

SEGMENTATION VECTORIELLE DE SIGNAUX EEG

G. CARRAULT, J.J. BELLANGER, J.M. BADIER*

Laboratoire Traitement du Signal et de l'Image, Université de Rennes I, 35042 RENNES CEDEX, France.

* Unité d'épileptologie clinique, Service de Neurologie, CHRU Rennes, Rue H. Le Guillou, 35033 RENNES CEDEX, France

RÉSUMÉ

Dans le cadre d'étude de l'épilepsie, le signal EEG présente dans certaines phases caractéristiques des changements brusques qui surviennent sur une ou plusieurs voies et doivent être détectés. La corrélation entre les voies liée à certains phénomènes de synchronisation mène à envisager une segmentation basée sur des modèles vectoriels. Cet article compare une méthode de segmentation vectorielle à l'utilisation voie par voie, de traitements scalaires dont on utilise les résultats au moyen d'une règle de fusion simple. Les premiers résultats obtenus dans différentes situations réalistes issues de signaux réels montrent que l'approche vectorielle est supérieure. Il s'avère cependant, que sur l'ensemble des cas étudiés le gain en performance apporté par celle-ci n'est parfois que peu sensible.

INTRODUCTION

Le signal EEG - lorsqu'il est recueilli au moyen d'électrodes profondes - est une observation multi-voies qui présente dans certaines phases caractéristiques telles qu'une crise d'épilepsie, des variations brutales de statistique (sur l'ensemble ou une partie des voies) survenant en des instants qu'il est important sur le plan clinique d'estimer dans l'absolu et relativement les uns aux autres ; on peut ainsi espérer mettre en évidence des chemins de propagation privilégiés correspondant à la contamination d'autres aires cérébrales par le foyer paroxystique initiateur. La solution de ce problème passe naturellement par l'utilisation d'algorithmes de segmentation en ligne pouvant être en version scalaire ou, plus rarement, vectorielle. Plusieurs auteurs ont déjà tenté de formuler une solution pour ce problème, soit sous la forme d'un traitement voie par voie [1][4], soit, beaucoup plus récemment, sous la forme d'un traitement global [5]. Cet article cherche donc à évaluer dans quelle mesure une approche de traitement global est préférable au traitement voie par voie. Ceci repose sur la définition d'un critère de comparaison ; celui classique [2] que nous avons retenu est le retard moyen de détection pour un temps moyen entre fausses alarmes fixé. Plusieurs situations cliniques peuvent être envisagées : la seconde partie tente de les présenter de manière formelle. La troisième et quatrième partie résument les outils méthodologiques : d'une part la méthode d'identification des modèles (utilisation de l'algorithme des moindres carrés exacts en treillis dans sa version pré-fenêtrée, écrit sous forme vectorielle décomposée) et d'autre part l'approche retenue pour la segmentation des signaux (la méthode dite des deux modèles [3], l'un sur un horizon long croissant, l'autre court fixe glissant).

ABSTRACT

In an epileptic context, the multidimensional EEG signals show, on some characteristic segments in one or more channels, abrupt changes of dynamics which must be detected. Segmentation methods based on a vectorial modelling and related algorithms can be designed to take account of the channels cross correlations due to synchronisation phenomena. In this paper, a vectorial detection method is compared to an independent scalar processing scheme coupled to a simple fusion rule.

Enfin, la dernière section présente des situations cliniques simulées et discute des performances obtenues pour les approches vectorielle et scalaire.

POSITION DU PROBLEME

On considère ici le cas général des signaux vectoriels (Enregistrements multi-voies) pour lesquels les changements ne concernent pas obligatoirement exhaustivement toutes les voies et ne sont pas obligatoirement synchrones.

Plus précisément en considérant par exemple un signal vectoriel comportant M voies, on peut distinguer le cas où une rupture des statistiques marginales (voie par voie) intervient simultanément sur 2, 3, où les M voies (les voies restantes ne présentant pas de rupture sur l'intervalle d'observation) du cas où interviennent des ruptures non situées au même instant (asynchrones) sur des voies distinctes, le décalage temporel les séparant étant petit devant l'intervalle temporel moyen entre deux ruptures successives sur une voie quelconque (ce qui permet de les associer au même événement physiologique).

Plus formellement, on considère un signal discret vectoriel $Y(t) = [Y_1(t) \dots Y_M(t)]^T$ comportant M voies généralement corrélées et l'apparition possible d'une rupture de statistique sur un sous ensemble I de l'ensemble $\{1..M\}$ des numéros de voie. Ceci est formalisé en introduisant un paramètre vectoriel θ et la collection d'instant $\{t_i, i \in I\}$:



- $\forall i \in I, r_i$ est l'instant de rupture du modèle scalaire de la voie i ;
- quand t atteint r_i , on considère passer du modèle $\theta_{i,0}$ au modèle $\theta_{i,1}$ pour la voie i seule ;
- les r_i peuvent être considérés comme égaux (ruptures simultanées) ou bien distincts dans le cas le plus général (ruptures décalées) avec des écarts $\Delta_{ij} = r_i - r_j$ ne pouvant excéder une valeur fixée ; dans ces deux cas, on parlera de "rupture vectorielle" ;
- $\forall t < \inf\{r_i\} : \theta = \theta_0 ; \forall t > \sup\{r_i\} : \theta = \theta_1 ; \forall t \in [\inf\{r_i\} \dots \sup\{r_i\}] : \theta$ évolue de θ_0 vers θ_1 .

Il s'agit bien ici de détecter la "rupture vectorielle" tout en estimant I et les r_i pour $i \in I$. Ceci a été envisagé en comparant l'utilisation d'une statistique à valeur scalaire, exploitant au mieux et simultanément les M voies, à celle de M statistiques locales obtenues séparément sur chaque voie avant d'être "fusionnées" pour obtenir les décisions requises. Dans le cas où les Δ_{ij} sont nuls et où le cardinal de I est égal à M , la première approche (sous réserve d'une modélisation adaptée) s'impose par son caractère optimal. Dans le cas contraire, la prise en compte des Δ_{ij} ne semble pas être possible simplement dans des algorithmes vectoriels qui, si le cardinal de I est trop inférieur à M , peuvent voir à priori leurs performances se dégrader.

On a expérimenté ici quelques algorithmes "en ligne" susceptibles de répondre au problème posé. Le critère de performance retenu pour comparer des algorithmes séquentiels de détection est défini comme suit :

1) On considère le temps écoulé moyen $E\{\tau_F\}$ pour atteindre, à partir de la fin T_D d'une période d'initialisation, le premier instant $T_D + \tau_F$ où le test (répété séquentiellement) accepte l'hypothèse $H1$ (une rupture présente entre T_D et l'instant courant) alors qu'on se trouve sous l'hypothèse $H0$ (il n'y a pas eu, en réalité de rupture entre T_D et l'instant courant). On admet que la probabilité pour que ce premier instant soit fini est égale à 1.

2) En considérant qu'on est sous l'hypothèse $H1$ à partir de $T_R > T_D$, T_R correspondant donc à l'instant de rupture, on considère le premier instant supérieur à T_D d'acceptation de l'hypothèse $H1$. En notant T_A cet instant, on introduit la moyenne du temps de retard à la détection conditionnellement à ce que $\tau_F > T_R - T_D$:

$$E\{T_A - T_R / \tau_F > T_R - T_D\} = E\{\tau_D / \tau_F > T_R - T_D\}$$

Un critère couramment adopté dans la littérature [2] pour traduire le fait qu'un algorithme 1 est meilleur qu'un algorithme 2 est alors :

$$E\{\tau_{F1}\} = E\{\tau_{F2}\} \Rightarrow E\{T_{A1} - T_R / \tau_{F1} > T_R - T_D\} < E\{T_{A2} - T_R / \tau_{F2} > T_R - T_D\}$$

Enfin, nous ne chercherons pas ici à développer plus en avant une forme de modèle vectoriel qui caractérise une bonne représentation de l'asynchronisme des ruptures et de l'intercorrélation entre voies. Nous ne considérerons en conséquence qu'une seule forme de représentation du signal à partir de modèles autorégressifs scalaire ou vectoriels : ceux que des études respectivement antérieures [1] ou récentes [5] ont validés.

REPRESENTATION DES SEGMENTS STATIONNAIRES

La modélisation A.R. (autorégressive) classique pour une voie est donnée par :

$$y(t) = - \sum_{i=1}^p a_i \cdot y(t-i) + e(t), \quad a_i \in \mathbb{R}$$

où p ici est l'ordre du modèle et où $e(t)$ est une suite décorrélée centrée. L'extension retenue ici de la modélisation A.R pour un signal vectoriel d'ordre p comportant M voies scalaires consiste à écrire :

$$\sum_{i=0}^p \Delta_i \cdot Y_{t-i} = E_t$$

où Δ_0 une matrice triangulaire strictement inférieure (régression spatiale sur les capteurs d'indice(s) inférieur(s) et les $\Delta_i, 1 \leq i \leq p$ sont des matrices carrées d'ordre M qui paramétrisent le modèle, Y_t et E_t sont les vecteurs des observations et des résidus de prédiction de dimension M . En pratique, les coefficients matriciels $\Delta_i, 0 \leq i \leq p$ peuvent être estimés au moyen des algorithmes de moindres carrés exacts suivant les méthodes prewindowed (PW), de covariance (Growing Memory Covariance (GMC), et de la fenêtre glissante (Sliding Window Covariance (SWC). Les deux algorithmes (GMC, SWC) ayant une complexité double en termes de calcul, notre choix pour cette étude s'est porté sur l'algorithme PW, écrit sous sa forme normalisée vectorielle décomposée [7], avec un facteur d'oubli λ . L'intérêt de la méthode est de ne faire intervenir aucune opération matricielle et revient à associer des cellules élémentaires (représentant la transformée de Yule) pour réaliser la minimisation.

METHODE RETENUE DE DETECTION DE RUPTURE DE MODELES

Quelques soient les cas de figure monodimensionnel ou multidimensionnel, l'approche qui consiste à comparer deux modèles, l'un estimé sur une fenêtre globale et l'autre évalué sur un horizon court glissant, a été retenue [3].

a) Cas monodimensionnel.

Partant de l'idée de comparaison entre un modèle de filtre long et un modèle de filtre court, on aboutit au calcul de la distance entre les deux modèles. C'est la divergence de Kullback entre les lois de probabilités conditionnelles qui a été retenue. L'ensemble des données initiales est transformé par :

$$w_i = - \int p^0(y_i | y^{t-1}) \cdot \text{Log} \frac{p^1(y_i | y^{t-1})}{p^0(y_i | y^{t-1})} dy_i + \text{Log} \frac{p^1(y_i | y^{t-1})}{p^0(y_i | y^{t-1})}$$

où p^0 est la densité de probabilité conditionnelle avant la rupture, p^1 est la densité de probabilité conditionnelle après la rupture, et $y^{t-1} = (y_{t-1}, y_{t-2}, \dots, y_1)$ est le vecteur des observations passées. L'espérance conditionnelle étant notée E_0 avant la rupture et E_1 après la rupture, on a :

$$E0(w_t|y^{t-1})=0, \quad E1(w_t|y^{t-1})=J(p^0, p^1) > 0$$

où $J(p_0, p_1)$ représente la divergence Kullback entre p_0 et p_1 .

Tout changement de modèle influence la statistique w_t qui, de valeur moyenne nulle avant la rupture, devient à moyenne positive après la rupture. Le changement de comportement de la statistique sera donc traduit par w_t ; cette dernière étant une suite blanche, un algorithme du type somme cumulée peut être utilisé pour détecter le changement de modèle [3].

Pour un modèle autorégressif gaussien, cette statistique devient :

$$w_t = \frac{1}{2} \left[-2 \cdot \frac{e_t^0 \cdot e_t^1}{\sigma_1^2} + \left(1 + \frac{\sigma_0^2}{\sigma_1^2}\right) \cdot \left(\frac{e_t^0}{\sigma_0}\right)^2 - 1 + \frac{\sigma_0^2}{\sigma_1^2} \right] \quad (1)$$

e_t^0 , σ_0^2 sont respectivement le résidu et la puissance du modèle long et e_t^1 , σ_1^2 le résidu et la puissance du modèle court. En pratique, les deux modèles sont estimés sur deux fenêtres (l'une longue croissante, l'autre courte glissante) et les quantités e_t^0 , σ_0^2 , e_t^1 , σ_1^2 sont remplacées dans (1) par leurs estimations.

Par rapport à l'approche originale proposée par M. Basseville, les deux algorithmes d'identification des modèles sont ici récursifs et exacts. Le glissement du modèle court est obtenu en choisissant un facteur d'oubli $\lambda_1 < 1$, le modèle global ou long est lui identifié en retenant $\lambda_0 > \lambda_1$ et égal à 1.

b) Cas vectoriel (instants de ruptures simultanés)

Le cas scalaire se généralise facilement au cas vectoriel. Partant du même principe que dans le cas monodimensionnel et dans l'hypothèse gaussienne, la densité de probabilité conditionnelle est ici telle que :

$$p^i(Y_t | Y^{t-1}) = \frac{1}{\sqrt{2 \cdot \Pi \cdot \det(R_t^i)}} \cdot e^{-\frac{1}{2} E_t^i T \cdot R_t^i^{-1} \cdot E_t^i}$$

où $E_t^i = Y_t - \hat{Y}_t^i$ est le vecteur des résidus (des innovations) du filtre correspondant au modèle i ($=0, 1$) et R_t^i est la matrice de covariance des résidus.

En suivant la même approche que dans le cas scalaire, la statistique w_{vt} sous l'hypothèse d'un modèle autorégressif gaussien vectoriel, s'écrit :

$$w_{vt} = \frac{1}{2} E_t^0 T \left[(R_t^0)^{-1} + (R_t^1)^{-1} \right] \cdot E_t^0 - E_t^1 T (R_t^1)^{-1} \cdot E_t^0 + \frac{1}{2} \ln \left[(R_t^1)^{-1} \cdot (R_t^0) \right] \cdot \frac{M}{2}$$

où M est la dimension de l'espace signal, c'est à dire le nombre de composantes du vecteur des résidus. i.e. le nombre de capteurs. L'orthogonalité entre les composantes du vecteur des résidus permet de construire, en multidimensionnel, une statistique pour la segmentation sans aucune inversion de matrice. La statistique w_{vt} calculée à l'aide des matrices de covariances R_t^0 et R_t^1 des résidus qui sont diagonales dans le cas de l'algorithme vectoriel utilisé ici peut donc s'écrire :

$$w_{vt} = \frac{1}{2} \sum_{i=1}^M \left[\frac{e_i^0(t)^2 \cdot [\sigma_i^2 + \sigma_i^2]}{\sigma_i^2 \cdot \sigma_i^2} - 2 \frac{e_i^1(t) \cdot e_i^0(t)}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_i^2} - 1 \right] \quad (2)$$

ce qui montre que w_{vt} est en fait de la forme $\sum_{i=1}^M w_i(t)$ où les $w_i(t)$ sont des quantités scalaires calculées par la formule (1), composante par composante.

c) Fusion de statistiques monodimensionnelles

On remarque que dans le cas où les M voies sont statistiquement indépendantes, la régression vectorielle se ramène à M régressions scalaires et la formule ci dessus permet alors de fusionner de manière exacte les M statistiques scalaires indépendantes w_{sit} $i=1..M$ - correspondant à l'utilisation de M algorithmes scalaires en introduisant la statistique :

$$w_{Ft} = \frac{1}{2} \sum_{i=1}^M w_{sit} \quad (3)$$

Une généralisation possible de (2) et (3) est :

$$w_{FIt} = \sum_{i \in I} w_{sit}, \quad w_{vIt} = \sum_{i \in I} w_{it}$$

correspondant à la fusion des seules voies d'indices i dans I . Le choix de I peut relever d'informations a priori issues d'un module de traitement de niveau plus élevé. Nous n'étudierons pas ici le problème de l'estimation de I .

Même dans le cas où les M voies ne sont pas décorréelées on peut envisager d'utiliser la statistique (3), plus simple à calculer que (2). Il est clair, d'après [3], qu'avant la rupture et pour un ensemble I donné :

- 1) w_{vIt} est un bruit blanc centré,
- 2) w_{FIt} est un bruit blanc centré si les voies $i \in I$ sont décorréelées,
- 3) w_{FIt} est centré mais non exactement blanc a priori si les voies $i \in I$ ne sont pas toutes décorréelées.

La sommation pour $t \geq t_D$ de ces quantités conduit à des suites aléatoires qui peuvent être approchées par un mouvement brownien dans les cas 1) et 2). Bien que cela puisse être sensiblement inexact, on propose dans le cas 3) de retenir cette hypothèse simplificatrice.

d) Utilisation d'un algorithme du type Hinkley

Pour les trois cas, on a choisi de manière classique, d'utiliser l'algorithme de Page-Hinkley [2] pour détecter le saut de moyenne à l'instant de rupture (dans le signal original) pour les statistiques correspondantes. Dans [6], on trouve une évaluation théorique de $Eo\{w_t\}$ et $Eo\{w_t^2\}$ dans le cas scalaire et où le modèle θ_1 est estimé sur un horizon de N points. Pour certaines approximations raisonnables, on aboutit à $Eo\{w_t\} = 0$. On constate expérimentalement qu'en fait $Eo\{w_t\} = d_0 > 0$, où d_0 est d'autant plus important que la dimension de θ est élevé, pour un horizon du modèle court donné. On a choisi en conséquence l'approche qui consiste à estimer $Eo\{w_t\}$ et $VAR\{w_t\}$ pour tout $t < t_D$ afin d'appliquer l'algorithme de Hinkley comme suit (w_t désignant w_{vIt} , w_{FIt} avec $I = \{1..M\}$ où w_{sit}) :



$$\left[\sum_{t=t_D}^n (w_{.t} - \widehat{E}_0\{w_{.t}\} - \delta) - \min_{r \leq n} \sum_{t=t_D}^r (w_{.t} - \widehat{E}_0\{w_{.t}\} - \delta) \right] > h$$

où δ est un paramètre positif qui s'interprète comme le plus petit écart que l'on désire détecter sur $E\{w_{.t}\}$, T_D étant la fin de la période d'initialisation ; h est le seuil de détection qui, conjointement avec $E_0\{w_{.t}\}$, $VAR\{w_{.t}\}$ et δ , détermine le temps moyen F s'écoulant entre l'initialisation et la première fausse alarme [2] :

$$F = \frac{[e^{2\gamma_0 h} - 1] / \gamma_0 - h}{\gamma_0} \quad \gamma_0 = [\mu_0] / \widehat{VAR}\{w_{.t}\} \quad \text{où } \mu_0 = \widehat{E}_0\{w_{.t}\} + \delta$$

Cette formule permet de déterminer une valeur h correspondant à une valeur donnée de F avec une précision liée à celle des estimateurs \widehat{E}_0 et \widehat{VAR} qui est en principe suffisamment bonne ici.

RESULTATS ET DISCUSSION

a) expérimentation

Les performances que l'on peut attendre, à partir des algorithmes proposés, ont été étudiées. Dans tous les cas, l'instant de rupture est situé en $T_R=4096$ et le facteur d'oubli du modèle court θ est égal à $\lambda = 1-1/256$ pour les algorithmes scalaires et vectoriels. Quatre types de rupture de modèle ont été simulées à partir de modèles identifiés sur des signaux réels, segmentés manuellement. Ceci permet de reproduire des situations réalistes relativement au domaine d'étude envisagé, l'EEG. Ces situations réelles ont ensuite été simulées en utilisant des modèles estimés avant et après les ruptures observées. Ces quatre situations, construites sur un ensemble de 3 voies (V1, V2, V3) sont :

- S1** : V1, V2 et V3 corrélées pour $t < T_R$ et pour $t > T_R-1$;
- S2** : V1, V2 et V3 décorréliées pour $t < T_R$ et pour $t > T_R-1$;
- S3** : V1, V2 et V3 corrélées pour $t < T_R$ et corrélées pour $t > T_R-1$;

Dans les trois cas **S1**, **S2** et **S3**, une rupture simultanée est présente sur V1, V2 et V3.

S4 : V1, V2, V3 corrélées pour $t > T_R$ et $t > T_R-1$; les voies V1 et V3 sont inchangées quand t atteint T_R ; la voie V2 est modifiée pour $t > T_R-1$ par ajout d'un signal décorrélié de V1 et V3.

Cliniquement, **S1** correspond à un changement de rythme en cours de crise, **S2** à un changement de rythme dans une activité de fond, **S3** à un début de crise et **S4** a été inventée pour obtenir un résultat lorsque seul un sous ensemble I des voies est le siège d'une rupture.

De manière classique [2], les retards à la détection ont été considérés conditionnellement à l'absence de fausse alarme entre T_D et T_R . Toute fausse alarme survenant entre T_D et T_R a mené à une réinitialisation de l'algorithme de Hinkley. Différentes valeurs de F ont été imposées théoriquement par la relation indiquée plus haut. δ a été choisi heuristiquement en fonction de l'ordre de grandeurs des distances entre modèles rencontrés sur les signaux étudiés. En assimilant, ces dernières à la différence Δ entre les valeurs moyennes avant et après T_R pour les statistiques $w_{.t}$ utilisées, le paramètre δ a été fixé empiriquement à $\delta = \Delta_{min}/3$ pour les algorithmes scalaires et Δ_{min} pour les autres ; Δ_{min} est la plus petite valeur de D rencontrée sur l'ensemble des w_{sjt} , $i=1, 2, 3$, toutes situations confondues.

b) Resultats obtenus

En notant D_{Vi} , D_V , D_F les temps de retard moyen à la détection obtenus respectivement pour l'algorithme scalaire sur la voie i , l'algorithme vectoriel et la fusion des algorithmes scalaires et en introduisant $D_S = \min\{D_{Vi}\}$, la comparaison de ces quantités à taux de fausse alarme fixé montre que :

- pour **S1** : $D_V < D_F < D_S$, pour **S2** : $D_V \# D_F < D_S$
- pour **S3** : $D_V \# D_F < D_S$, pour **S4** : $D_V < D_S < D_F$

Ces résultats montrent que dans les cas où les voies sont corrélées avant et/ou après T_R , l'algorithme vectoriel se comporte mieux que les 2 autres. Comme on pouvait le prévoir, la fusion des algorithmes scalaires conduit à des performances quasi-identiques à celles obtenues avec l'algorithme vectoriel, lorsque les voies sont décorréliées, mais deviennent inférieures lorsque la rupture n'a lieu que sur une seule voie. La figure 1 reproduit les résultats obtenus pour les situations **S3** et **S4**. On peut remarquer que sur ces exemples les performances les moins bonnes, obtenues quand on utilise les algorithmes scalaires avec des voies corrélées, restent cependant acceptables en pratique.

REFERENCES

- [1] Appel U., Brandt A., A comparative study of three sequential time series segmentation algorithms, Signal Processing, N° 6, 1984, pp.45-60.
- [2] Basseville M. Edge detection using sequential methods for change in level - Part II Sequential detection of change in mean, IEEE ASSP, Vol 29, N° 1, FEB 1981, pp 32-50.
- [3] Basseville M., Benveniste A., Detection of abrupt changes in signals and dynamical systems, LNCIS n° 77, Springer-Verlag, Berlin, 1986.
- [4] Bodenstern G., Praetorius H.L., Feature extraction from the electroencephalogram by adaptive segmentation, Proc. IEEE, Vol 65, N° 5, may 1977, pp 642-652.
- [5] Gath I., Feuerstein C., Pham D. T., Rondouin G., On the tracking of rapid dynamic changes in seizure EEG., IEEE BME, Vol 30, N° 9, Sept 92, pp 952-958.
- [6] Basseville M., Sequential detection of abrupt changes in spectral characteristics of digital signals, Rapport de recherche IRISA.
- [7] Jazaerli S., Contribution à la décomposition des algorithmes de moindres carrés vectoriels normalisés ; application au traitement d'antennes, Thèse de l'Université de Rennes I, Juin 1987.

Figure 1 : Retard à la détection en fonction du temps moyen entre Fausses alarmes

