

MODELE DE MARKOV CACHE HARMONIQUE POUR L'ETUDE DES SIGNAUX EEG

Emilie VILLARON, Bruno TORRESANI,
LATP, Marseille

September 13, 2010

Plan

- 1 Introduction
- 2 Modèle de Markov caché harmonique
- 3 Inference
- 4 Application sur des données réelles

Une famille de signaux vectoriels non-stationnaires

Les signaux électrophysiologiques présentent les caractéristiques suivantes :

- Ils sont **non stationnaires**, avec des changements lisses ou plus abruptes dans leur contenu spectral
- Ce sont des signaux **vectoriels** présentant des corrélations significatives entre les différents composants, ces dernières pouvant varier avec le temps

Les représentations temps-fréquence sont un choix naturel de description de tels signaux. Comment modéliser ensuite

- les changements abruptes décrits ci-dessus
- les corrélations entre les composants

Notre choix : base de Fourier locale (MDCT) n'introduisant pas de corrélations "parasites" (à l'inverse des repères de Gabor par exemple).

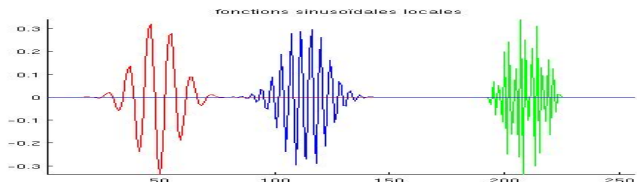
Représentation d'un signal monocapteur dans le plan temps-fréquence

Soit $\mathcal{B} = \{u_{tf}, (t, f) \in \Lambda\}$ une base de Fourier locale de dimension finie :

$$u_{tf}[n] = \sqrt{\frac{2}{L}} w[n - tL] \cos\left(\pi(f + 1/2)\frac{n - tL}{L}\right)$$

$t = 0, \dots, N_t - 1$ (resp. $f = 0, \dots, N_f - 1$) représentant l'indice temporel (resp. fréquentiel).

Ici w est une fenêtre régulière permettant la localisation fréquentielle des atomes de Fourier.



Représentation d'un signal monocapteur dans le plan temps-fréquence (2)

Décomposition d'un signal mono-capteur Tout signal $x \in \mathbb{R}^L$ se décompose sur la base \mathcal{B} :

$$x = \sum_{t=1:N_t-1, f=0:N_f-1} \langle x, u_{tf} \rangle u_{tf}$$

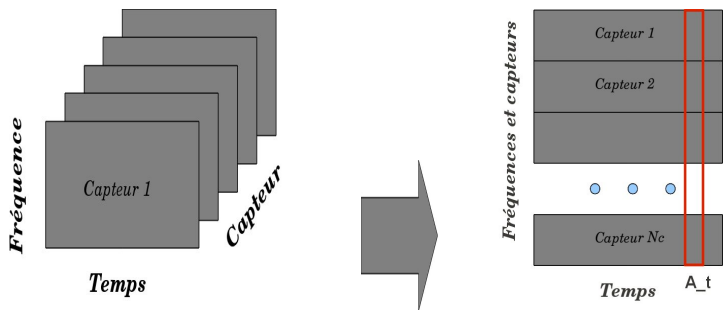
où $L = N_t \times N_f$

Cas d'un signal multicapteur: Pour chaque canal c ,

$$x^c = \sum_{t=1:N_t-1} \sum_{f=0:N_f-1} \langle x^c, u_{tf} \rangle u_{tf}$$

et le signal complet x est donc caractérisé par un tableau à 3 dimensions de coefficients

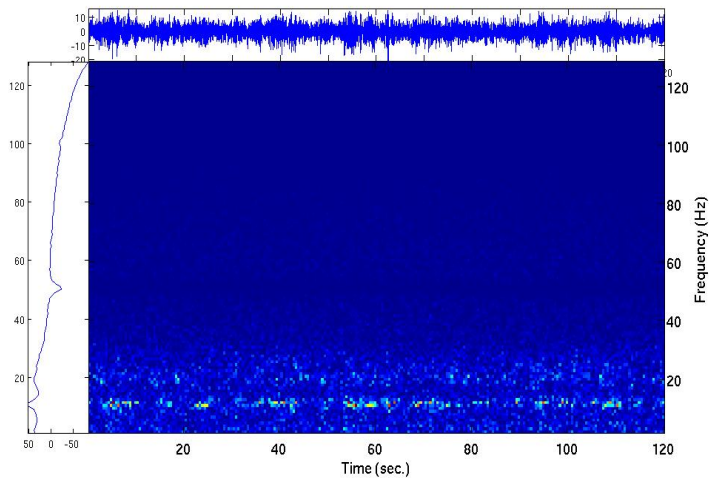
$$\alpha_{tf}^c = \langle x^c, u_{tf} \rangle$$



Le modèle de Markov caché harmonique que nous allons développer est un modèle statistique de l'évolution temporelle de ces coefficients.

$$\{\langle x^c, u_{\bullet f} \rangle, c = 1, \dots, N_c, f = 1, \dots, N_f\}$$

Exemple de décomposition d'un signal EEG mono-capteur



Plan

- 1 Introduction
- 2 Modèle de Markov caché harmonique
- 3 Inference
- 4 Application sur des données réelles

Le modèle

On considère N_c signaux (representant les observations de N_c capteurs) de la forme :

$$\underline{x}^c = \sum_{t,f} a_{tf}^c u_{tf}$$

On définit $\underline{A}_t = \{(a_{tf}^c, f = 0, \dots, N_f - 1), c = 1, \dots, N_c\}$. La distribution de ces vecteurs est gouvernée par un état caché variant en temps

$X_t \in \{1, \dots, N_s\}$:

- $\psi_{\underline{A}_t}(\underline{a}|X_t = k) = \mathcal{N}(\underline{a}; 0, \Sigma_k)$: **corrélations entre fréquences et capteurs**
- $\mathbb{P}(\underline{A}_0 \in E_0, \underline{A}_1 \in E_1, \dots, \underline{A}_{N_t-1} \in E_{N_t-1} / X_0 = k_0, \dots, X_{N_t-1} = k_{N_t-1}) = \prod_{t=0}^{N_t-1} \mathbb{P}(\underline{A}_t \in E_t / X_t = k_t)$
- Les états cachés (qui dépendent uniquement de l'indice temporel et sont communs à tous les capteurs) forment une chaîne de Markov caractérisée par sa matrice de transition π :

$$\pi_{i,j} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

Comment éviter la “malédiction de la dimension”

Habituellement, $N_f \simeq 30$ et $N_c \simeq 20$: $\Sigma_k \in \mathcal{M}_{600 \times 600}(\mathbb{R})$.

▷ Nécessité d'*hypothèses simplificatrices* pour réduire la dimension de l'espace.

On utilise des matrices de covariance sous forme de produits de Kronecker :

$$\Sigma_k = \Sigma_k^{(c)} \otimes \Sigma_k^{(f)}$$

où

- ▶ $\Sigma_k^{(c)} \in \mathcal{M}_{N_c \times N_c}(\mathbb{R})$ représente les corrélations entre plages de fréquences, et
- ▶ $\Sigma_k^{(f)} \in \mathcal{M}_{N_f \times N_f}(\mathbb{R})$ regroupe les covariances entre capteurs

Composante par composante :

$$\mathbb{E} \left\{ a_f^c \cdot a_{f'}^{c'} \right\} = \Sigma_{cc'}^{(c)} \Sigma_{ff'}^{(f)}$$

Plan

- 1 Introduction
- 2 Modèle de Markov caché harmonique
- 3 Inference
- 4 Application sur des données réelles

Quelques questions ... et quelques réponses

- Etant donné un jeu de coefficients de synthèse $(\underline{A}_t)_{t=0,\dots,N_t-1}$, estimer, au sens du maximum de vraisemblance
 - ▶ la matrice de transition de la chaîne de Markov Π
 - ▶ les matrices de covariance Σ_k , ou les couples $(\Sigma_k^{(c)}, \Sigma_k^{(f)})$ quand Σ_k est un produit de Kronecker

▷ *algorithme EM*

- Une fois connus ces paramètres MAP, quelle est la séquence d'états cachés la plus vraisemblable ?

$$(X_0^{MAP}, \dots, X_{N_t-1}^{MAP}) = \arg \max_{l=(l_0, \dots, l_{N_t-1})} \mathbb{P}(X = l | \underline{A}_t, t = 0, \dots, N_t - 1)$$

▷ *algorithme de Viterbi*

Etape “Espérance” (1)

Pour $t \in 0, \dots, N_t - 1$, on définit les quantités suivantes:

$$\alpha_t^k = \mathbb{P}(X_t = k / (\underline{A}_r)_{r=0, \dots, t}) \times L_t$$

où L_t est la vraisemblance des observations jusqu'à l'instant t .

$$\beta_t^k = \mathbb{P}(\underline{A}_{t+1}, \underline{A}_{t+2}, \dots, \underline{A}_{N_t-1} | X_t = k)$$

pour $k = 1, \dots, N_s$ et $t = 0, \dots, N_t - 1$.

Ces quantités peuvent être calculées à partir des observations grâce aux équations “forward” et “backward”.

Etape “Espérance” (2)

Une fois calculées ces quantités, on estime la **distribution des transitions** (X_t, X_{t+1}) pour $t = 0, \dots, N_t - 2$ conditionnellement aux observations jusqu'à l'instant $N_t - 1$

$$\mathbb{P}(X_t = k, X_{t+1} = l / (\underline{A}_t)_{t=0:N_t-1}) = \frac{1}{L_{N_t-1}} \alpha_t^k \pi_{k,l} \psi(\underline{A}_{t+1}) \beta_{t+1}^l$$

où

$$L_{N_t-1} = \sum_{k=1}^{N_s} \alpha_t^k \beta_t^k \text{ pour tout instant } t$$

La **distribution de l'état caché** X_t , $0 \leq t \leq N_t - 2$ conditionnellement aux observations jusqu'à l'instant $N_t - 1$ vérifie

$$\mathbb{P}(X_t = k / (\underline{A}_r)_{r=0:N_t-1}) = \frac{1}{L_{N_t-1}} \alpha_t^k \beta_t^k$$

Formules de re-estimation de Baum-Welch (1)

Estimation des paramètres de la chaîne

Dans les deux cas,

$$\hat{\nu}_k = \mathbb{P}(X_0 = k / (\underline{A}_t)_{t=0:N_t-1})$$

$$\hat{\pi}_{k,l} = \pi_{k,l} \frac{\sum_{t=0}^{N_t-2} \alpha_t^k \times \psi_l(\underline{A}_{t+1}) \times \beta_{t+1}^l}{\sum_{t=0}^{N_t-2} \alpha_t^k \beta_t^k}$$

Quand les matrices de covariance sont données sous la forme d'un produit de Kronecker, le calcul des densités conditionnelles est simplifié par le fait que $\Sigma_k^{-1} = (\Sigma_k^{(c)})^{-1} \otimes (\Sigma_k^{(f)})^{-1}$ et $\det(\Sigma_k) = \det(\Sigma_k^{(c)})^{N_f} \times \det(\Sigma_k^{(f)})^{N_c}$

Estimation des matrices de covariance (forme globale)

$$\hat{\Sigma}_k = \frac{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k / (\underline{A}_r)_{r=0:N_t-1}) \underline{A}_t \ ^t \underline{A}_t}{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k / (\underline{A}_r)_{r=0:N_t-1})}$$

Formules de re-estimation de Baum-Welch (2)

Estimation des matrices de covariance sous forme d'un produit tensoriel: $\Sigma = \Sigma^{(c)} \otimes \Sigma^{(f)}$

La re-estimation des matrices de covariance dans l'étape M est obtenue par une procédure de maximisation alternée :

on considère les matrices $\underline{M}_t = (\underline{a}_{t,f}^c, f = 0, \dots, N_f - 1 \quad c = 1, \dots, N_c)$ de taille $N_f \times N_c$ et

on itère les pas suivants :

- Estimation de $\Sigma^{(c)}$ étant donné $\Sigma_k^{(f)}$:

$$\widehat{\Sigma_k^{(c)}} = \frac{1}{N_f} \frac{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k) \quad {}^t \underline{M}_t \quad (\Sigma_k^{(f)})^{-1} \quad \underline{M}_t}{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k)}$$

- *Normalisation:*

$$\widehat{\Sigma}_k^{(c)} = \widehat{\Sigma}_k^{(c)} / \|\widehat{\Sigma}_k^{(c)}\|_F ,$$

où $\|\cdot\|_F$ désigne la norme de Frobenius.

- *Estimation de $\Sigma^{(f)}$ pour $\widehat{\Sigma}_k^{(c)}$ fixée:*

$$\widehat{\Sigma}_k^{(f)} = \frac{1}{N_c} \frac{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k) \underline{M}_t (\Sigma_k^{(c)})^{-1} {}^t \underline{M}_t}{\sum_{t=0}^{N_t-1} \mathbb{P}(X_t = k)}$$

Validation sur des simulations numériques

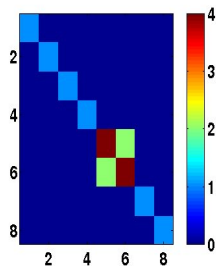
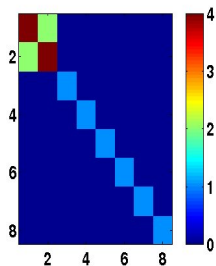
Des vecteurs de coefficients de synthèse ont été simulés selon le modèle décrit (matrices de covariance sous forme de produit de Kronecker) avec les paramètres $N_c = 8$, $N_t = 2000$, $N_f = 10$ and $N_s = 2$

pour différentes valeurs de la probabilité de transition $p_{00} = P(X_{t+1} = 0 | X_t = 0)$ in $[0.75; 0.95]$.

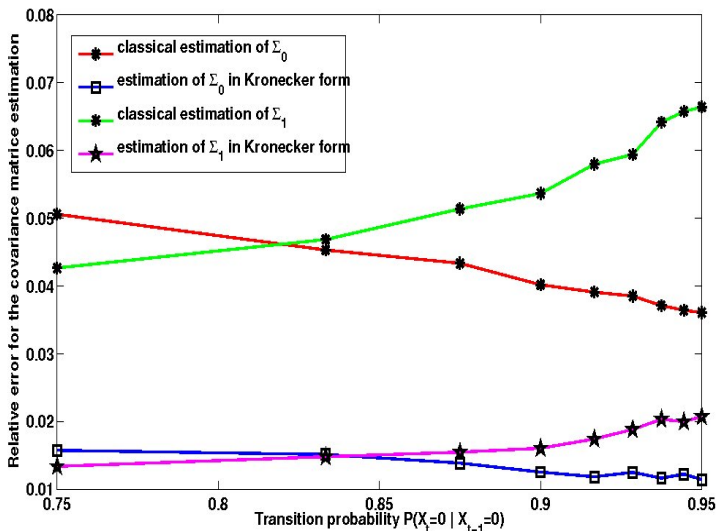
et nous avons calculé les erreurs relatives d'estimation $\frac{\|\hat{\Sigma} - \Sigma\|_F}{\|\Sigma\|_F}$

Remarque: p_{00} contrôle la longueur moyenne des plages d'état 0 .

Matrices de covariance "capteurs" pour l'état '0' (gauche) et '1' (droite)



Evolution des erreurs d'estimation relatives en fonction de la probabilité de transition p_{00}

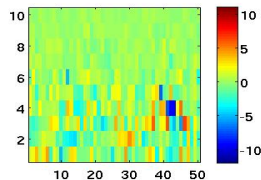
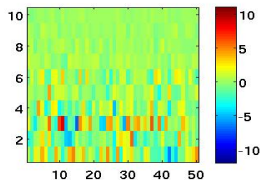
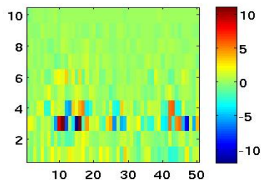
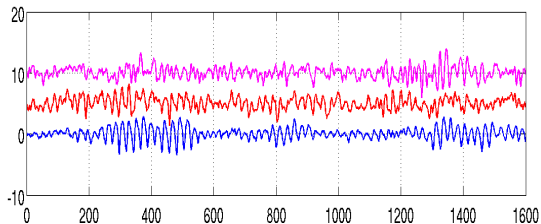
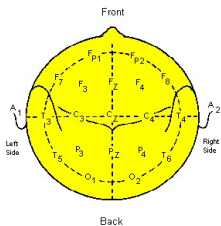


Plan

- 1 Introduction
- 2 Modèle de Markov caché harmonique
- 3 Inference
- 4 Application sur des données réelles

Les ondes alpha

- Durée moyenne d'une plage : 250-500 ms
- 8-12 Hz
- Région occipitale du scapl
- Traduit généralement un état de perte d'attention ou de relaxation

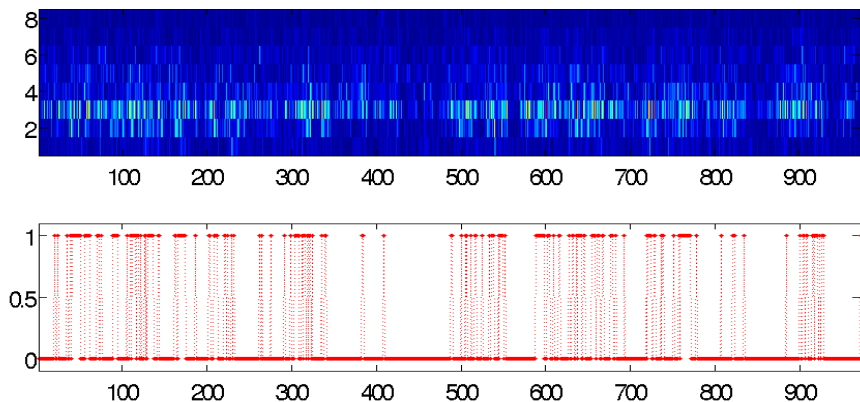


Une étude de cas

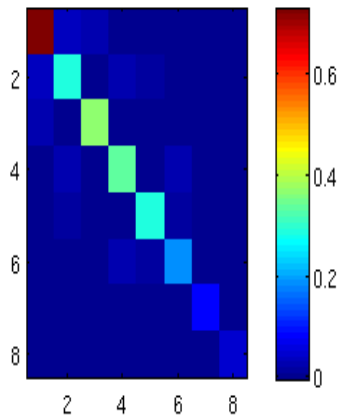
- Données : EEGs de repos issus du projet CODYSEP : patients atteints de sclérose en plaques (SEP) et sujets témoins (TE)
- Fréquence d'échantillonnage : $f_e = 256\text{Hz}$
- Nombre de plages de fréquences : $N_f = 8$
- Largeur des plages de fréquences : 4 Hz
- Résolution temporelle : 125 ms

- **Buts :**
 - ▶ Détection automatique et caractérisation des ondes alpha dans les signaux d'EEG de repos : $N_s = 2$ alpha ($k = 1$) / non-alpha ($k = 0$)
 - ▶ Discrimination entre les patients et les sujets sains

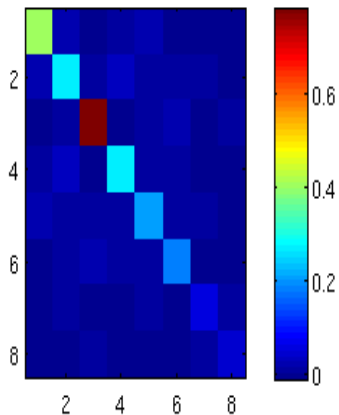
Coefficients MDCT pour un capteur fixé (haut) et séquence des états cachés estimée par l'algorithme de Viterbi à partir de l'ensemble des capteurs (bas)



Matrices de covariance "fréquences" pour l'état 0 (gauche) et l'état 1 (droite)

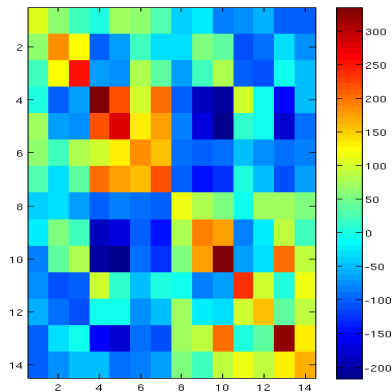
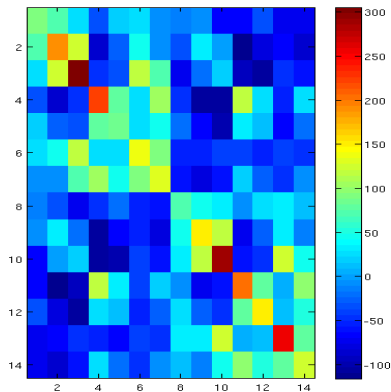


'état non alpha'

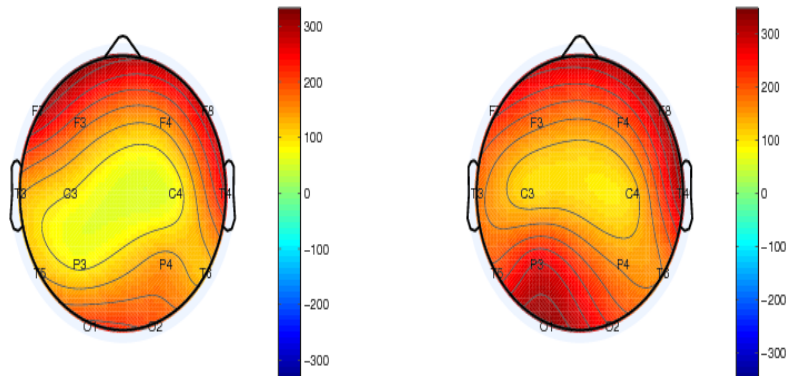


'état alpha'

Matrices de covariance "capteurs" pour les états non-alpha (gauche) et alpha (droite)



Représentations des variances des capteurs pour les états non- alpha (gauche) et alpha (droite)



Le problème clinique : la désynchronisation des ondes alpha

“in multiple sclerosis, both axonal damage and demyelination occur. Therefore it can be expected that the connectivity between the different regions in the brains of MS patients will be impaired compared to healthy controls”

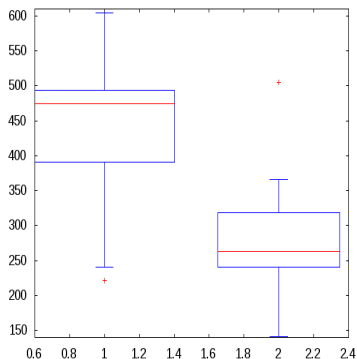
Cover et al. Multiple sclerosis patients show a highly significant decrease in alpha band interhemispheric synchronization measured using MEG

14 patients (SEP) and 16 sujets témoins (TEM) ont été sélectionnés

- ▶ les séquences d'états cachés sont estimées pour tous les sujets, séparément pour les capteurs gauches et les capteurs droits
- ▶ Pour chaque sujet, on calcule la distance de Hamming entre les séquences obtenues pour les capteurs gauches $X_t^{(L)}$ et droits $X_t^{(R)}$

$$D = \sum_t \left| X_t^{(L)} - X_t^{(R)} \right|$$

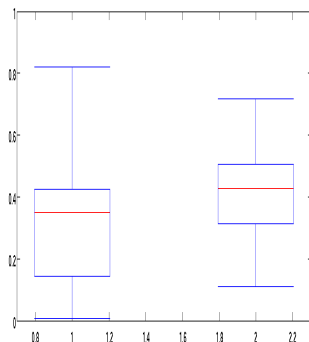
Boxplots des distances de Hamming entre les états cachés gauche et droite pour les patients (left) et les contrôles (SEP)



Pour comparer les deux distributions, on effectue un test de Mann-Whitney. La P-valeur $P = 0.0384$ correspondante confirme quantitativement l'hypothèse de deux distributions distinctes.

Approche basée sur la cohérence

A titre de comparaison, nous calculons aussi la cohérence entre les signaux gauche et droite (après filtrage passe-bande pour se focaliser sur la bande alpha).



- Les patients présentent des valeurs de cohérence plus faibles, ... mais
- la différence est beaucoup moins significative

Conclusion

- Nouveau modèle avec modélisation explicite de la non-stationnarité par l'utilisation d'un HMM
- Utilisation de matrices de covariance sous forme de produits de Kronecker : gain de temps et de stabilité
- Validation du modèle sur des données réelles : segmentation non-supervisée de signaux pertinente
- Validation de l'hypothèse de désynchronisation des ondes alpha de façon plus significative qu'avec les approches classiques.

Perspectives

- Application à d'autres types de signaux EEG : oscillations gamma, P300 (en vue d'applications BCI)
- Amélioration de la résolution temps-fréquence : utilisation de repères temps-fréquences.

Problème: l'indépendance conditionnelle des coefficients n'est plus vérifiée , il faut un nouvel algorithme.