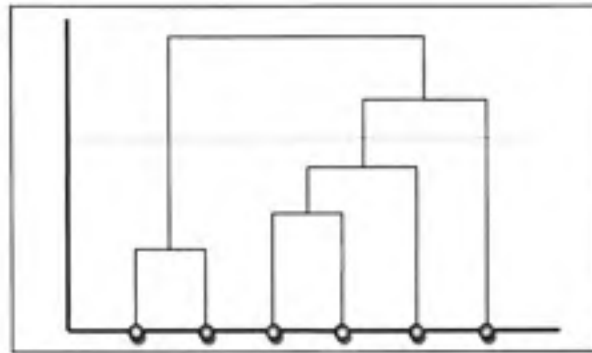
Dans le même ordre d'idée, une autre méthode est proposée par Van't Ent (Van 't Ent et al., 2003). La première étape de l'algorithme consiste à recalculer les pointes en utilisant les signaux élevés à la puissance quatre et de calculer la moyenne de tous les capteurs. Les pointes sont recalées sur le point d'amplitude maximum de cette fonction. Pour chaque pointe, une matrice de mesures contenant une ligne par capteurs (151 capteurs MEG dans le cas présent) et 11 échantillons (5 de part et d'autre du marqueur) est retenue. Les mesures sont ensuite normalisées en utilisant la norme de Frobenius défini par l'équation (1.2). Finalement, la distance euclidienne entre les pointes est utilisée comme caractéristique de classification dans un algorithme hiérarchique proposée par Ward (Ward, 1963). L'originalité de cette publication est l'utilisation de mesures MEG pour évaluer la performance de la méthode, ce qui est rare dans le domaine de la classification. Appliqué sur les mesures MEG de 4 patients épileptiques cet algorithme de classification a généré des groupes ayant tous une projection distincte sur le scalp. La résolution du problème inverse sur les groupes obtenus était en accord avec les lésions observées en IRM.

Wilson (Wilson et al., 1999) propose une approche qui utilise un indice de perception des pointes. Si l'on demande à plusieurs spécialistes de marquer les pointes d'un signal, il est fort probable que les résultats obtenus varient d'une personne à l'autre. À partir d'un ensemble

d'experts, il est ainsi possible de calculer la probabilité qu'une pointe soit identifiée par un expert. C'est cette probabilité qui est utilisée comme indice de perception. L'approche proposée par les auteurs utilise un réseau de neurones pour calculer l'indice de perception associé à une nouvelle pointe. Parce que ce type d'algorithme nécessite un entraînement, une base de données a été créée à cet effet. Cinq experts ont marqué les pointes de cette base de données en leur attribuant un indice de perception. Le réseau de neurones est ensuite entraîné sur cette base de données pour mettre en relation l'indice de perception des experts avec les questions suivantes:

1. La pointe ressort-elle de l'activité physiologique?
2. La pointe ressort-elle de l'activité rythmique locale?
3. La pointe ressort-elle de l'activité mesurée sur l'électromyogramme (EMG)?
4. La pointe a-t-elle une onde lente?
5. Y a-t-il des pointes au même moment sur d'autres électrodes?

Les pointes à classifier sont ensuite catégorisées à l'aide d'un algorithme hiérarchique. Un algorithme hiérarchique est initialisé en attribuant un numéro de groupe différent à chaque élément. Il y a donc autant de groupes que d'éléments. Ensuite, les deux éléments les plus proches l'un de l'autre selon une certaine fonction de distance sont regroupés en leur donnant le même numéro de groupe. Cette étape est répétée jusqu'à ce que tous les éléments portent le même numéro de groupe. Le résultat final peut être observé sous la forme d'un dendrogramme, comme l'illustre la Figure 1.3. Pour obtenir la classification finale, il suffit de couper le dendrogramme à une hauteur choisie en fonction du nombre de groupe désiré.



**Figure 1.3 Exemple de dendrogramme coupé pour classifier six éléments en trois groupes.**

Afin de rendre la classification plus robuste, les caractéristiques morphologiques telles que l'amplitude et la largeur de la pointe sont utilisées pour la classification en plus de l'indice de perception. Dans ce cas, la classification sert principalement à éliminer les artéfacts.

La validation de leur démarche a été faite sur une base de données constituée de pointes réelles. La sensibilité et la sensibilité de l'algorithme étaient acceptables (entre 0.8 et 0.9) et les auteurs observent que la performance est très similaire à celle d'un expert. Malheureusement, puisque leur approche n'a pas été testée sur des mesures simulées, il est impossible d'avoir une mesure objective de sa performance.

D'autres algorithmes similaires à ceux présentés plus haut ont aussi été proposés par Lantz (Lantz et al., 1998), Wahlberg (Wahlberg et Salomonsson, 1996).

### 1.3.2 Les algorithmes de classification de type II

Les travaux de Flanagan (Flanagan et Agawal, 2002) proposent une méthode simple permettant de classifier visuellement les pointes. Au lieu d'extraire une information morphologique des signaux, les auteurs utilisent la position d'une source dipolaire qui explique le mieux les mesures. La position du dipôle a été obtenue à l'aide de la méthode de résolution du problème inverse proposée par Sun (Sun, 1997) qui utilise une matrice de gain

obtenu à l'aide d'un modèle à quatre sphères concentriques. Tous les dipôles (un pour chaque pointe) sont ensuite placés sur une sphère et visuellement regroupés par un expert. Les auteurs indiquent que dans certains cas, leur algorithme permet d'obtenir rapidement la forme moyenne de certains types de pointe. Par contre, ils mentionnent aussi que leur technique n'est pas fiable pour la classification de pointe de faible intensité puisqu'il y a une grande variation dans la position du dipôle.

Ossadtchi (Ossadtchi et al., 2004) propose aussi une technique utilisant la position et l'orientation d'un dipôle équivalent comme caractéristiques de classifications. Par contre, les auteurs proposent d'utiliser l'analyse par composantes indépendantes (ICA) pour améliorer la reconstruction des sources. Ils utilisent donc un indice  $I_m$  de qualité de pointes pour le signal  $s_m$  d'une pointe définie par:

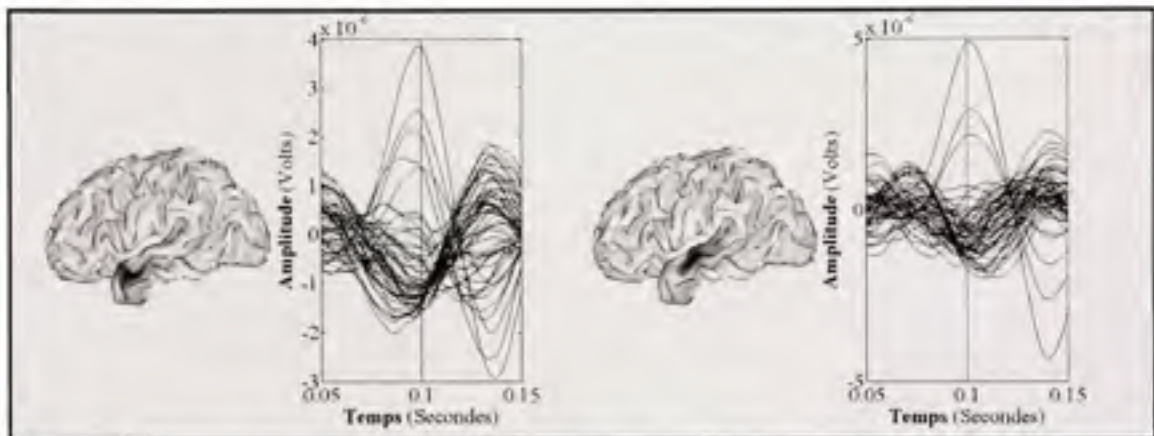
$$I_m = \frac{\max(s_m(t))}{\frac{1}{T} \int |s_m(t)| dt} \quad (1.4)$$

où  $T$  est le nombre d'échantillons du signal. Les composantes maximisant l'indice de qualité de pointes sont choisies pour effectuer la localisation de sources à l'aide de la technique *Recursively Applied and Projected Multiple Signal Classification* (RAP-MUSIC) (Mosher et Leahy, 1999). Les dipôles sont ensuite regroupés en utilisant leur position et la morphologie de leur activation comme paramètre. Lorsqu'appliquée sur quatre jeux de données réels, cette technique a identifié des classes de pointes à proximité des zones épileptiques identifiées lors de chirurgies.

#### 1.4 Problématique et objectifs

Dans le cadre de ce mémoire, la classification des pointes épileptiques permet de regrouper les pointes qui sont suffisamment similaires pour en faire la moyenne. Les algorithmes de classification de type I présentés dans la Section 1.3.1 tentent d'atteindre cet objectif, mais possèdent des faiblesses. Si les auteurs n'ont pas recours à l'anatomie du sujet pour poser des hypothèses sur la provenance des mesures, ils sont contraints d'utiliser uniquement

l'information des mesures comme critère de classification. Dans ce cas, on suppose que deux pointes qui possèdent une morphologie similaire ont une représentation dans l'espace des sources qui sont similaires. Toutefois, il est très facile de montrer à l'aide d'une simulation que cette hypothèse est fautive. La Figure 1.4 présente deux pointes ayant des morphologies similaires, mais provenant de sources différentes. Les algorithmes de la première catégorie ne réussiront pas à séparer ce type de pointes.



**Figure 1.4 Deux pointes EEG simulées ayant des morphologies similaires, mais des générateurs différents.**

La solution du problème de classification doit utiliser la représentation d'une pointe dans l'espace des sources, donc être de type II. Comme on l'a vu précédemment, les seuls algorithmes connus dans cette catégorie utilisent des géométries sphériques et un nombre de dipôles limités. La résolution spatiale de ces techniques est donc très approximative. De plus, la technique d'optimisation pour trouver le ou les dipôles associés aux mesures est connue pour être instable à cause de minimums locaux. Dans le cas où deux générateurs de pointes sont proches l'un de l'autre, comme ceux de la Figure 1.4, les dipôles trouvés auront fort probablement une position et une orientation similaire. Encore une fois, il sera impossible de différencier les deux pointes.

Afin de pallier les faiblesses des algorithmes de classification de pointes épileptiques proposés par d'autres auteurs, un nouvel algorithme a été développé dans ce mémoire. Pour obtenir une meilleure résolution spatiale et augmenter la spécificité de la classification, l'algorithme de classification utilise un modèle distribué plutôt que dipolaire. En plus, la géométrie du modèle utilisée est fidèle à l'anatomie du sujet, contrairement à un modèle sphérique.



## CHAPITRE 2

### MÉTHODOLOGIE : CLASSIFICATION DES POINTES ÉPILEPTIQUE

Comme présenté dans le chapitre précédent, l'algorithme de classification proposé dans ce mémoire utilise un modèle distribué et une géométrie réaliste de l'anatomie du sujet. Deux hypothèses doivent être respectées pour que cette technique soit valide :

1. La propagation des courants électriques et des champs magnétiques dans le cerveau peut être modélisée par le problème direct, soit l'équation (2.1).

$$M = GJ + \varepsilon \quad (2.1)$$

2. La matrice de gain  $G$  du problème direct, calculé à l'aide de la méthode BEM, est valide et relie correctement l'activation des sources aux mesures des capteurs EEG ou MEG.

Comme toutes les techniques décrites dans le Chapitre 1, l'algorithme de classification des pointes épileptiques proposé dans ce mémoire peut être divisé en deux étapes : l'extraction des caractéristiques et la classification. Lors de l'extraction des caractéristiques, les sources qui peuvent expliquer les mesures sont identifiées. Pour atteindre cet objectif, on utilise la méthode *Multivariate Source Prelocalization* (MSP), développée par Mattout (Mattout et al., 2005). Les pointes expliquées par les mêmes sources sont ensuite regroupées à l'aide de l'algorithme de classification *k-means*.

La première section de ce chapitre présente en détail la technique MSP et indique comment les sources qui peuvent expliquer les mesures sont identifiées. Dans la deuxième section, le fonctionnement de l'algorithme *k-means* est décrit.

#### 2.1 Extraction des caractéristiques

La première étape de l'algorithme a comme objectif d'obtenir une représentation des pointes dans l'espace des sources. Contrairement aux techniques présentées dans la section

précédente, l'amplitude de l'activation des sources en fonction du temps n'est pas calculée. On considère plutôt la probabilité qu'une source participe aux mesures caractérisant une pointe. On utilise la méthode MSP proposée par Mattout (Mattout et al., 2005) pour estimer cette probabilité sur l'ensemble des sources distribuées sur le cortex.

Avant de procéder, il est essentiel de bien décrire le contenu de la matrice de gain  $G$  qui relie les sources aux mesures. Chaque colonne  $g_i$  de cette matrice contient  $N_c$  valeurs qui représentent les mesures obtenues si une source est activée avec une amplitude de 1. Si chaque colonne est normalisée à l'aide l'équation (2.2), alors les  $g_i$  ne correspondent plus aux mesures observées pour le dipôle  $i$ , mais au *ratio* des mesures observées.

$$\tilde{g}_i = \frac{g_i}{\|g_i\|^2} \quad (2.2)$$

Cette normalisation est intéressante parce qu'elle permet de donner un poids équivalent à toutes les sources. En d'autres mots, les sources profondes seront aussi bien représentées sur les capteurs que les sources de surface. Puisque les données expliquées par  $G$  ont été ramenées sur une sphère de rayon 1, les mesures  $M$  sont aussi ramenées sur cette sphère. La matrice de gain et les mesures normalisées selon la norme  $L_2$  sont respectivement nommé  $\tilde{G}$  et  $\tilde{M}$ .

On considère la décomposition en valeurs singulières (SVD) de la matrice de gain normalisée :

$$\tilde{G} = B\Lambda^{1/2}C^T \quad (2.3)$$

Dans cette décomposition  $B$  et  $C$  sont des matrices orthogonales de dimension  $N_c \times N_c$  et  $N_s \times N_s$  respectivement. Chaque colonne de ces matrices, notée  $b_i$  et  $c_i$ , correspond à une orientation principale de la matrice de gain : les  $b_i$  dans l'espace des capteurs et les  $c_i$  dans l'espace des sources. La matrice diagonale  $\Lambda^{1/2}$ , de dimension  $N_c \times N_s$ , contient les valeurs propres ( $\sqrt{\lambda_i}$ ) qui pondèrent chacune des orientations de  $B$  et  $C$ . Plus le  $\lambda_i$  est élevée, plus les



colonnes  $b_i$  et  $c_i$  qui lui sont associées sont importantes. L'équation (2.3) peut donc être réécrite :

$$\bar{G} = \sum_{i=1}^{N_c} \sqrt{\lambda_i} b_i c_i^T \quad (2.4)$$

Les  $b_i$  et les mesures devrait théoriquement être corrélés, puisque la matrice de gain explique les mesures. La corrélation entre le vecteur  $b_i$  et les mesures est évalué en projetant  $M$  sur  $b_i$  pour obtenir  $\Gamma$  :

$$\Gamma_i = b_i^T \bar{M} \quad (2.5)$$

et en calculant coefficient de corrélation multiple défini par :

$$R_i^2 = \frac{\Gamma_i \Gamma_i^T}{N_t} \quad (2.6)$$

La valeur de  $R_i^2$  mesure donc la corrélation entre les mesures et la direction  $i$  de la matrice de gain. Si le calcul est repris pour toutes les colonnes de  $B$ ,  $R^2$  devient alors un vecteur de  $N_c$  éléments. Le vecteur  $\lambda$ , composé des  $N_c$  valeurs  $\sqrt{\lambda_i}$ , mesure l'importance de chaque composante de la matrice de gain.

Dans la prochaine étape, on crée le projecteur qui filtre les mesures en enlevant l'information qui n'est pas expliquée par la matrice de gain. Ce projecteur est construit de telle manière qu'il permet aussi d'éliminer les composantes de la matrice de gain qui ne sont pas corrélées avec les mesures. Tout d'abord, les vecteurs  $R^2$  et  $\lambda$  sont ordonnés simultanément afin que les valeurs de  $R^2$  soient en ordre croissant. Ensuite, la somme cumulative des éléments de  $\lambda$  est calculée jusqu'à un seuil nommé  $T_{MSP}$ . Ceci correspond à filtrer les mesures qui ne peuvent être expliquées par la matrice de gain, mais en tenant compte du poids de chaque composante éliminée. Dans ce mémoire, la valeur de  $T_{MSP}$  a été choisie empiriquement et est 0.95. On souhaite donc filtrer les mesures en gardant 95 % des composantes de la matrice de gain.

Une sous-matrice de  $B$ , identifiée  $B_s$ , est construite en ne gardant que les colonnes de  $B$  qui correspondent aux valeurs propres  $\lambda_l$  incluses dans la somme cumulative. Comme l'indique Mattout, cette sous-matrice est fortement corrélée avec les mesures, mais est issue de  $\bar{G}$ . L'intérêt de cette matrice est que le projecteur  $B_s B_s^T$  permet de projeter dans un espace qui peut être expliqué par la matrice de gain. Les mesures qui peuvent être expliquées par la matrice de gain sont donc obtenues à l'aide de :

$$M_s = B_s B_s^T \bar{M} \quad (2.7)$$

Pour identifier les sources qui sont corrélées avec les mesures, le projecteur  $P_s$  est créé à l'aide de  $M_s$  :

$$P_s = M_s (M_s^T M_s)^* M_s^T \quad (2.8)$$

où l'étoile dans l'équation (2.8) représente la pseudo-inverse. Les scores MSP sont définis comme les valeurs se trouvant sur la diagonale de la matrice obtenue en projetant  $\bar{G}$  sur  $P_s$ .

$$S_{MSP} = \text{diag}(\bar{G}^T P_s \bar{G}) \quad (2.9)$$

Pour une matrice  $M$  de dimension  $N_c \times N_t$ , un score entre 0 et 1 est obtenu pour chaque source de la matrice de gain. Cette valeur peut être interprétée comme la probabilité que la source participe aux mesures. Une pointe épileptique peut donc être représentée par un vecteur de scores MSP qui contient  $N_s$  éléments.

La Figure 2.1 présente l'histogramme des scores MSP d'un signal. Puisqu'on observe relativement peu de source avec un score élevé, la MSP permet d'identifier quelques sources fortement liées aux mesures.

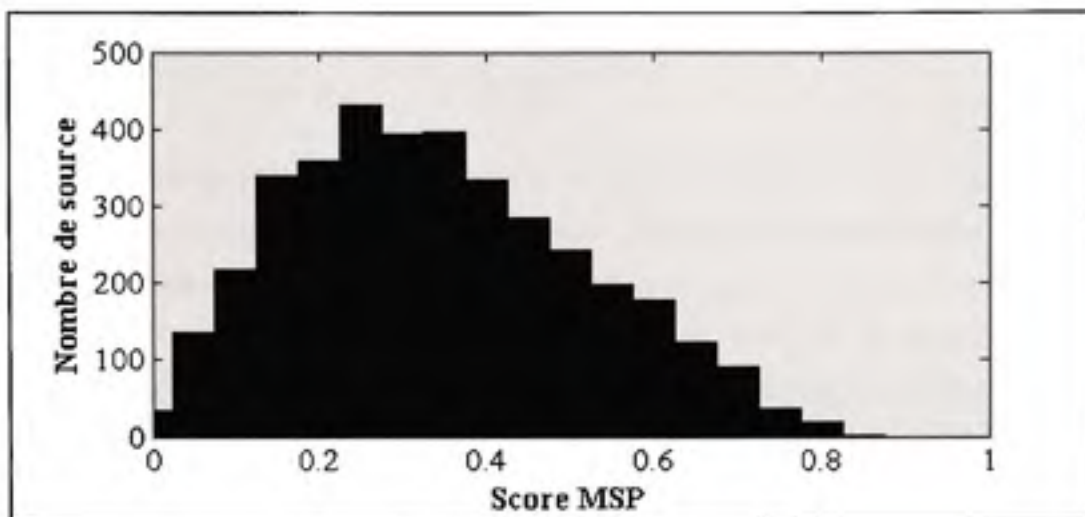


Figure 2.1 Histogramme des scores MSP.

## 2.2 Classification des pointes

La prochaine consiste à associer un numéro de groupe à chaque point en utilisant la représentation des mesures dans l'espace des sources obtenues à l'aide de la MSP. L'algorithme de classification utilisé dans ce mémoire est le *k-means*. Ses paramètres d'entrées sont une liste de  $X$  vecteurs à  $V$  éléments chacun et un nombre de groupe  $N$ . Dans cet algorithme, chaque vecteur est considéré comme un point dans un espace à  $V$  dimensions. L'étape d'initialisation consiste à placer  $N$  centres de groupe dans l'espace. Chaque point est ensuite associé à un seul groupe, celui dont le centre est le plus proche selon la distance euclidienne. La position du centre de chaque groupe est modifiée pour correspondre à la position moyenne des points appartenant au groupe. L'appartenance de chaque point à un groupe et la position du centre des groupes sont recalculées jusqu'à ce que les points ne changent plus de groupe. D'une manière plus mathématique, si l'on définit  $X$  vecteurs  $v_i$  appartenant chacun à un groupe  $S_j$  ayant son centre à la position  $c_j$  alors le *k-means* revient à trouver les groupes  $S_j$  qui minimise  $d$  :

$$d = \sum_{i=1}^N \sum_{v_i \in S_j} \|v_i - c_j\|^2 \quad (2.10)$$

La solution trouvée dépend évidemment de la position des centres de chaque groupe, ce qui rend le *k-means* sensible aux erreurs d'initialisation. De plus, il est possible que l'algorithme trouve un minimum local de la fonction  $d$ . Le risque que ce scénario se produise est principalement une fonction du nombre de dimensions du problème de classification, soit de  $V$ . Par contre, la vitesse et la simplicité de cet algorithme justifient qu'on l'utilise.

Les vecteurs de scores MSP obtenus dans la Section 2.1 peuvent être classifiés directement à l'aide du *k-means*. Cependant, leur grande taille augmenterait le risque que la fonction de classification trouve un minimum local. La solution proposée pour diminuer le nombre de dimensions du problème est de choisir quelques régions du cerveau pour représenter chaque pointe. Ici, une région correspond à un ensemble de sources qui sont voisines les unes des autres. La proximité entre deux sources est fonction de leur distance géodésique : deux sources sont voisines si elles sont reliées par une arête du modèle<sup>1</sup>.

Puisque les scores MSP peuvent être interprétés comme la probabilité qu'une source participe à la solution, les plus grandes valeurs du vecteur de scores correspondent à des sources très probables que l'on choisit pour créer des régions d'activation. Les 10 sources ayant les plus hauts scores et leurs sources voisines sont ainsi choisies pour résumer une pointe. Le nombre 10 utilisé dans cet algorithme a été déterminé empiriquement, il n'a aucune signification physique et il est discutable. La quantité d'information requise pour synthétiser une fenêtre de mesures est ainsi dramatiquement diminuée. Le nombre de valeurs requises passe de  $N_s$ , qui vaut entre 6000 et 10000, à 10 régions soit environ 50 sources.

Pour mesurer le niveau de similitude entre deux pointes, on propose de compter le nombre de sources que celles-ci partagent. Cette mesure est utilisée comme critères de classification. Un vecteur  $V_i$  est construit pour chaque pointe de telle sorte que  $V_i(j)$  contiennent le nombre de

<sup>1</sup> Les arêtes sont les liens qui existent entre les sources pour former la surface matière blanche/matière grise.

sources que les pointes  $i$  et  $j$  ont en commun. Ce sont ces vecteurs qui seront utilisés comme paramètres d'entrée pour le *k-means*.

Appliqué à un ensemble de pointes représentées par une fenêtre de mesures, cet algorithme attribue une étiquette à chaque pointe. Si deux pointes portent la même étiquette, elles sont dans le même groupe.

Toute la méthodologie présentée dans cette section a été implémentée sous la forme d'une librairie MATLAB nommée *BrainEntropy*. Les détails de conception sont présentés dans l'annexe I à la page 70.

## CHAPITRE 3

### ÉVALUATION DE LA PERFORMANCE ET SIMULATION DE SIGNAUX

Avant d'appliquer l'algorithme proposé sur de vrais signaux, il faut d'abord en évaluer la performance. Étant donné que la vraie classification des signaux épileptiques n'est jamais connue, ceux-ci ne peuvent pas être utilisés pour une validation. Une méthodologie a donc été développée pour simuler des signaux dont la classe est connue. Les signaux simulés sont idéals pour la phase de test puisqu'ils permettent d'évaluer de façon objective la performance de l'algorithme. De plus, des mesures générées avec des niveaux de variables permettent de tester les algorithmes sur des problèmes faciles à résoudre, puis d'augmenter la difficulté pour trouver les limites de leur performance.

Tout le code qui a été utilisé pour simuler et classifier les signaux présentés dans cette section est offert dans le paquet *BrainEntropy*. Une description du contenu de ce paquet est disponible dans l'annexe I à la page 70. Le fichier « `tutorial_classification.m` » présente les différentes étapes du processus de classification.

Les deux premières sections de ce chapitre présentent la méthodologie utilisée pour générer des mesures semblables à celles enregistrées en épilepsie. La section suivante présente la performance de l'algorithme sur des signaux EEG et MEG simulés. La dernière section présente une approche simple pour combiner les classifications obtenues sur des mesures simultanées EEG et MEG.

#### 3.1 Simulation de signaux cérébraux simple

Les mesures les plus simples qui peuvent être créées respectent l'équation suivante:

$$M = GJ \tag{3.1}$$

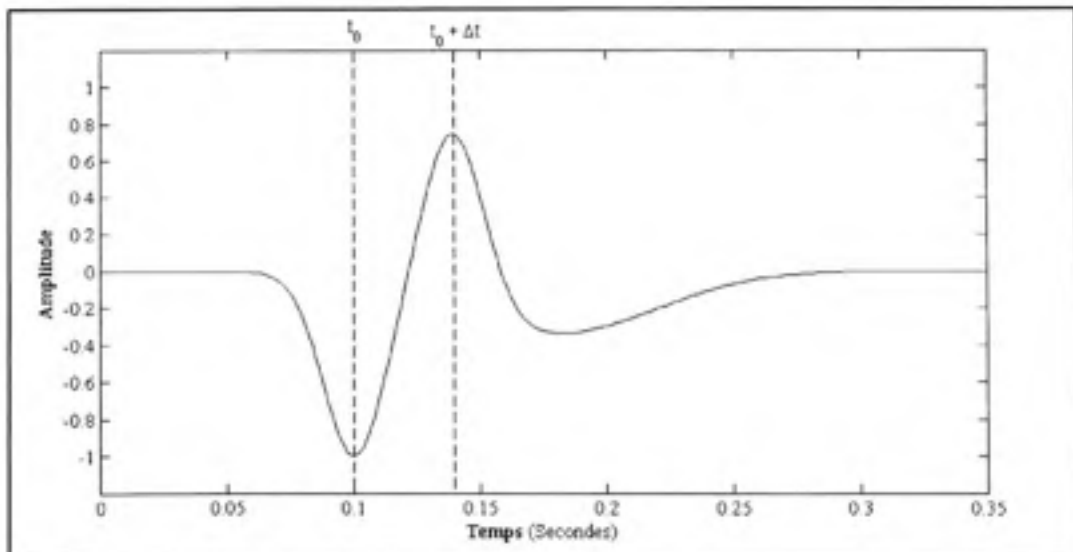
où  $G$  est une matrice de gain connue. C'est la version sans bruit de l'équation (1.1) du problème direct. Ici, les mesures générées à l'aide de cette équation sont simples:  $J$  ne contient que le signal qui génère une pointe.

Il n'est pas nécessaire que les signaux artificiels aient toutes les caractéristiques des vraies mesures EEG ou MEG. Par contre, une certaine similitude devrait être respectée. Les pointes épileptiques EEG sont généralement composées d'une oscillation rapide suivie d'une oscillation plus lente. Pour obtenir cette forme d'onde on considère un signal de source composé d'une somme de trois fonctions gaussiennes comme l'illustre les équations (3.2) et (3.3).

$$g(t, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) \quad (3.2)$$

$$p(t, t_0, \Delta t) = g(t, t_0, \alpha\Delta t) - g(t, \Delta t + t_0, \alpha\Delta t) + \gamma \cdot g(t, \beta\Delta t + t_0, \Delta t) \quad (3.3)$$

Dans les simulations de ce travail,  $\alpha$ ,  $\beta$  et  $\gamma$  sont respectivement fixés à 0.3, 2 et 1.2. La Figure 3.1 présente la forme générale du signal qui en résulte.



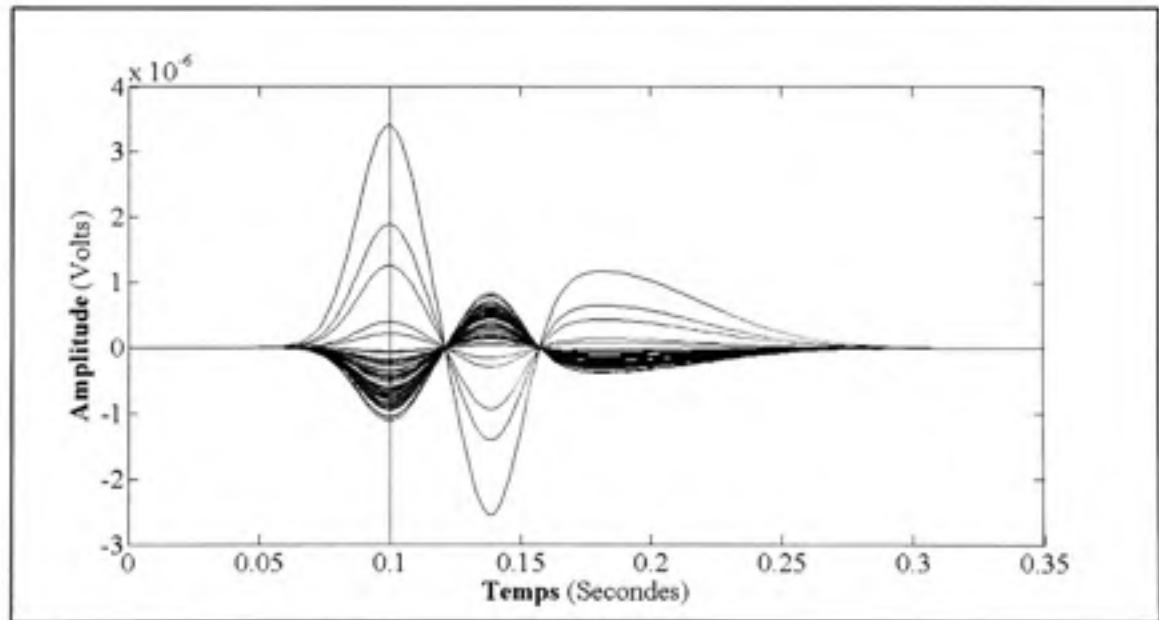
**Figure 3.1 Générateur de pointe épileptique simulée à l'aide de trois fonctions gaussiennes.**

La fonction  $p$  modélise à l'activation d'une source du modèle qui génère la pointe. Dans une simulation simple, les autres sources du modèle ne participent pas aux mesures. L'activation des sources  $J$  de l'équation (3.1) est donc obtenue à l'aide de :

$$J_i(t) = \begin{cases} p(t, t_0, \Delta t) & \text{si } i \in k \\ 0 & \text{si } i \notin k \end{cases} \quad (3.4)$$

où  $J_i$  correspond à l'activation de la source  $i$  et  $k$  correspond aux numéros des sources qui génèrent la pointe. Dans cette équation,  $k$ ,  $t_0$  et  $\Delta t$  sont des paramètres fixés.

Étant donné une configuration de sources actives, les mesures obtenues en faisant le produit de la matrice de gain avec le  $J$  simulé. La Figure 3.2 présente un exemple de mesures EEG obtenues avec une source dont l'activité est représentée à la Figure 3.1.



**Figure 3.2 Un exemple d'une pointe EEG simulé à l'aide d'une source unique.**

Quoique pertinentes, les mesures résultantes sont loin de la réalité. En utilisant une seule source, la zone du cortex générant la pointe est beaucoup trop petite puisqu'ici une source représente une surface de moins de  $1 \text{ cm}^2$ . Afin de rendre les mesures plus réalistes, il est



important de distribuer l'activité sur une plus grande zone du cortex. Pour atteindre cet objectif, les voisins de la source associée à une pointe seront eux aussi activés. Le concept de voisin est identique à celui présenté dans le Chapitre 2. Les voisins sont en effet choisis en fonction du maillage du modèle et non de leurs distances.

En supposant que la pointe ait une source focale, les sources voisines de la source activée auront une activité plus faible, pondérée par un coefficient  $\varphi$  entre 0 et 1. Il est donc possible d'améliorer la simulation en utilisant une nouvelle définition de  $J$  :

$$J_i(t) = \begin{cases} p(t, t_0, \Delta t) & \text{si } i \in k \\ \varphi(i) \cdot p(t, t_0, \Delta t) & \text{si } i \text{ est voisin d'un élément de } k \\ 0 & \text{sinon} \end{cases} \quad (3.5)$$

### 3.2 Simulation de signaux cérébraux avec de l'activité physiologique

L'activité électrique du cerveau mesurée chez le patient ne contient pas uniquement le signal de pointes épileptiques. Elle contiendra aussi l'activité physiologique du cerveau qui englobe toute l'activité cérébrale qui n'est pas reliée à l'épilepsie. Par contre, elle n'inclut pas les artéfacts causés par la respiration et le rythme cardiaque. La prochaine étape dans l'amélioration de la simulation des mesures est d'inclure de l'activité physiologique à la simulation.

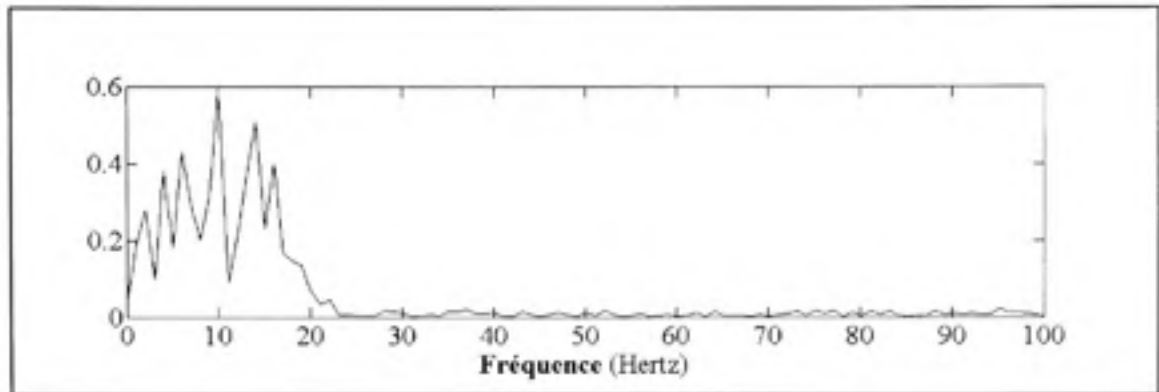
On suppose ici que l'activité physiologique du cerveau se situe normalement entre 8 et 12 Hz (activité alpha). Pour simuler cette composante, on utilise un spectre gaussien de moyenne 5 Hz et d'écart type 5 Hz, représenté par l'équation (3.6).

$$g(f) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{(f-5)^2}{50}} \quad (3.6)$$

Le spectre est ensuite modifié à l'aide de :

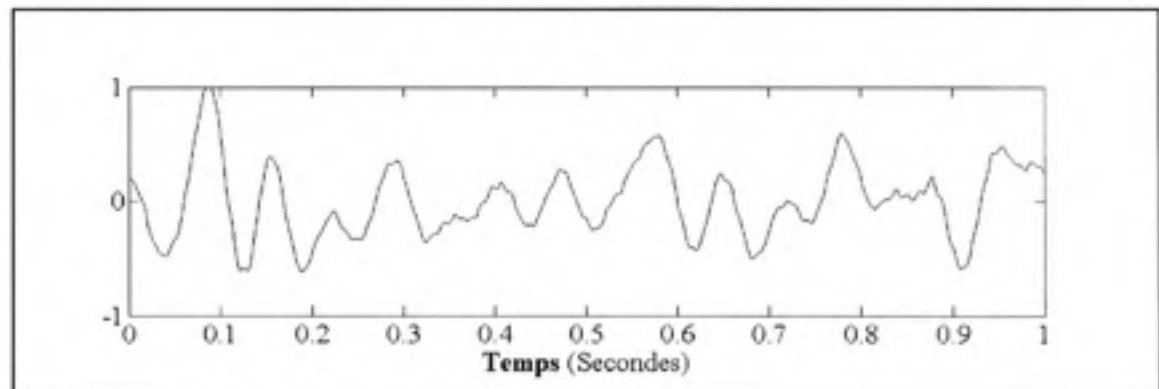
$$S(s) = B(f) \cdot g(f) + \xi(f) \quad (3.7)$$

où  $B(f)$  et  $\xi(f)$  sont des bruits blancs. Le résultat obtenu est présenté à la Figure 3.3.



**Figure 3.3 Spectre bruitée de l'activité physiologique.**

Lorsqu'une transformée de Fourier inverse est appliquée à la fonction présentée plus haut, le signal de la Figure 3.4 est obtenu.

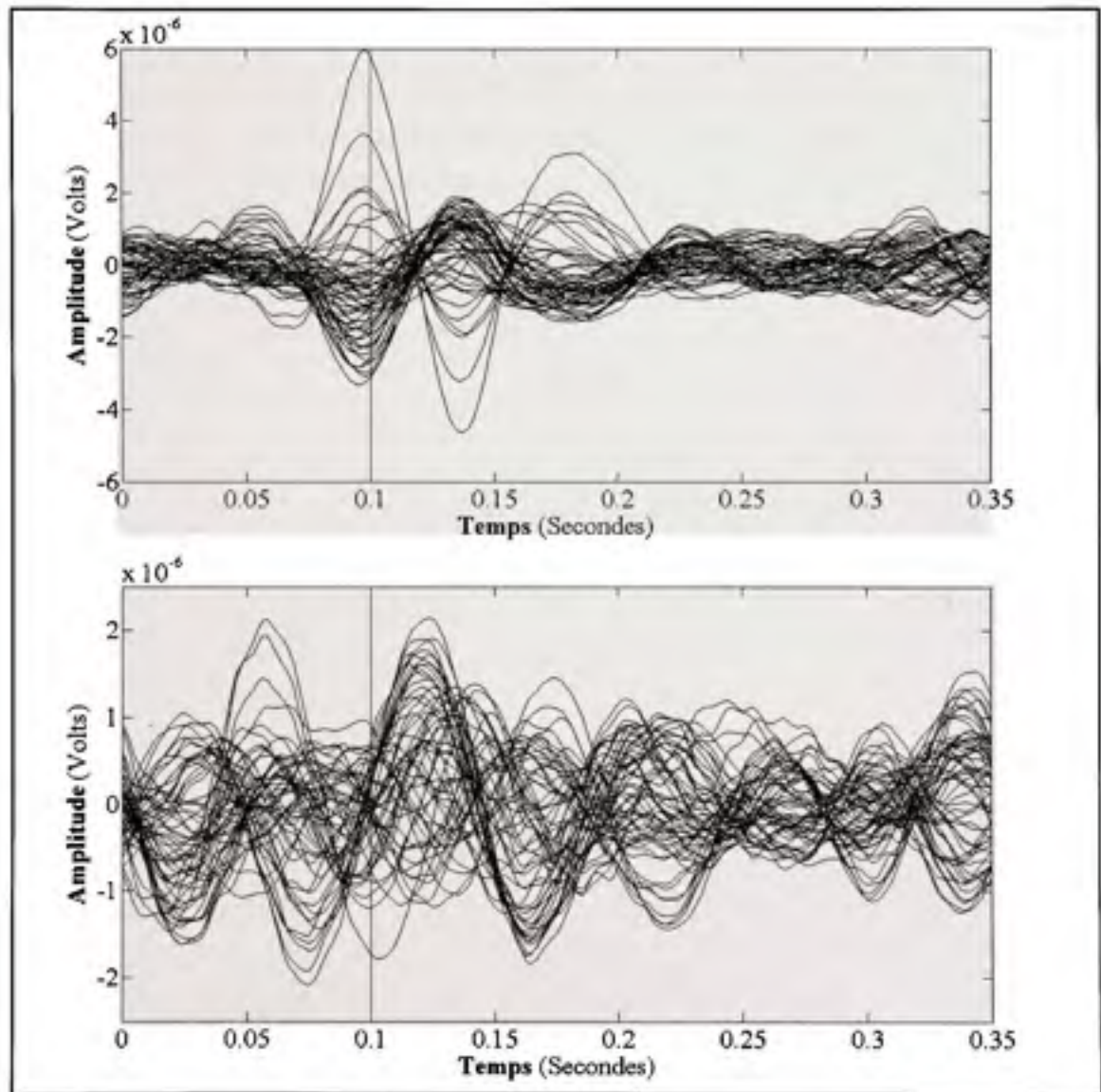


**Figure 3.4 Le signal obtenu lorsqu'une transformée de Fourier inverse est appliquée au spectre de la Figure 3.3.**

L'intérêt d'utiliser cette technique pour la génération d'activité physiologique est que le signal final contient principalement de l'activité entre 8 et 12 Hz, mais aussi un bruit aléatoire généré par les bruits blancs.

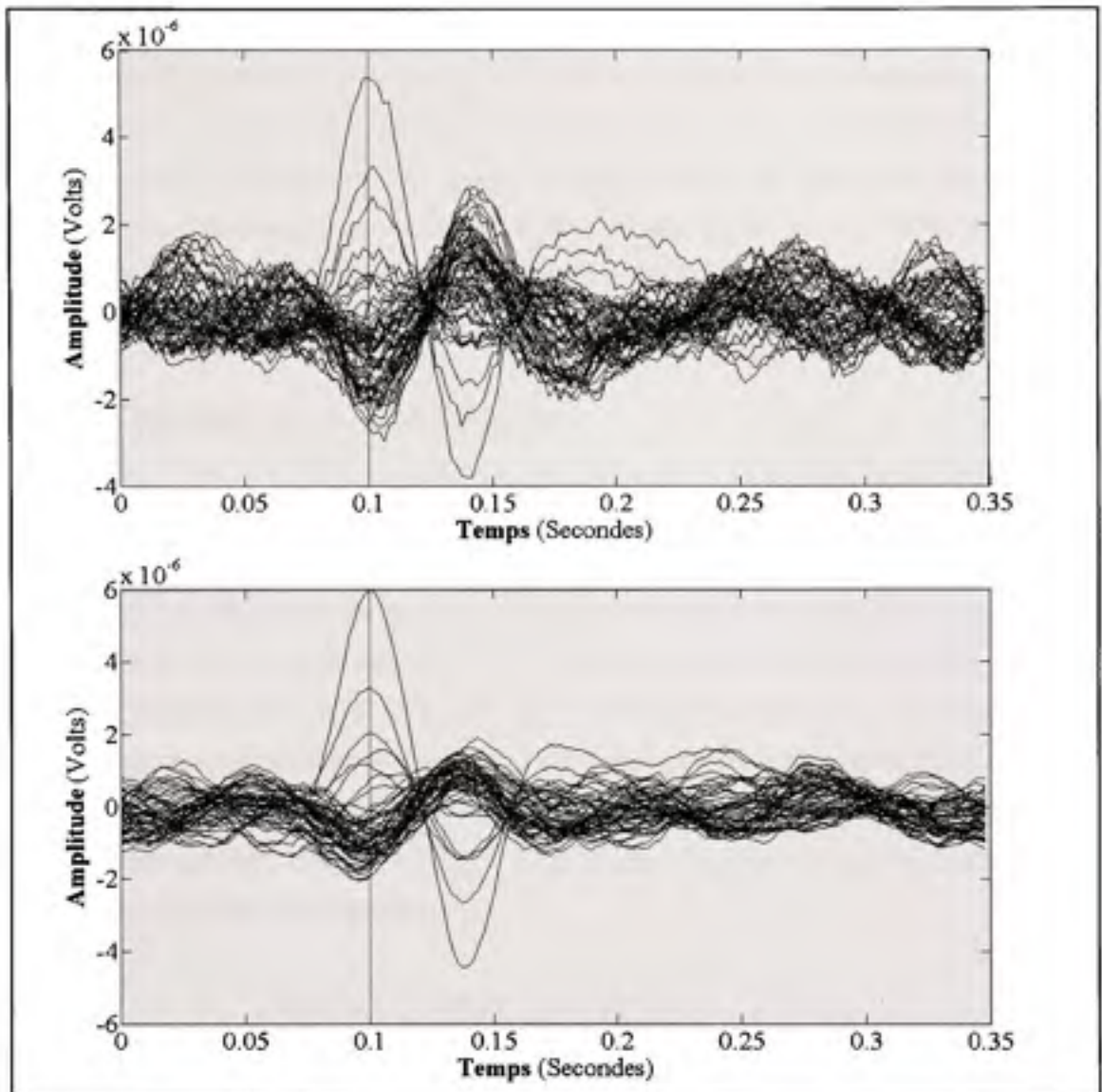
À partir de l'activité épileptique et l'activité physiologique simulées, on considère des signaux combinant la puissance de chacune des composantes pour obtenir une très grande variété de signaux. La Figure 3.5 présente deux pointes EEG contenant de l'activité

physiologique et de l'activité épileptique. Dans le graphique du haut, le ratio entre l'amplitude de la pointe et l'activité physiologique est de 2. Dans le graphique du bas, le ratio d'amplitude est de 6. Dans les deux graphiques, la ligne verticale indique la position de la première oscillation de la pointe.



**Figure 3.5** Deux exemples de signaux EEG simulés contenant de l'activité physiologique en plus d'une pointe. La pointe du graphique du haut a une amplitude de 2 et celle du bas a une amplitude de 6.

Ici, les mesures obtenues sont réalistes, mais parfaitement expliquées par la matrice de gain. En réalité, plusieurs phénomènes extérieurs ont une influence sur les mesures obtenues. On note entre autres : le clignement des yeux, les mouvements du patient, les imprécisions du système de mesures, les interférences magnétiques et électriques, la différence d'impédance entre les électrodes. Tous ces phénomènes sont modélisés dans les mesures simulées en leur ajoutant un bruit blanc  $\varepsilon$ . On obtient ainsi l'équation (1.1) du problème direct. Encore une fois, la puissance du bruit peut être gérée pour obtenir différents signaux. La Figure 3.6 présente deux signaux contenant du bruit de mesures. Le signal du haut à un rapport signal sur bruit (SNR) de 10 dB et celui du bas est de 15 dB.



**Figure 3.6** Deux exemples de signaux EEG simulés contenant de l'activité physiologique en plus d'une pointe d'amplitude 6. Le signal du haut a un SNR de 10 dB. Le signal du bas a un SNR de 15 dB.

### 3.3 Simulation de signaux EEG et MEG mesurés simultanément

La méthodologie présentée dans les sections précédentes permet de simuler des mesures EEG ou MEG. Il suffit de changer la matrice de gain pour changer de modalité. Cependant, il peut être intéressant de simuler les signaux enregistrés simultanément en EEG et MEG. Dans cette





Toutefois, il est peu probable que la classification obtenue soit parfaite, l'algorithme mélangera probablement quelques étiquettes. Le Tableau 3.2 présente le résultat d'une classification avec cinq erreurs affichées en caractère gras. La performance de l'algorithme pour une classification spécifique peut être quantifiée à l'aide de la variance moyenne des étiquettes des groupes.

$$\frac{1}{N_g} \sum_{i=1}^{N_g} \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2 \quad (3.8)$$

Par exemple, la classification idéale présentée dans le Tableau 3.1 obtient une variance moyenne de 0 alors que celle présentée dans le Tableau 3.2 obtient une variance moyenne de 0.185. La pire variance qui peut être obtenue pour une classification à 2 groupes est 0.25. Cette valeur est obtenue quand les étiquettes sont réparties également à travers les groupes.

Tableau 3.2 Exemple d'une classification avec quelques erreurs

<b>Numéro de pointe</b>	1	2	3	4	5	6	7	8	9	10
<b>Générateur</b>	A	A	A	A	A	A	A	A	A	A
<b>Numéro de groupe</b>	1	2	1	1	2	1	1	1	1	2
<b>Numéro de pointe</b>	11	12	13	14	15	16	17	18	19	20
<b>Générateur</b>	B	B	B	B	B	B	B	B	B	B
<b>Numéro de groupe</b>	2	2	1	2	2	2	1	2	2	2

La performance de l'algorithme a été évaluée en fonction de 3 paramètres : l'amplitude de la pointe par rapport à l'activité physiologique, le SNR et la distance entre les générateurs. Les valeurs de l'amplitude testées sont 2, 4, 6, 8 et 10. Trois SNR ont été utilisés soit 10, 20 et 30 dB. La distance entre les générateurs est de 2, 4, 6, 8, et 10 niveau de voisins. Pour chaque combinaison des trois paramètres, 10 simulations ont été effectuées en choisissant les générateurs au hasard, mais en respectant le critère de distance. On obtient donc un total de 750 jeux de données simulés conjointement en EEG et MEG. Pour chaque simulation,



l'algorithme de classification est appliqué aux mesures EEG et MEG indépendamment. La variance moyenne des groupes est calculée et moyennée pour les 10 simulations d'un jeu de paramètres.

Pour toutes les simulations, le modèle utilisé pour la génération de signaux possède 6013 sources disposées sur une surface représentant un cortex réel. Ils possèdent 56 capteurs EEG et 272 capteurs MEG. Comme il a été indiqué dans le Chapitre 2, la matrice de gain a été obtenue à l'aide de la méthode BEM et d'une IRM anatomique.

Afin de déterminer la difficulté du problème de classification posé par ce jeu de donnée, les pointes sont classifiées à l'aide d'un algorithme similaire à celui proposé par Van't Ent (Van't Ent et al., 2003). La seule modification qui a été apportée à l'algorithme est le type d'algorithme de classification. Les numéros de groupe sont obtenus à l'aide d'un algorithme de type *k-means* plutôt que hiérarchique. La Figure 3.7 présente les résultats obtenus sur les signaux EEG avec un SNR de 30 dB.

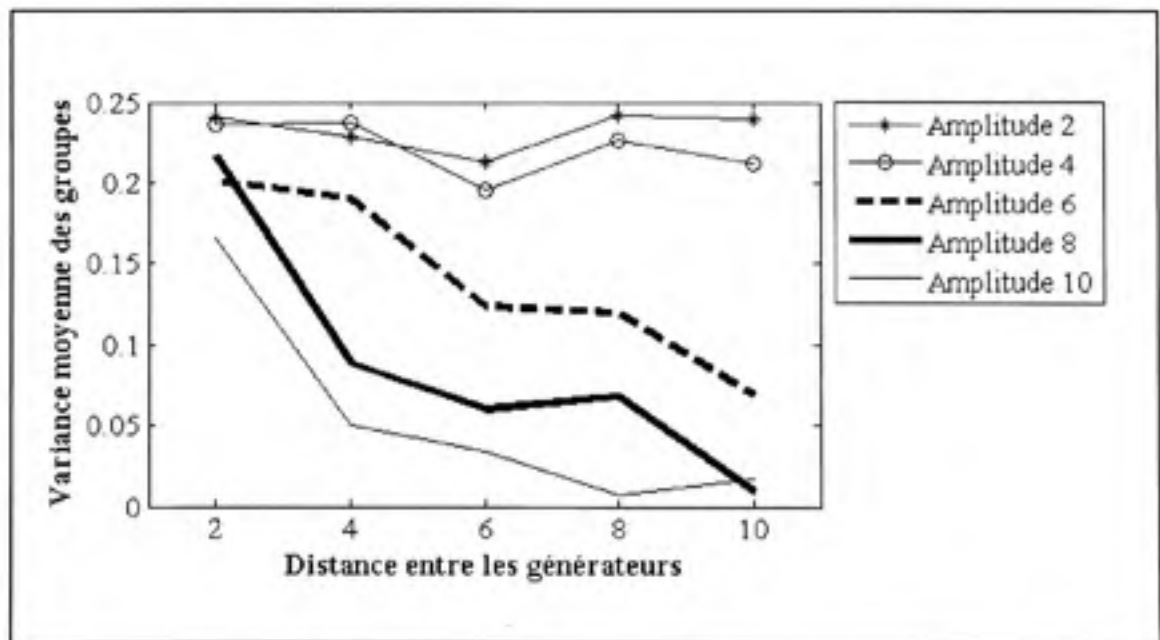


Figure 3.7 Performance de l'algorithme de comparaison sur les mesures EEG possédant un SNR de 30 dB.

Les résultats indiquent que l'algorithme est capable de classer les pointes correctement si l'amplitude de la pointe est supérieure à 8 et si la distance entre les générateurs est supérieure à 6. Cependant, pour des amplitudes ou des distances plus faibles les groupes formés sont aléatoires puisque la variance est proche de 0.25. À titre de comparaison, les pointes présentées à la Figure 1.4 ont une distance de quatre niveaux de voisins entre leurs générateurs et une amplitude de 6. Pour ces paramètres, l'algorithme de Van't Ent modifié obtient une variance moyenne de 1.9.

### 3.4.1 Les résultats EEG

Les résultats en EEG obtenue à l'aide de l'algorithme proposé dans ce mémoire sont présentés dans la Figure 3.8, la Figure 3.9 et la Figure 3.10. Chaque graphique représente un niveau de bruit spécifique. Les différentes courbes d'un graphique illustrent la performance pour une amplitude de pointe spécifique. Les ratios d'amplitudes 6 et 8 sont particulièrement intéressants parce qu'ils correspondent à l'amplitude des pointes épileptiques des jeux de données réelles.

En comparant les trois graphiques, on observe que le niveau de bruit des signaux a relativement peu d'impact sur la classification. Cette conclusion n'est pas surprenante puisque la MSP filtre les mesures en éliminant l'information qui n'est pas corrélée avec la matrice de gain, soit le bruit additif. En comparant les courbes d'un graphique, on note que le ratio d'amplitude de la pointe par rapport à la physiologique joue un rôle important dans la classification. Pour obtenir de bon résultat, il faut un ratio d'amplitude d'au moins 4, mais 6 est idéal. Encore une fois, cette conclusion était prévisible. Plus l'activité physiologique est forte par rapport au signal générant la pointe, plus les scores MSP tenteront d'expliquer l'activité physiologique. Parce que l'activité physiologique est générée aléatoirement, les sources choisies pour chaque pointe seront aussi aléatoires. Finalement, on observe que la performance de l'algorithme s'améliore quand la distance entre les générateurs augmente. Cette situation était prévisible puisqu'il est possible que deux générateurs qui sont très proches l'un de l'autre partagent des sources. En bref, l'algorithme proposé offre des

performances idéales lorsque les générateurs de pointes sont à une distance de quatre voisins ou plus et que la pointe a un ratio d'amplitude d'au moins quatre.

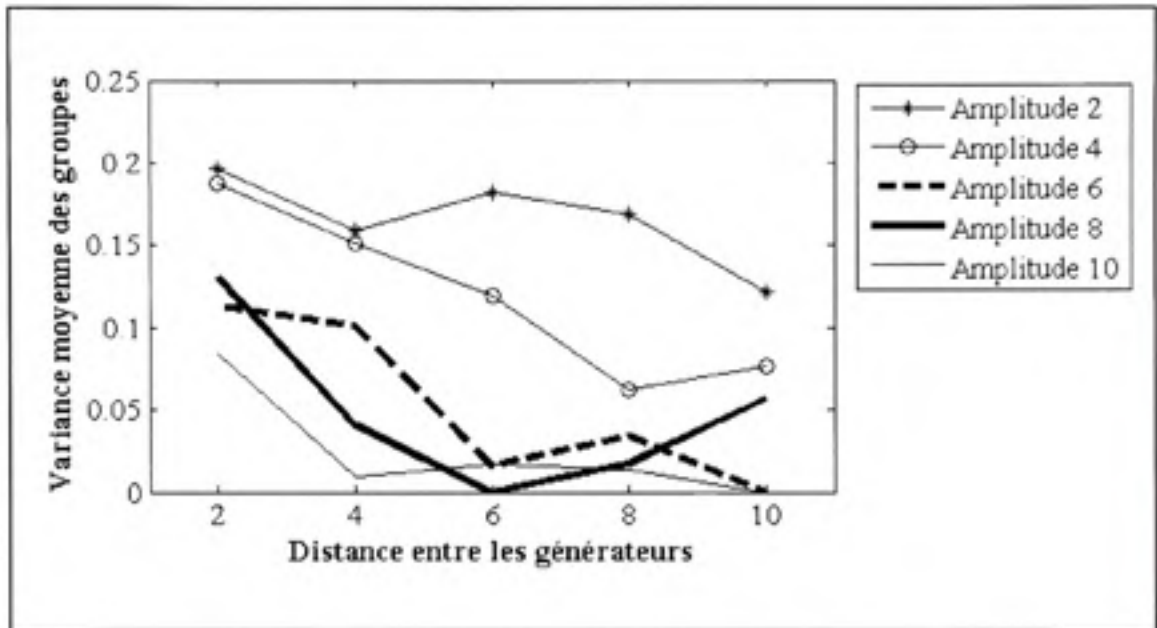


Figure 3.8 Performance de la classification EEG pour un SNR de 10 dB.

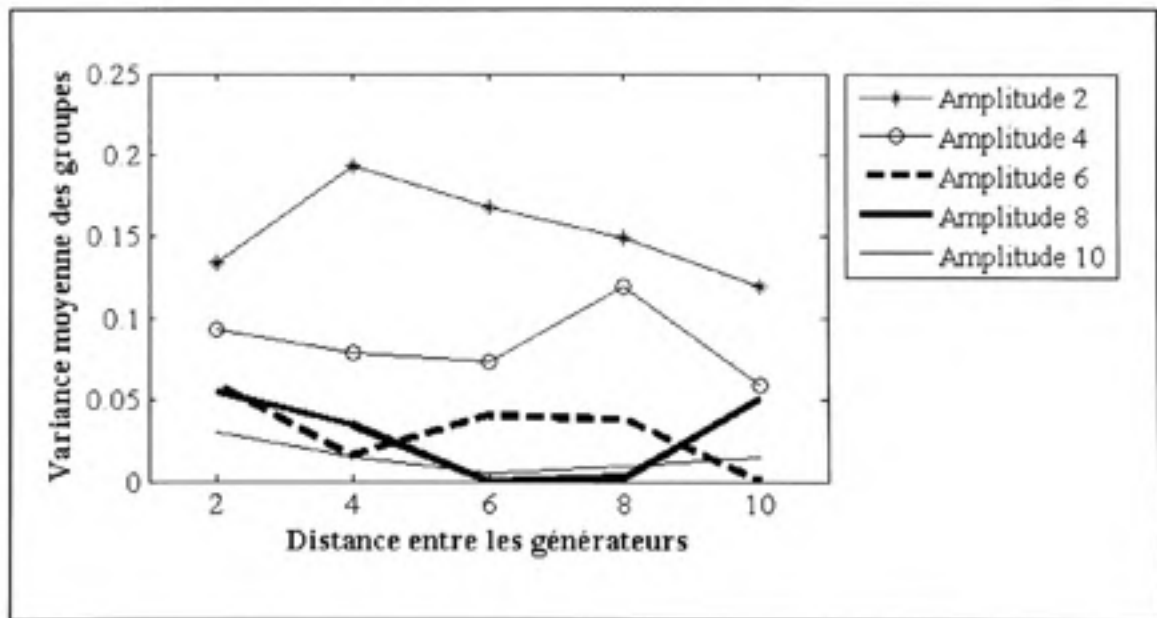


Figure 3.9 Performance de la classification EEG pour un SNR de 20 dB.

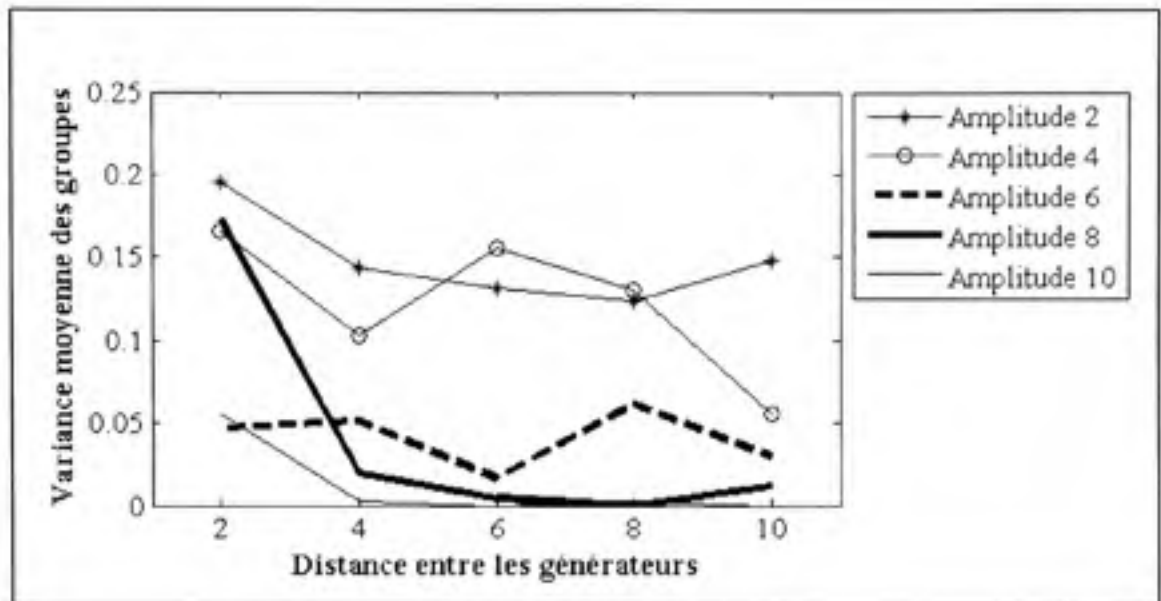


Figure 3.10 Performance de la classification EEG pour un SNR de 30 dB.

#### 3.4.2 Les résultats MEG

Les résultats en MEG obtenus sont présentés dans la Figure 3.11, la Figure 3.12 et Figure 3.13. Comme pour les résultats EEG, chaque graphique représente un niveau de bruit spécifique et les différentes courbes d'un graphique illustrent la performance pour une amplitude de pointe spécifique.

Contrairement à la classification EEG, le niveau de bruit a un impact important sur la performance de la classification MEG. La performance est nettement inférieure pour un SNR de 10 dB que pour 20 et 30 dB. Encore une fois, l'amplitude de la pointe et la distance entre les générateurs affectent les résultats d'une manière significative. L'augmentation de la variance lorsque la distance entre les générateurs augmente peut être expliquée par la position des générateurs. Si les générateurs, choisis aléatoirement, sont positionnés sur la partie ventrale du cerveau, la classification sera difficile, peu importe le ratio d'amplitude ou la distance entre les générateurs. La conclusion est pratiquement la même pour la classification EEG : l'algorithme offre de très bonnes performances si la pointe à un ratio d'amplitude d'au moins quatre et que les générateurs sont à une distance d'au moins quatre voisins.

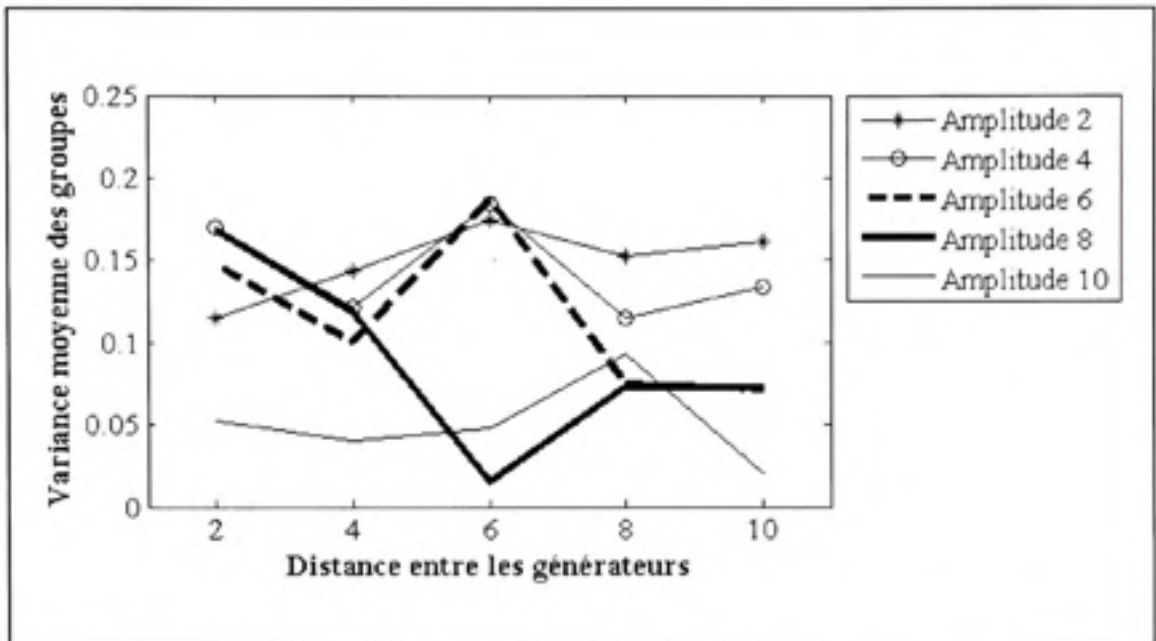


Figure 3.11 Performance de la classification MEG pour un SNR de 10 dB.

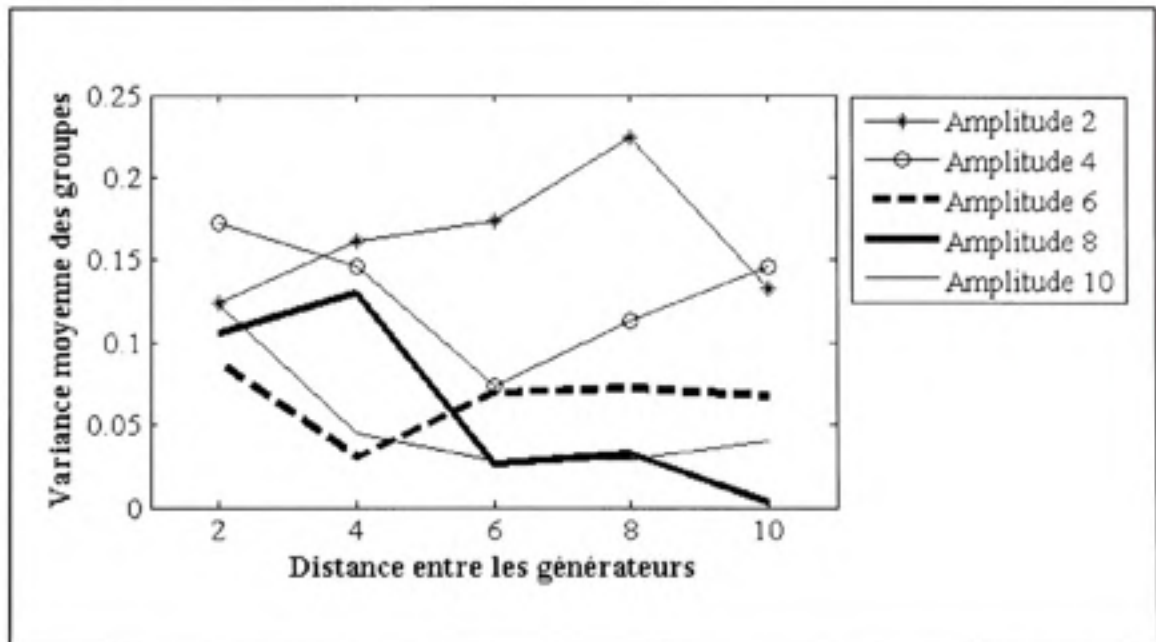


Figure 3.12 Performance de la classification MEG pour un SNR de 20 dB.

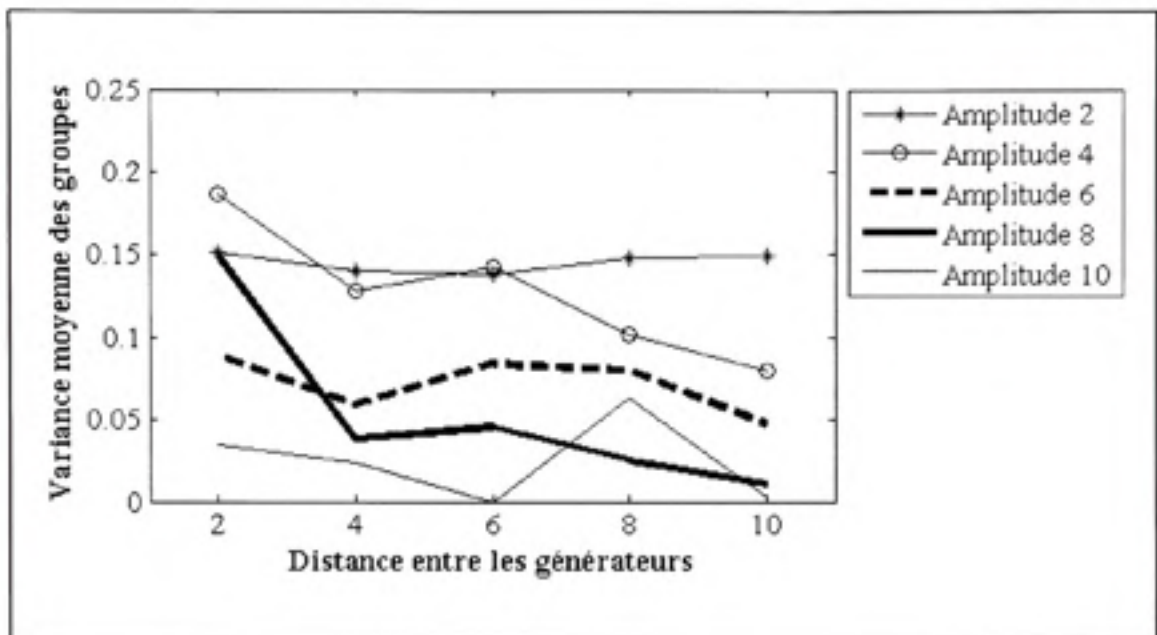


Figure 3.13 Performance de la classification MEG pour un SNR de 30 dB.

### 3.5 Classification combinée EEG et MEG

Toutes les pointes traitées dans ce mémoire, qu'elles soient simulées ou réelles, ont été enregistrées à l'aide de deux modalités simultanément, l'EEG et la MEG. Il est donc intéressant de combiner les deux classifications obtenues. Cette combinaison devrait permettre d'améliorer la performance de l'algorithme.

Par exemple, le Tableau 3.3 présente les résultats pour une classification effectuée indépendamment en EEG et MEG avec quelques erreurs de classification.

Tableau 3.3 Un exemple de résultats pour une classification EEG et MEG indépendante

<b>Numéro de pointe</b>	1	2	3	4	5	6	7	8	9	10
<b>Générateur</b>	A	A	A	A	A	A	A	A	A	A
<b>Numéro de groupe EEG</b>	1	2	2	2	2	2	1	2	2	2
<b>Numéro de groupe MEG</b>	1	1	1	2	1	1	1	1	1	1
<b>Numéro de pointe</b>	11	12	13	14	15	16	17	18	19	20
<b>Générateur</b>	B	B	B	B	B	B	B	B	B	B
<b>Numéro de groupe EEG</b>	1	1	1	1	1	1	1	2	1	1
<b># de groupe MEG</b>	2	2	2	2	1	2	2	2	2	2

Si une seule des deux modalités est prise en compte, les erreurs se glisseront dans la classification. Par contre, le nombre d'erreurs peut être diminuées en réorganisant l'information un nouveau tableau. Les pointes portant l'étiquette  $i$  en EEG et  $j$  en MEG sont placées dans la cellule  $(i, j)$  du tableau. Le Tableau 3.4 présente le résultat obtenu.

Tableau 3.4 Classification des pointes EEG et MEG simultanée

		MEG									
EEG	1	7 15	11 12 13 14 16 17 19 20								
	2	3 5 6 8 9 10	4 18								

La classification EEG a permis de filtrer les erreurs de la classification MEG et vice versa. Les pointes de la cellule (1,2) proviennent toutes du générateur A et les pointes de la cellule (2,1) proviennent toutes du générateur B. Les deux cellules (1,1) et (2,2) contiennent les erreurs de classification qui sont considérées comme des valeurs aberrantes.



## CHAPITRE 4

### CLASSIFICATION DE POINTES ÉPILEPTIQUES RÉELLES

Lors de l'évaluation de la performance de l'algorithme de classification, celui-ci a été traité comme une boîte noire. Cette approche était acceptable parce que la classification idéale était connue et la qualité des résultats pouvait être évaluée à posteriori. Toutefois, lorsque l'on tente de trier des pointes réelles, la vraie solution n'est jamais disponible. Il est donc important d'observer les résultats obtenus après chaque étape de la classification afin d'en évaluer l'intérêt.

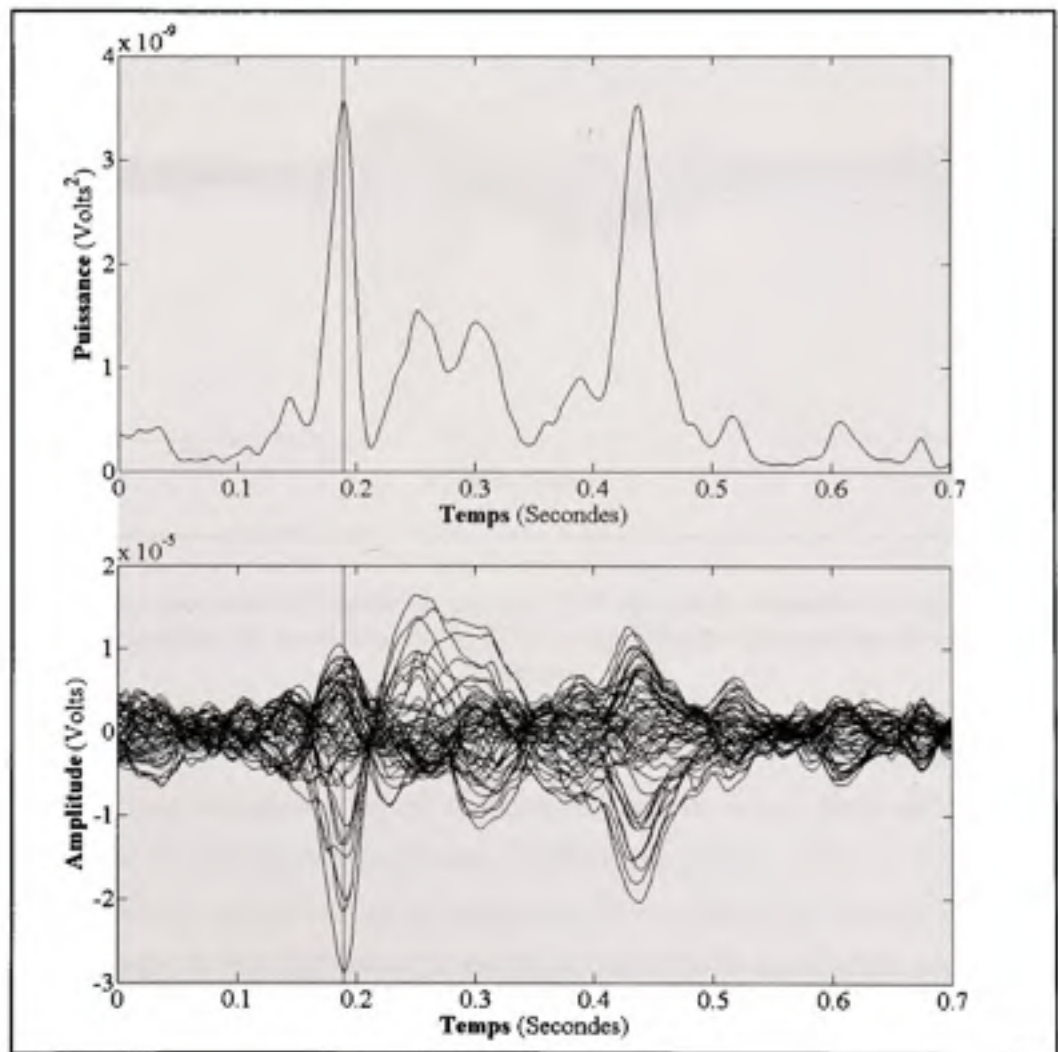
La première section de ce chapitre présente les résultats obtenus lorsque la méthodologie est appliquée à des pointes épileptiques mesurées en EEG. La deuxième section présente les résultats obtenus pour les mêmes pointes mesurées en MEG. La dernière section présente la classification obtenue lorsque les classifications EEG et MEG sont combinées.

#### 4.1 Classification de pointes épileptiques réelles en EEG

Le jeu de données traité dans ce mémoire contient 86 pointes enregistrées simultanément en EEG et en MEG. Cependant, les pointes ont été marquées par un neurologue en observant uniquement les mesures EEG. Lors de l'analyse habituelle, la résolution du problème inverse est effectuée sur la moyenne de toutes les pointes. Parce que les pointes ont été marquées à la main, il faut s'assurer que les pointes sont bien recalées avant de calculer la moyenne. Le recalage est essentiel, parce que si les signaux utilisés lors du calcul de la moyenne sont décalés, alors le SNR ne sera pas amélioré. Il faut donc replacer les marqueurs de chaque pointe à une position plus objective que celle identifiée visuellement par les neurologues. Le critère de recalage choisi est le point d'amplitude maximum de la fonction de puissance globale (FPG) sur une petite fenêtre autour du marqueur original. En définissant  $s_i(t)$  le signal du capteur  $i$ , la FPG est alors donnée par :

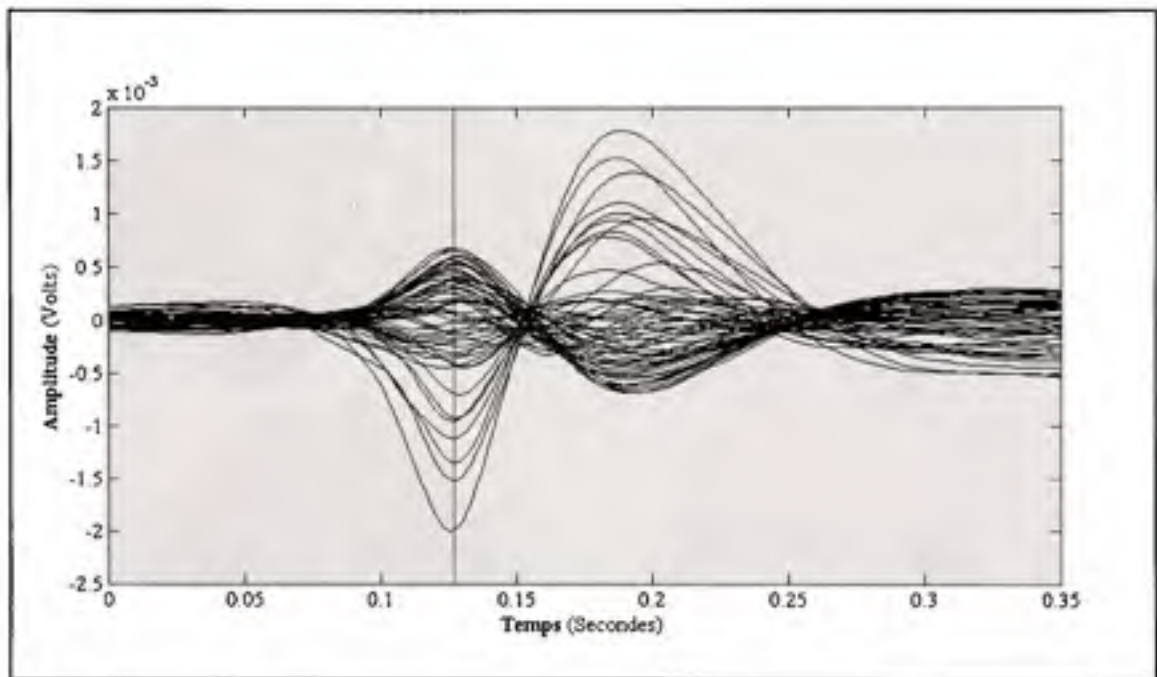
$$f(t) = \sum_{i=1}^{N_c} s_i(t)^2 \quad (4.1)$$

La Figure 4.1 présente la position du maximum de la FPG, à l'aide d'une ligne verticale, pour une fenêtre de 40 échantillons autour du marqueur original (haut). La position du marqueur ajouté par les neurologues n'est pas illustrée dans la figure, mais était 0.20 seconde. La figure présente encore la position du marqueur recalé, mais sur les mesures EEG (bas).



**Figure 4.1** L'image du haut montre la position du maximum de la fonction de puissance globale de la première pointe du jeu de données. L'image du bas présente la même position sur les mesures EEG.

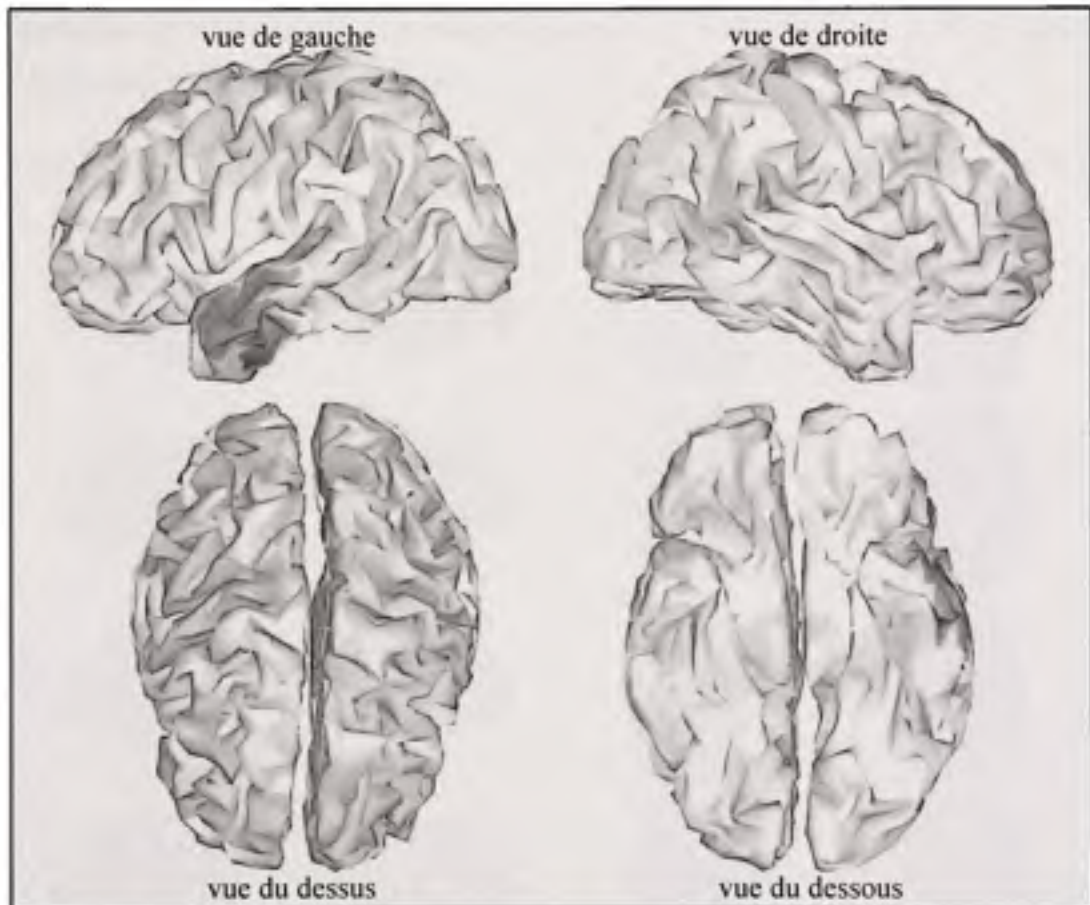
Une fois la position des marqueurs ajustée d'une façon objective, la moyenne des pointes peut être calculée. La moyenne des 86 pointes du jeu de données est présentée dans la Figure 4.2. À titre d'information, la position du maximum de la FPG pour le signal moyen est représentée par la ligne verticale.



**Figure 4.2** La moyenne de toutes les pointes EEG du jeu de données. La ligne verticale présente la position du maximum de la FPG lorsqu'elle est calculée sur la moyenne des mesures.

Puisque le signal de la Figure 4.2 a été obtenu en calculant la moyenne de plusieurs pointes épileptiques, l'activité des sources qui explique ces mesures devrait permettre d'identifier la région du cerveau qui génère les pointes. L'activité des sources qui a été trouvée avec la technique MEM est présentée dans la Figure 4.3. Les vues présentées sont : la vue de droite, la vue de gauche, la vue du dessus et la vue du dessous. Toutes les solutions présentées dans ce mémoire seront affichées de cette manière. Les zones rouges indiquent une activation des sources (une activation positive) alors que les zones bleues indiquent une désactivation (une activation négative). Les zones grises sont celles où il n'y a pas ou peu d'activité. Il faut noter que les résultats illustrés ici représentent un seul échantillon de temps. Cet échantillon

est proche de la ligne verticale de la Figure 4.1, qui correspond au maximum de la FPG, mais pas nécessairement superposé à celle-ci. L'échantillon choisi peut varier puisque la position est choisie manuellement en fonction des résultats observés. La zone d'activation trouvée est donc temporale gauche. Il n'y a pas de désactivation apparente.

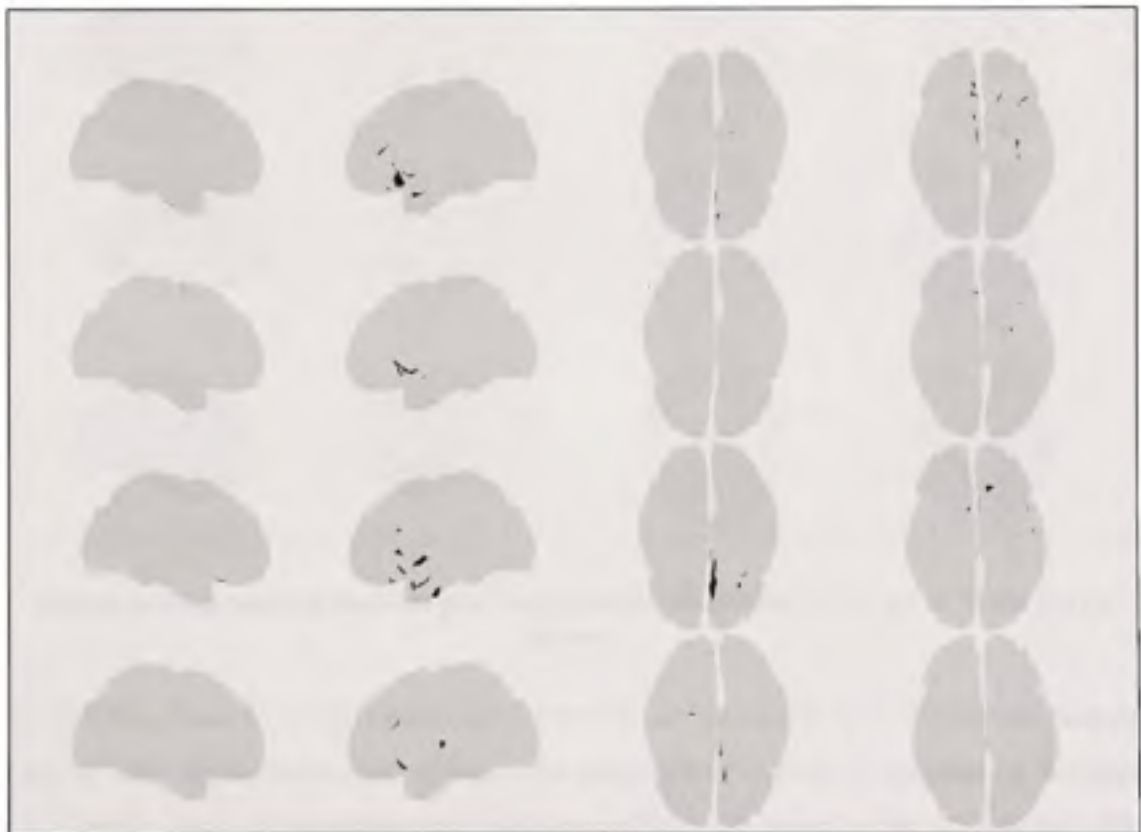


**Figure 4.3** La solution du problème inverse effectué sur la moyenne de toutes les pointes EEG du jeu de données.

La solution trouvée est intéressante, mais est peut-être inexacte. Si dans les 86 pointes moyennées il existe en réalité plusieurs types de pointes, des détails pourraient avoir été perdus à cause de la moyenne. Les pointes seront donc classifiées afin de voir si la résolution du problème inverse sur chaque groupe présentera des activations différentes.



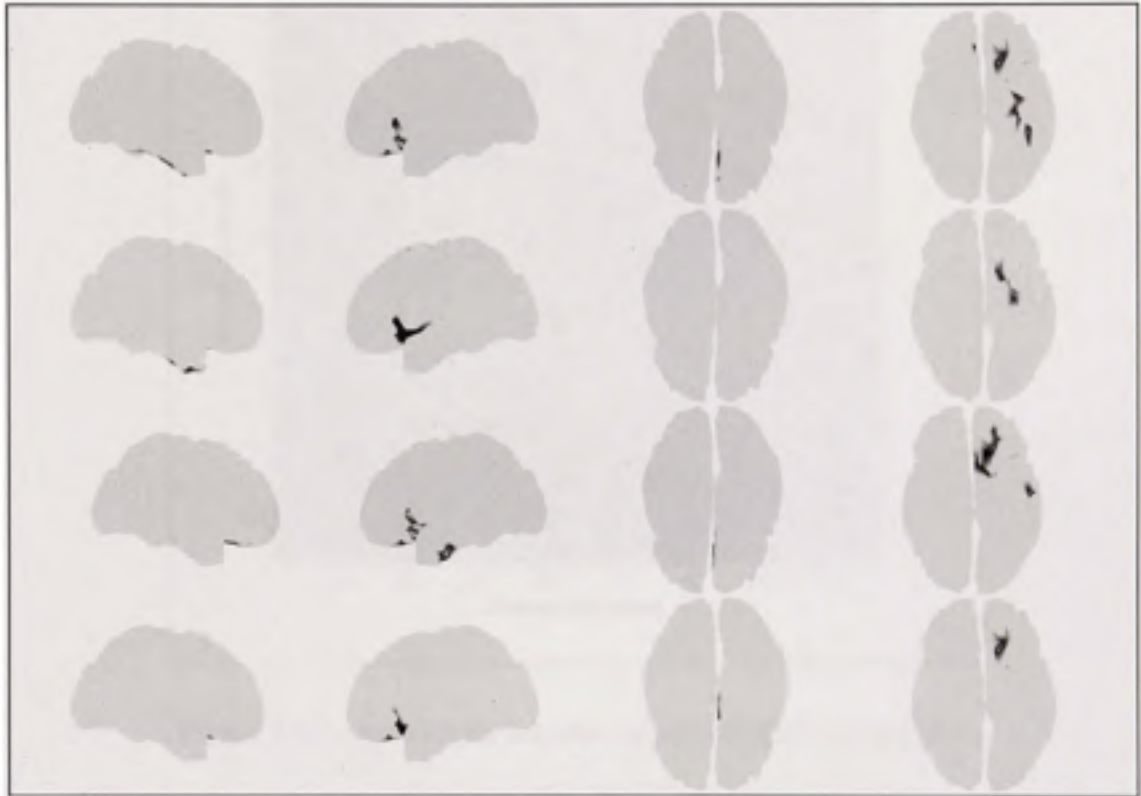
La première étape de la classification est de calculer les scores MSP pour une fenêtre de temps centrée sur les nouveaux marqueurs. La Figure 4.4 présente les scores MSP pour les pointes 3, 23, 44 et 70 sur la surface du cortex. Il ne faut pas confondre cette représentation des scores avec celle des solutions de problème inverse. Ici, les zones noires indiquent que les sources ont une forte probabilité d'appartenir à la solution, mais ne donnent aucune information sur leur intensité. La plupart des pointes ont donc des scores élevés dans les zones frontale et temporale du cerveau.



**Figure 4.4** Les scores MSP sur la surface du cortex pour les pointes 3, 23, 44 et 70 du jeu de données.

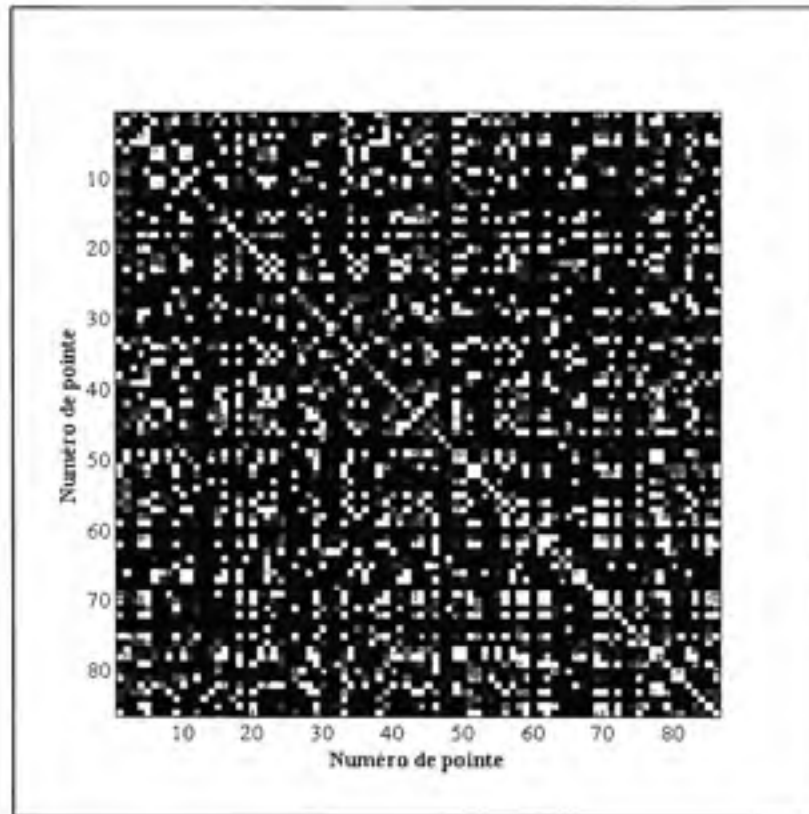
La prochaine étape de la méthodologie est de choisir quelques sources qui représentent bien les pointes. Comme il a été présenté dans la section Chapitre 2, les 10 sources ayant les plus hauts scores pour chaque pointe sont sélectionnées. En plus, les sources dans le voisinage de celles choisies sont aussi retenues. Cette réduction permet d'obtenir environ 50 sources pour

représenter chaque pointe. La Figure 4.5 présente les sources sélectionnées pour les mêmes pointes que la Figure 4.4.



**Figure 4.5** Les sources choisies pour représenter les pointes 2, 23, 44 et 70 du jeu de donnée.

La troisième étape de la classification est d'obtenir une mesure de la similitude des pointes. Pour ce faire, les vecteurs  $v_i(j)$  associés aux pointes sont calculés et présentés à la Figure 4.6. Chaque ligne de la matrice représente un vecteur. On rappelle que ces vecteurs sont construits de telle sorte que  $v_i(j)$  contienne le nombre de sources que les pointes  $i$  et  $j$  ont en commun. Les pixels bleus indiquent une faible quantité de sources communes et les pixels rouges indiquent une grande quantité de sources communes.



**Figure 4.6** Le nombre de sources communes entre les pointes.

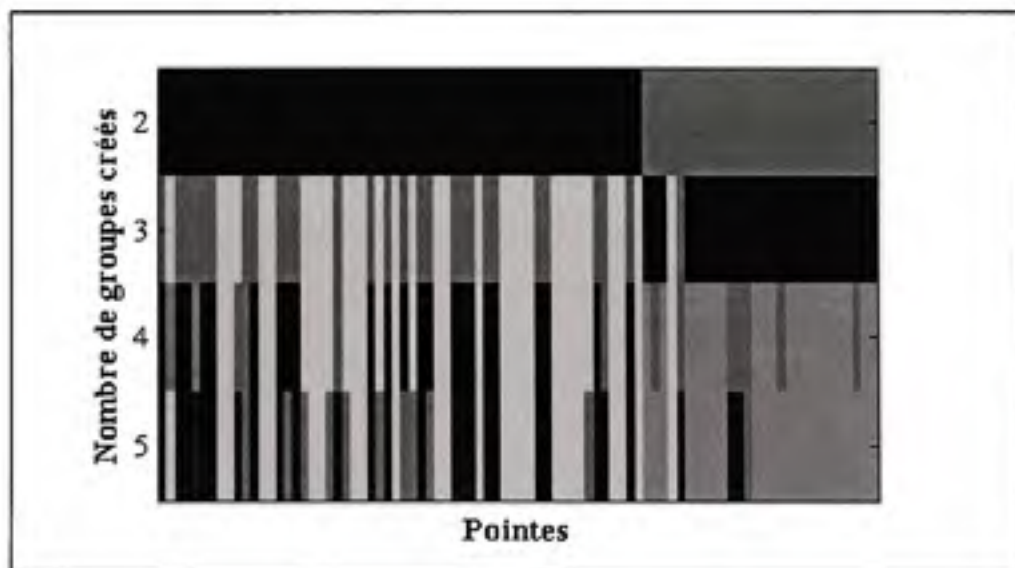
La présence de valeurs élevées sur la diagonale n'est pas surprenante, une pointe a nécessairement toutes les sources en commune avec elle-même. Par contre, la présence de haute valeur à d'autres positions dans la matrice est intéressante parce que ceci indique que plusieurs pointes ont une signature similaire. À titre indicatif, le nombre moyen de sources communes est environ 8.

La dernière étape de l'algorithme est d'associer un numéro de groupe à chaque pointe. Chaque ligne de la matrice des sources communes est alors considérée comme un vecteur décrivant une pointe. L'algorithme *k-means* permet de classifier les vecteurs et d'obtenir les groupes. Cependant, le *k-means* nécessite le nombre de groupes à créer comme paramètre d'entrée. Comme le vrai nombre de groupe est inconnu, la classification est effectuée pour deux à cinq groupes. Afin de choisir la classification la plus intéressante, on observe la



distribution des pointes dans les différents groupes créés. La Figure 4.7 présente cette distribution. Pour une ligne du graphique, une couleur est associée à un groupe, donc une étiquette. Sur la première ligne, toutes cases bleu foncé représentent les pointes portant l'étiquette 1. Les cases bleu pâle représentent les pointes portant l'étiquette 2. Attention, les couleurs ne peuvent pas être comparées ligne par ligne puisque la couleur est arbitraire. Dans ce graphique, les pointes ont été réordonnées en fonction de la classification à 2 groupes. C'est-à-dire que les premières colonnes (bleu foncé de la première ligne) du graphique représentent les pointes du groupe 1 de la classification à 2 groupes. Les colonnes suivantes (bleu pâle de la première ligne) représentent les pointes du groupe 2 de la classification à 2 groupes.

Pour choisir le nombre de groupe, on opère de la façon suivante. Lorsque le bon nombre de groupe est atteint, les classifications suivantes (celles à plus de groupes) ne devraient que subdiviser les groupes actuels. Ce phénomène est observé entre la première et la deuxième ligne de la Figure 4.7, ce qui indique que le nombre de groupe idéal est 2.



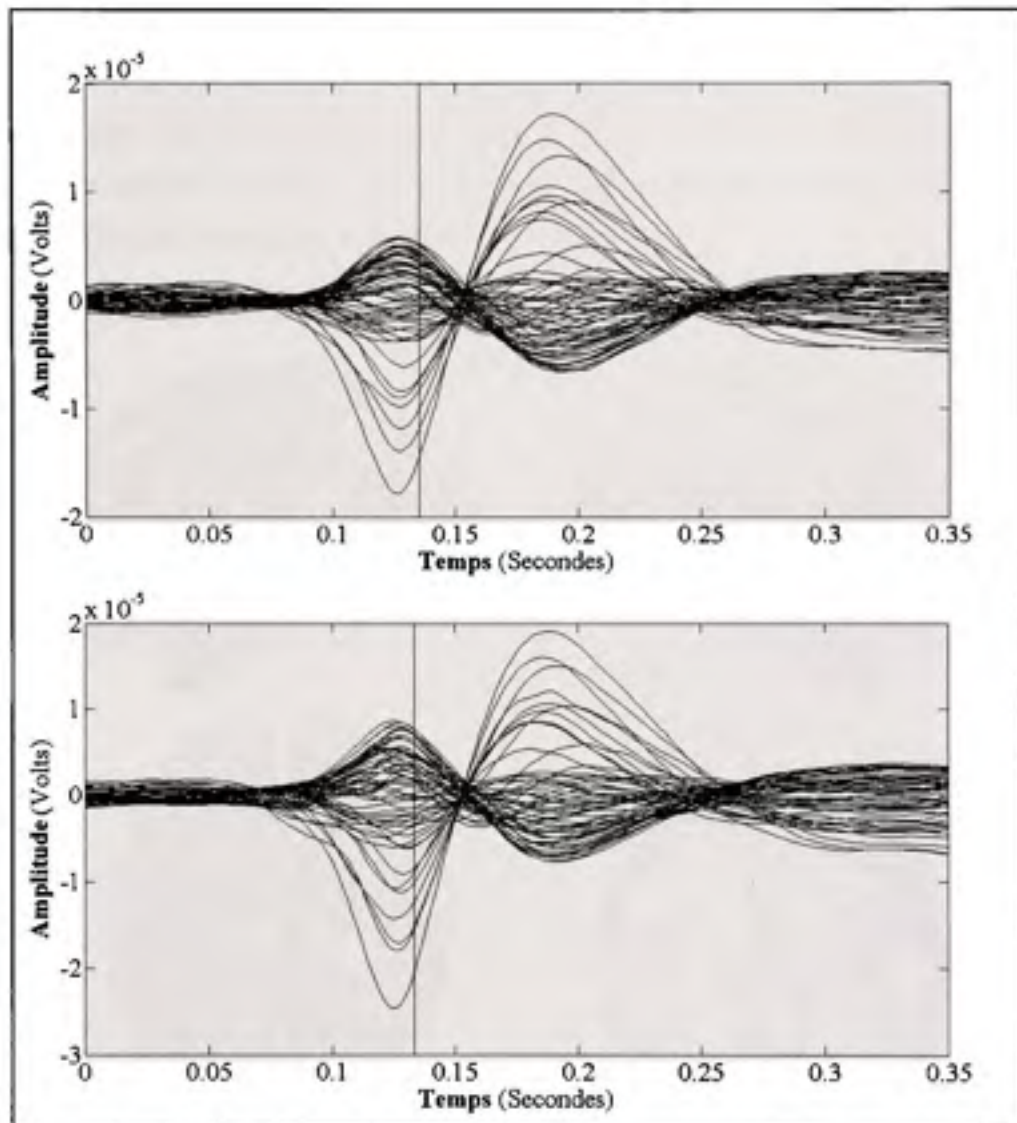
**Figure 4.7** L'organisation des groupes EEG. Les pointes sont organisées en fonction de la classification à 2 groupes.

Le Tableau 4.1 présente les résultats de la classification EEG à 2 groupes. Les pointes sont réparties en groupes inégaux (58 et 28) et il ne semble pas y avoir d'ordre précis dans la distribution.

Tableau 4.1 Résultat de la classification EEG à 2 groupes

Numéro de groupe (nombre de valeurs)	Numéro des pointes appartenant au groupe													
1 (58)	02	03	06	07	08	10	11	12	13	14	15	16	17	19
	21	22	23	24	25	26	27	28	30	31	32	34	35	36
	37	40	41	42	45	47	48	49	50	53	54	55	57	60
	63	64	65	66	67	68	71	73	74	76	77	78	80	81
	82	84												
2 (28)	01	04	05	09	18	20	29	33	38	39	43	44	46	51
	52	56	58	59	61	62	69	70	72	75	79	83	85	86

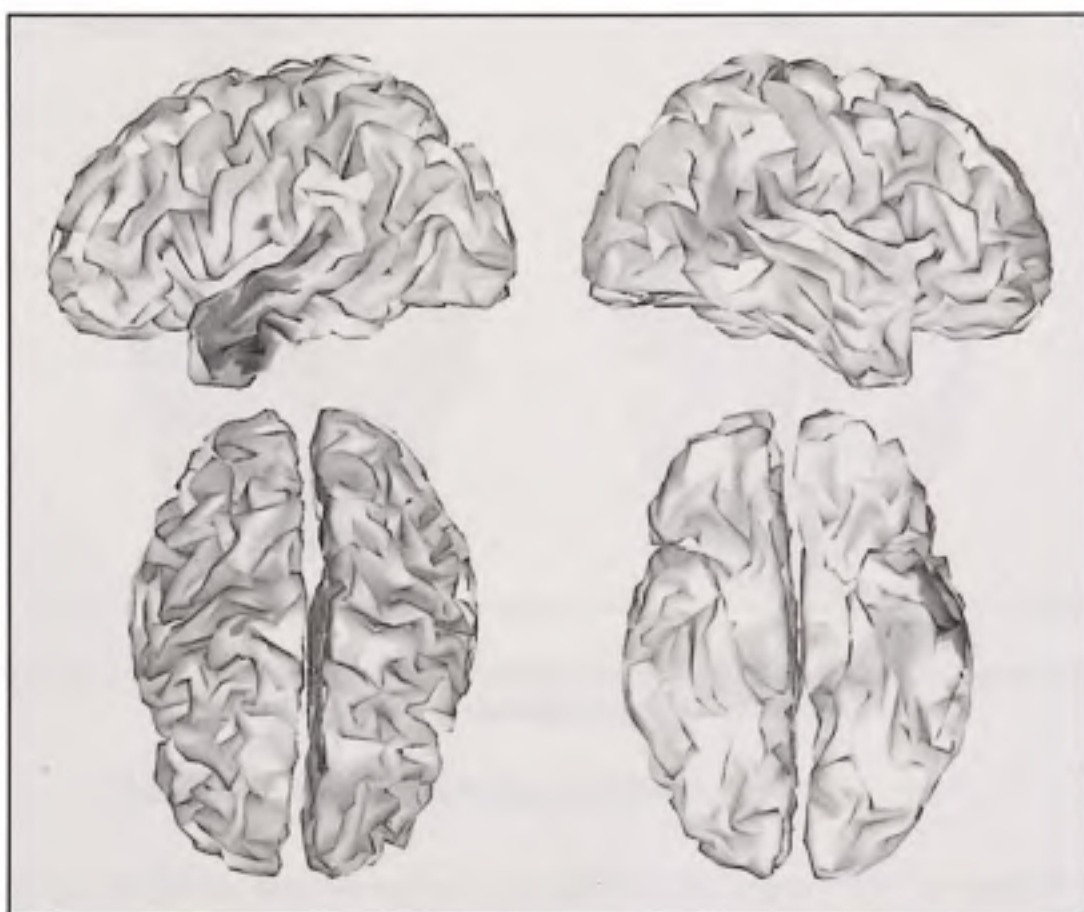
La Figure 4.8 présente la moyenne des signaux EEG des pour chaque groupe obtenu. La forme des signaux obtenus est très similaire, à l'exception d'une légère différence d'amplitude. Cette situation est tout à fait normale puisque la morphologie des signaux n'a pas été utilisée lors de la classification. Pour différencier les classes, il faut plutôt observer l'activation des sources qui leur est associée.



**Figure 4.8** La moyenne des signaux EEG du groupe 1 (haut) et du groupe 2 (bas) d'une classification EEG à 2 groupes.

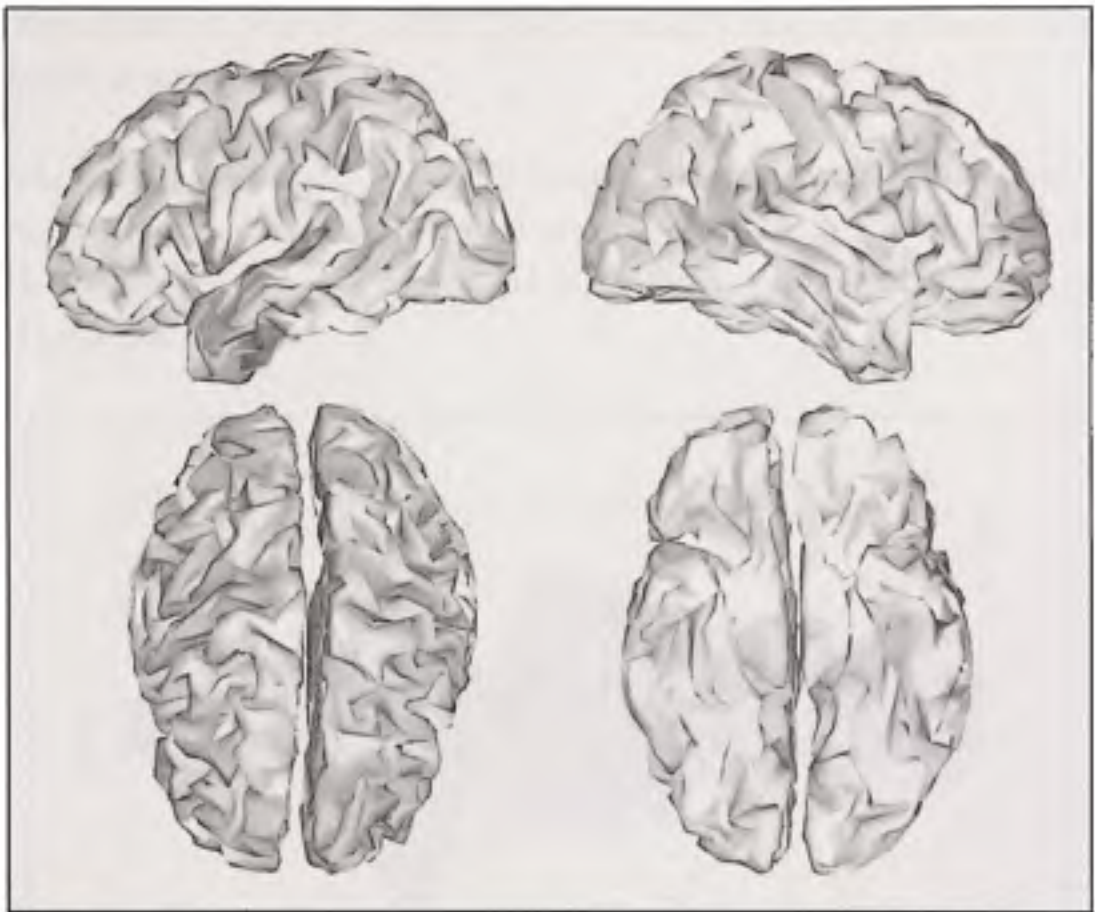
Le résultat du problème inverse trouvé pour la moyenne des pointes EEG du groupe 1 est présenté dans la Figure 4.9. La classification ne semble pas avoir amélioré la résolution du problème puisque l'activation des sources trouvées est pratiquement identique à celle observée pour la moyenne de toutes les pointes du jeu de données. La seule distinction visible entre les deux solutions est l'apparition d'une légère désactivation dans la zone pariétale. L'activation des sources associées à la moyenne des pointes du deuxième groupe,

illustrée par la Figure 4.10, est beaucoup plus intéressante. On note l'apparition d'une nouvelle zone d'activation sur le cortex insulaire. L'apparition de cette activation est importante parce qu'elle correspond à la zone trouvée lors des tests IRMf (Grova, 2009). Parce que le groupe 2 contient moins d'individus que le groupe 1, cette zone d'activation était perdue lors de la moyenne de toutes les pointes.



**Figure 4.9** La solution du problème inverse obtenue pour la moyenne des pointes EEG du groupe 1 de la classification EEG à 2 groupes.





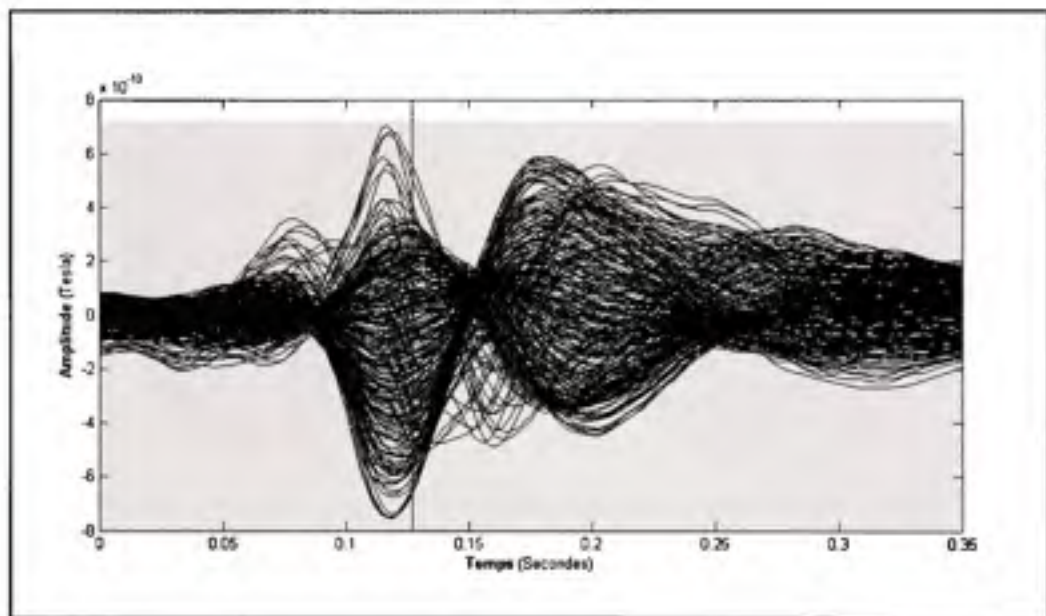
**Figure 4.10** La solution du problème inverse obtenue pour la moyenne des pointes EEG du groupe 2 d'une classification EEG à 2 groupes.

#### 4.2 Classification de pointes épileptiques réelles en MEG

Bien que les pointes aient été marquées en observant les mesures EEG, la résolution du problème inverse en MEG est tout aussi importante pour les neurologues. Parce que les deux modalités observent l'activité des sources d'une manière bien différente, la résolution des problèmes inverses en EEG et MEG offre généralement des solutions différentes. Il est donc tout à fait pertinent d'appliquer la méthodologie de classification aux signaux MEG. Il faut cependant noter que la matrice de gain MEG disponible pour le jeu de données contient quelques erreurs détectées lors de son calcul. Puisque l'impact d'une matrice de gain erronée

sur la classification n'a pas été évalué, les résultats présentés dans cette section doivent être interprétés prudemment.

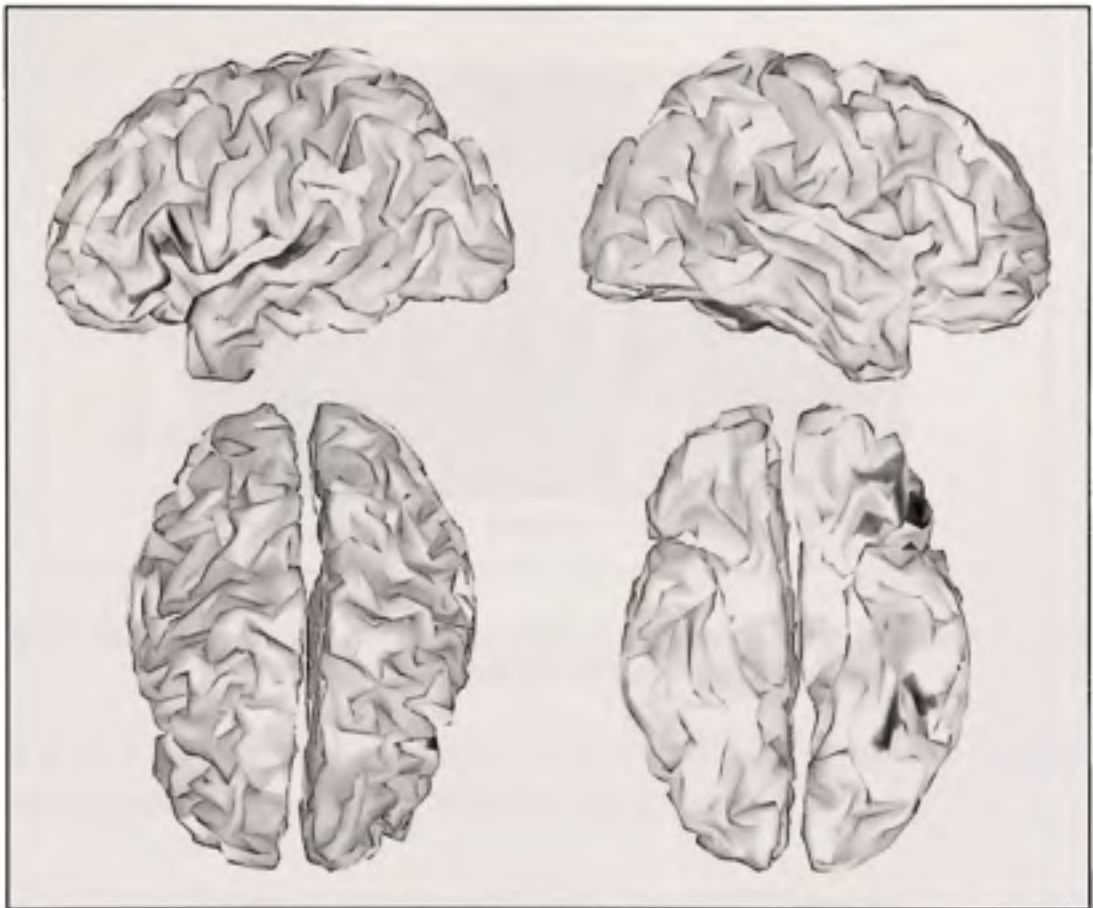
Encore une fois, puisque les pointes ont été marquées à la main, elles doivent être recalées. Le même critère de recalage est utilisé que pour les pointes EEG, soit le maximum de la FPG. La Figure 4.11 présente la moyenne des 86 MEG après le recalage ainsi que la position du maximum de la FPG.



**Figure 4.11** La moyenne de toutes les pointes MEG du jeu de données. La ligne verticale présente la position du maximum de la FPG lorsqu'elle est calculée sur la moyenne des mesures.

La solution du problème inverse calculé sur la moyenne des pointes MEG est présentée à la Figure 4.12. On observe une zone d'activation sur le dessus du sillon latéral gauche et une désactivation dans la zone frontal inférieure et dans la zone ventral.

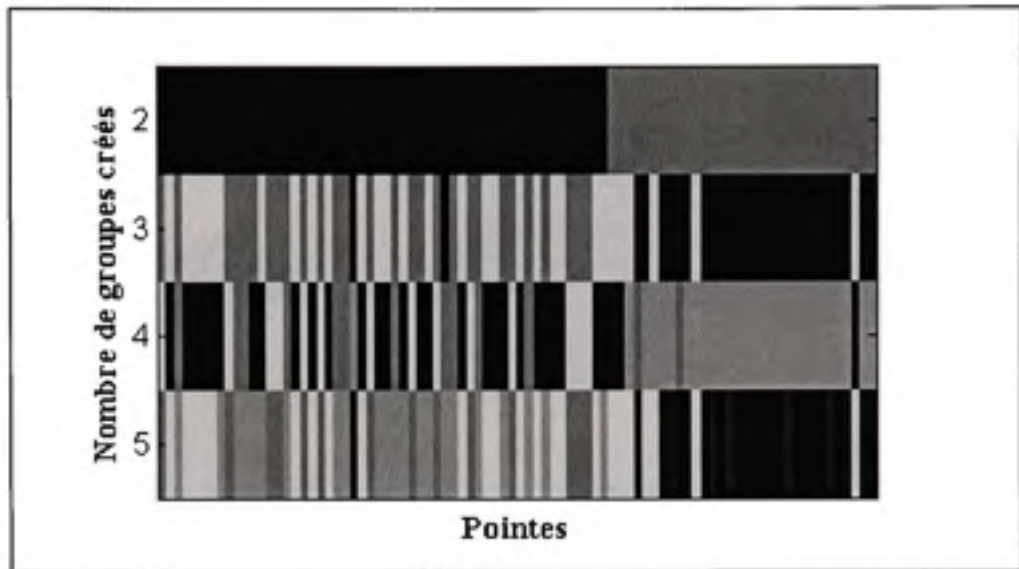
Encore une fois, la solution trouvée est intéressante, mais a la possibilité d'être améliorée par la classification des pointes. Pour cette classification, les détails des scores MSP obtenus, les sources choisies et les vecteurs  $v_l(j)$  ne sont pas présenté en détail.



**Figure 4.12** La solution du problème inverse effectué sur la moyenne de toutes les pointes MEG du jeu de données.

Bien que les mesures EEG et MEG contiennent de l'information sur les mêmes événements, il n'est pas certain que le nombre de groupe EEG corresponde au nombre de groupes MEG. Les pointes sont donc classifiées pour 2, 3, 4 et 5 groupes. L'organisation des groupes est présentée dans la Figure 4.13. Le bon nombre de groupes semble, encore une fois, être deux puisque la classification à trois groupes n'est qu'une subdivision de celle à deux groupes.





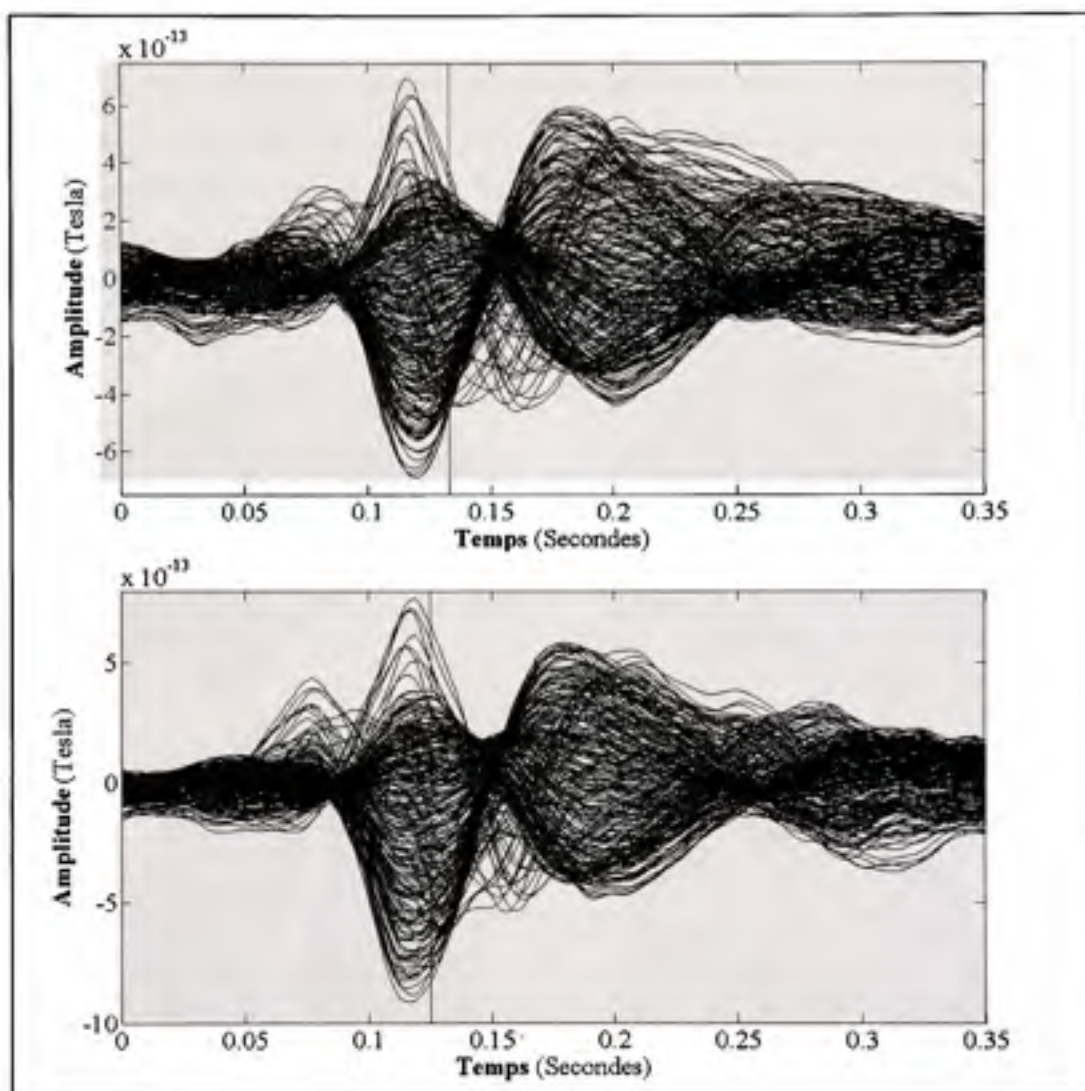
**Figure 4.13 L'organisation des groupes MEG. Les pointes sont organisées en fonction de la classification à 2 groupes.**

Le Tableau 4.1 présente les résultats de la classification MEG à 2 groupes. Les pointes sont divisées en groupes inégaux de 54 et 32 pointes respectivement.

**Tableau 4.2 Résultat de la classification MEG à 2 groupes**

<b>Numéro de groupe (nombre de valeurs)</b>	<b>Numéro des pointes appartenant au groupe</b>													
1 (54)	01	03	04	06	07	08	10	11	12	13	14	16	17	19
	21	22	23	24	25	26	27	28	30	31	32	33	34	35
	36	37	38	41	42	45	47	48	51	52	54	56	59	60
	64	69	70	74	76	78	79	82	83	84	85	86		
2 (32)	02	05	09	15	18	20	29	39	40	43	44	46	49	50
	53	55	57	58	61	62	63	65	66	67	68	71	72	73
	75	77	80	81										

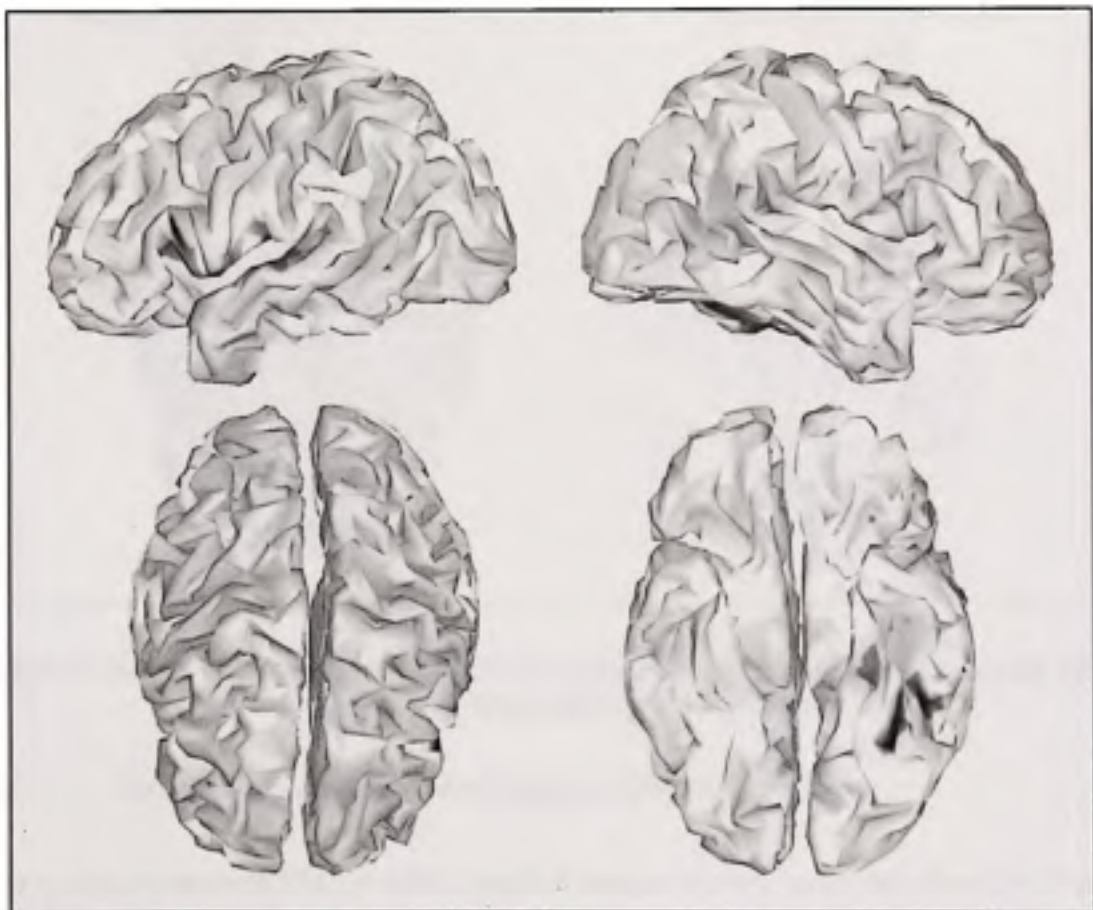
Les signaux moyens pour chaque groupe sont présentés à la Figure 4.14. Ceux-ci présentent encore une morphologie très similaire.



**Figure 4.14 La moyenne des signaux MEG du groupe 1 (haut) et du groupe 2 (bas) d'une classification à 2 groupes.**

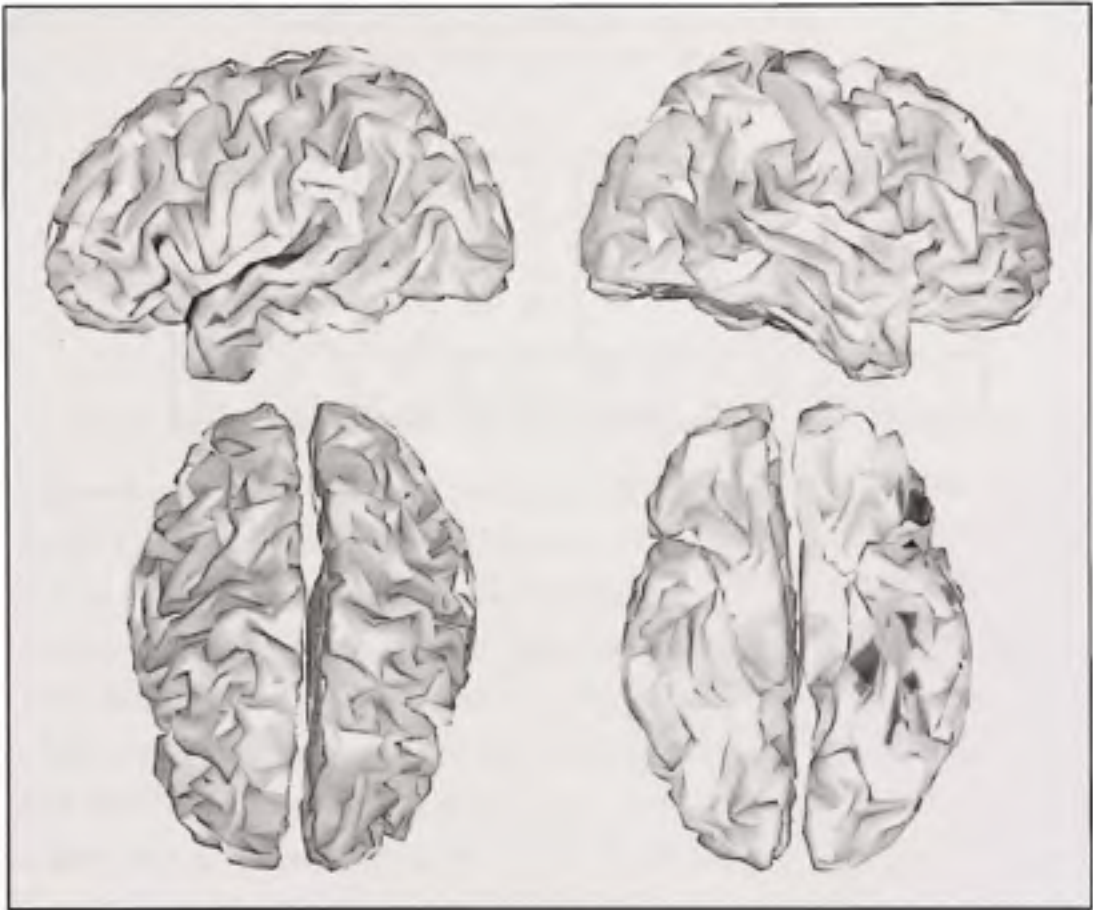
La Figure 4.15 présente la solution du problème inverse obtenu pour la moyenne des pointes du groupe 1. Les activations et désactivations observées sont très similaires à celle de la solution pour toutes les pointes. L'activation se trouve encore sur le sillon latéral gauche et la désactivation sur le cortex insulaire et dans la zone ventrale. La solution du problème inverse pour les pointes du groupe 2 est illustrée dans la Figure 4.16. Cette solution met l'emphase sur l'activation du sillon latéral gauche et diminue l'important de la désactivation dans le

lobe frontal. La classification en MEG n'a pas permis d'identifier de nouvelle zone d'activation. Cependant, les deux groupes formés ont séparé deux zones d'activations. Cette conclusion est intéressante pour les neurologues puisque ceci voudrait dire que certaines pointes sont liées à une activation du sillon latéral gauche alors que d'autres sont associées à une désactivation dans la zone frontale.



**Figure 4.15** La solution du problème inverse obtenue pour la moyenne des pointes MEG du groupe 1 de la classification à 2 groupes.





**Figure 4.16** La solution du problème inverse obtenue pour la moyenne des pointes MEG du groupe 2 de la classification à 2 groupes.

#### 4.3 Classification multimodales de pointes épileptiques réelles

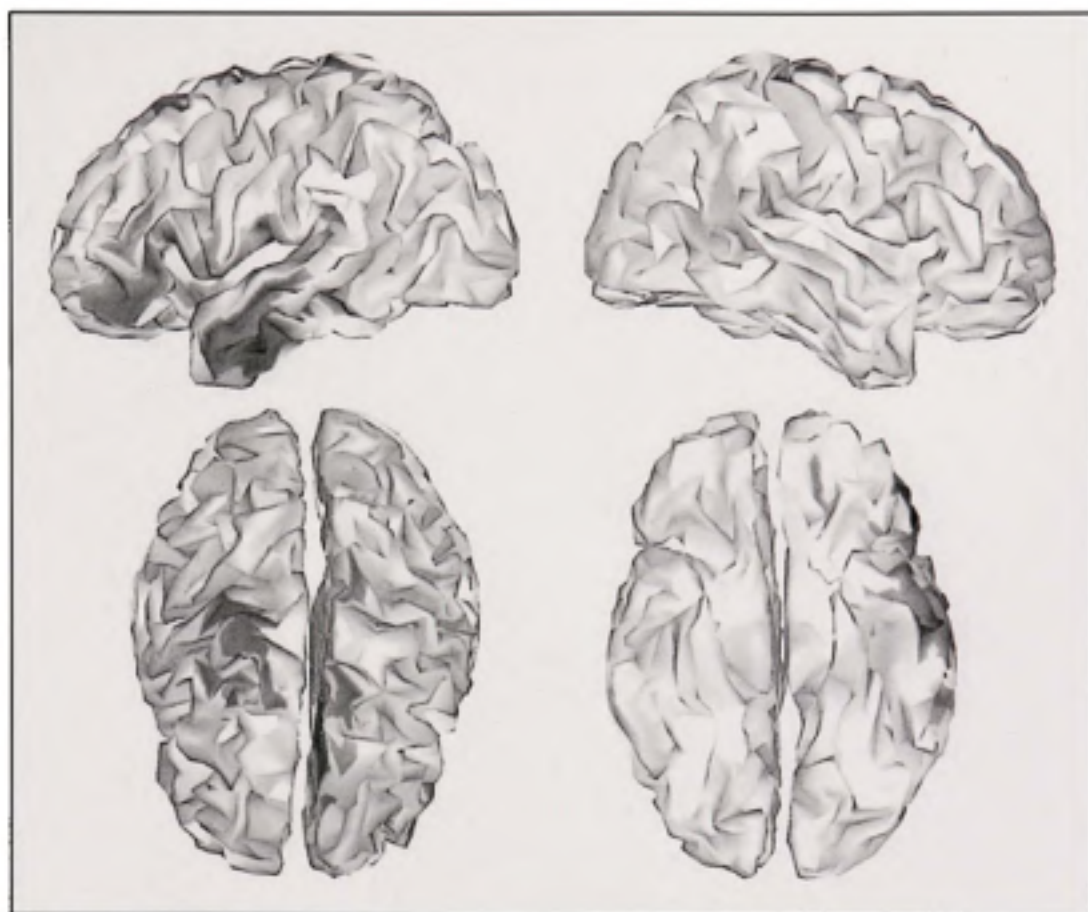
Les groupes trouvés en EEG et MEG semblent indiquer qu'il y aurait effectivement 2 types de pointes dans le jeu de données. Par contre, il est possible que certaines erreurs de classification viennent corrompre la solution des problèmes inverses. Pour vérifier cette hypothèse, la technique de combinaison des classifications proposée dans la section 3.5 est appliquée aux mesures.

Le Tableau 4.3 présente le résultat obtenu. Les pointes sont séparées en quatre groupes, dont 3 petits de 14, 14 et 18 pointes et un grand de 40 pointes.

Tableau 4.3 Le résultat de la combinaison des classifications EEG et MEG

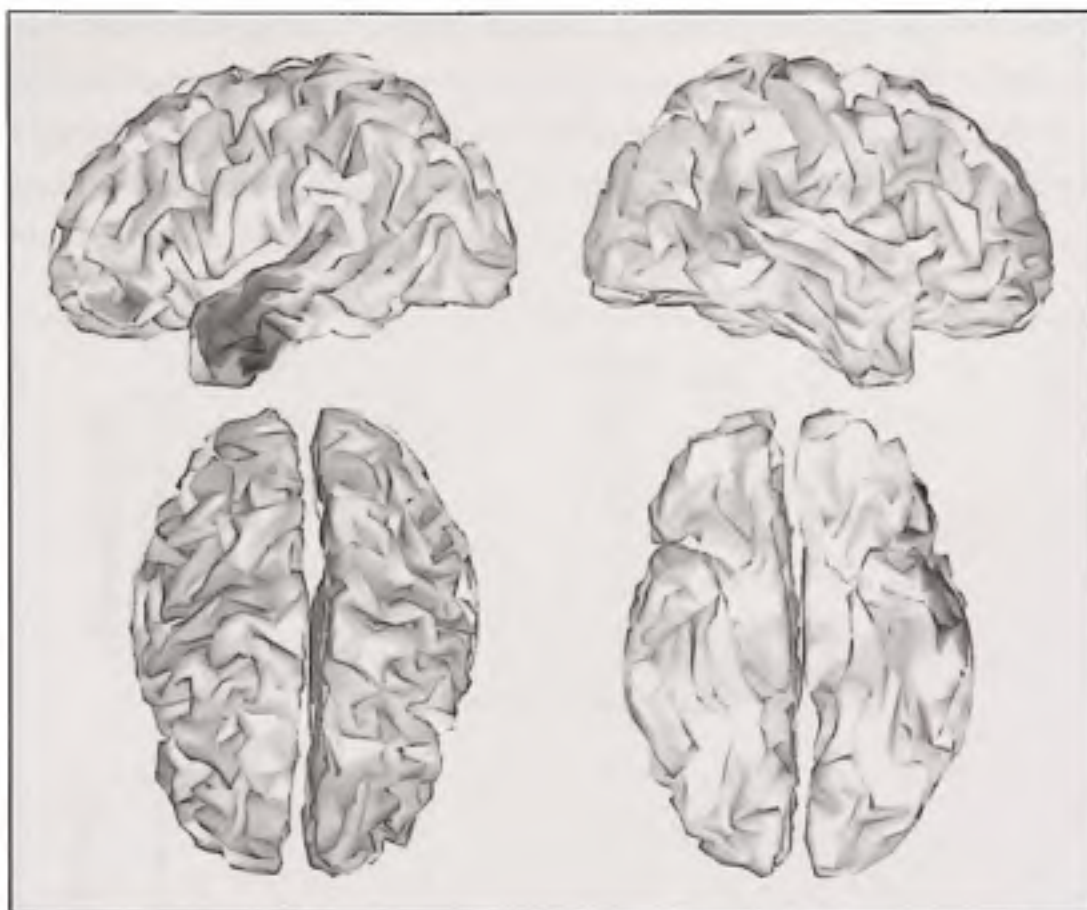
		MEG																						
EEG		03	06	07	08	10	11	12																
		13	14	16	17	19	21	22																
		23	24	25	26	27	28	30	02	15	40	49	50	53	55									
		31	32	34	35	36	37	41	57	63	65	66	67	68	71									
		42	45	47	48	54	60	64	73	77	80	81												
		74	76	78	82	84																		
		1	4	33	38	51	52	56	05	09	18	20	29	39	43									
		59	69	70	79	83	85	86	44	46	58	61	62	72	75									

La solution trouvée pour le plus grand groupe, celui formé des pointes du groupe 1 en EEG et du groupe 1 en MEG, est présentée à la Figure 4.17. Étrangement, la solution associée à ces pointes est beaucoup moins focale que celle obtenue à l'aide des groupes EEG et MEG indépendants. De plus, une désactivation apparaît dans la zone pariétale. Puisque ce groupe est formé des 40 pointes qui étaient communes au groupe 1 en EEG (58 pointes) et au groupe 1 en MEG (54 pointes) on s'attendait à une activation similaire à celle de la Figure 4.9. La solution trouvée peut cependant être expliquée par le retrait des 18 pointes. Si celles-ci sont associées à une activation temporelle très forte, alors il est possible que la solution de la moyenne corresponde à la Figure 4.9. La solution pour les 18 pointes du groupe 2 est présentée à la Figure 4.18 et confirme notre hypothèse.



**Figure 4.17** La solution du problème inverse pour le premier groupe de la classification combinée EEG et MEG.



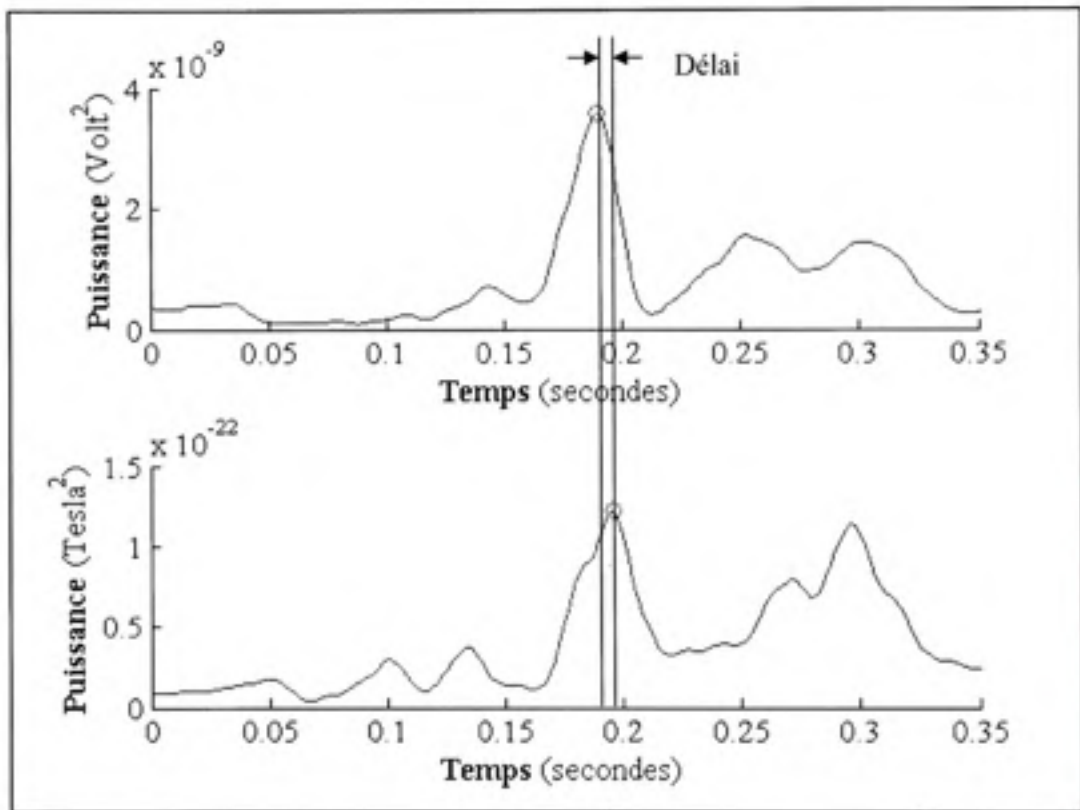


**Figure 4.18** L'activation des sources pour les 18 points du groupe 2 de la classification EEG et MEG combinée.

#### 4.4 Délai entre les pointes EEG et MEG

Une théorie intéressante pour expliquer les classifications différentes obtenues en EEG et en MEG a récemment été proposée. Cette théorie veut que les sources générant les pointes épileptiques se déplacent dans le temps. Parce que la MEG est pratiquement aveugle aux sources radiales, les mesures EEG pourraient présenter une pointe claire alors que les mesures MEG ne présenteraient rien de particulier. Au fur et à mesure que l'activité épileptique se déplace, elle deviendrait graduellement visible pour la MEG. Concrètement, cette situation causerait un décalage entre la pointe EEG et la pointe MEG.

Pour vérifier cette hypothèse, la différence entre la position du maximum de la FPG en EEG et en MEG est observée pour chaque pointe. Un exemple du calcul de ce délai est présenté à la Figure 4.19. La courbe du haut présente la FPG pour les mesures EEG et la courbe du bas présente la FPG pour les mesures MEG. Les cercles illustrent la position du maximum des FPG. Le délai est la différence en secondes entre les maximums.



**Figure 4.19 Exemple de la mesure du délai entre la position du maximum de la FPG en EEG et en MEG.**

Une fois que le délai est calculé pour toutes les pointes, la moyenne et l'écart-type des délais pour chaque groupe sont calculés. Les résultats obtenus sont présentés dans le Tableau 4.4.

Tableau 4.4 Moyenne et écart-type des délais pour chaque groupe

	<b>Groupe 1</b>	<b>Groupe 2</b>	<b>Groupe 3</b>	<b>Groupe 4</b>
<b>Moyenne (ms)</b>	-15.521	-9.259	-1.786	-2.530
<b>Écart-type (ms)</b>	32.043	29.525	18.505	22.795

La moyenne de chaque groupe est effectivement différente, mais les écarts-types élevés ne permettent pas de confirmer l'hypothèse. La théorie proposée ne peut évidemment pas être rejetée par un seul exemple, mais les résultats obtenus semblent indiquer que les groupes de pointes formés à l'aide de l'algorithme proposé dans ce mémoire n'ont pas de lien avec le délai observé en EEG et MEG.

## CONCLUSION

Lors du diagnostic de l'épilepsie, l'EEG apparait comme l'outil de prédilection choisi par les neurologues. En plus de permettre la détection des pointes épileptiques, cet outil peut aussi être utilisé pour localiser la source de la pathologie dans le cerveau. Cette tâche, nommée résolution du problème inverse, n'est pas triviale parce que le modèle mathématique sous-jacent admet une infinité de solutions. Plusieurs techniques ont toutefois été développées pour résoudre ce problème. Par contre, pour offrir des résultats fiables, les algorithmes résolvant le problème inverse nécessitent un rapport signal sur bruit très élevé. Une technique commune pour augmenter ce rapport est de faire la moyenne de plusieurs événements similaires, donc de plusieurs pointes épileptiques. Par contre, si les pointes retenues ne sont pas similaires, la qualité de la solution trouvée sera douteuse. Une astuce efficace pour pallier cette difficulté est la classification des pointes épileptiques. Un algorithme de classification de pointes efficace devrait permettre de regrouper les événements qui sont suffisamment similaires pour en calculer la moyenne et par la suite résoudre le problème inverse.

La plupart des algorithmes de classification se fondent sur la morphologie des pointes pour en évaluer la similitude. Bien qu'intéressantes, ces techniques ne permettent pas de différencier deux événements provenant de sources différentes, mais ayant une morphologie similaire. Une autre approche est d'associer à chaque pointe un dipôle équivalent. La position et l'orientation du dipôle sont alors utilisées comme critère de classification. Cette approche considère la représentation des mesures dans l'espace des sources, mais offre une résolution spatiale limitée.

La contribution principale de ce mémoire est de proposer un nouvel algorithme de classification se basant sur la représentation des mesures dans l'espace des sources d'un modèle distribué. Cette approche offre une meilleure résolution spatiale que le modèle dipolaire et permet de différencier deux événements morphologiquement similaires, mais provenant de sources distinctes. Les tests effectués sur des mesures simulées ont permis

d'évaluer la performance de ce nouvel algorithme. En général, la qualité de classifications obtenues est acceptable. Les meilleures performances étaient atteintes lorsque les pointes avaient un ratio d'amplitude de 6 par rapport au signal physiologique et lorsque les sources qui génèrent les pointes sont distinctes pour chaque groupe.

La technique de classification proposée dans ce mémoire a aussi été appliquée à des mesures réelles. Dans cette situation, l'objectif était d'identifier certaines zones actives du cerveau qui aurait été perdues lors du calcul de la moyenne. Les résultats du problème inverse sur les mesures EEG, obtenu à l'aide de la technique du MEM, ont effectivement permis de trouver une nouvelle zone d'activation cohérente avec les résultats de l'IRMf. Dans le cas des mesures MEG, la classification a permis de séparer les pointes responsables d'une désactivation dans la zone frontal inférieure de celles responsables d'une activation dans le sillon latéral gauche.

La dernière analyse sur des données réelles qui a été effectuée a tenté de combiner les classifications EEG et MEG. Les solutions du problème inverse sur les nouveaux groupes formés n'ont pas apporté de nouvelles informations. Une théorie qui voulait que les groupes soient liés à un délai entre les pointes EEG et MEG a aussi été testée. Cependant, les délais observés n'ont pas permis de confirmer cette hypothèse.

Globalement, la technique de classification présentée dans ce mémoire semble corriger les lacunes des algorithmes disponibles dans la littérature, même si une comparaison directe n'a pas été effectuée. Cependant, pour obtenir des résultats fiables, il faut respecter les hypothèses sur lesquelles cette technique se fonde. Tout d'abord, la propagation des courants électriques et des champs magnétiques dans le cerveau peut être modélisée par le problème direct. Ensuite, la matrice de gain du problème direct est valide et relie correctement l'activation des sources aux mesures des capteurs EEG ou MEG. Si ces conditions ne sont pas respectées, les résultats offerts par l'algorithme risquent d'être erronés.

Dans ce mémoire, les pointes épileptiques classifiées et utilisées lors de la résolution du problème inverse ont été marquées par un neurologue. La lecture des signaux EEG et le

marquage des pointes est une tâche longue et fastidieuse. Il est donc intéressant de faire usage d'un outil de détection automatique des pointes épileptiques. Dans ce contexte, l'algorithme de classification présenté dans ce mémoire pourrait être utilisé pour améliorer la qualité de la détection en regroupant les événements similaires. De plus, si le nombre de groupes à former lors de la classification pouvait être détecté automatiquement, l'utilisation de cette technique serait grandement simplifiée.



## ANNEXE I

### BRAINENTROPY

L'objectif principal du paquet *BrainEntropy* est de fournir un outil cohérent de traitement des signaux cérébraux. Cet outil devra relier les mesures aux sources qui les génèrent à l'aide de la matrice de gain et regrouper toute cette information d'une manière intuitive. Il servira aussi de plateforme de développement pour de nouvelles techniques d'analyse. La programmation orientée objet (POO) est donc tout indiquée comme paradigme de développement parce qu'elle donne un sens physique aux entités du problème et est facilement extensible. Un avantage secondaire de la POO est qu'elle permet à de nouveaux utilisateurs d'exploiter des méthodologies complexes sans se soucier des détails de l'implémentation. Parce que cet outil informatique sera utilisé sous plusieurs systèmes d'exploitation et que l'utilisation d'outils mathématiques validés s'avère capitale, le langage choisi pour sa conception est MATLAB.

Bien que l'architecture initiale ait été développée pour traiter les signaux électroencéphalographiques et magnétoencéphalographiques, elle peut être utilisée pour des signaux temporels quelconques. À titre d'exemple, le paquet a récemment été utilisé pour l'analyse temps-fréquences de signaux optiques. Les méthodes actuelles du paquet permettent de résoudre des problèmes inverses, d'analyser les plans temps-fréquences et de classifier des signaux.

La description de *BrainEntropy* contenue dans ce mémoire n'est pas exhaustive, plusieurs propriétés et méthodes ne sont pas présentées. Cette description a pour objectif d'illustrer la philosophie avec laquelle le paquet a été conçu et de justifier l'existence des différents éléments. Pour des renseignements détaillés, le lecteur devra se référer à l'aide disponible sous MATLAB.

L'architecture de *BrainEntropy* est fondée sur le patron de conception modèle-vue-contrôleur (MVT). Dans les sections suivantes, les diverses composantes de cette architecture seront présentées.

Les classes du paquet *BrainEntropy* sont relativement indépendantes. C'est-à-dire qu'elles peuvent être utilisées à l'extérieure de la hiérarchie. Cependant, lorsque les objets sont utilisés indépendamment certaines fonctionnalités ne sont pas disponibles. La description suivante suppose que les objets font partie d'un modèle complet. Afin de les identifier rapidement, toutes les classes qui font partie du modèle portent le préfixe *c*.

Un modèle *BrainEntropy* encapsule les renseignements suivants :

1. L'anatomie du patient, sous la forme de surfaces et de maillages.
2. Les caractéristiques de l'instrument de mesure, comme la position des capteurs et les unités de mesure.
3. Les matrices de gain qui relient l'information vue par les capteurs aux signaux source qui la génère.
4. Les mesures et leurs fréquences d'échantillonnages.
5. Les intensités des sources qui expliquent les mesures.

La hiérarchie *BrainEntropy* permet d'ordonner toutes ces données d'une manière intuitive tout en évitant la redondance.

Le modèle d'un sujet possède deux branches principales. La première contient l'information propre à l'anatomie du sujet tandis que la deuxième contient l'information des modalités et les mesures.

La propriété `MODALITIES` du modèle permet d'accéder aux différentes modalités associées à un modèle, sous la forme d'un tableau d'objets de la classe `c_MODALITY`. Une modalité représente tous les enregistrements effectués avec un équipement spécifique, par exemple un électroencéphalogramme. Les capteurs sont donc propres à la modalité et sont enregistrés dans la propriété `SENSORS` de la classe. Parce que la matrice de gain représente le lien entre les sources et les capteurs, elle est aussi liée à une modalité. Une des contraintes de *BrainEntropy* est que chaque modalité ne peut posséder qu'une seule matrice de gain. De plus, celle-ci doit être compatible avec l'information des sources et des capteurs. Par

exemple, si une instance de la classe `C_MODALITY` possède 56 capteurs, la matrice de gain doit posséder 56 lignes.

Parce que plusieurs mesures peuvent être effectuées dans le cadre d'une modalité, les signaux sont gardés sous la forme d'un tableau d'objets de la classe `C_DATA`. Le tableau est contenu dans la propriété `SIGNALS` de la classe `C_MODALITY`. Les objets de la classe `C_DATA` donnent accès à toute l'information du problème direct. Ils contiennent les mesures temporelles ( $M$ ), les intensités des sources qui expliquent les mesures ( $J$ ) et donnent accès à la matrice de gain ( $G$ ). Il faut cependant noter que la matrice de gain ne sera disponible que si l'objet de la classe `C_DATA` se retrouve dans une hiérarchie *BrainEntropy*, donc que l'objet supérieur est de la classe `C_MODALITY`.

Les méthodes de la classe `C_DATA` s'appliquent normalement à la totalité des mesures contenues dans l'instance. Il faut donc un autre mécanisme pour appliquer une méthode à une sous-section des données. La classe `C_MARKER` remplira cette tâche.

La classe `C_MARKER` est un outil qui permet d'extraire une section des mesures d'un signal. Les propriétés d'un marqueur sont sa position et sa durée. Lorsqu'un marqueur est ajouté à un signal, il est possible d'extraire les mesures à l'aide de la méthode `SUBSIGNALS` de la classe `C_DATA`. Afin de rendre cet outil plus puissant, la section des mesures est retournée sous la forme d'une instance de la classe `C_DATA`.

L'information sur l'anatomie du patient est gardée dans un tableau d'instances de la classe `C_MESH`. Ce tableau est contenu dans la propriété `MESHES` du modèle. Un objet de la classe `C_MESH` représente le maillage d'une surface du modèle, comme le cortex ou la peau. Il contient donc la position des nœuds et la matrice qui indique comment les relier pour créer la surface.

Un des objets contenus dans la propriété `MESHES` du modèle est particulier: il est de la classe `C_SOURCE` plutôt que `C_MESH`. Cette classe représente une surface particulière, celle de l'espace des sources. La différence entre la surface des sources et les autres surfaces est

qu'elle est composée de dipôles plutôt que de nœuds. Une autre restriction de *BrainEntropy* est qu'il ne peut y avoir qu'une instance de la classe `C_SOURCE` dans le tableau des maillages. Donc, il ne peut y avoir qu'un espace des sources par modèle.

L'objet `C_SOURCE` a un impact important sur le reste du modèle, il impose le nombre de sources du modèle. Donc, si un modèle possède 6000 sources, toutes les matrices de gain doivent posséder 6000 colonnes, peu importe leurs modalités.

D'une manière homologue aux signaux, il est possible d'obtenir une section de l'espace des sources. Tous les dipôles de l'espace des sources possèdent un numéro de parcelle. En utilisant la méthode `GET_CLUSTER` de la classe `C_SOURCE`, il est possible d'extraire les dipôles qui portent une valeur particulière. Ici aussi, pour rendre cet outil plus puissant, les dipôles sont retournés sous la forme d'un objet de la classe `C_SOURCE`.

La visualisation des mesures et surfaces du modèle est une partie importante de l'analyse des données. Plusieurs classes graphiques ont donc été développées pour permettre d'observer le contenu d'un modèle. Afin de les identifier rapidement, tous les objets de visualisation portent le préfixe `CG`.

Il est possible que certains objets graphiques possèdent des contrôles qui interagissent avec le modèle. Dans ce cas, les `CG` font aussi office de contrôleurs. Les objets de la classe graphique n'affectent jamais l'information du modèle, même dans le cas où ils contiennent des contrôles. Ils ne font que déléguer les tâches (en appelant une ou des méthodes du modèle, par exemple) et illustrer les changements.

Les classes graphiques actuelles permettent de visualiser les signaux (`CG_SIGNAL`) et les surfaces (`CG_SOURCE3`, `CG_MAIILLAGE`) du modèle.

Comme l'illustre le digramme de classe de la Figure-A 1-1, la plupart des classes contenues dans le paquet *BrainEntropy* sont dérivées de la classe `C_LATIS`. Les seules exceptions à cette règle sont les classes `C_COLLECTION` et `C_COORDINATE`, ainsi que leurs enfants. La

classe `C_LATIS` sert de classe universelle (un peu comme la classe `OBJECT` du langage Java) et est elle-même dérivée de la classe `HANDLE` de MATLAB. Ce schéma d'héritage indique que toutes les classes du paquet *BrainEntropy* sont copiées par référence.

Les propriétés de la classe `C_LATIS` sont très générales. Elles contiennent une description de l'objet, un nom et une date de création. De plus, elle contient la propriété `ROOT` qui donne accès à l'objet supérieur dans la hiérarchie de *BrainEntropy*.

La classe `C_COLLECTION` est une structure de données qui offre plus de fonctionnalité que les tableaux MATLAB. Une des exigences des tableaux MATLAB est que tous les éléments doivent être du même type. Il est donc impossible de garder plusieurs instances de classes différentes dans un seul tableau. De plus, les éléments d'un tableau ne sont accessibles que par leurs indices de case. En utilisant une instance de la classe `C_COLLECTION` comme conteneur, plusieurs objets de classes différentes peuvent être gardés dans un seul tableau (en apparence). De plus, les éléments contenus dans une instance de la classe `C_COLLECTION` sont accessibles à l'aide d'une chaîne de caractères. Par exemple, dans une instance de la classe `C_SIGNALS` qui contient plusieurs instances de `C_DATA`, on peut accéder à chacun de ceux-ci en utilisant le contenu de la propriété `NAME`.

Les classes `C_COORDINATES`, `C_MARKERS`, `C_MESHES`, `C_MODALITIES`, `C_NODES`, `C_SENSORS`, `C_SIGNALS` sont toutes des alias de la classe `C_COLLECTION`. La seule différence entre ces classes et la classe `C_COLLECTION` est que celles-ci imposent des restrictions sur le type des objets qui peuvent y être ajoutés. Par exemple, la classe `C_MODALITIES` est une `C_COLLECTION` dans laquelle on ne peut ajouter que des instances de la classe `C_MODALITY` (ou une sous-classe de `C_MODALITY`). La classe `C_SENSORS` est une `C_COLLECTION` dans laquelle on ne peut ajouter que des instances de la classe `C_SENSOR` (ou une sous-classe de `C_SENSORS`).

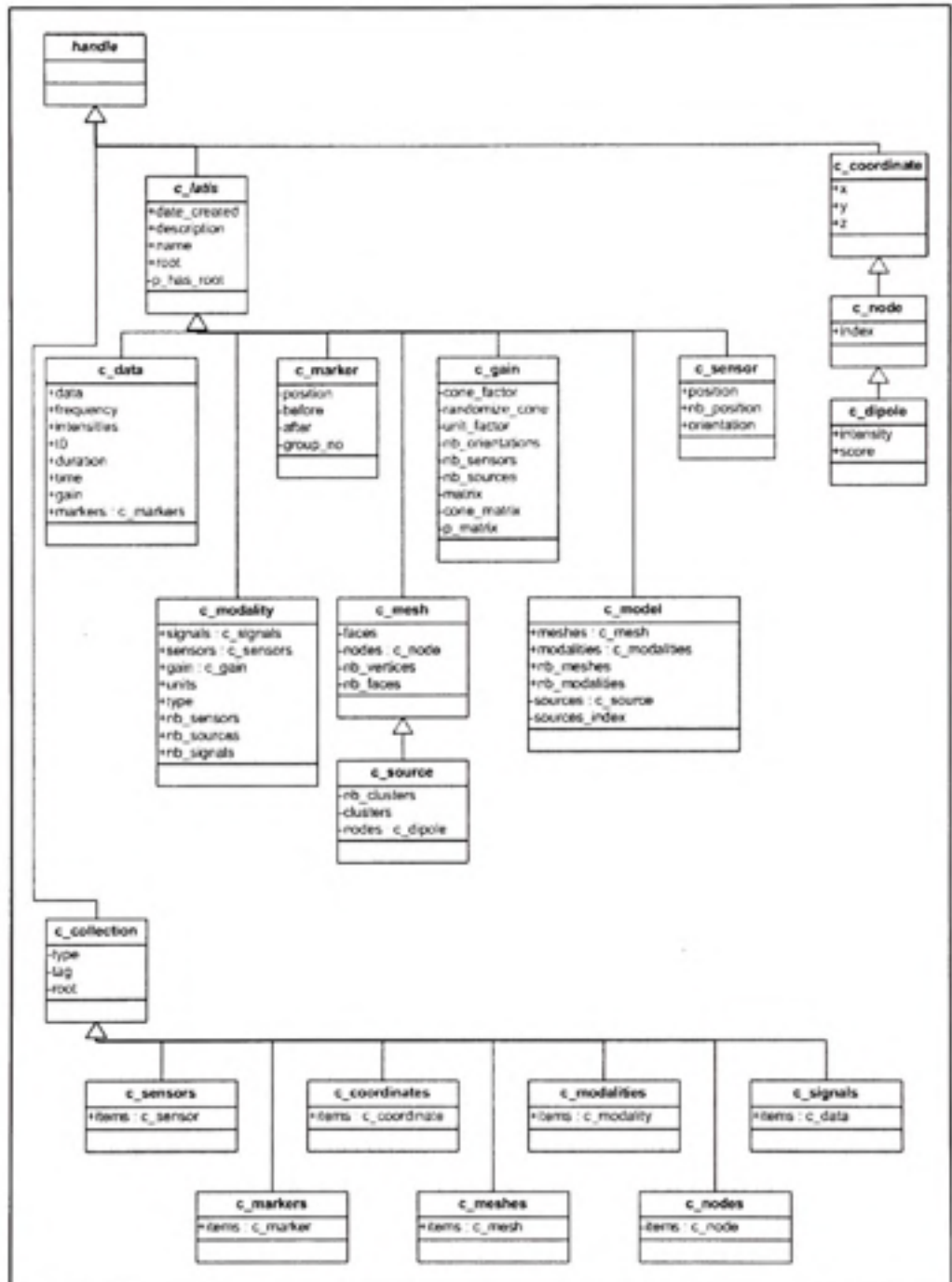


Figure-A I-1 Diagramme des classes de la librairie BrainEntropy.



## BIBLIOGRAPHIE

Amblard, C., E. Lapalme et J. M. Lina. 2004. « Biomagnetic source detection by maximum entropy and graphical models ». *Biomedical Engineering, IEEE Transactions on*, vol. 51, n° 3, p. 427-442.

Baillet, S., J. C. Mosher et R. M. Leahy. 2001a. « Electromagnetic brain mapping ». *Signal Processing Magazine, IEEE*, vol. 18, n° 6, p. 14-30.

Baillet, Sylvain, J.C. Mosher et R.M. Leahy. 2001b. « Electromagnetic brain mapping ». *IEEE Signal Processing Magazine*, vol. 18, n° 6, p. 14-30.

Bar-Hillel, Aharon, Adam Spiro et Eran Stark. 2006. « Spike sorting: Bayesian clustering of non-stationary data ». *Journal of Neuroscience Methods*, vol. 157, n° 2, p. 303-316.

« BESA: Brain Electrical Source Analysis », 2009. <[www.besa.de](http://www.besa.de)>.

« Centre Hospitalier Universitaire Vaudois Lausanne: Épilepsie ». 2009, 26 juin. En ligne.

Clarke, C.J.S.;Janday B.S. 1989. « The solution of the biomagnetic inverse problem by maximum statistical entropy ». *Inverse problems*, vol. 4, p. 483.

Deslauriers-Gauthier, Samuel, Jean-Marc Lina, Étienne Lemay et Christophe Grova. 2009. *Epileptic spike classification and source localization*. Montréal: Centre de Recherches Mathématiques.

Ermer, J.J., J.C. Mosher, S. Baillet et R.M. Leahy. 2001. « Rapidly re-computable EEG forward models for realistic head shapes ». *Physics in Medicine and Biology*, vol. 46, n° 4, p. 1265-1281.

Flanagan, D., et R. Agawal. 2002. « Computer-aided Spatial Classification of Epileptic Spikes ». *Journal of Clinical Neurophysiology*, vol. 19, n° 2, p. 125-135.

Geddes, L., et L. Baker. 1967. « The specific resistance of biological materials - A compendium of data for the biomedical engineer and physiologist ». *Medical and Biological Engineering*, vol. 5, p. 271-293.

Grova, Christophe. 2009. « Localisation de source du jeu de donnée PA15 ».

Hulata, Eyal, Ronen Segev et Eshel Ben-Jacob. 2002. « A method for spike sorting and detection based on wavelet packets and Shannon's mutual information ». *Journal of Neuroscience Methods*, vol. 117, n° 1, p. 1-12.

Lantz, G., P. Wahlberg, G. Salomonsson et I. Rosn. 1998. « Categorization of interictal epileptiform potentials using a graph-theoretic method ». *Electroencephalography and Clinical Neurophysiology*, vol. 107, n° 5, p. 323-331.

*Le petit Larousse illustré*. 1998. Paris.

Lemay, Étienne. 2009. « Localisation de sources à l'aide de la fusion multimodale EEG et MEG ». École de technologie supérieure.

Marieb, Elaine N. 2005. *Anatomie et physiologie humaines*. Éditions du renouveau pédagogique inc.

Mattout, J., M. Pélégriani-Issac, L. Garnero et H. Benali. 2005. « Multivariate source prelocalization (MSP): Use of functionally informed basis functions for better conditioning the MEG inverse problem ». *NeuroImage*, vol. 26, n° 2, p. 356-373.

Mosher, J., et R. Leahy. 1999. « Source localization using recursively applied and projected (RAP) MUSIC ». *IEEE Transactions on signal processing*, vol. 47, n° 2, p. 332-340.

Ossadtchi, A., S. Baillet, J.C. Mosher, D. Thyerlei, W. Sutherling et R.M. Leahy. 2004. « Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering ». *Clinical Neurophysiology*, vol. 115, n° 3, p. 508-522.

Quiroga, R. Quian, Z. Nadasdy et Y. Ben-Shaul. 2004. « Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering ». *Neural Computation*, vol. 16, p. 1661-1687.

« Stellate ». 2009. <[www.stellate.com](http://www.stellate.com)>.

Sun, M. 1997. « An efficient algorithm for computing multishell spherical volume conductor models in EEG dipole source localization ». *IEEE Transactions on Biomedical Engineering*, vol. 44, p. 1243-1252.

Van 't Ent, D., I. Manshanden, P. Ossenblok, D. N. Velis, J. C. de Munck, J. P. A. Verbunt et F. H. Lopes da Silva. 2003. « Spike cluster analysis in neocortical localization related epilepsy yields clinically significant equivalent source localization results in magnetoencephalogram (MEG) ». *Clinical Neurophysiology*, vol. 114, n° 10, p. 1948-1962.

Wahlberg, P., et G. Lantz. 2000. « Methods for robust clustering of epileptic EEG spikes ». *Biomedical Engineering, IEEE Transactions on*, vol. 47, n° 7, p. 857-868.

Wahlberg, P., et G. Salomonsson. 1996. « Feature Extraction and Clustering of EEG Epileptic Spikes ». *Computers and Biomedical Research*, vol. 29, n° 5, p. 382-394.

Ward, J.H. 1963. « Hierarchical grouping to optimise an objective function. ». *Journal of American Statistics Associated*, vol. 58, p. 236-244.

Wheeler, Bruce C., et William J. Heetderks. 1982. « A Comparison of Techniques for Classification of Multiple Neural Signals ». *Biomedical Engineering, IEEE Transactions on*, vol. BME-29, n° 12, p. 752-759.

Wilson, Scott B., Christine A. Turner, Ronald G. Emerson et Mark L. Scheuer. 1999. « Spike detection II: automatic, perception-based detection and clustering ». *Clinical Neurophysiology*, vol. 110, n° 3, p. 404-411.