

# Classification de signaux EEG et filtre CSP sparse.

**Marie Chavent**, Frédérique Faïta-Aïnseba, Pierrick Legrand,  
Laurent Vézard

Université de Bordeaux,  
Inria Bordeaux Sud-Ouest, équipe CQFD

Thèse de [Vézard, 2013 ]

# Plan

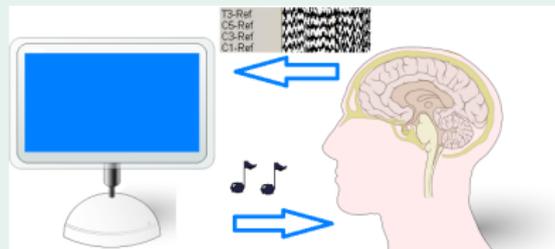
- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 Classification de signaux EEG avec filtre CSP
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

# Plan

- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 Classification de signaux EEG avec filtre CSP
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

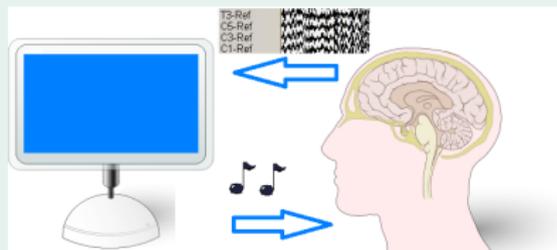
## Objectifs :

- Caractériser l'état de vigilance d'une personne à partir de son électro-encéphalogramme (EEG).
- Créer une interface cerveau-ordinateur.
- Influencer l'état de vigilance d'une personne à l'aide de musiques synthétisées.



## Objectifs :

- **Caractériser l'état de vigilance d'une personne à partir de son électro-encéphalogramme (EEG).**
- Créer une interface cerveau-ordinateur.
- Influencer l'état de vigilance d'une personne à l'aide de musiques synthétisées.

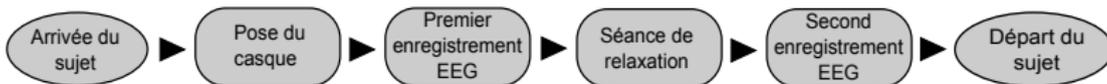


# Plan

- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 Classification de signaux EEG avec filtre CSP
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

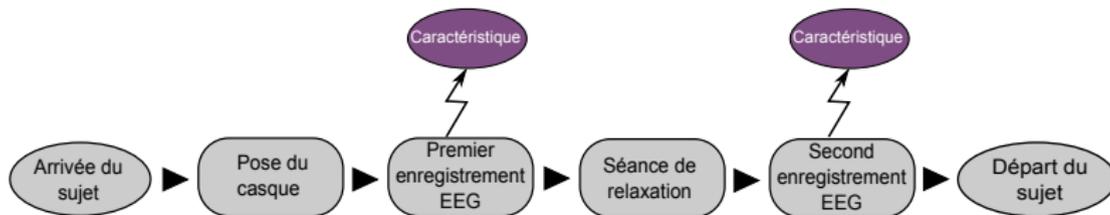
# Protocole d'acquisition

- Premier enregistrement EEG  $\implies$  sujet dans un état de vigilance haut (état dit "normal")
- Second enregistrement EEG  $\implies$  sujet dans un état de vigilance bas (état dit "relaxé")



# Protocole d'acquisition

- Premier enregistrement EEG  $\implies$  sujet dans un état de vigilance haut (état dit "normal")
- Second enregistrement EEG  $\implies$  sujet dans un état de vigilance bas (état dit "relaxé")

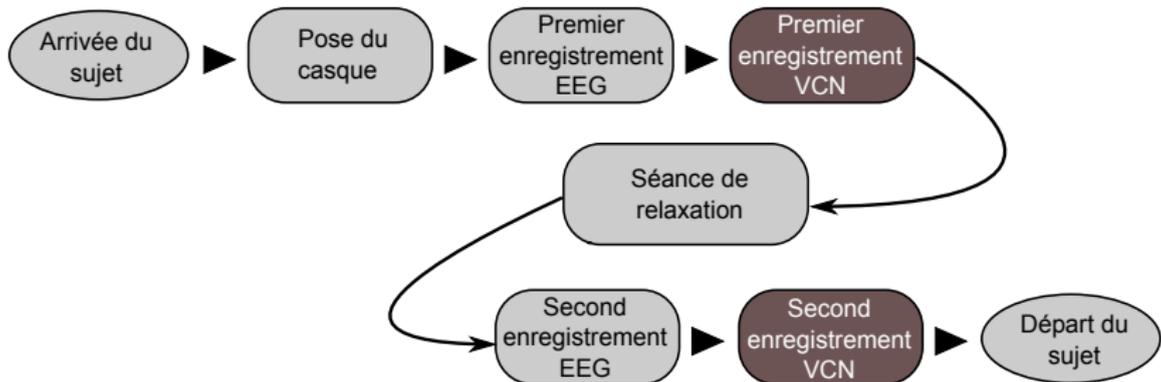


# Protocole d'acquisition

## Validation des données

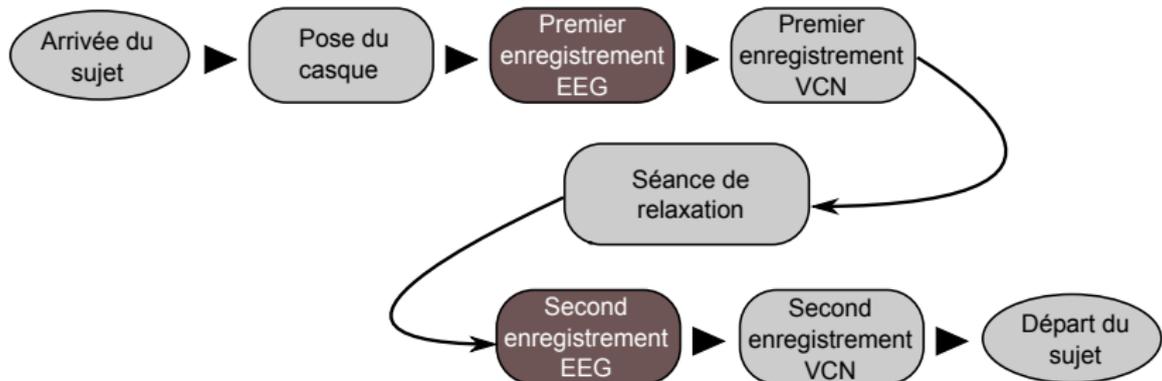
La séance de relaxation est elle efficace ?

# Protocole d'acquisition

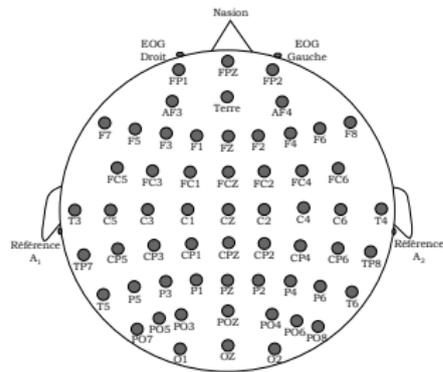


Test VCN

# Enregistrement EEG

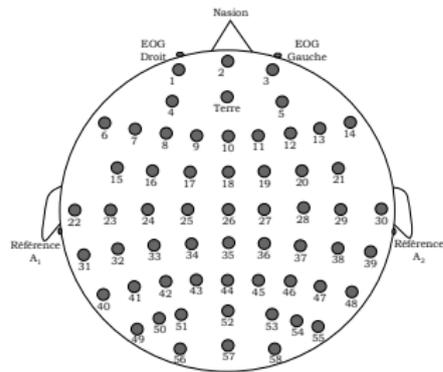


## Enregistrement EEG



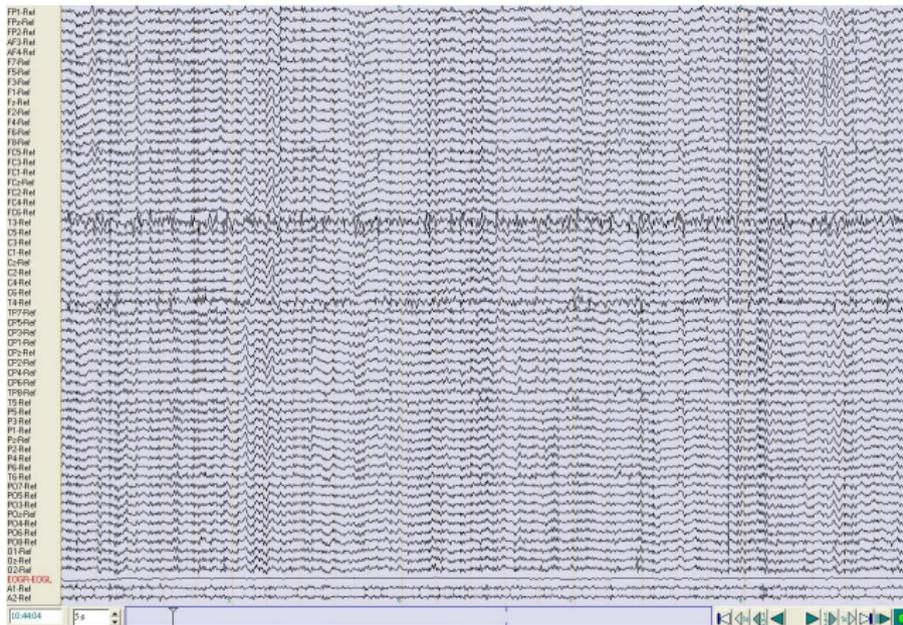
- Durée de pose du casque EEG : 45 minutes.
- Enregistrement de l'EEG d'un sujet dans l'état de vigilance "normal" (avant la séance de relaxation) et "relaxé" (après).
- Sujet immobile, avec les yeux ouverts.
- Fréquence d'échantillonnage de 256Hz.
- Durée d'enregistrement : 3 minutes (47000 points de mesure).

## Enregistrement EEG

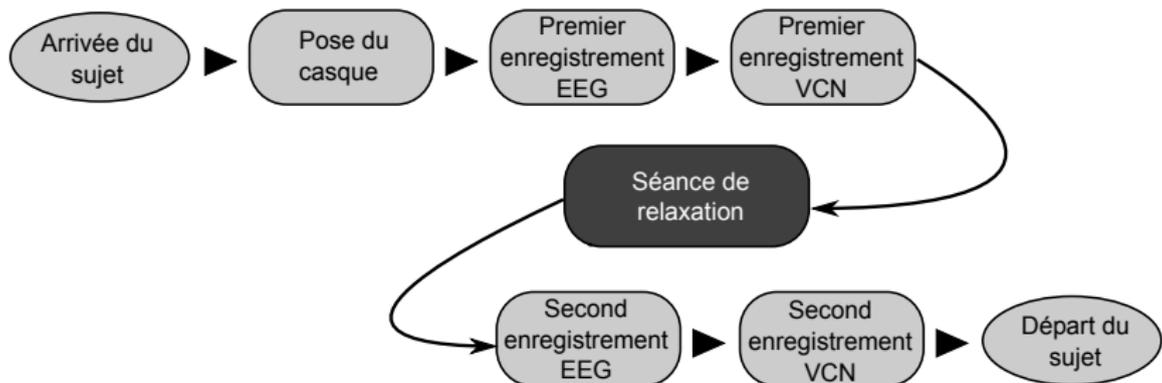


- Durée de pose du casque EEG : 45 minutes.
- Enregistrement de l'EEG d'un sujet dans l'état de vigilance "normal" (avant la séance de relaxation) et "relaxé" (après).
- Sujet immobile, avec les yeux ouverts.
- Fréquence d'échantillonnage de 256Hz.
- Durée d'enregistrement : 3 minutes (47000 points de mesure).

# Enregistrement EEG



# Séance de relaxation

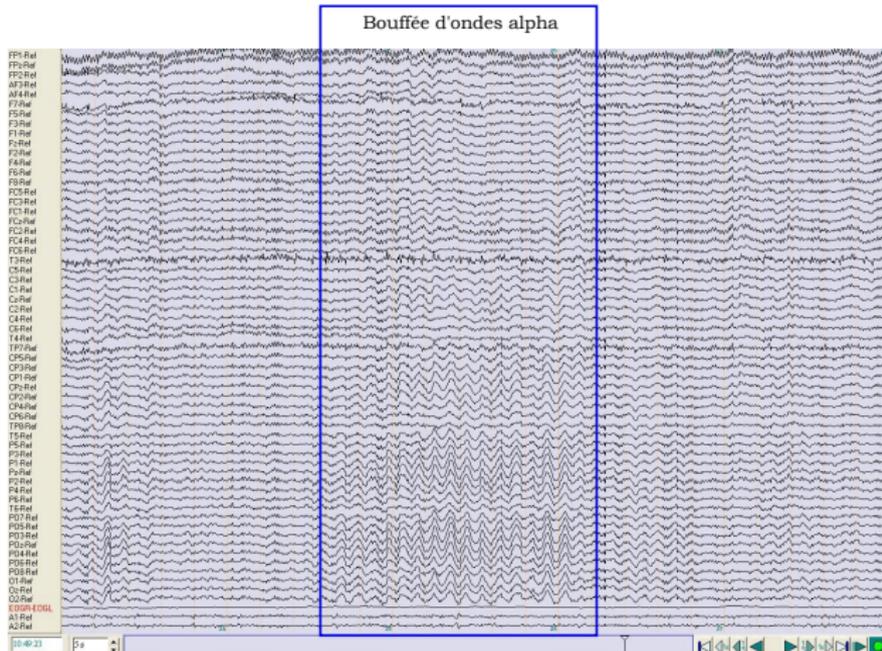


# Séance de relaxation

Environ 20 minutes

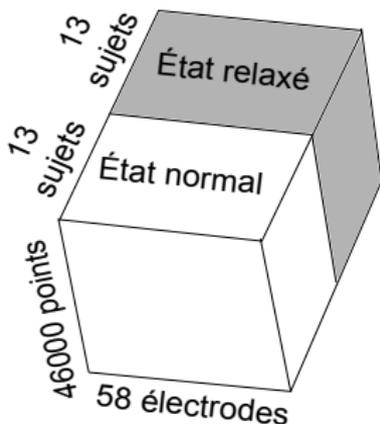
- Training autonome [Schultz, 1958]
- Relaxation musculaire progressive [Jacobson, 1974]
- Visualisation

# Séance de relaxation

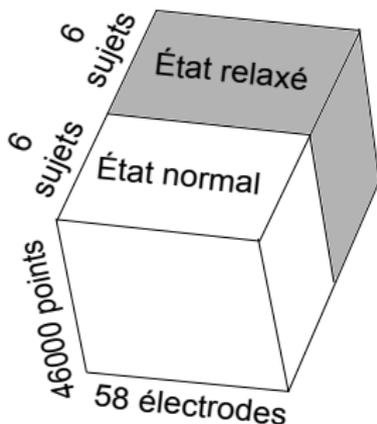


Un signal par  **sujet, électrode, état psycho-physiologique.**

### Première campagne d'acquisition



### Seconde campagne d'acquisition

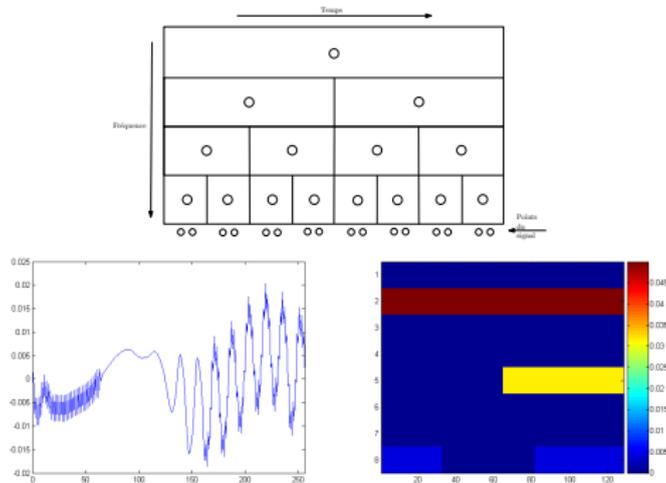


# Plan

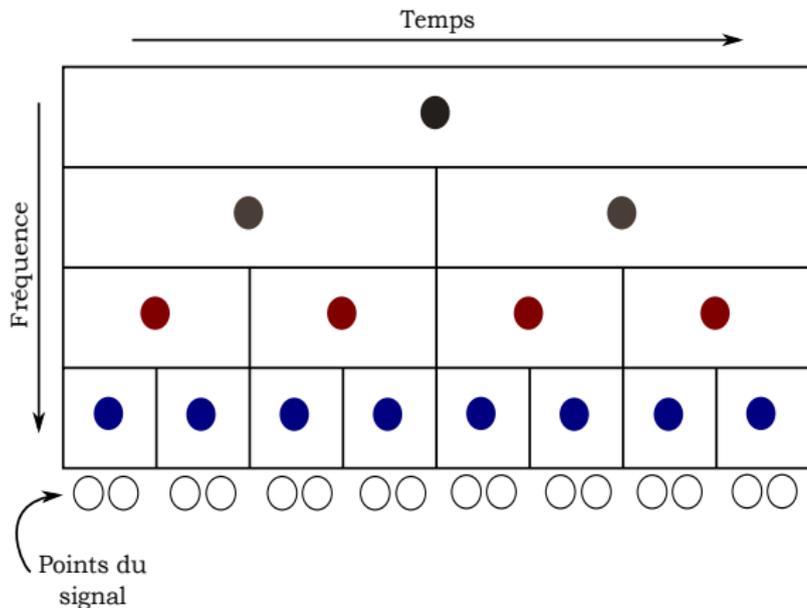
- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 Classification de signaux EEG avec filtre CSP
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

# Décomposition en ondelettes discrète dyadique

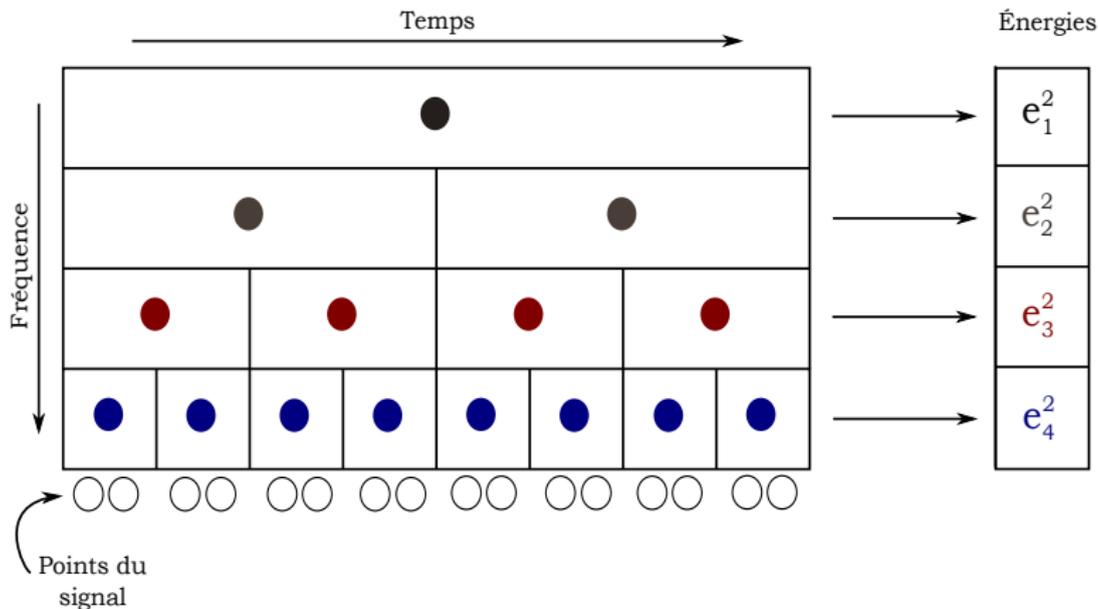
La **grille dyadique** donne une représentation spatio-fréquentielle de la décomposition en ondelettes discrète dyadique



# Énergie du signal

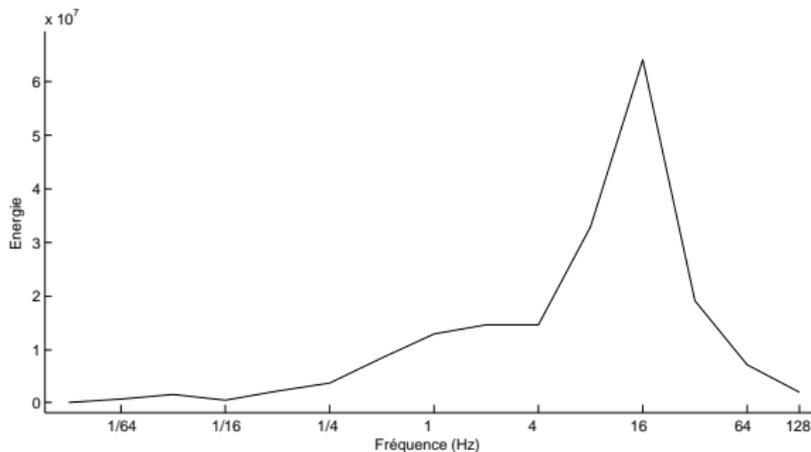


# Énergie du signal



## Calcul de la pente

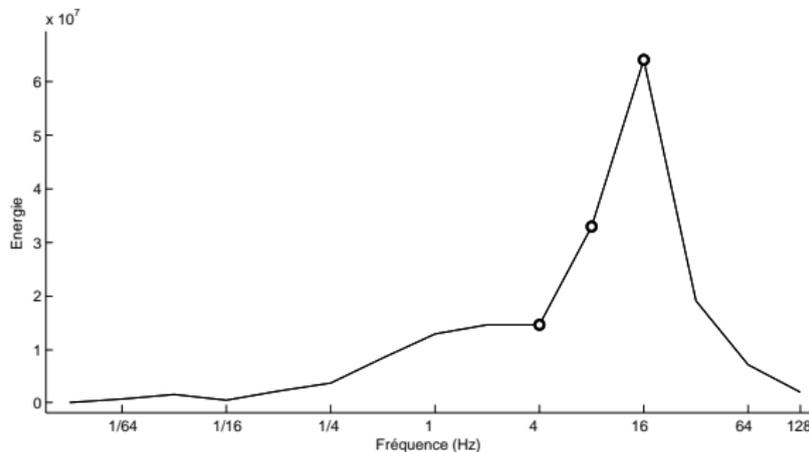
- Alpha : 8 – 12Hz.
- Ondes caractéristiques d'un état relaxé.



Régression linéaire entre 4 et 16Hz.

## Calcul de la pente

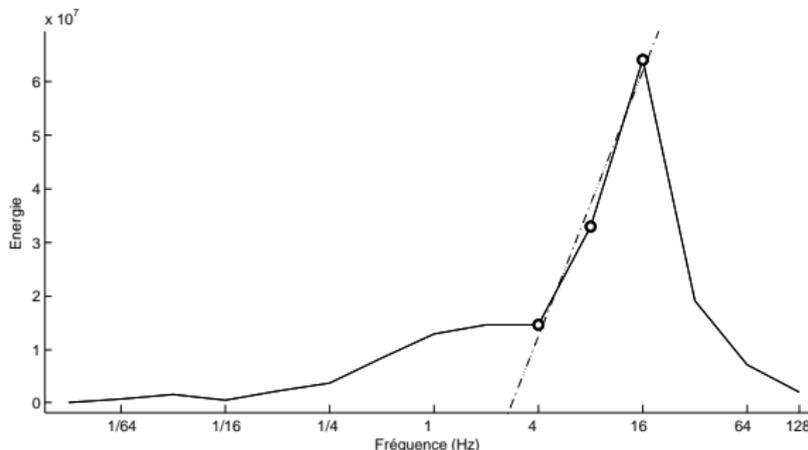
- Alpha : 8 – 12Hz.
- Ondes caractéristiques d'un état relaxé.



Régression linéaire entre 4 et 16Hz.

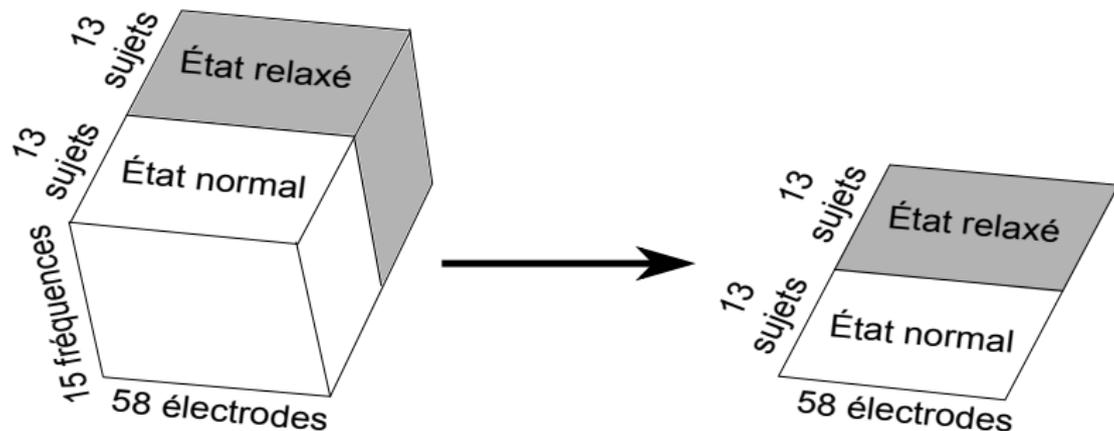
## Calcul de la pente

- Alpha : 8 – 12Hz.
- Ondes caractéristiques d'un état relaxé.

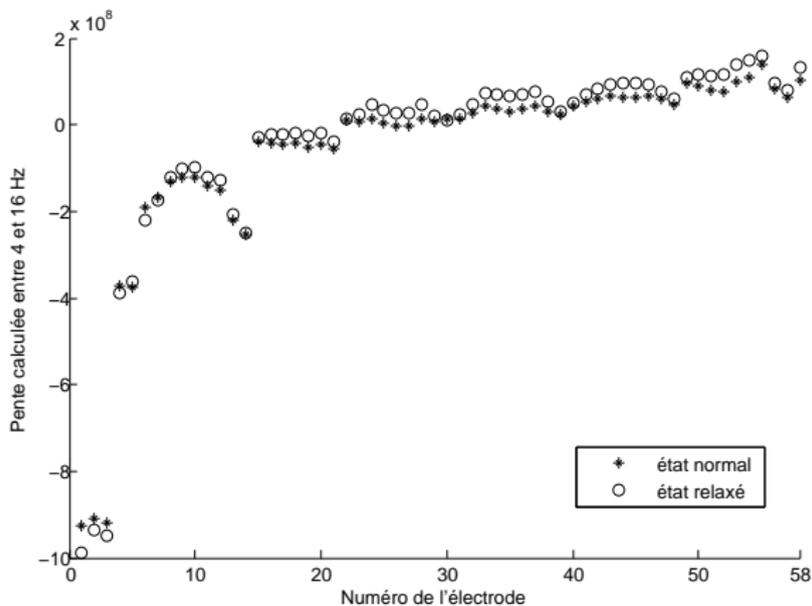


Régression linéaire entre 4 et 16Hz.

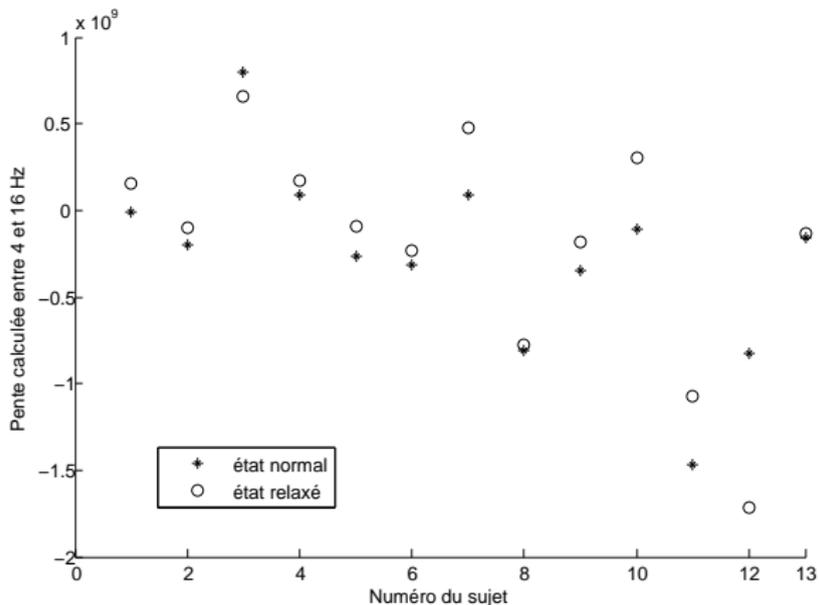
# Réduction de dimensions



# Résultats



# Résultats



## Filtrage spatial

Signaux synthétiques obtenus par combinaison linéaire des variables initiales.

$$\mathbf{f} = \mathbf{X}\mathbf{w}$$

où :

- $\mathbf{X}$  est la matrice de données avec les signaux en colonne,
- $\mathbf{w}$  est le **filtre spatial**,
- $\mathbf{f}$  est le **signal filtré** (projeté).

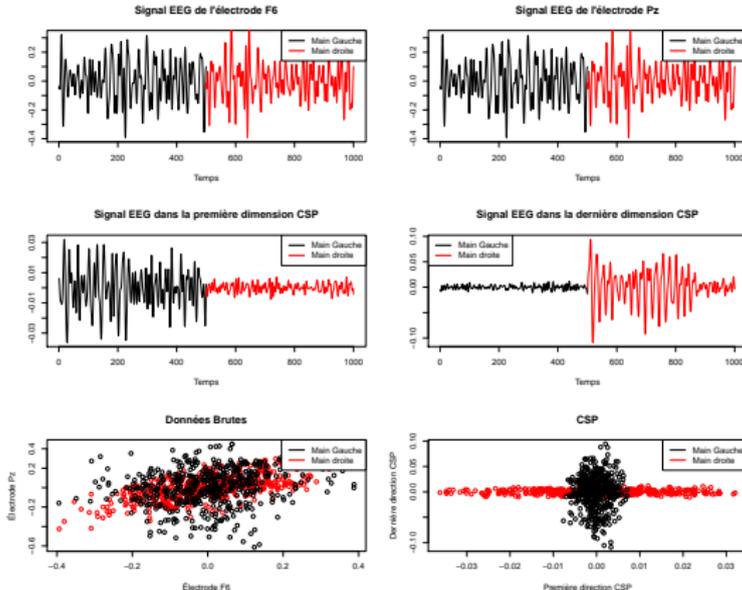
## L'objectif du CSP

Trouver des **signaux synthétiques** (filtrés) de **variance maximale** dans une condition et simultanément de **variance minimale** dans l'autre condition [Fukunaga and Koontz, 1970], [Koles et al., 1990].

- ⇒ Permet de construire des signaux synthétiques qui **discriminent les deux conditions**.
- ⇒ Utilisé pour résoudre des problèmes de **classification binaire**.

## Données

Jeux de données IIIa [Schlögl et al., 2005] de la compétition BCI III [Blankertz et al., 2006] représentant des signaux EEG enregistrés chez des sujets durant des mouvements imaginés de la main droite et de la main gauche.



## Définitions et notations

### Un essai

**Un essai** est un enregistrement EEG effectué durant une période de temps  $T$ .

### Données

- $\{1, \dots, i, \dots, n\} \implies$  Ensemble de  $n$  essais dont  $n_1$  dans la condition 1 (classe 1) et  $n_2$  dans la condition 2 (classe 2).
- $X_i, i = 1, \dots, n \implies$  **Matrices des données EEG** centrées, de dimensions  $T \times p$ , correspondant aux signaux EEG recueillis pendant ces essais.
- $\mathbf{X}_1$  et  $\mathbf{X}_2 \implies$  Matrice, de dimensions  $Tn_c \times p$ , des données EEG des essais de la classe  $c$  **concaténées en ligne**.

### Matrices de variances-covariances empiriques $\mathbf{C}_1$ et $\mathbf{C}_2$

$$\mathbf{C}_c = \frac{1}{Tn_c} \mathbf{X}_c^T \mathbf{X}_c = \frac{1}{n_c} \sum_{i \in \mathcal{I}_c} \frac{1}{T} X_i^T X_i = \frac{1}{n_c} \sum_{i \in \mathcal{I}_c} C_i$$

où  $c \in \{1, 2\}$ ,  $\mathcal{I}_c$  est l'ensemble des indices correspondant aux essais dans la classe  $c$ .

## Deux problèmes d'optimisation

### Premier ensemble de filtres spatiaux

Recherche des filtres  $\mathbf{w} \in \mathbb{R}^p$  qui maximisent  $\text{var}(\mathbf{X}_1 \mathbf{w})$  et minimisent  $\text{var}(\mathbf{X}_2 \mathbf{w})$  où  $\text{Var}(\mathbf{X}_c \mathbf{w}) = \frac{1}{T n_c} \mathbf{w}^T \mathbf{X}_c^T \mathbf{X}_c \mathbf{w} = \mathbf{w}^T \mathbf{C}_c \mathbf{w}$  avec  $c = 1, 2$ .

### Premier problème d'optimisation

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \quad (1)$$

### Solution

- $\mathbf{w}$  est le vecteur propre de  $\mathbf{C}_2^{-1} \mathbf{C}_1$  associé à la plus grande valeur propre.

Démonstration

- Obtention d'un ensemble de  $k$  filtres spatiaux  $\mathbf{w}_1, \dots, \mathbf{w}_k$  maximisant (1) sous la contrainte  $\mathbf{w}_j^T \mathbf{C}_2 \mathbf{w}_l = 0, \forall l \neq j$ .

## Second ensemble de filtres spatiaux

Recherche des filtres  $\mathbf{v} \in \mathbb{R}^p$  qui maximisent  $\text{var}(\mathbf{X}_2\mathbf{v})$  et minimisent  $\text{var}(\mathbf{X}_1\mathbf{v})$  où  $\text{Var}(\mathbf{X}_c\mathbf{v}) = \frac{1}{Tn_c} \mathbf{v}^T \mathbf{X}_c^T \mathbf{X}_c \mathbf{v} = \mathbf{v}^T \mathbf{C}_c \mathbf{v}$ , avec  $c = 1, 2$ .

## Second problème d'optimisation

$$\max_{\mathbf{v} \in \mathbb{R}^p} \frac{\mathbf{v}^T \mathbf{C}_2 \mathbf{v}}{\mathbf{v}^T \mathbf{C}_1 \mathbf{v}} \quad (2)$$

## Solution

- $\mathbf{v}$  est le **vecteur propre** de  $\mathbf{C}_2^{-1} \mathbf{C}_1$  associé à la plus grande valeur propre.
- Obtention d'un ensemble de  $k$  **filtres spatiaux**  $\mathbf{v}_1, \dots, \mathbf{v}_k$  maximisant (2) sous la contrainte  $\mathbf{v}_j^T \mathbf{C}_1 \mathbf{v}_l = 0, \forall l \neq j$ .

# Une unique décomposition en éléments propres

On a vu que :

- Le premier ensemble de filtres  $\mathbf{w}_1, \dots, \mathbf{w}_k$  est obtenu en calculant la décomposition en éléments propres de  $\mathbf{C}_2^{-1}\mathbf{C}_1$ .
- Le second ensemble de filtres  $\mathbf{v}_1, \dots, \mathbf{v}_k$  sont obtenus en calculant la décomposition en éléments propres de  $\mathbf{C}_1^{-1}\mathbf{C}_2$ .

Or les  $k$  premiers vecteurs propres de  $\mathbf{C}_2^{-1}\mathbf{C}_1$  sont aussi les  $k$  derniers vecteurs propres de  $\mathbf{C}_2^{-1}\mathbf{C}_1$ .

- ⇒ Les filtres  $\mathbf{w}_1, \dots, \mathbf{w}_k$  et les filtres  $\mathbf{v}_1, \dots, \mathbf{v}_k$  sont obtenus en conservant les  $k$  premiers et les  $k$  derniers vecteurs propres de la matrice  $\mathbf{C}_2^{-1}\mathbf{C}_1$ .
- ⇒ En sortie du CSP, on obtient une matrice  $W$  de dimensions  $p \times 2k$  où :

$$W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k]$$

# Transformation “logvar”

- 1 On filtre les matrices  $X_i$  des données EEG des  $n$  essais :

$$F_i = X_i W$$

$F_i$  est une matrice de dimensions  $T \times 2k$  dont les colonnes sont les  $2k < p$  signaux synthétiques (filtré) de l'essai  $i$ .

- 2 Pour chaque essai  $i$  on extrait une unique caractéristique pour signal filtré  $j$  :

$$z_{ij} = \log(\text{var}(X_i w_j)).$$

- ⇒ Obtention d'une matrice  $Z$  caractérisant les  $n$  essais en  $2k$  dimensions.
- ⇒ Cette matrice de caractéristiques est utilisée pour faire de la classification binaire des essais.

# Extraction de caractéristiques sur nos données

Les signaux EEG des sujets des campagnes 1 et 2 ont été :

- bande-passe filtrés (différent de filtré par CSP !),
- découpés en essais,
- normalisés.

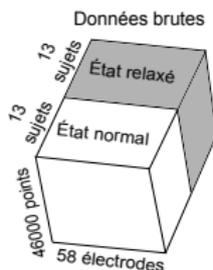
Il faudra donc **fixer** :

- Le filtre passe-bande  $[a, b]$  Hz.
- La longueur des essais soit le nombre  $T$  de points. Chaque signal est découpé en  $n^*$  essais où  $n^* = \lfloor \frac{46000}{T} \rfloor$ .
- Le **type de normalisation** (par l'inertie, par la variance).
- Le nombre  $k$  de paires de filtres CSP.

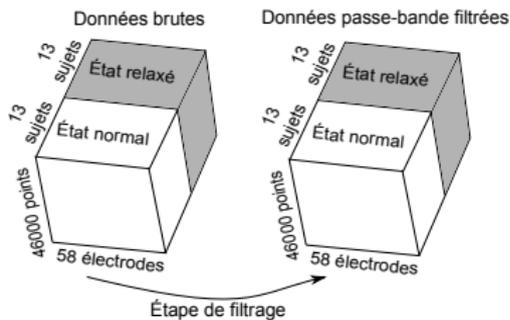
Attention au **changement d'unité statistique** :

Sujet  $\implies$  Essais

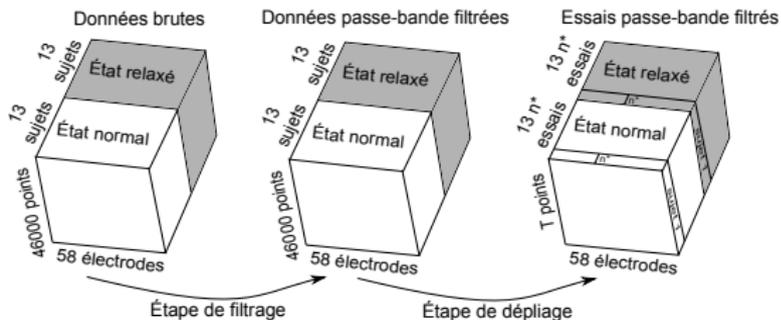
# Extraction de caractéristiques sur nos données



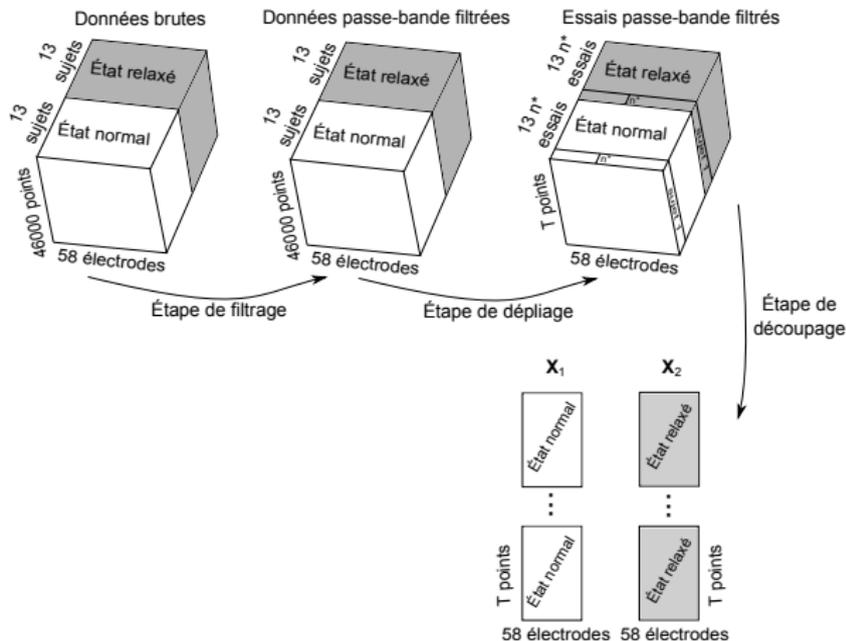
## Extraction de caractéristiques sur nos données



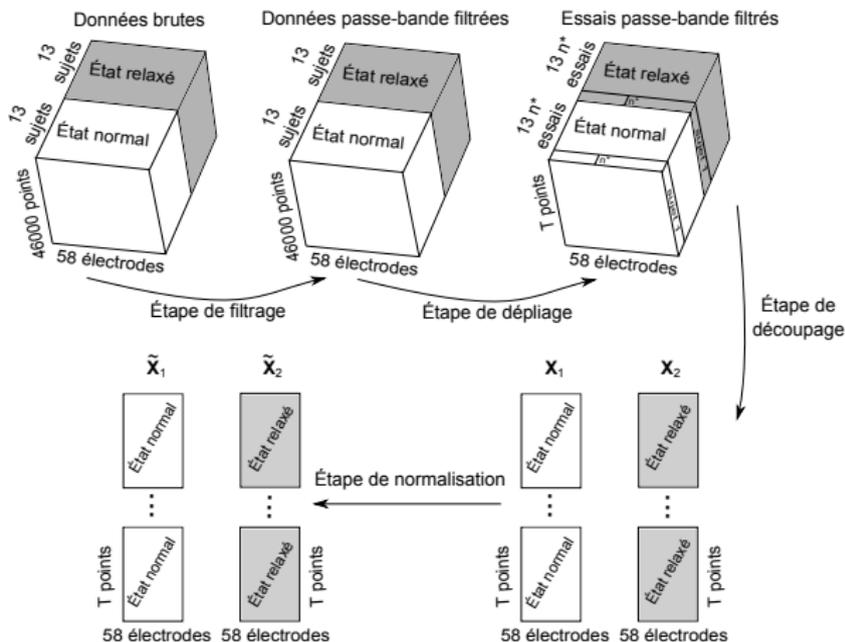
# Extraction de caractéristiques sur nos données



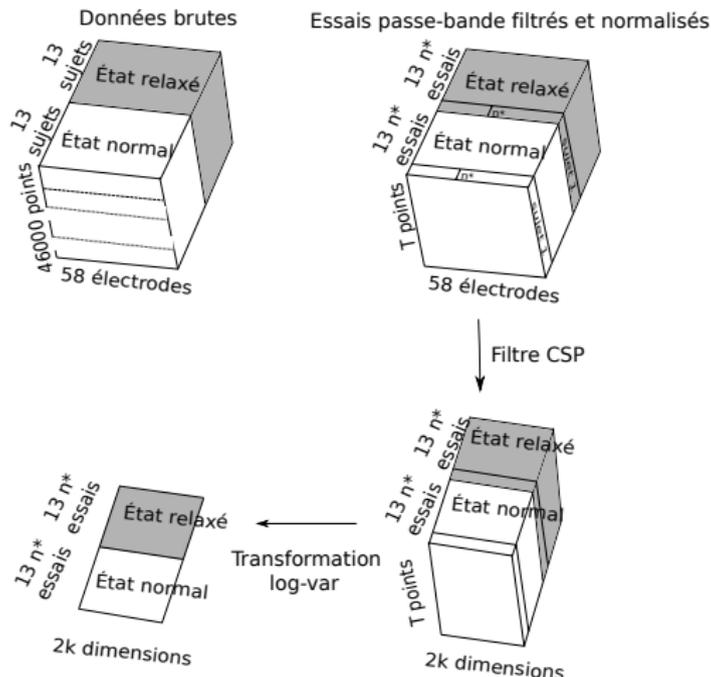
# Extraction de caractéristiques sur nos données



# Extraction de caractéristiques sur nos données



# Extraction de caractéristiques sur nos données

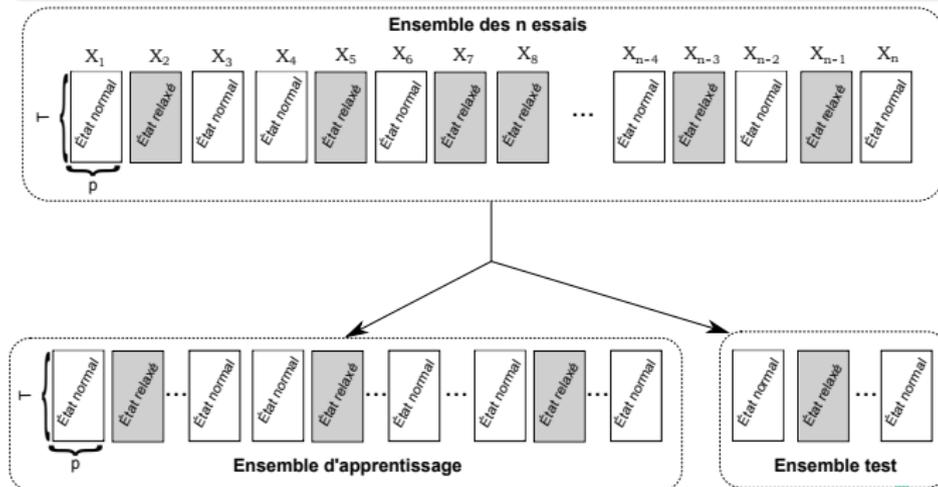


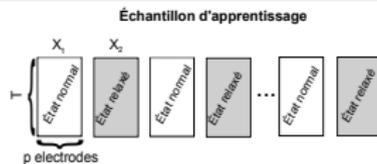
# Plan

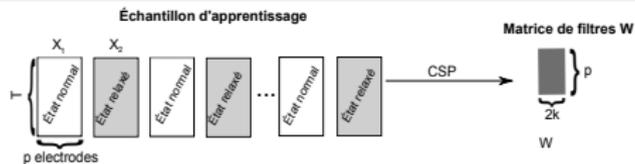
- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 **Classification de signaux EEG avec filtre CSP**
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

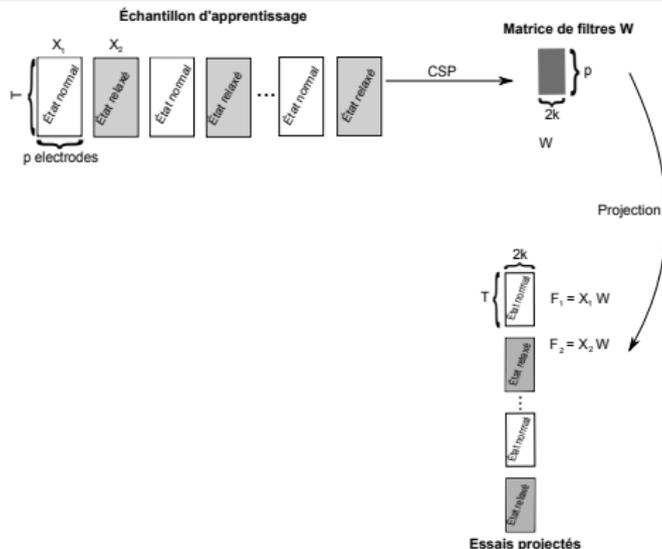
# La méthodologie générale

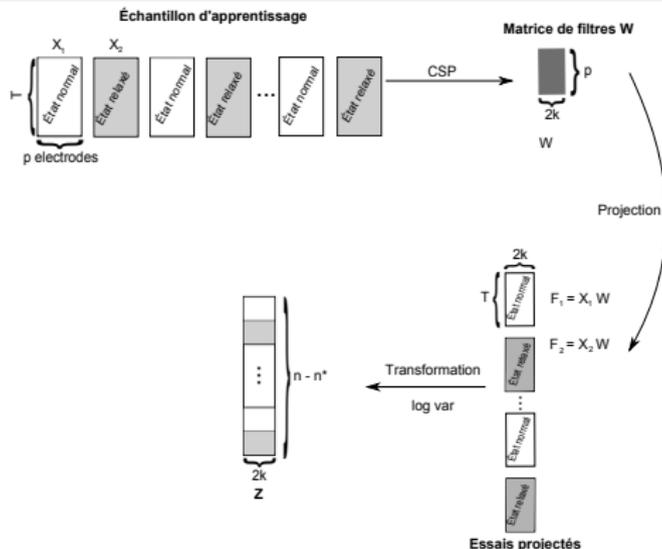
- Découpage en échantillon d'apprentissage et échantillon test.
- La règle de décision (filtres CSP + classifieur) est formée grâce aux essais de l'échantillon d'apprentissage.
- Taux de bons classements (TBC) estimé sur l'échantillon test.

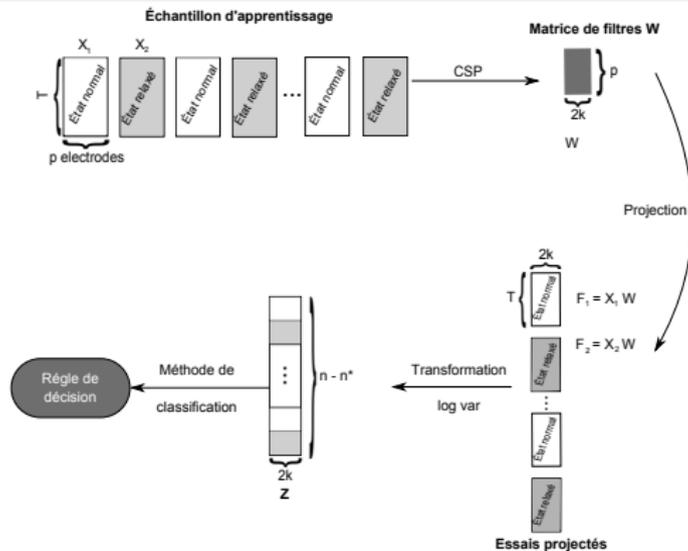


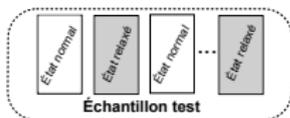
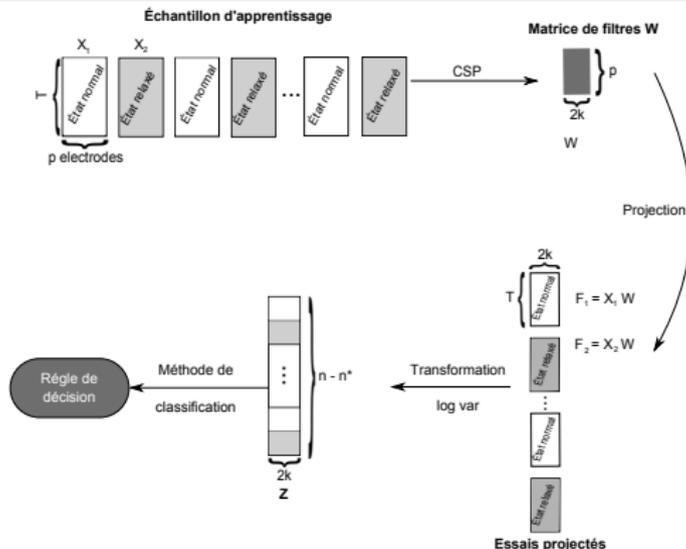


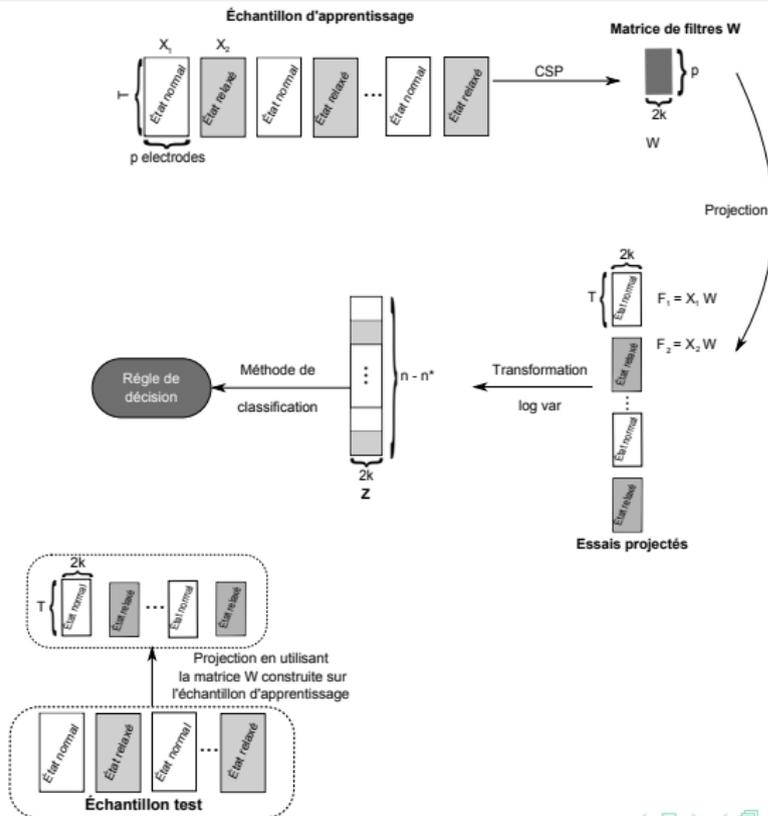


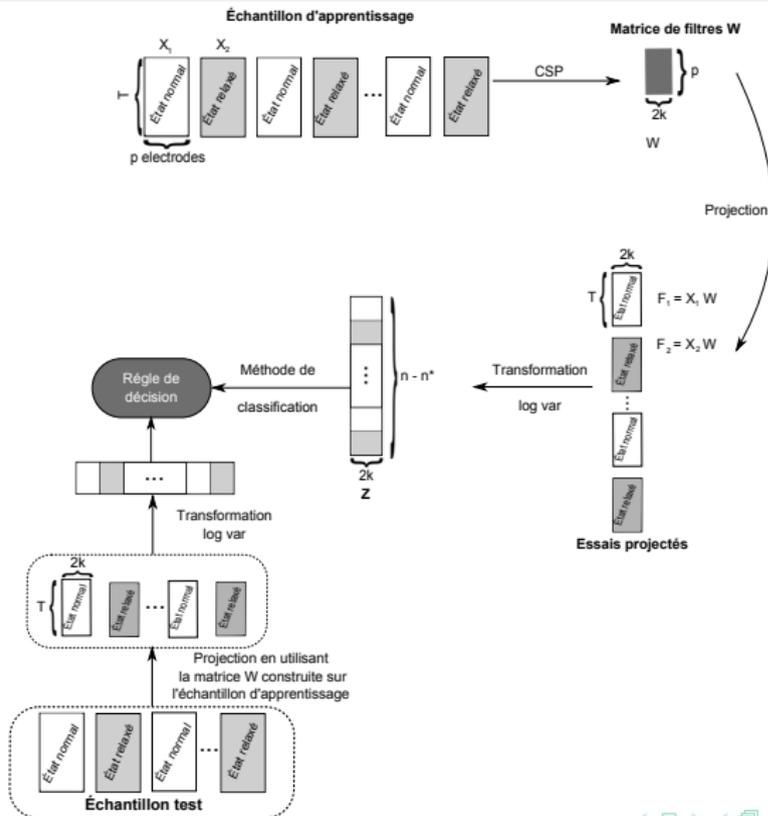


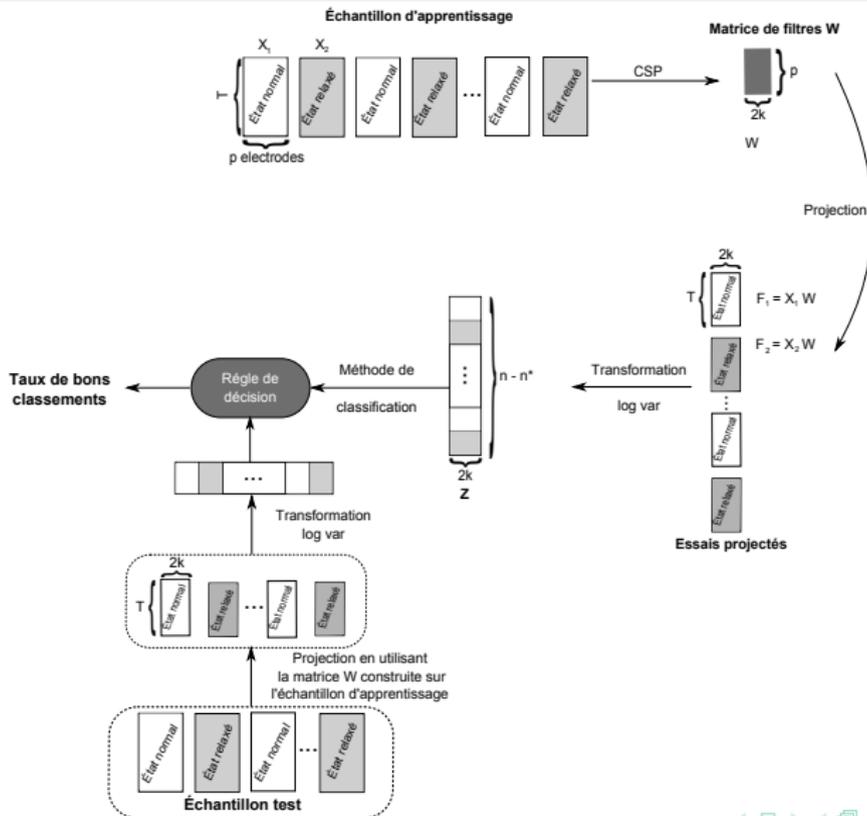








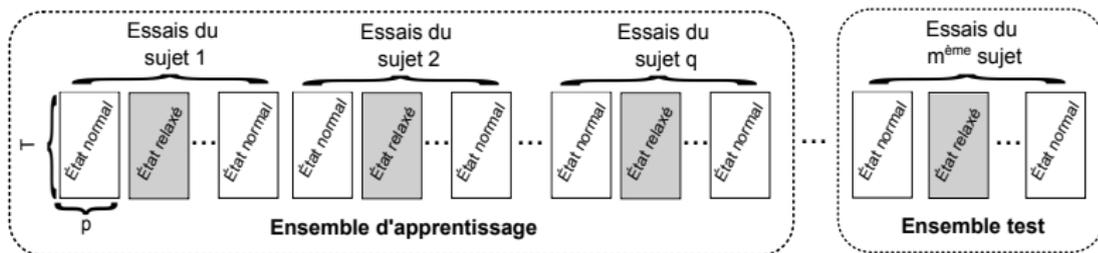




# Adaptation au problème multi-sujet

Un sujet = plusieurs essais donc :

- Deux taux de bon classement :
  - le TBC des essais,
  - le TBC des sujets obtenu par "vote majoritaire".
- Dans l'estimation du TBC par LOO (Leave One Out), on retire tous les essais d'un même sujet pour former l'ensemble test lors de la  $m$ -ème itération (avec  $m = 1, \dots, q$ ).



## Le choix des paramètres

Les paramètres à régler pour le pré-traitement sont :

- $a$  et  $b$  pour le filtrage passe-bande à  $[a, b]$  Hz.
- $T$ , le nombre de points des essais . On en déduit le nombre  $n^*$  avec  $n^* = \left( \lfloor \frac{46000}{T} \rfloor \right)$  .
- Le type de **normalisation**.
- $k$ , le nombre de paires de filtres CSP .

## Le choix des paramètres

En pratique :

- Méthode de classification utilisée : **CSP+logvar+LDA** (linear discriminant analysis)
- **Échantillon d'apprentissage/validation** : sujets de la campagne 1 pour estimer les paramètres
- **Échantillon test** : sujets de la campagne 2 pour estimer le TBC de la méthode de classification.

⇒ Choix des paramètres sur les 13 sujets de la campagne 1 en fonction du **TBC estimé par "LOO"** .

Réglages des paramètres

## Le choix des paramètres

La "meilleure" combinaison de valeurs des paramètres trouvée est :

- $[a, b] = [7, 30]$  pour le filtre bande-passe.
- $T = 2048$  points soit  $n^* = 22$  essais de 8 secondes.
- Normalisation par les écarts-types.
- $k = 3$  paires de filtre CSP.

# Plan

- 1 Introduction
- 2 Acquisition de données EEG
  - Protocole d'acquisition
  - Données
- 3 Extraction de caractéristiques
  - Critère des pentes
  - Common Spatial Pattern (CSP) et transformation logvar
- 4 Classification de signaux EEG avec filtre CSP
  - La méthodologie générale
  - Adaptation au problème multi-sujet
  - Le choix des paramètres
- 5 Sélection de variables CSP sparse
  - Quelques premiers résultats
  - Sélectionner des variables en CSP
  - Un algorithme de CSP sparse
  - Résultats de la classification avec filtre sCSP

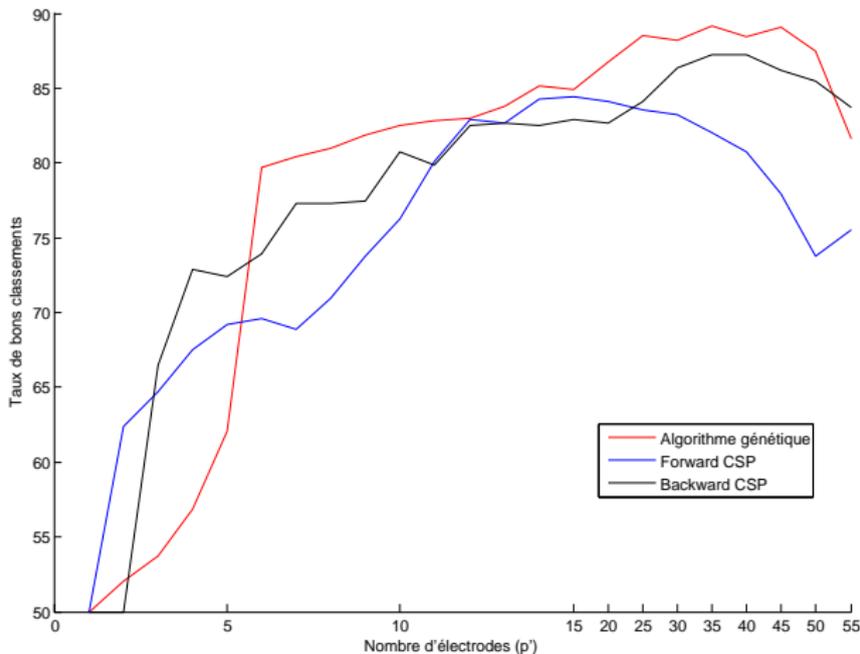
# Un problème de sélection de variables

## Objectif

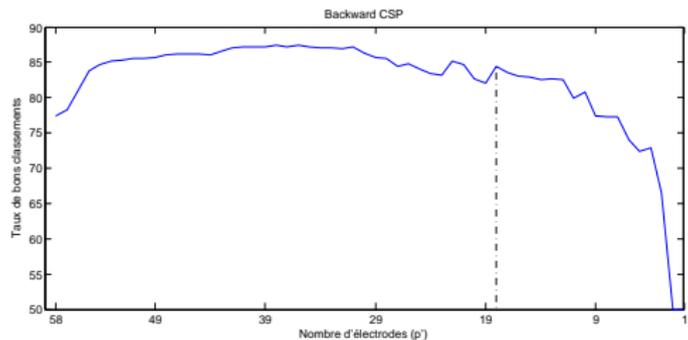
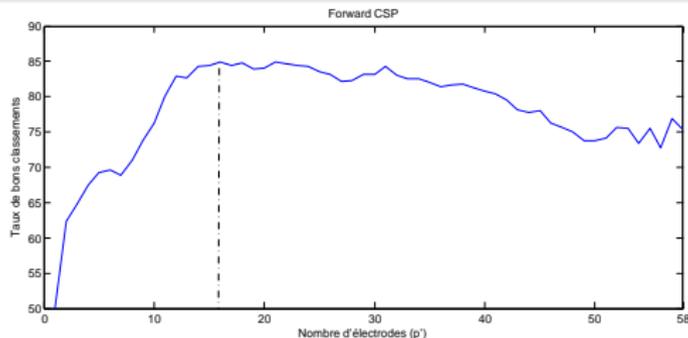
Trouver la meilleure combinaison de  $p'$  électrodes (où  $p = 58$  est le nombre d'électrodes et  $p' < p$ ).

- $C_{58}^{p'} = \frac{58!}{p'!(58 - p')!}$  possibilités.
- La recherche exhaustive peut être très coûteuse en temps de calcul.
- Solution testées :
  - **Algorithme génétique** [Vézard et al., 2015] (trop cher ?),
  - **Méthode séquentielle** de type backward et forward (trop simple ?),
  - **Sparsifier le CSP**.

## Quelques premiers résultats

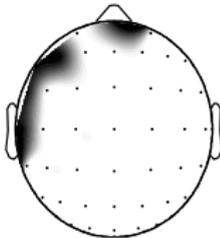


## Quelques premiers résultats

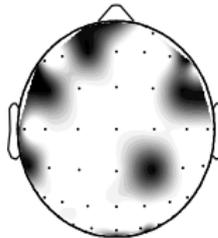


# Quelques premiers résultats

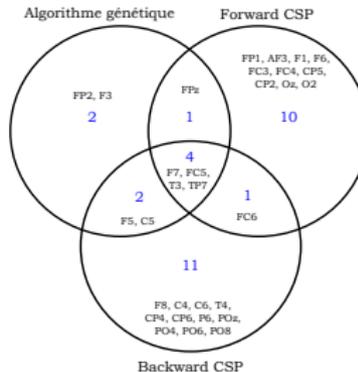
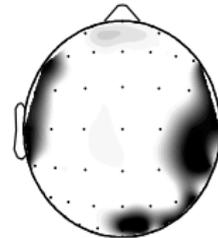
Sélection par algorithme génétique



Sélection par forward CSP



Sélection par backward CSP



# Sélectionner des variables en CSP

## L'idée générale

- 1 Ecrire le CSP sous la forme de **deux ACP avec métriques**
- 2 Ecrire ces deux ACP avec métriques comme **deux ACP classiques**
- 3 Sélectionner les variables avec un algorithme d'**ACP sparse**.

Plusieurs algorithmes d'ACP sparse dans la littérature, en particulier :

- **Zou, Hastie et Tibshirani (2006)** → approche de type "bloc"
- Shen et Huang (2006) → approche de type "déflation"
- Journee et al. (2010)

# 1. Le CSP à partir de deux ACP avec métriques

⇒ Réécriture des deux problèmes d'optimisation du CSP :

$$\max_{\mathbf{w} \in \mathbb{R}^P} \frac{\text{var}(\mathbf{X}_1 \mathbf{w})}{\text{var}(\mathbf{X}_2 \mathbf{w})} \iff \max_{\mathbf{w} \in \mathbb{R}^P} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \iff \max_{\mathbf{w} \in \mathbb{R}^P} \text{var}(\mathbf{X}_1 \mathbf{w})$$

s. c.  $\mathbf{w}^T \mathbf{C}_2 \mathbf{w} = 1$

$$\max_{\mathbf{v} \in \mathbb{R}^P} \frac{\text{var}(\mathbf{X}_2 \mathbf{v})}{\text{var}(\mathbf{X}_1 \mathbf{v})} \iff \max_{\mathbf{v} \in \mathbb{R}^P} \frac{\mathbf{v}^T \mathbf{C}_2 \mathbf{v}}{\mathbf{v}^T \mathbf{C}_1 \mathbf{v}} \iff \max_{\mathbf{v} \in \mathbb{R}^P} \text{var}(\mathbf{X}_2 \mathbf{v})$$

s. c.  $\mathbf{v}^T \mathbf{C}_1 \mathbf{v} = 1$

# 1. Le CSP à partir de deux ACP avec métriques

⇒ En posant  $\tilde{\mathbf{w}}_j = \mathbf{C}_2 \mathbf{w}_j$ , on a :

$$\begin{array}{l} \max_{\mathbf{w} \in \mathbb{R}^p} \text{var}(\mathbf{X}_1 \mathbf{w}) \\ \text{s. c. } \mathbf{w}^T \mathbf{C}_2 \mathbf{w} = 1 \end{array} \iff \begin{array}{l} \max_{\tilde{\mathbf{w}} \in \mathbb{R}^p} \text{var}(\mathbf{X}_1 \mathbf{C}_2^{-1} \tilde{\mathbf{w}}) \\ \text{s. c. } \tilde{\mathbf{w}}_j^T \mathbf{C}_2^{-1} \tilde{\mathbf{w}}_j = 1 \end{array} \iff \begin{array}{l} \text{ACP de } \mathbf{X}_1 \text{ avec} \\ \text{métrique } \mathbf{C}_2^{-1} \end{array}$$

⇒ En posant  $\tilde{\mathbf{v}}_j = \mathbf{C}_1 \mathbf{v}_j$ , on a :

$$\begin{array}{l} \max_{\mathbf{v} \in \mathbb{R}^p} \text{var}(\mathbf{X}_2 \mathbf{v}) \\ \text{s. c. } \mathbf{v}^T \mathbf{C}_1 \mathbf{v} = 1 \end{array} \iff \begin{array}{l} \max_{\tilde{\mathbf{v}} \in \mathbb{R}^p} \text{var}(\mathbf{X}_2 \mathbf{C}_1^{-1} \tilde{\mathbf{v}}) \\ \text{s. c. } \tilde{\mathbf{v}}_j^T \mathbf{C}_1^{-1} \tilde{\mathbf{v}}_j = 1 \end{array} \iff \begin{array}{l} \text{ACP de } \mathbf{X}_2 \text{ avec} \\ \text{métrique } \mathbf{C}_1^{-1} \end{array}$$

## 2. Le CSP à partir de deux ACP classiques

- 1 L'ACP de  $\mathbf{X}_1$  avec métrique s'obtient à partir de l'ACP classique de  $\mathbf{Z}_1 = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}}$  :
    - recherche des vecteurs propres  $\mathbf{u}_j$  de la matrice de covariance de  $\mathbf{Z}_1$ .
    - $\mathbf{w}_j = \mathbf{C}_2^{-\frac{1}{2}} \mathbf{u}_j$  donne le premier ensemble de filtres  $\mathbf{w}_1, \dots, \mathbf{w}_k$
  - 2 L'ACP de  $\mathbf{X}_2$  avec métrique à partir de l'ACP classique de  $\mathbf{Z}_2 = \mathbf{X}_2 \mathbf{C}_1^{-\frac{1}{2}}$  donne le second ensemble de filtres  $\mathbf{v}_1, \dots, \mathbf{v}_k$
- ⇒ Procédure équivalente à la procédure classique de CSP qui consiste à sélectionner les  $k$  premiers et derniers vecteurs propres de  $\mathbf{C}_2^{-1} \mathbf{C}_1$ .
- ⇒ On remplacera l'étape d'ACP classique par une ACP sparse pour sélectionner les électrodes.

### 3. Le CSP sparse à partir de deux ACP sparses

Le problème d'optimisation en ACP classique

$$\min_{F, \alpha} \|X - F\alpha^T\|^2 \quad (3)$$

- $F$  est la **matrice**  $n \times k$  des **scores** des  $n$  observations sur les  $k$  composantes principales ( $F = X\alpha$ )
- $\alpha$  est la **matrice**  $p \times k$  des **loadings** des  $p$  variables sur les  $k$  composantes principales.
- $\|A\| = \text{Tr}(AA^T)$  est la norme de Frobenius de la matrice  $A$ ,

D'après [Eckart et Young, 1936], si  $X = UDV^T$  est la **décomposition en valeurs singulières** de  $X$  avec

- $U_k$  et  $V_k$  sont les matrices respectivement composées des  $k$  premières colonnes de  $U$  et  $V$ ,
- $D_k$  est la matrices diagonale des  $s$  premières valeurs singulières de  $D$ ,

alors

$$F = U_k D_k \text{ et } \alpha = V_k.$$

### Problème d'optimisation de l'ACP sparse de Zou, Hastie et Tibshirani (2006)

$$\min_{\alpha, \beta} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda_j \|\beta_j\|_1 \quad (4)$$

s. c.  $\alpha^T \alpha = I_k$

où :

- $\beta$  est la **matrice** des **loadings spartes**,
- $\alpha$  est la matrice des loadings orthogonaux,
- $\|\cdot\|_1$  est la norme  $l_1$ ,
- $\beta_j \in \mathbb{R}^p$  est la  $j^{\text{ème}}$  colonne de  $\beta$ ,
- $\lambda_j$  est le **paramètre de régularisation**.

L'ACP sparse de Zou, Hastie et Tibshirani (2006) :

- Critère de type "block".  
↪ Recherche **simultanée** des  $k$  vecteurs de loadings sparses.
- Algorithme itératif d'optimisation  
↪ **Détails**
- Implémenté dans le package R **elasticnet**.
- Comment **fixer les  $k$  paramètres de régularisation** ( $\lambda^1, \dots, \lambda^k$ )?

L'idée proposée :

- Définir une grille matricielle de paramètres de régularisation qui est **une matrice**  
 $\Lambda = (\lambda_{ij}^j)_{r \times k}$ .  
↪ Comment **définir cette grille** ?
- Augmenter  $i$  de 1 à  $r$  augmente globalement la sparsité sur les  $k$  composantes principales de l'ACP.
- Choisir "**le pas**"  $i$  qui optimise **un critère** en validation croisée par exemple.

## Grille matricielle des paramètres de régularisation

Le principal général :

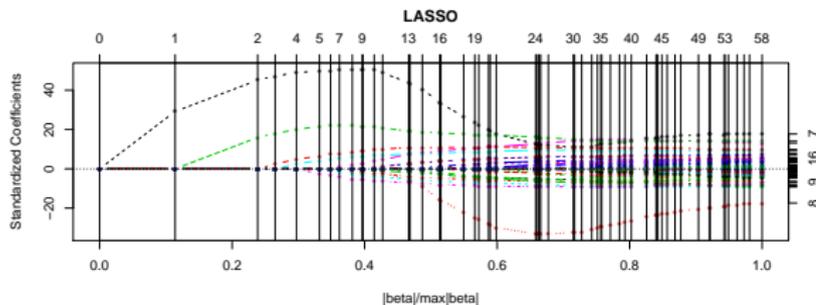
- 1 Pour chaque dimension  $j = 1, \dots, k$ , définir un vecteur de paramètres de régularisation  $\lambda^j = (\lambda_1^j, \dots, \lambda_r^j)$ .
  - ↔ Utilisation du **chemin de régularisation de la régression LASSO**.
  - ↔ Régression LASSO de la  **$j$ -ème composante principale de l'ACP de  $X$  sur  $X$** .
- 2 **Concaténer** ces vecteurs en colonne afin de construire la grille matricielle  $\Lambda$  de dimension  $r \times k$ .

Deux méthodes pour réaliser un régression LASSO et obtenir le chemin de régularisation :

- La méthode Least Angle Regression [Efron et al, 2003], implémentée dans le package R **lars**.
- La méthode proposée par [Friedman et al, 2010], implémentée dans le package R **glmnet**.

## Chemin de régularisation de la fonction $lars()$

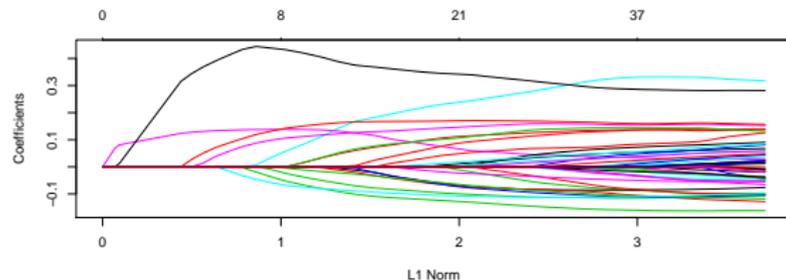
Un exemple :



↪ Chemin **complet** mais algorithme assez lent.

## Chemin de régularisation de la fonction $glmnet()$

Un exemple :



↪ Le chemin est tracé à partir d'une grille de  $n = 100$  valeurs (par défaut) obtenues en discrétisant un intervalle  $[\log(\lambda_{min}), \log(\lambda_{max})]$  en  $n$  intervalles égaux.

- $\lambda_{max} = \max_j |\text{cov}(x_j, y)|$  est la plus petite valeur de  $\lambda$  qui permette d'obtenir la nullité de tous les coefficients
- $\lambda_{min} = \epsilon \lambda_{max}$  où par défaut  $\epsilon = 0.01$  si  $n < p$  et  $\epsilon = 0.001$  sinon.

↪ Chemin "non complet" mais algorithme plus rapide rapide.

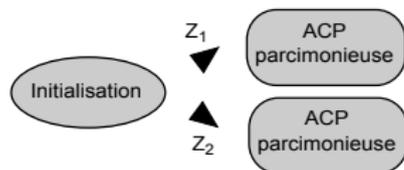
↪ Grille de 100 valeurs utilisées pour former la matrice de paramètres de régularisation  $\Lambda$  matrice l'ACP sparse.

# Un algorithme de CSP sparse

Initialisation

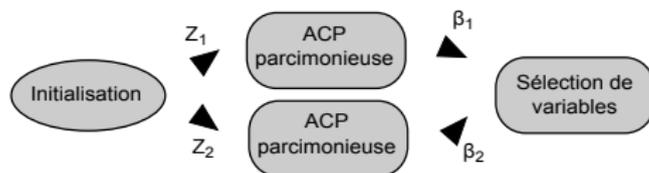
Algorithme détaillé

# Un algorithme de CSP sparse



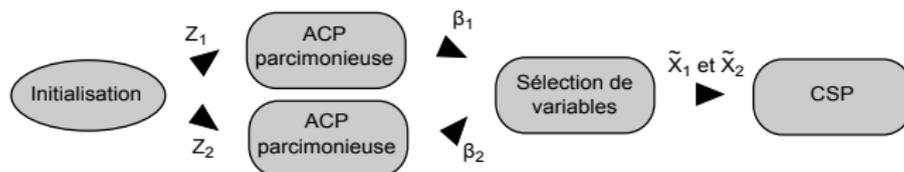
Algorithme détaillé

# Un algorithme de CSP sparse



Algorithme détaillé

# Un algorithme de CSP sparse



Algorithme détaillé

# Un algorithme de CSP sparse



Algorithme détaillé

# Choix des paramètres de régularisation en sCSP

En entrée de l'algorithme sCSP :

- $\mathbf{X}_1$  et  $\mathbf{X}_2$  les matrices des signaux dans les deux conditions,
- $k$  le nombre de paires de filtres,
- deux vecteurs  $(\lambda^1, \dots, \lambda^k)$  et  $(\phi^1, \dots, \phi^k)$  (un pour chaque ACP sparse).

Stratégie pour choisir ces deux vecteurs de paramètres :

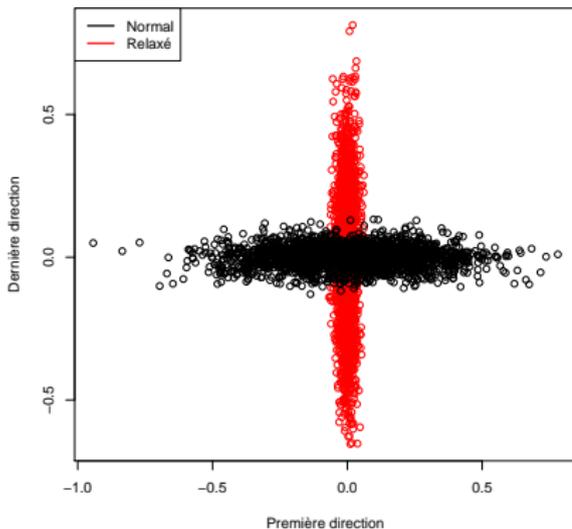
- 1 Construire les grilles matricielles  $\Lambda$  et  $\Phi$  de dimension  $r \times k$  à partir des matrices recodées  $Z_1$  et  $Z_2$ .
- 2 Parcourir ces grilles ligne à ligne.  
→ Plus le pas (la ligne)  $i$  augmente, plus le nombre de variables sélectionnées augmente.
- 3 Choisir le pas  $i$  :
  - en fonction du nombre de variables sélectionnées.
  - qui fournit le meilleur TBC lorsqu'on fait suivre le sCSP d'une étape "logvar+classification".

## Illustration sur un exemple

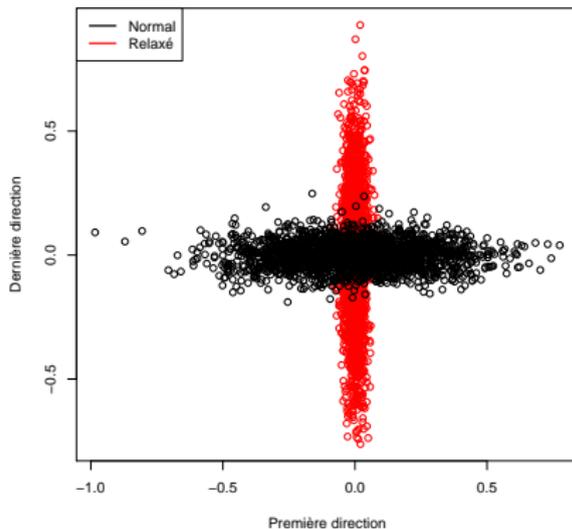
- Les 8 premières secondes d'enregistrement EEG du sujet 58 de l'expérience avant et après relaxation sont considérées.
  - ↪ Deux matrices  $\mathbf{X}_1$  et  $\mathbf{X}_2$  de dimension  $2059 \times 58$  (signaux échantillonnés à 256 Hz)
- Le troisième pas de la grille de régularisation a été choisi.
  - ↪ 7 variables (électrodes) sélectionnées pour construire  $k = 2$  paires de composantes CSP.
- On compare donc les signaux synthétiques obtenus avec la première et la seconde paire de filtres CSP
  - avec les 58 électrodes
  - avec les 7 électrodes sélectionnées.

## Illustration sur un exemple

Paire 1 de composantes CSP

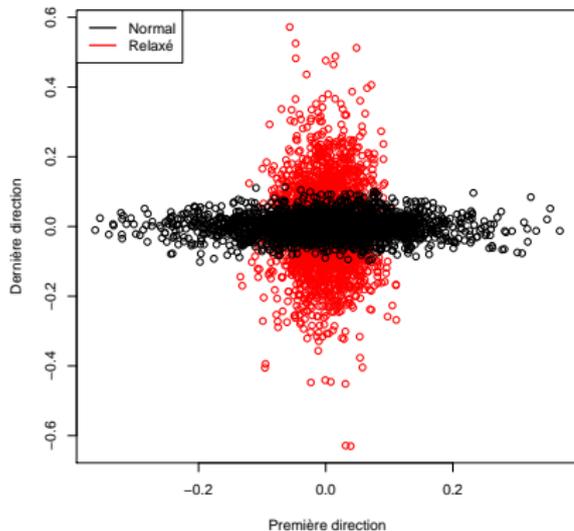


Paire 1 de composantes CSP parcimonieuses

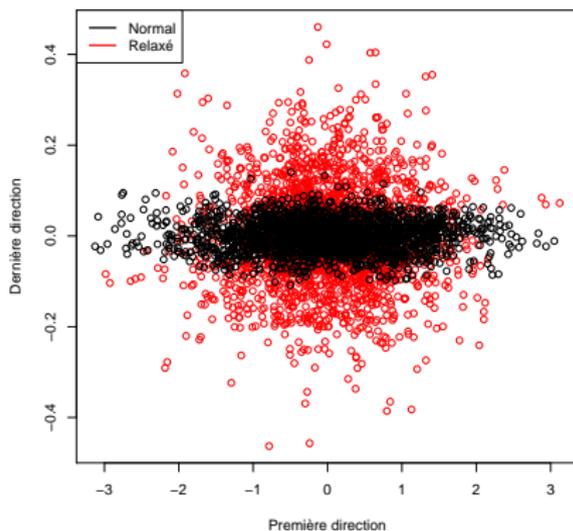


## Illustration sur un exemple

Paire 2 de composantes CSP



Paire 2 de composantes CSP parcimonieuses



# Résultats de la classification avec filtre sCSP

## Échantillons

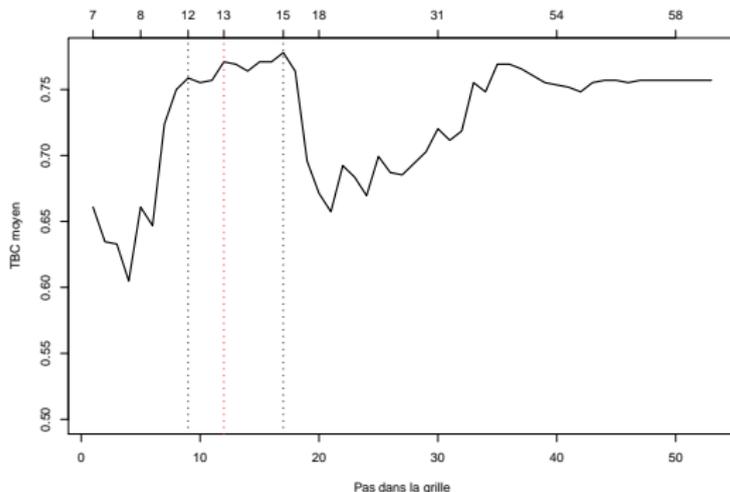
- Échantillon d'apprentissage : Essais des sujets de la campagne 1.
- Échantillon test : Essais des sujets de la campagne 2.

## Procédé expérimental

- Réglage du pas de la grille matricielle des paramètres de régularisation.
- Sélection des variables.
- Formation de la règle de décision.
- Calcul d'un taux de bons classement sur l'échantillon test.

## Sélection de variables

Le TBC est calculé par LOO (de type sujet/essai) sur les **essais des sujets de la campagne 1**, avec  $k = 3$  paires de filtres.



↔ On choisit de conserver **13 électrodes**.

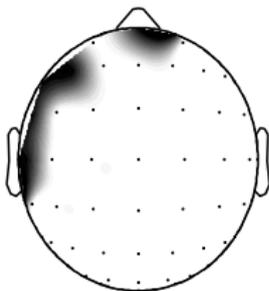
## TBC sur l'échantillon test

- Sur l'échantillon d'apprentissage :
  - le CSP sur 13 électrode donne la matrice de filtres  $W$  de taille  $13 \times 6$ .
  - une transformation log-var et un LDA fournissent la règle de décision.
- Sur l'échantillon test, on applique le filtre  $W$ +logvar+règle de décision pour obtenir le TBC.

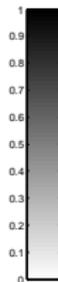
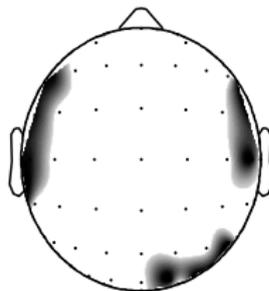
	Méthode d'évaluation	Nombre d'électrodes dans le modèle prédictif	TBC estimé sur l'échantillon test (campagne 2)	
			essais	sujets
Algorithme génétique : CSP et transformation log-var	LDA	9	71.59	75
sCSP et transformation log-var	LDA	13	73.11	75

⇒ Résultats comparables mais sCSP est beaucoup plus rapide.

Sélection par algorithme génétique

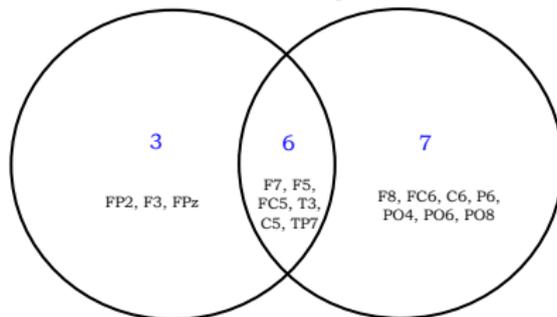


Sélection par CSP parcimonieux

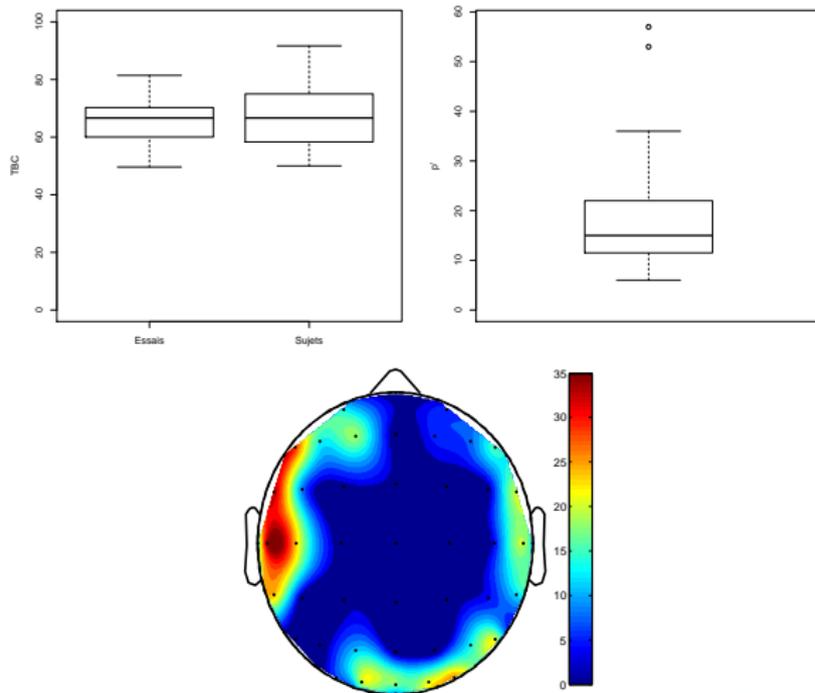


Algorithme  
génétique

CSP  
parcimonieux



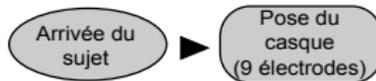
## Résultats pour 35 découpages



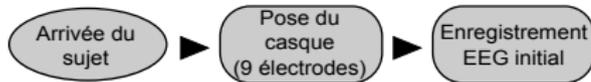
## La suite....

Arrivée du  
sujet

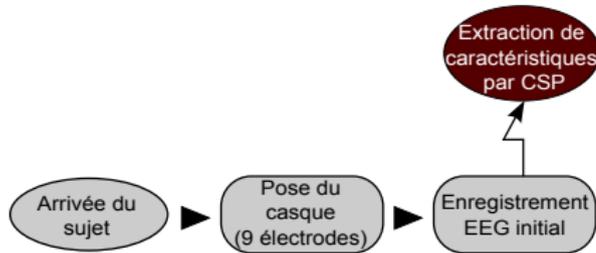
## La suite....



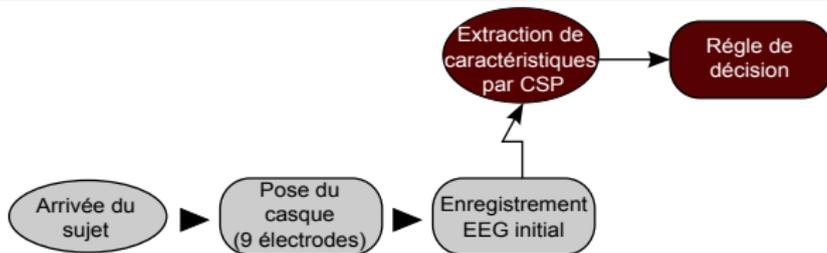
## La suite....



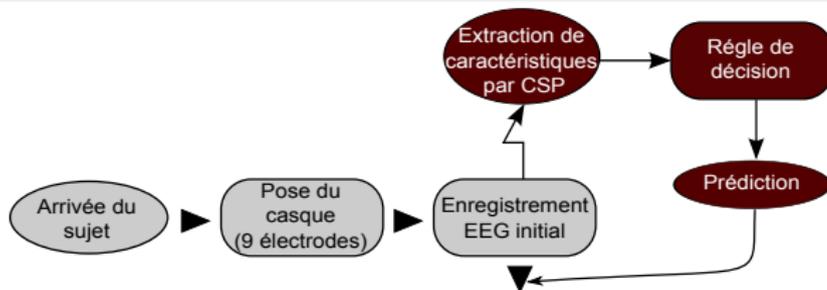
## La suite....



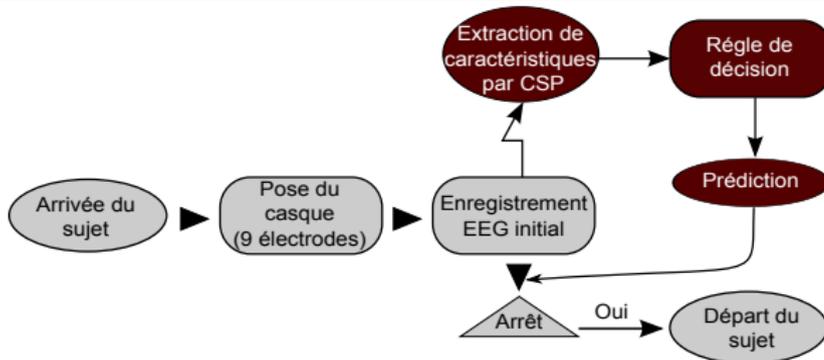
## La suite....



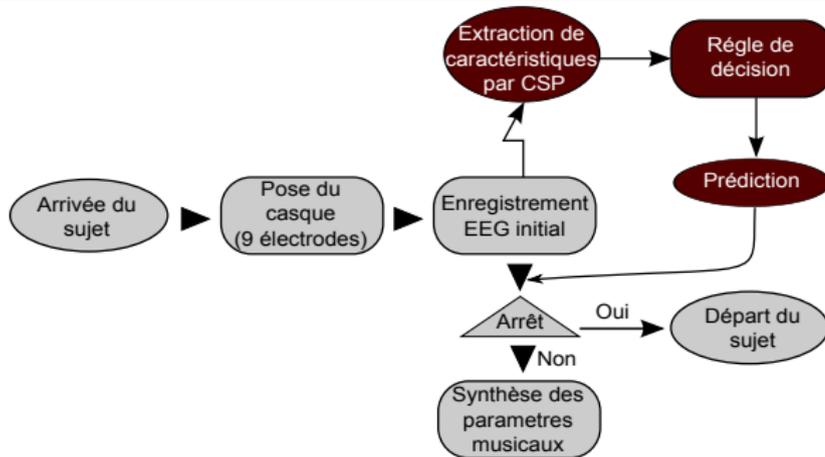
## La suite....



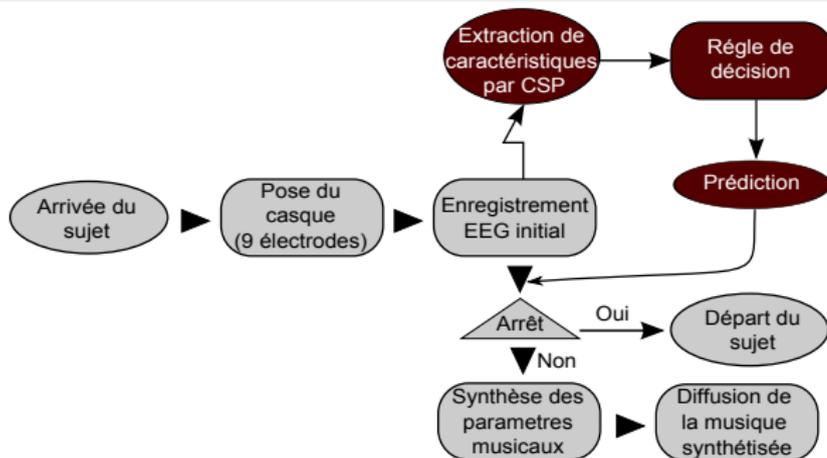
## La suite....



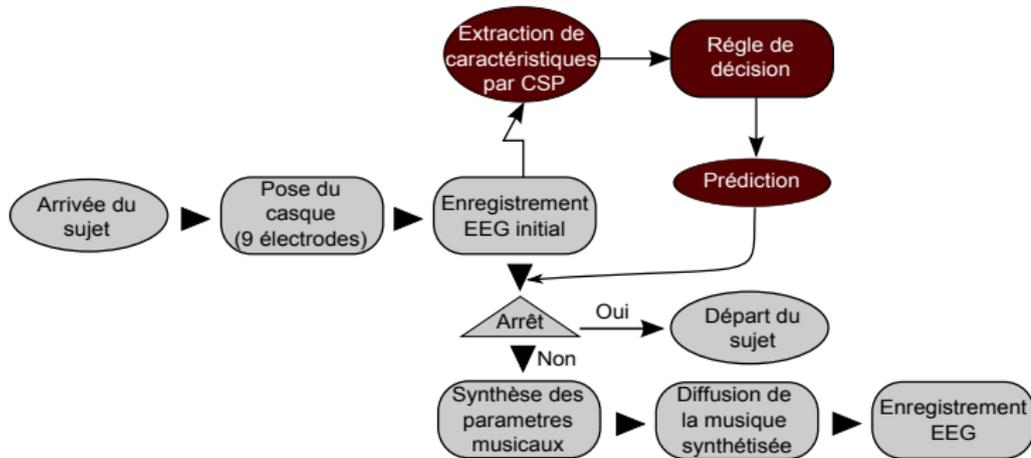
## La suite....



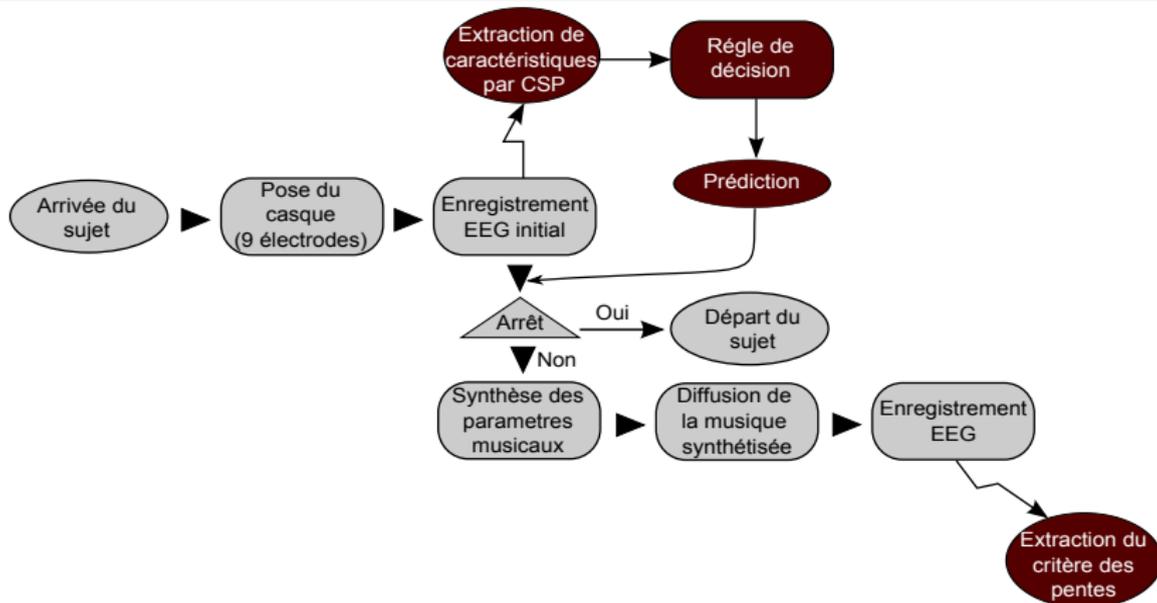
## La suite....



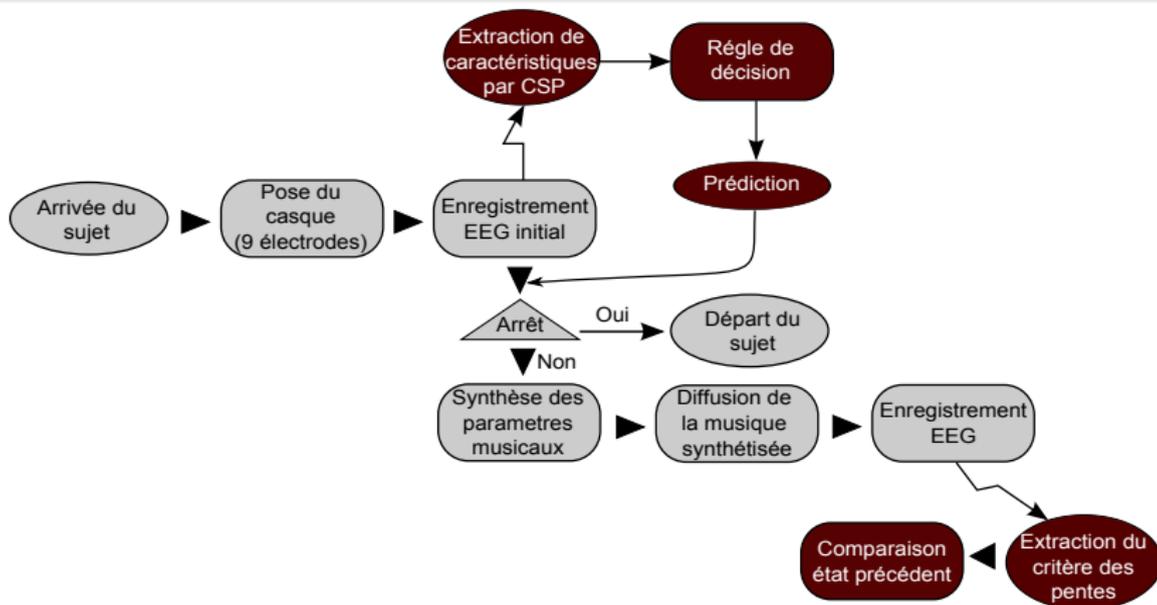
## La suite....



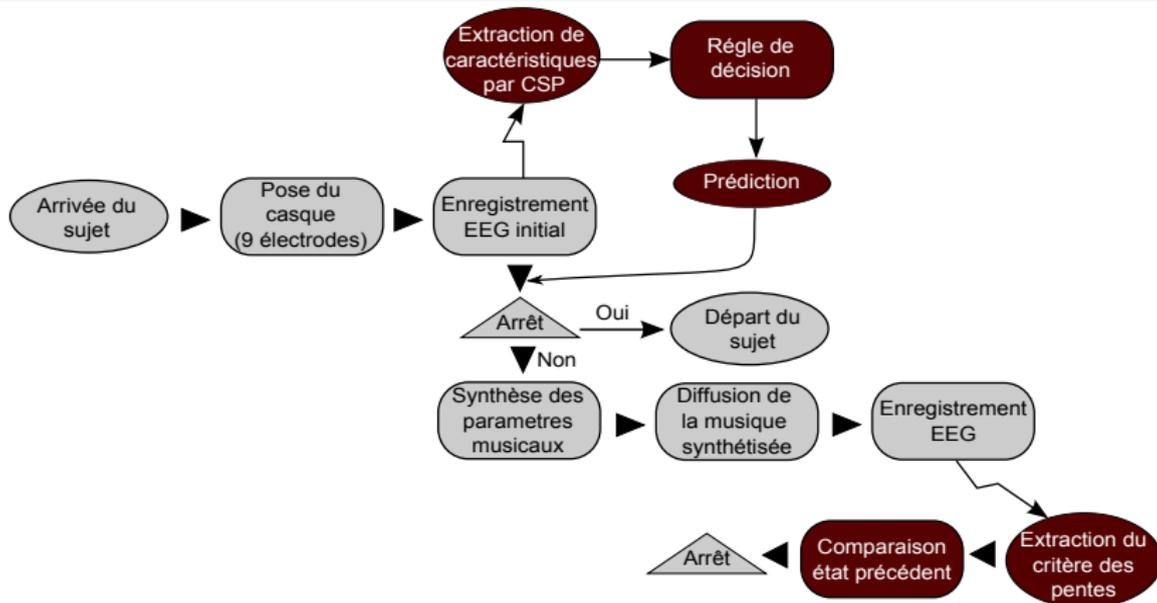
## La suite....



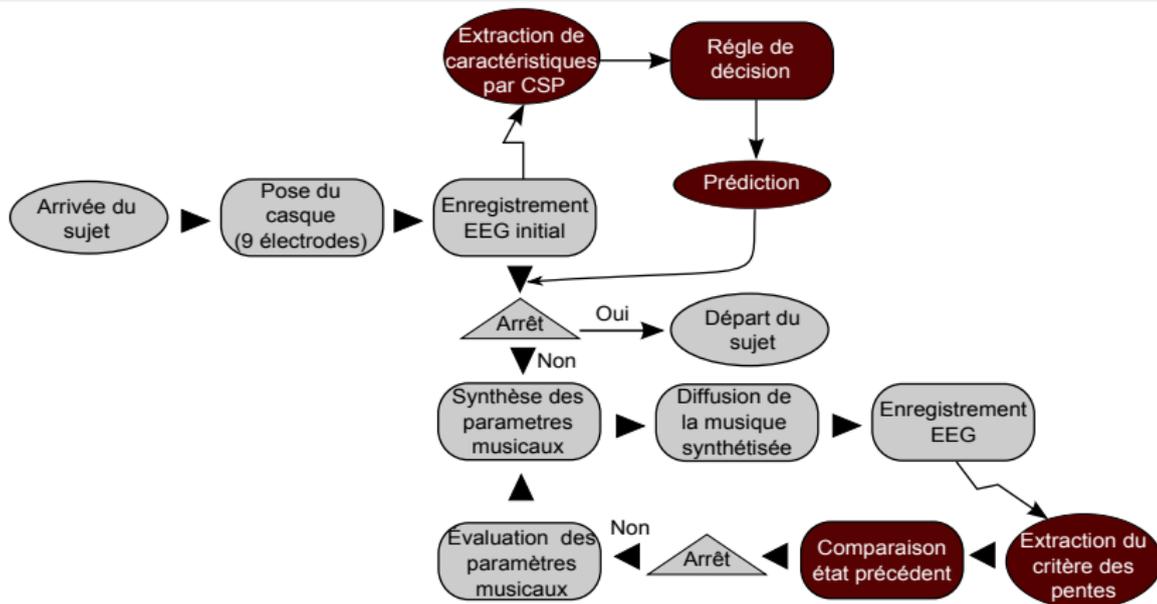
## La suite....



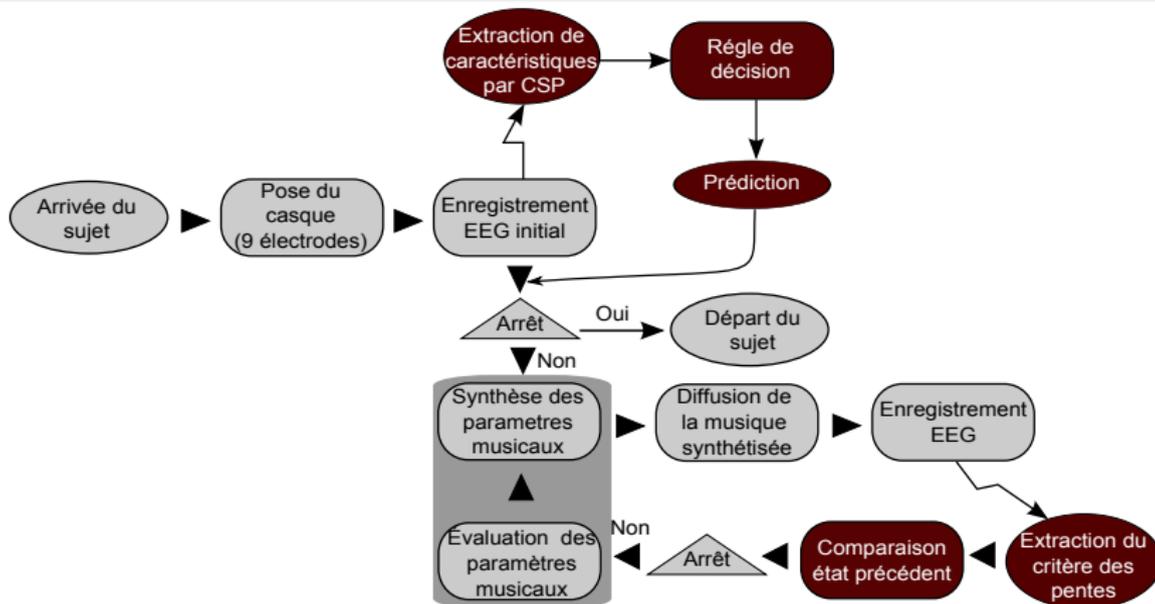
## La suite....



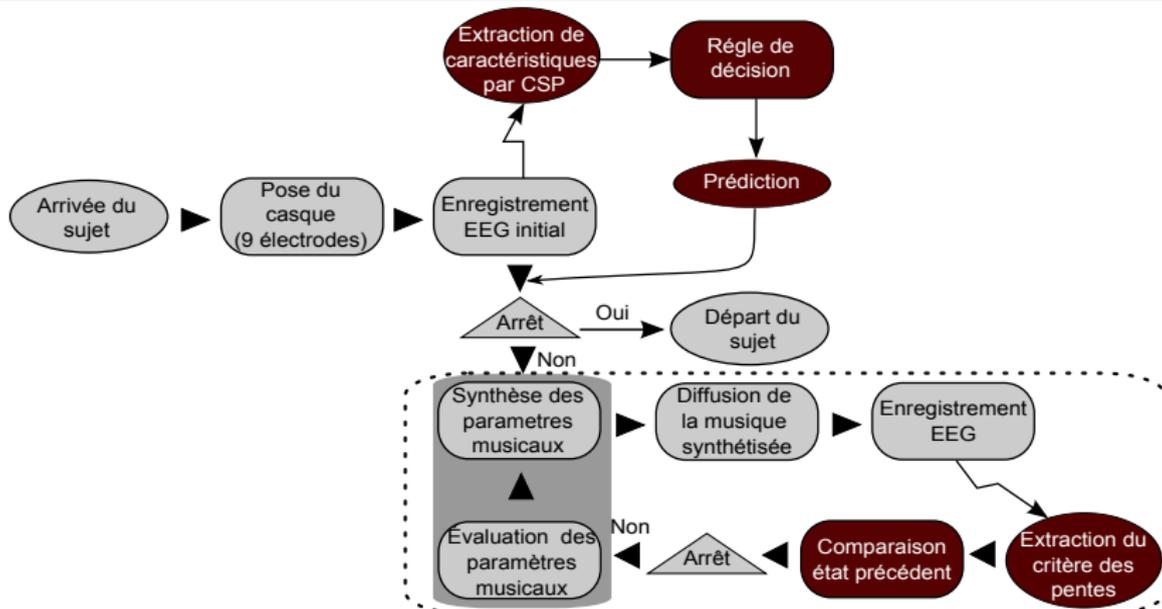
## La suite....



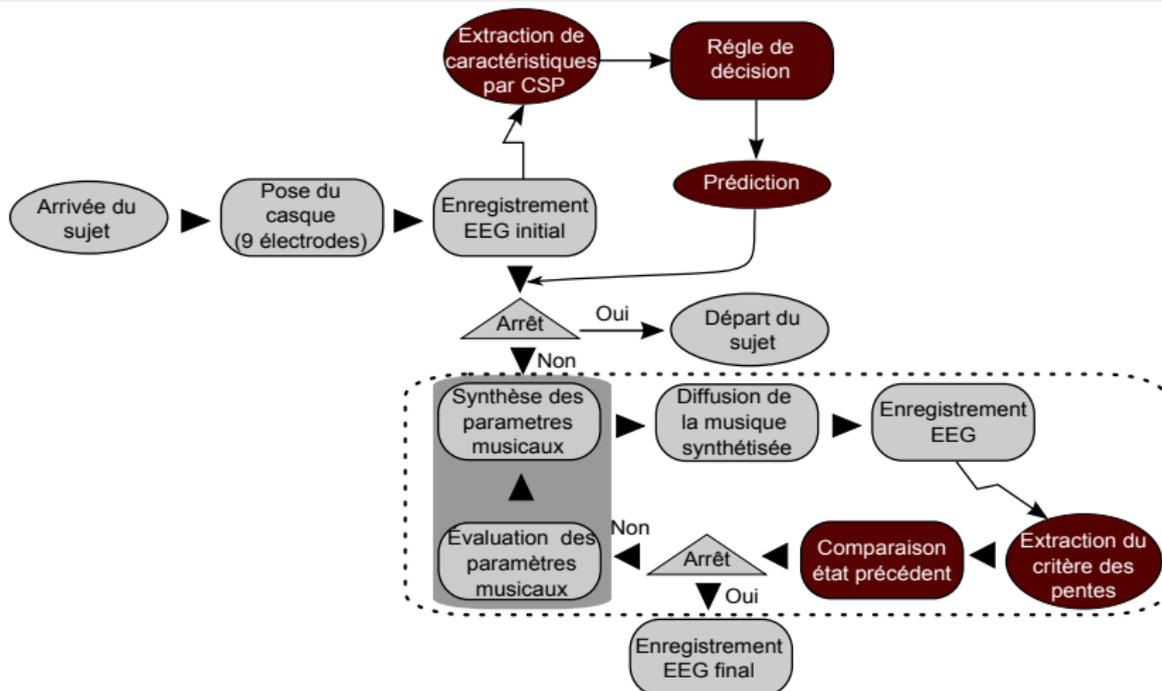
## La suite....



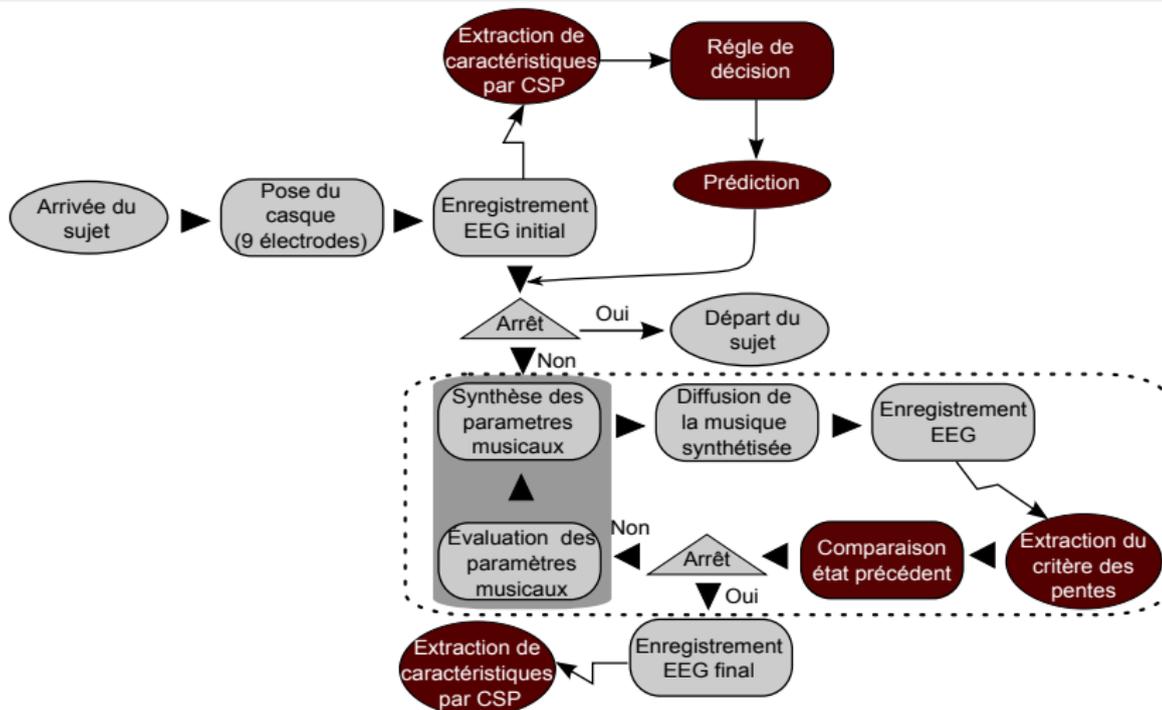
## La suite....



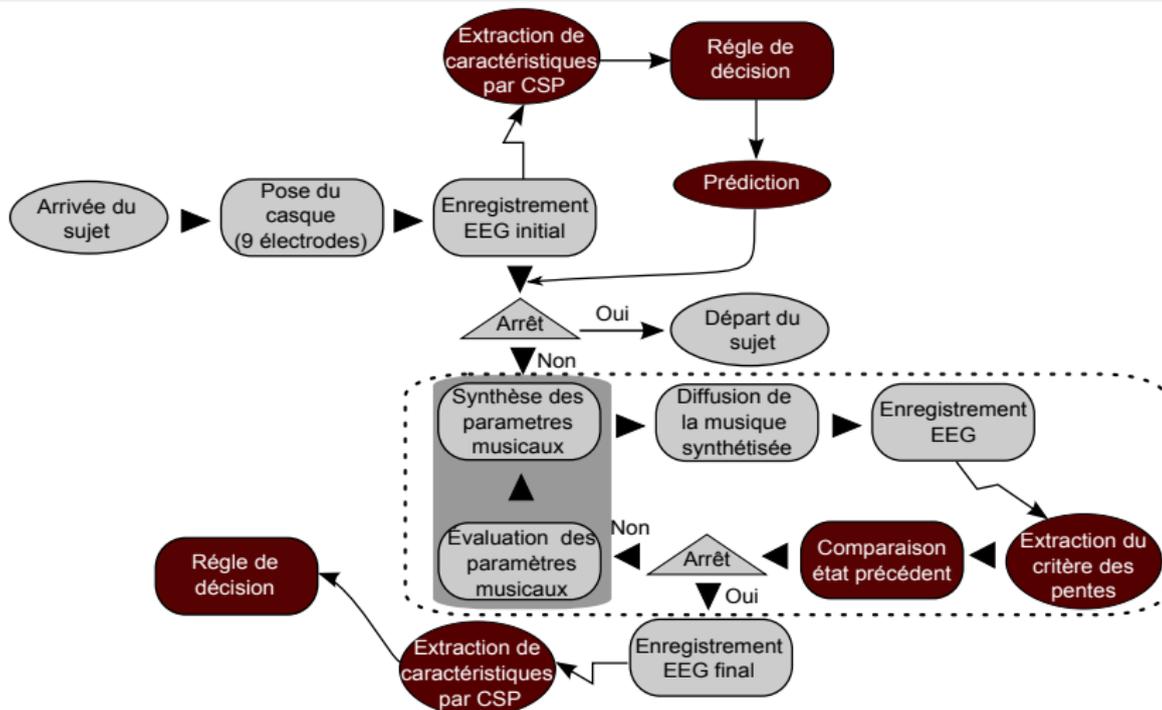
## La suite....



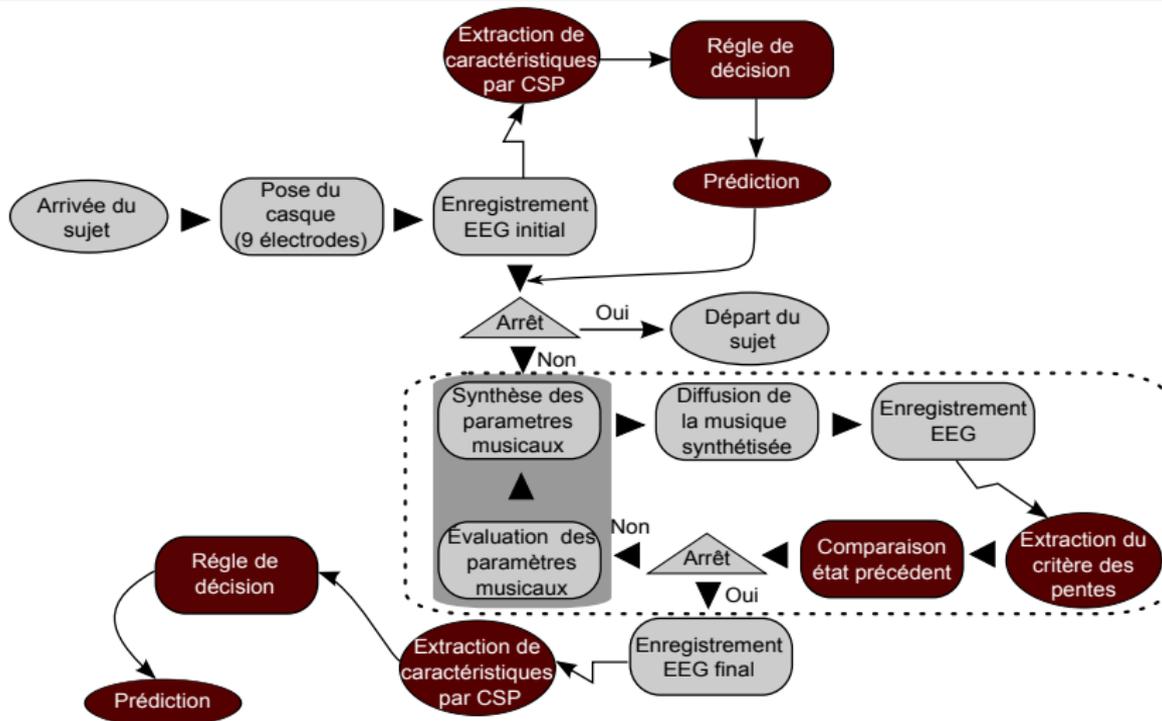
## La suite....



## La suite....



## La suite....



# Références I



B. Blankertz, K.R. Müller, D. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlögl, G. Pfurtscheller, J.R. Millán, M. Schröder and N. Birbaumer  
*The BCI competition III: Validating alternative approaches to actual BCI problems*  
Transactions on Neural Systems and Rehabilitation Engineering, 2006, 14, 153-159



C. Eckart and G. Young  
*The approximation of one matrix by another of lower rank.*  
Psychometrika, Springer, 1936, 1, 211-218



B. Efron, T. Hastie, I. Johnstone and R. Tibshirani,  
*Least angle regression.*  
The Annals of Statistics, 2003, 32, 407-499



J. Friedman, T. Hastie and R. Tibshirani,  
*Regularization Paths for Generalized Linear Models via Coordinate Descent.*  
Journal of Statistical Software, 2010, 33, 1-22



K. Fukunaga and W. L. G. Koontz,  
*Application of the Karhunen-Loève Expansion to Feature Selection and Ordering.*  
IEEE Transactions on Computers, 1970, C-19, 311-318



E. Jacobson.  
*Biologie des émotions. Les bases théoriques de la relaxation.*  
E.S.P., Paris, 1974

## Références II



M. Journée, Y. Nesterov, P. Richtarik and R. Sepulchre,  
*Generalized power method for sparse PCA.*  
Journal of Machine Learning Research, 2010, 11, 517-553



Z. J. Koles, M. S. Lazar and S. Z. Zhou  
*Spatial patterns underlying population differences in the background EEG.*  
Brain topography, 1990, vol. 2, 275-284



A. Schlögl, F. Lee, H. Bischof and G. Pfurtscheller  
*Characterization of four-class motor imagery EEG data for the BCI-competition 2005.*  
Journal of Neural Engineering, 2005, vol. 2, no. 4, pp. L14-L22



J. H. Schultz.  
*Le Training Autogène.*  
PUF, Paris, 1958



L. Vézard, P. Legrand, M. Chavent, F. Faïta-Aïnseba, and L. Trujillo  
*EEG classification for the detection of mental states.*  
To appear in Applied Soft Computing, 2015



L. Vézard,  
*Réduction de dimension en apprentissage supervisé : applications à l'étude de l'activité cérébrale.*  
Thèse de l'Université de Bordeaux, 2013

## Références II



H. Zou, T. Hastie and R. Tibshirani,

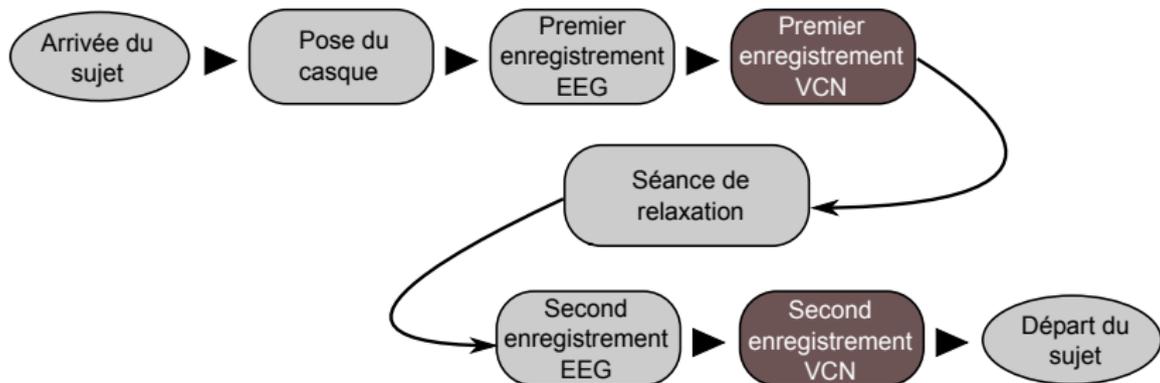
*Sparse Principal Component Analysis.*

Journal of Computational and Graphical Statistics, 2006, vol. 15, no. 2, pp. 265-286

# Contenu supplémentaire

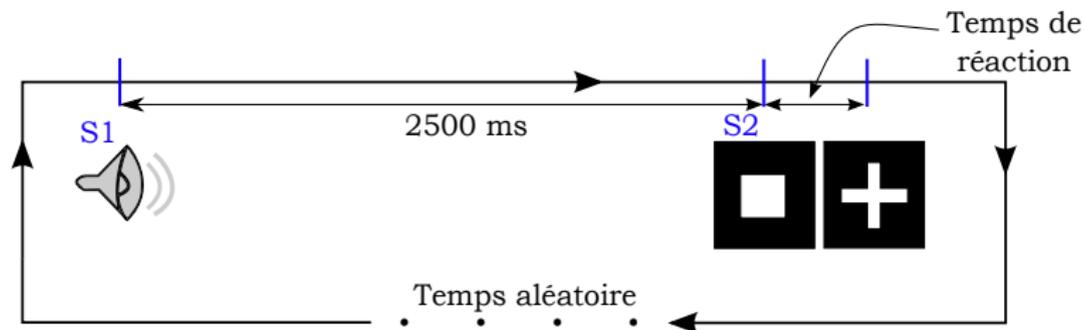
- Test VCN
- Réglages des paramètres
- Backward et Forward CSP
- Algorithme génétique et pentes
- Algorithme génétique et CSP

# Test VCN



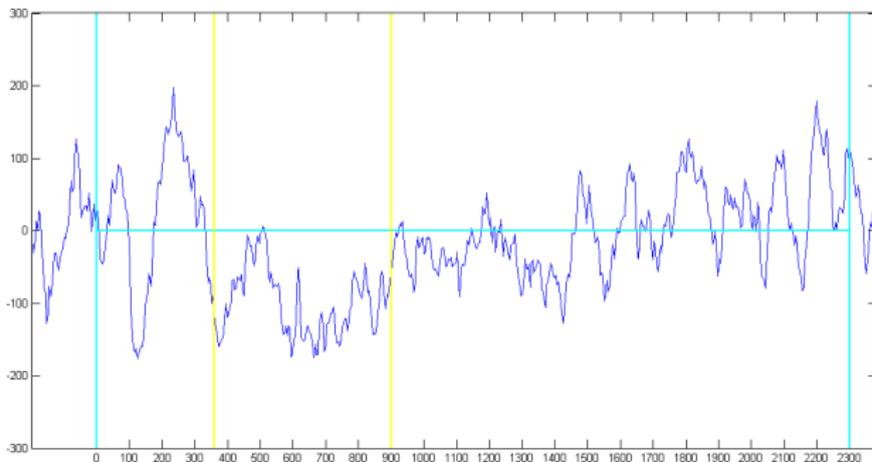
# Test VCN

- Test pour vérifier l'efficacité de la séance de relaxation
- Basé sur des données électro-physiologique : amplitude d'un potentiel évoqué



# Traitement des données VCN

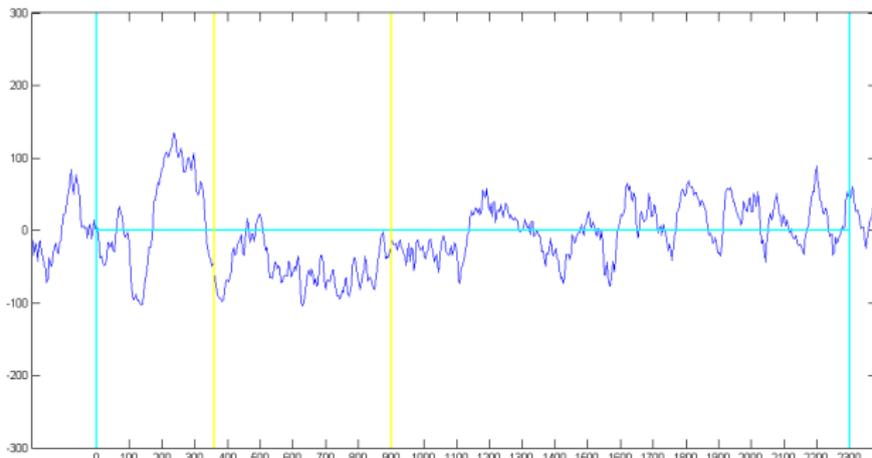
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



1 essai

# Traitement des données VCN

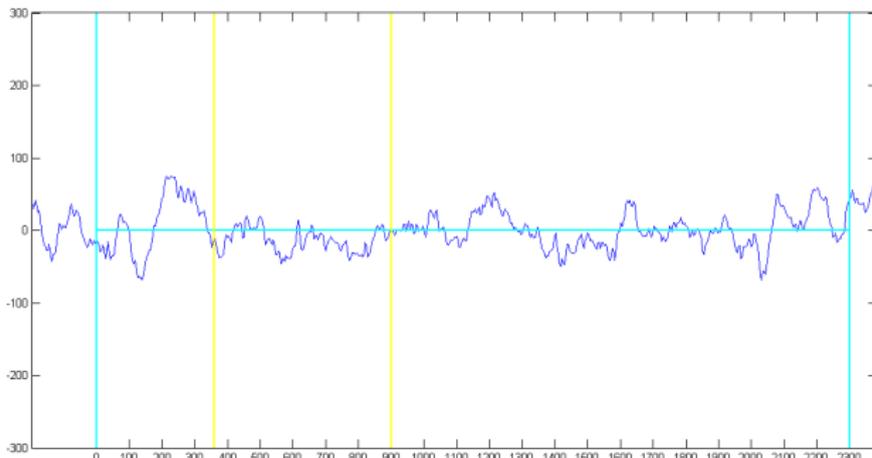
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 2 essais

# Traitement des données VCN

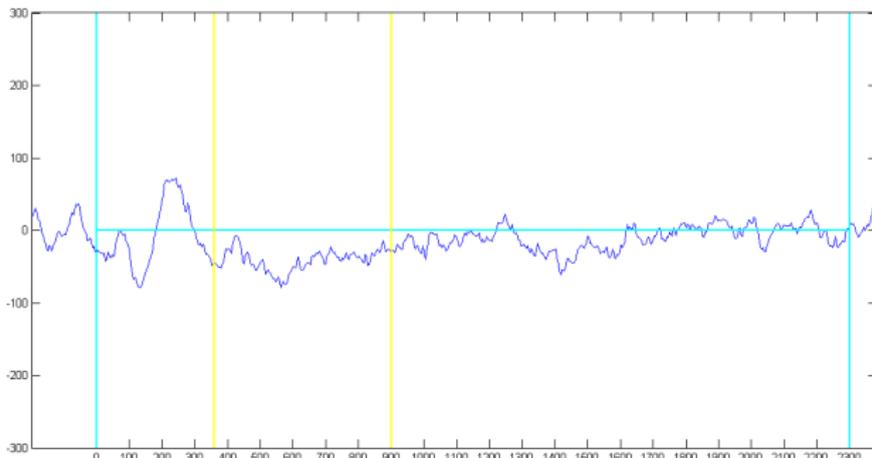
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 5 essais

# Traitement des données VCN

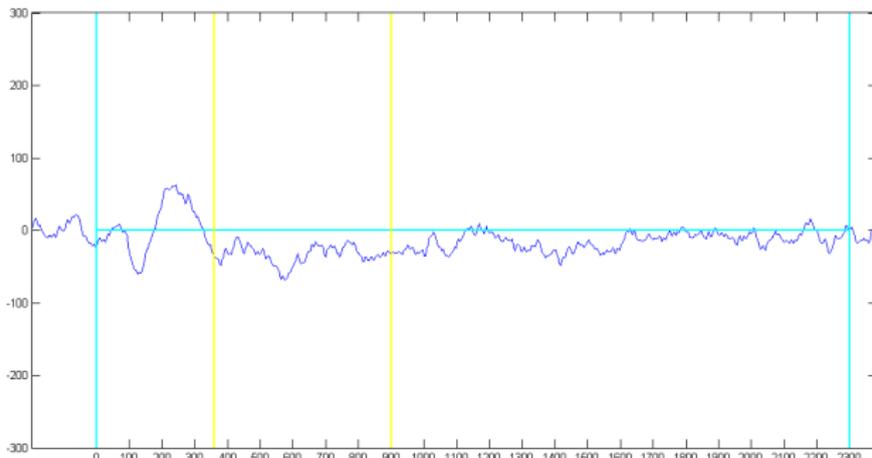
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 10 essais

# Traitement des données VCN

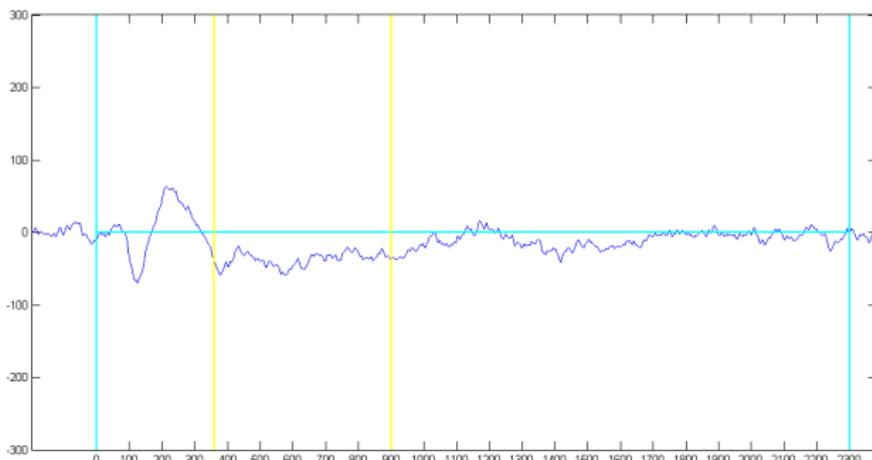
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 20 essais

# Traitement des données VCN

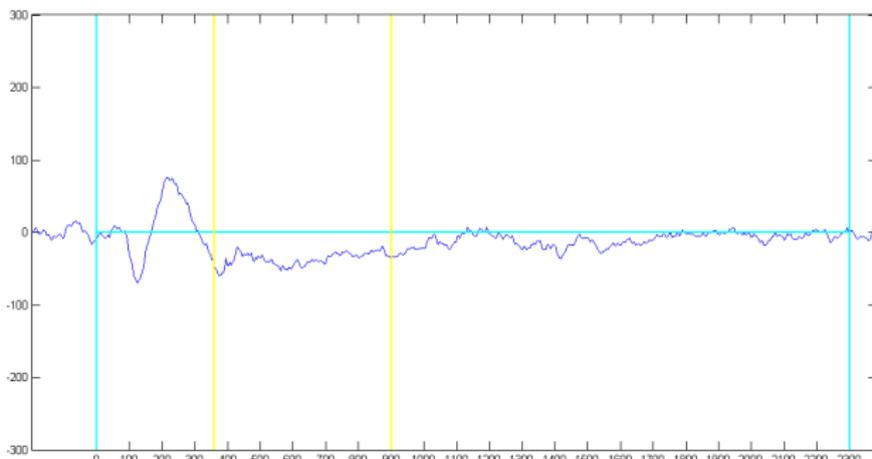
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 30 essais

# Traitement des données VCN

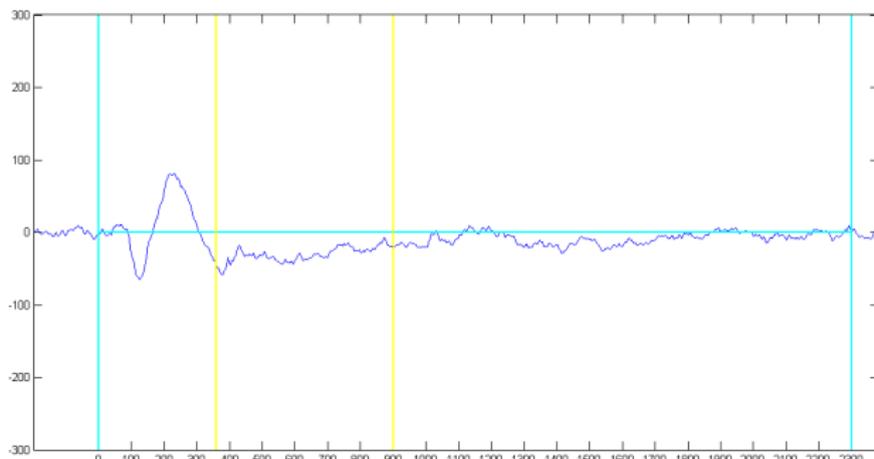
- Potentiels évoqués indécélables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).



Moyenne sur 40 essais

# Traitement des données VCN

- Potentiels évoqués indétectables tels quels  $\implies$  technique du moyennage
- 50 essais par session VCN (50 avant relaxation, 50 après).

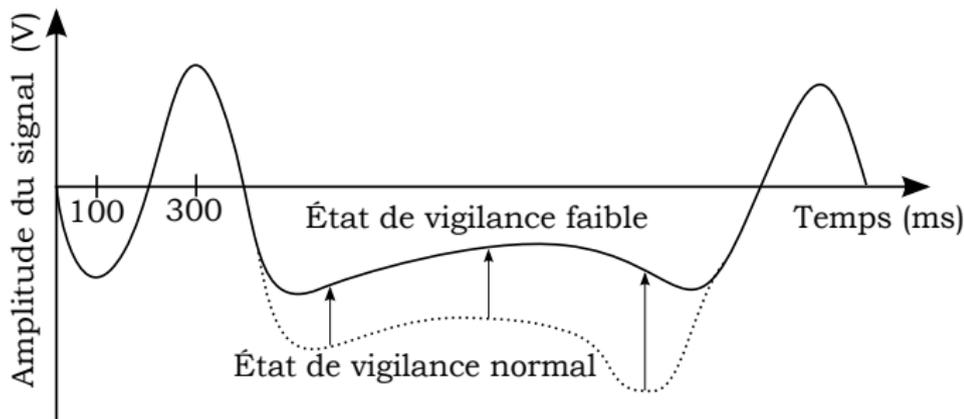


Moyenne sur 50 essais

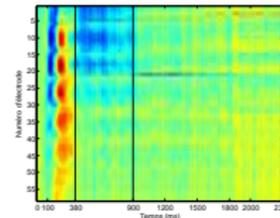
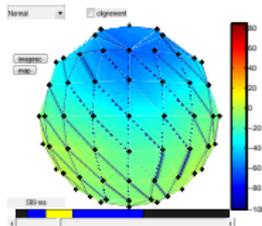
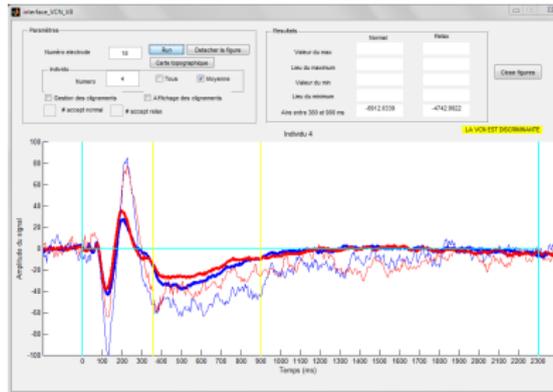
# La VCN comme indicateur des changements d'état de vigilance

Participant less alert  $\Leftrightarrow$  CNV amplitude  $\searrow$

- Lapse in concentration [Tecce et al, 1978]
- Sleep deprivation [Naitoh et al, 1970]
- Drowsiness [Timsit-Berthier et al, 1981]



# La VCN comme indicateur des changements d'état de vigilance

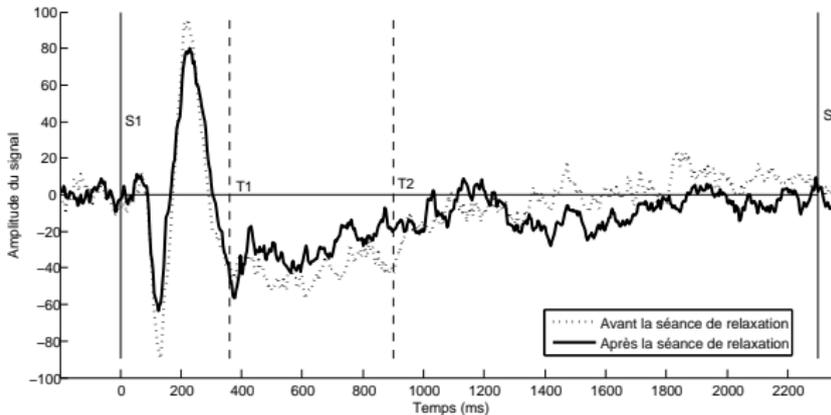


# Traitement des données VCN

## Résultats de l'analyse VCN

19 des 58 sujets ont été conservés

Exemple de tracés de VCN recueillis avant et après la séance de relaxation pour un sujet conservé :

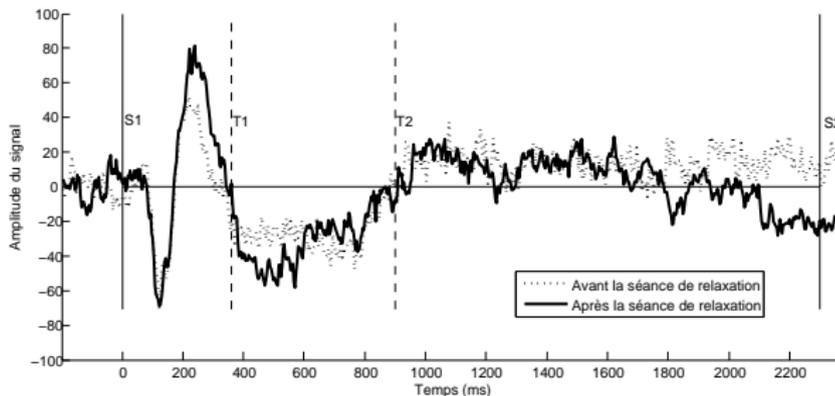


# Traitement des données VCN

## Résultats de l'analyse VCN

19 des 58 sujets ont été conservés

Exemple de tracés de VCN recueillis avant et après la séance de relaxation pour un sujet rejeté :



Retour

# Réglages des paramètres

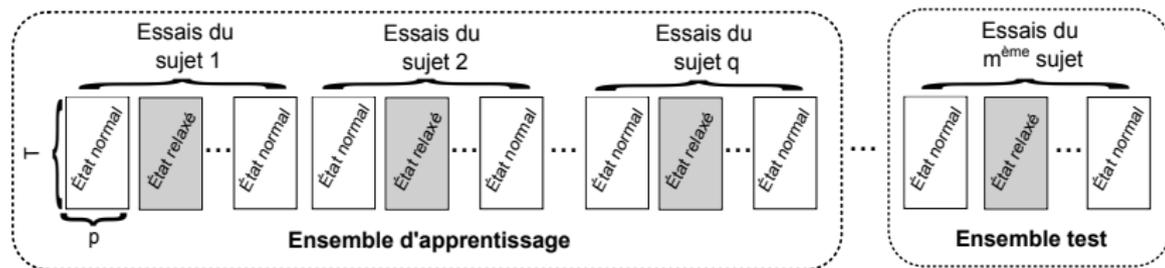
## Paramètres

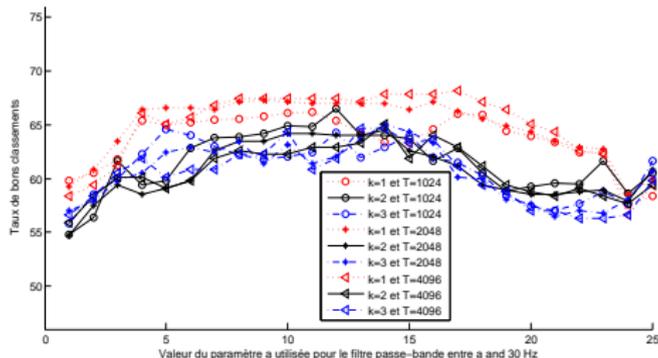
- Filtre passe-bande :  $[a, b]$  Hz où  $a = 1, \dots, 25$  Hz et  $b = a + 3, 40$  Hz.
- Longueur des essais :  $T = 1024, 2048$  ou  $4096$  points
- Nombre de paires de filtres CSP :  $k = 1, 2$  ou  $3$ .

## Normalisation

- $\tilde{X}_i = X_i$  i.e. les données ne sont pas normalisées.
- $\tilde{X}_i = \frac{X_i}{\text{trace}(C_i)}$  i.e. les données sont normalisées par l'inertie de  $X_i$ .
- $\tilde{X}_i = X_i D$ , où  $D = \begin{pmatrix} \frac{1}{S_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{S_p} \end{pmatrix}$ , et où  $S_j, j = 1, \dots, p$ , est l'écart type empirique de la  $j^{\text{ème}}$  colonne de  $X_i$ .

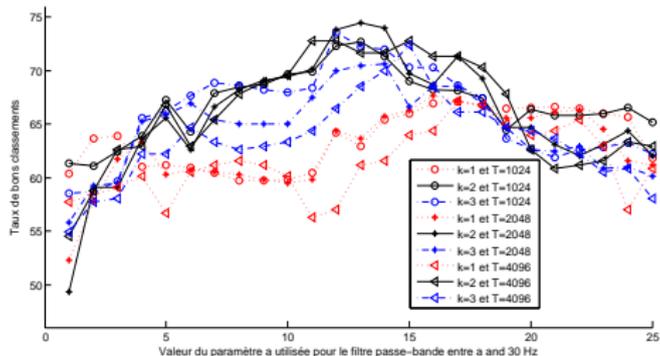
# Estimation du TBC par Leave One Out





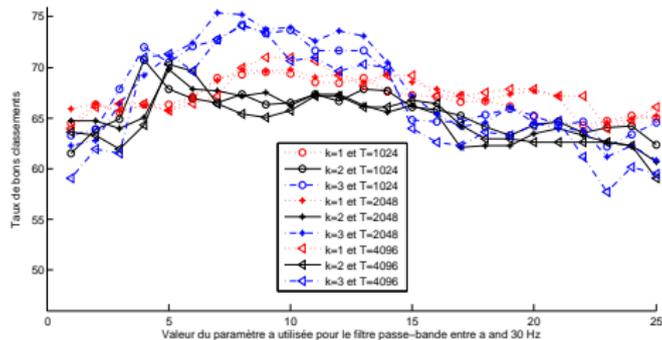
## Résultats

- Filtre passe-bande :  $[7, b]$  Hz.
- Filtre passe-bande :  $[7, 30]$  Hz.
- Longueur des essais :  $T = 2048$  points, d'où  $n^* = 22$  essais  $\left(\lfloor \frac{46000}{2048} \rfloor = 22\right)$ .
- Nombre de paires de filtres CSP : 3.
- Normalisation:  $\tilde{X}_i = X_i D$ .



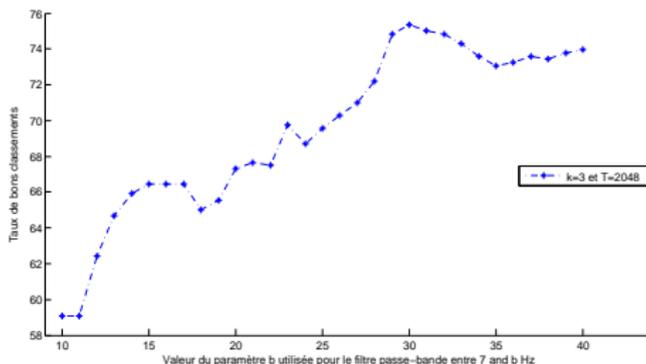
## Résultats

- Filtre passe-bande :  $[7, b]$  Hz.
- Filtre passe-bande :  $[7, 30]$  Hz.
- Longueur des essais :  $T = 2048$  points, d'où  $n^* = 22$  essais  $\left(\lfloor \frac{46000}{2048} \rfloor = 22\right)$ .
- Nombre de paires de filtres CSP : 3.
- Normalisation:  $\tilde{X}_i = X_i D$ .



## Résultats

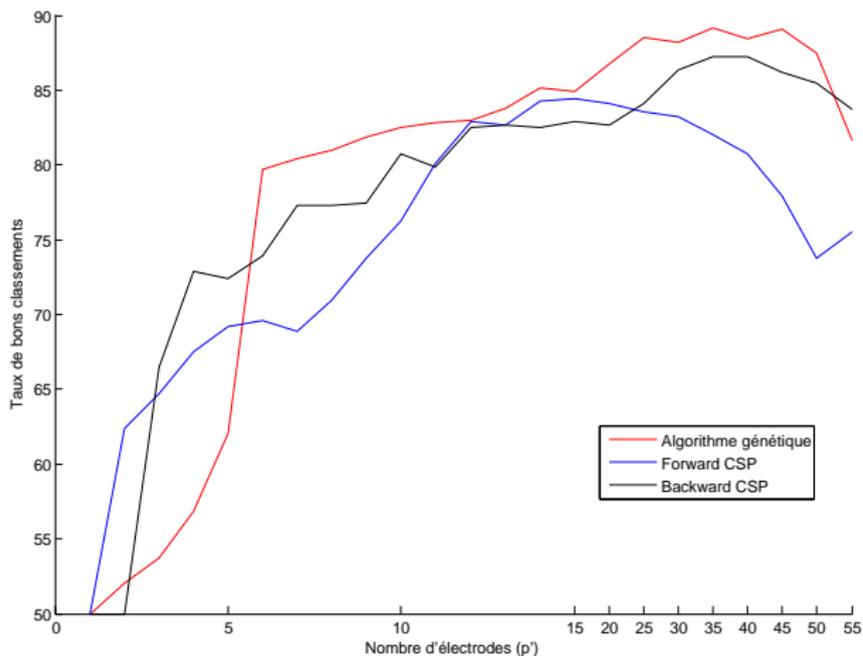
- Filtre passe-bande :  $[7, b]$  Hz.
- Filtre passe-bande :  $[7, 30]$  Hz.
- Longueur des essais :  $T = 2048$  points, d'où  $n^* = 22$  essais ( $\lfloor \frac{46000}{2048} \rfloor = 22$ ).
- Nombre de paires de filtres CSP : 3.
- Normalisation:  $\tilde{X}_i = X_i D$ .



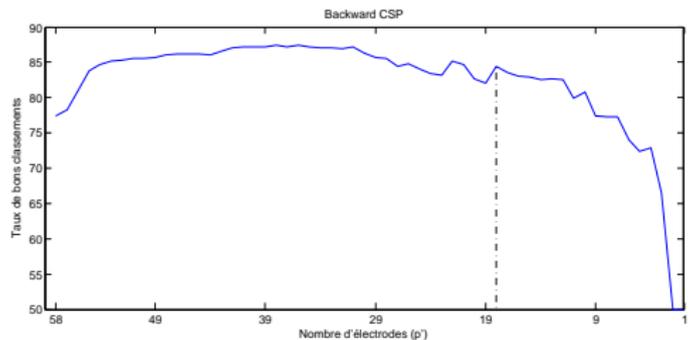
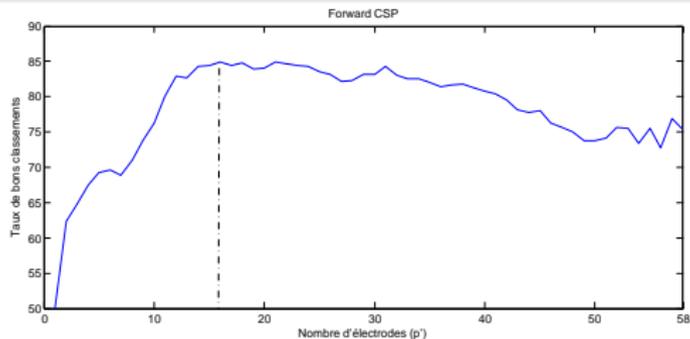
## Résultats

- Filtre passe-bande :  $[7, b]$  Hz.
- Filtre passe-bande :  $[7, 30]$  Hz.
- Longueur des essais :  $T = 2048$  points, d'où  $n^* = 22$  essais  $\left(\lfloor \frac{46000}{2048} \rfloor = 22\right)$ .
- Nombre de paires de filtres CSP : 3.
- Normalisation:  $\tilde{X}_i = X_i D$ .

# Backward et Forward CSP

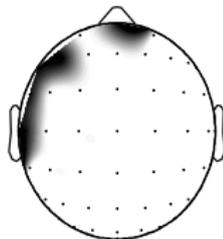


# Backward et Forward CSP

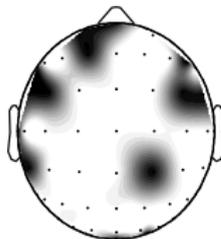


# Comparaison des sélection d'électrodes

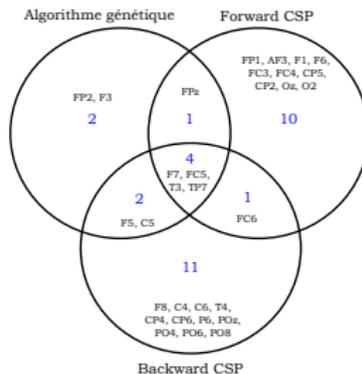
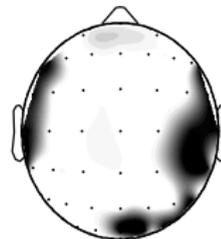
Sélection par algorithme génétique



Sélection par forward CSP



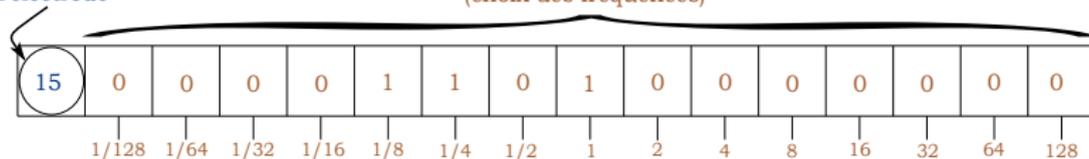
Sélection par backward CSP



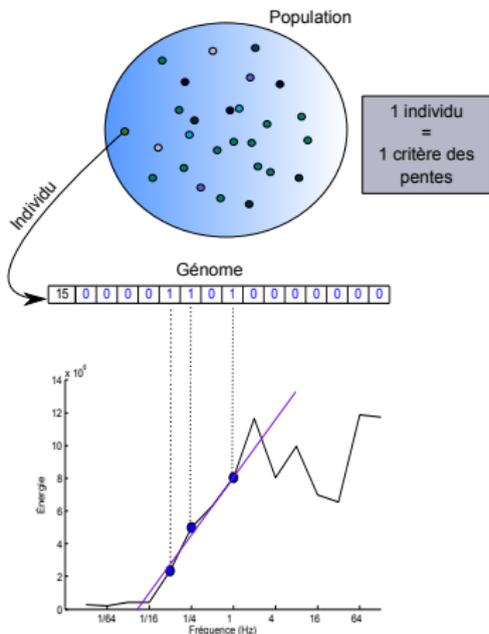
# Algorithme génétique pour l'optimisation du critère des pentes

Numéro de l'électrode

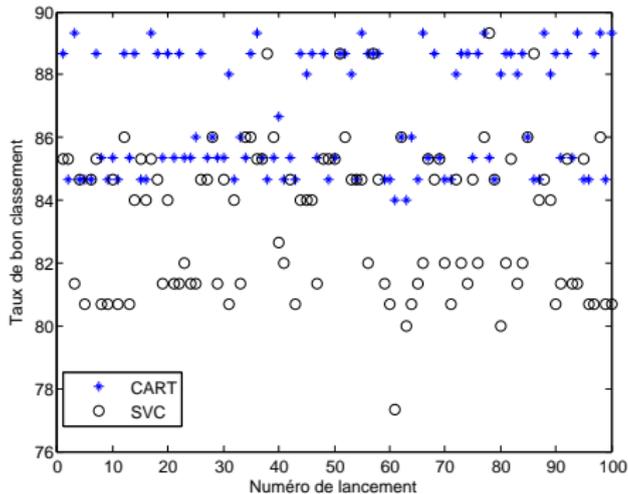
Partie binaire  
(choix des fréquences)



# Algorithme génétique pour l'optimisation du critère des pentes



# Algorithme génétique pour l'optimisation du critère des pentes



Retour

# Algorithme génétique et CSP

## Objectif

Optimisation itérative d'une fonction fitness.

## Idées générales

- Abstraction de la théorie de l'évolution de Darwin.
- Utilisation d'opérations génétiques (sélection, mutation, croisement).

## Terminologie

- Individu = Solution potentielle.
- Génome = Série de chiffres binaires ou réels.

Initialisation  
aléatoire de  
la population

Initialisation  
aléatoire de  
la population

58 composantes où  
 $p'$  sont égales à 1

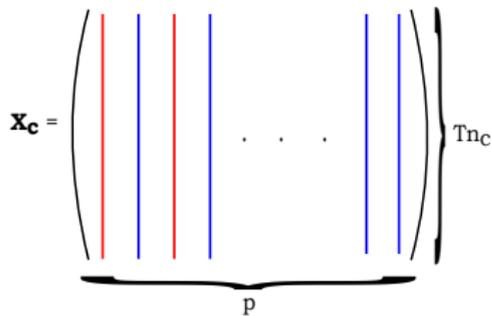
Génome 

0	1	0	1	.	.	.	1	1
---	---	---	---	---	---	---	---	---

Initialisation  
aléatoire de  
la population

58 composantes où  
 $p'$  sont égales à 1

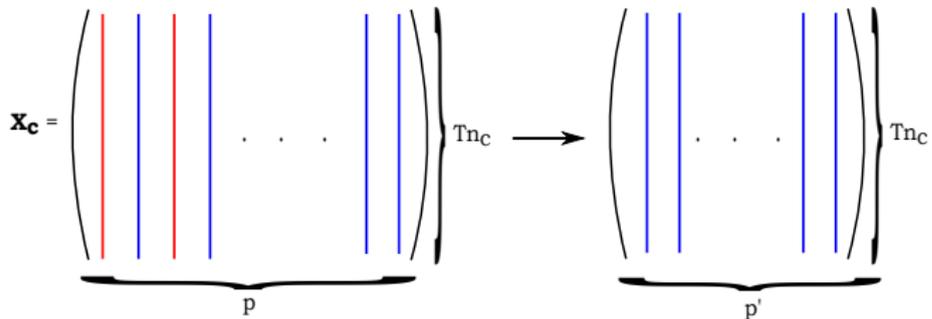
Génome

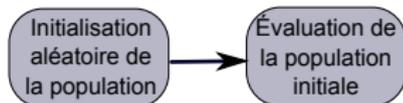


Initialisation  
aléatoire de  
la population

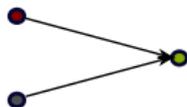
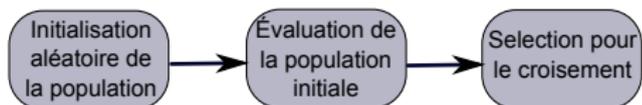
58 composantes où  
 $p'$  sont égales à 1

Génome

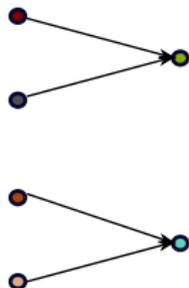
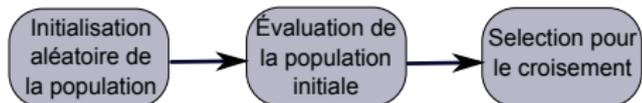




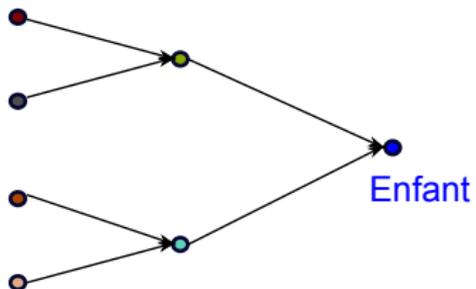
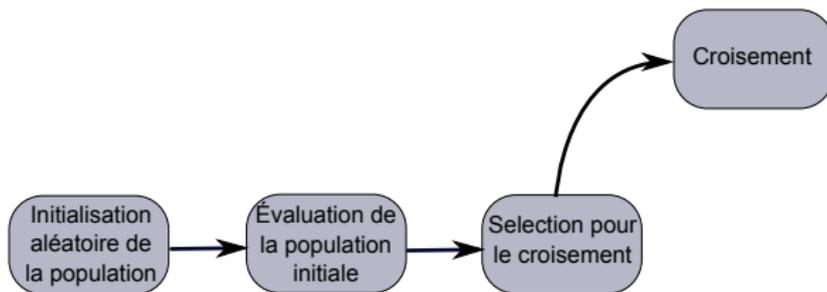
Méthode statistique

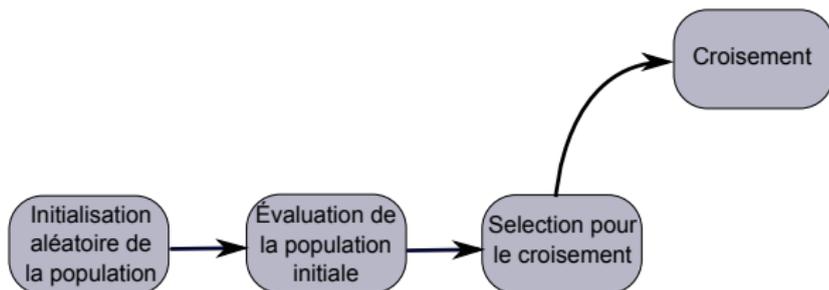


Tournoi

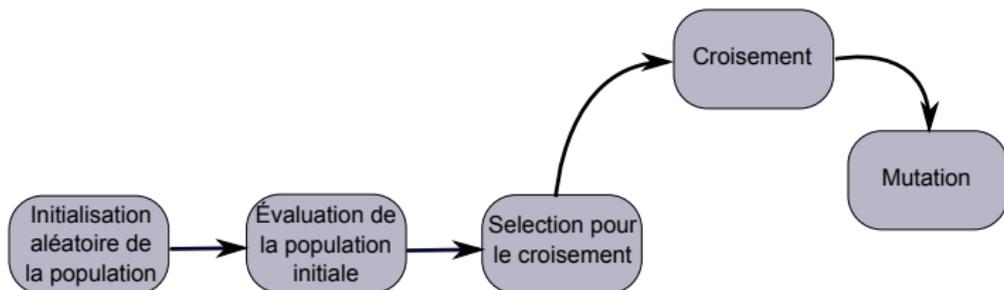


Tournoi

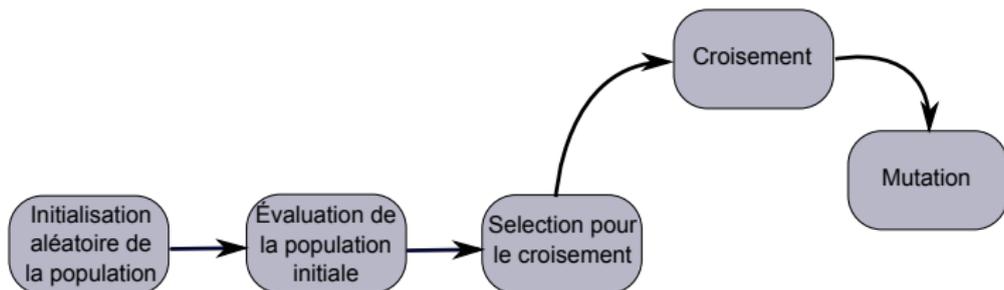




P1	P2	Enfant
0	0	0
0	1	1
1	0	0
0	0	1

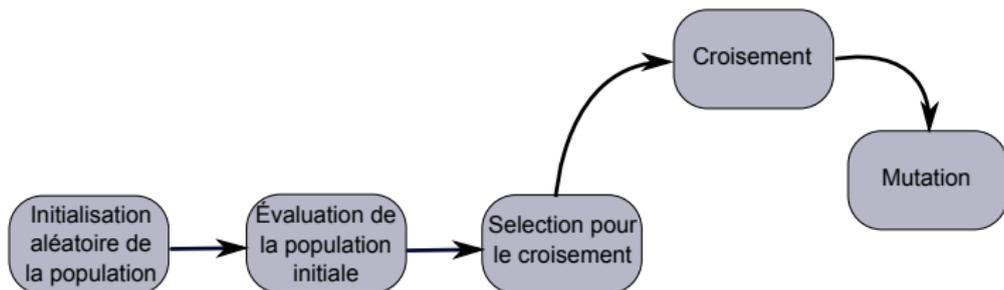


Chaque composante  
du génome mute avec  
une probabilité de  $2/58$

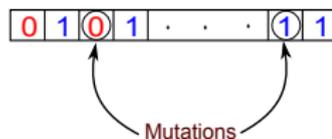


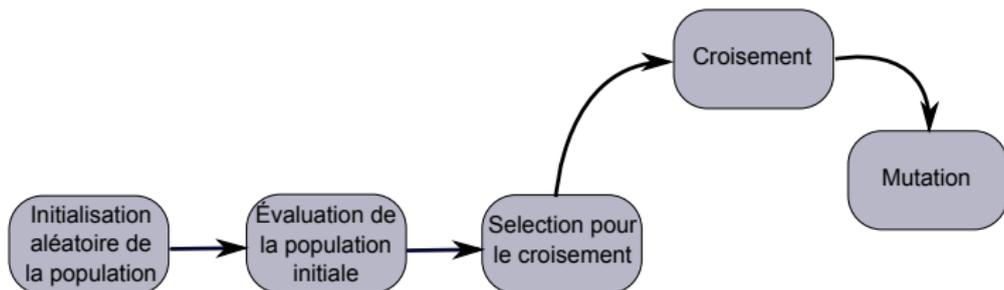
Chaque composante  
du génome mute avec  
une probabilité de  $2/58$

0 1 0 1 . . . 1 1



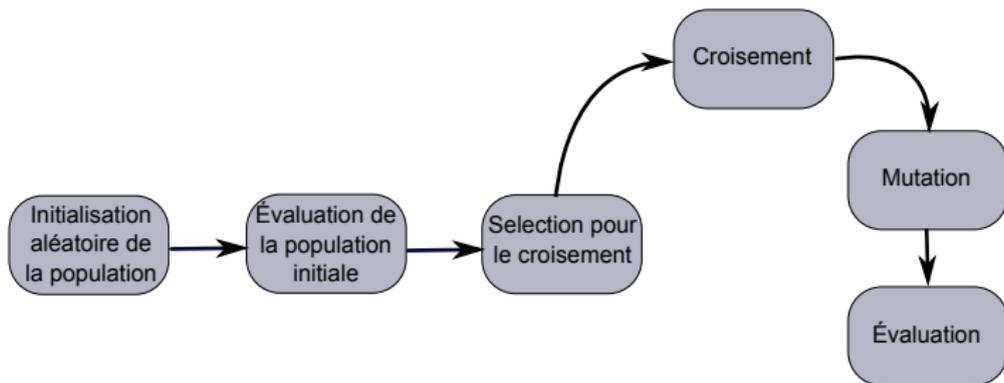
Chaque composante  
du génome mute avec  
une probabilité de  $2/58$



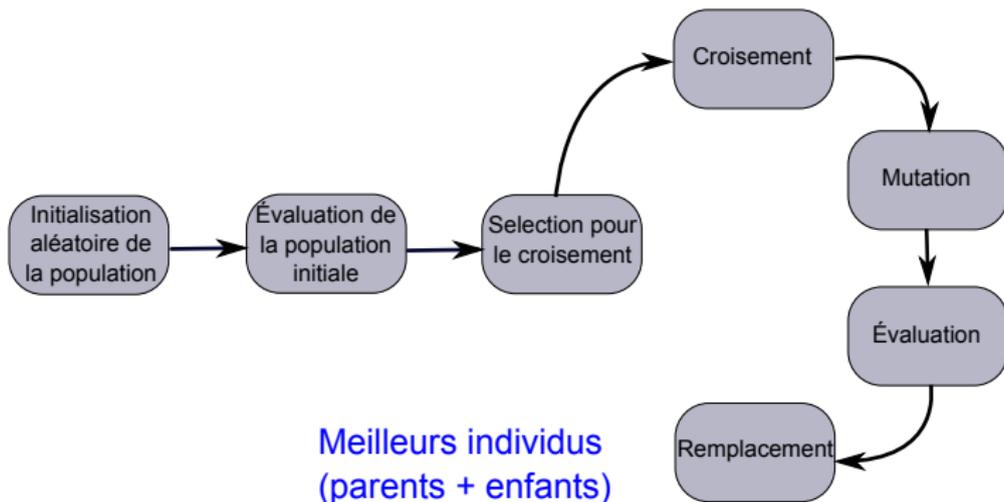


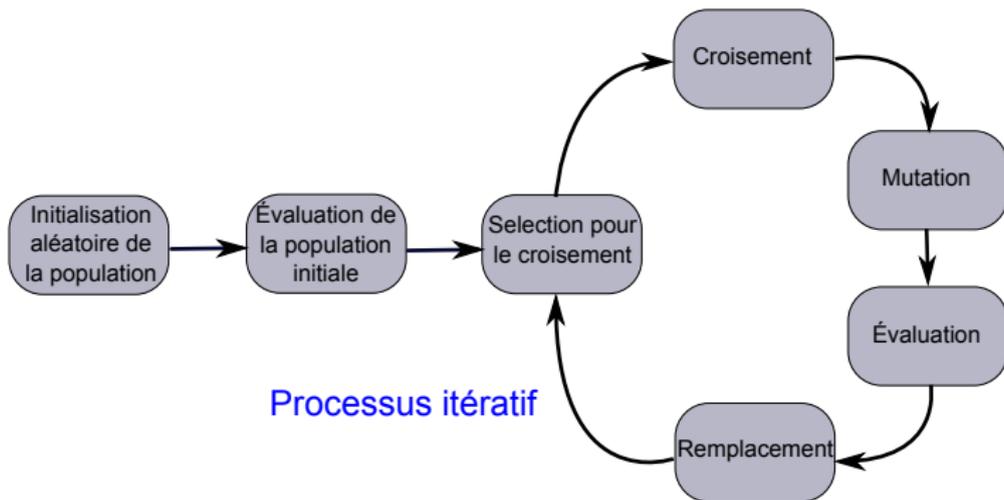
Chaque composante  
du génome mute avec  
une probabilité de  $2/58$

0 1 1 1 . . . 0 1

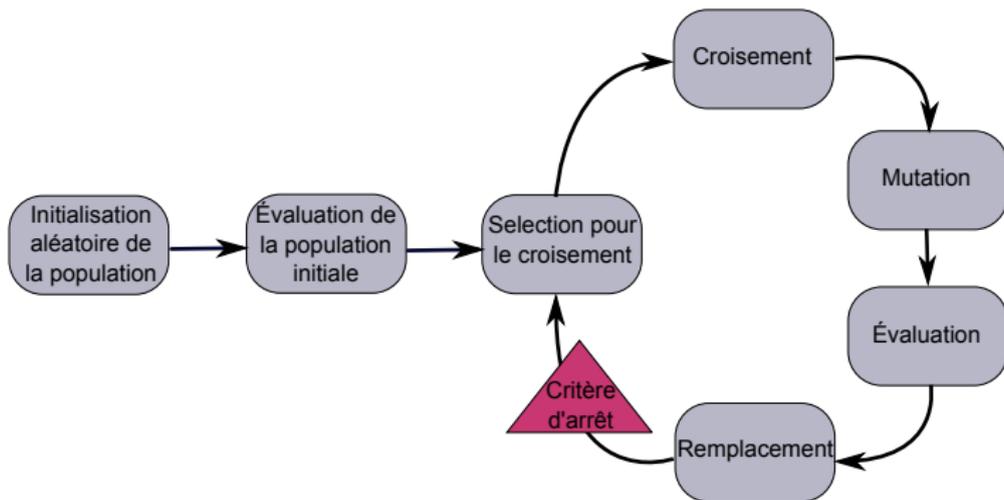


Méthode statistique

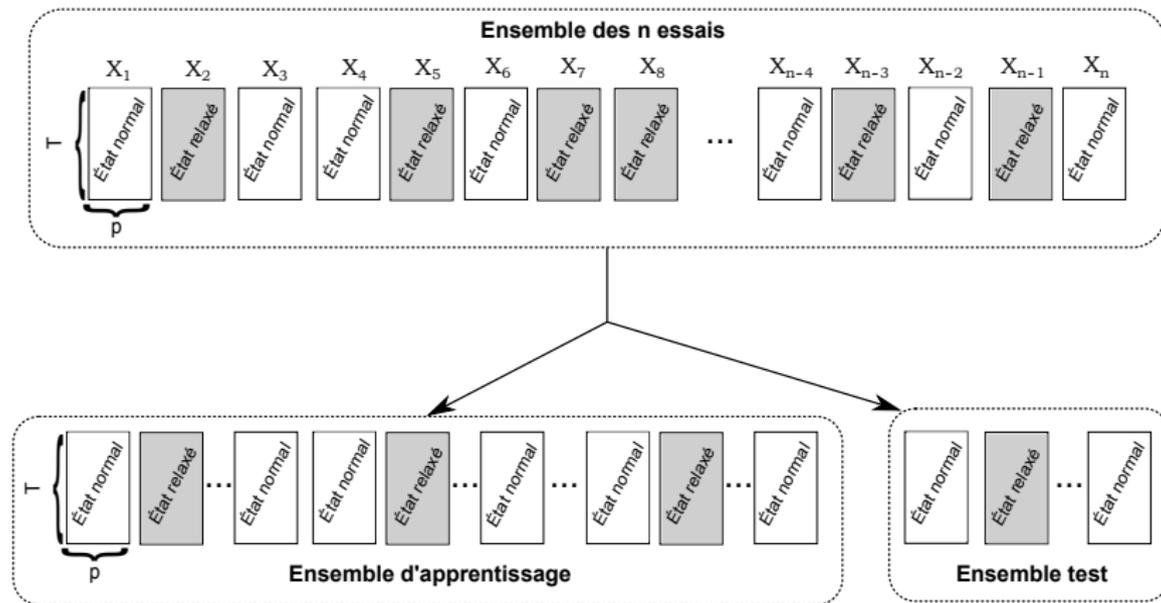




Processus itératif



# Estimation du TBC par Leave One Out sur les essais des sujets de la campagne 1



Génomés

1 0 0 1 | 1 0

0 0 1 1 | 0 1

⋮

0 1 0 1 | 1 1

58 composantes où  
p' sont égales à 1

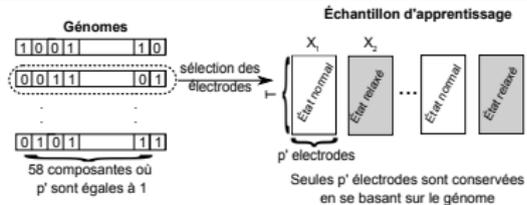
Géomes

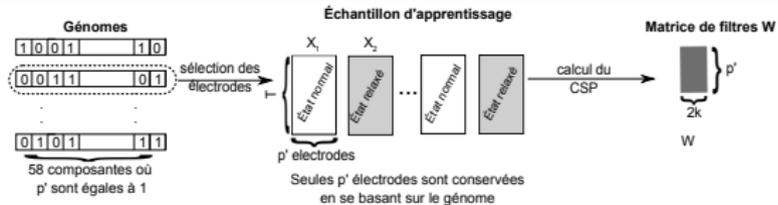
1 0 0 1 1 0

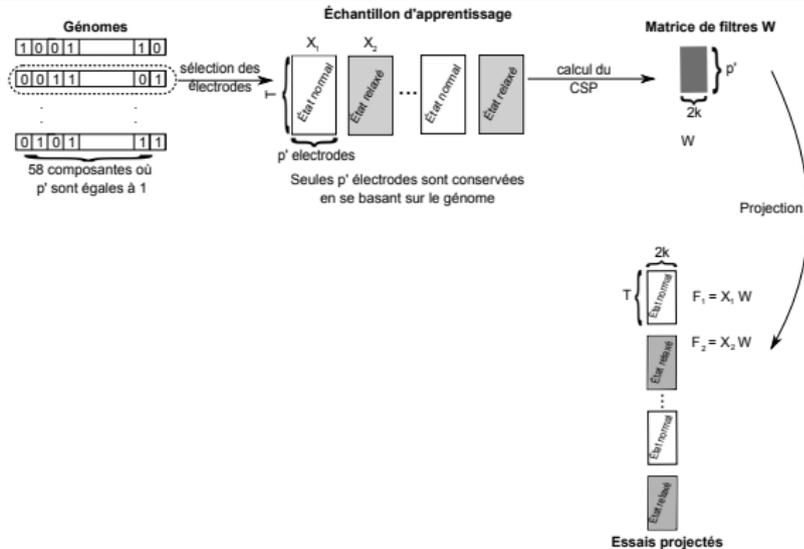
0 0 1 1 0 1

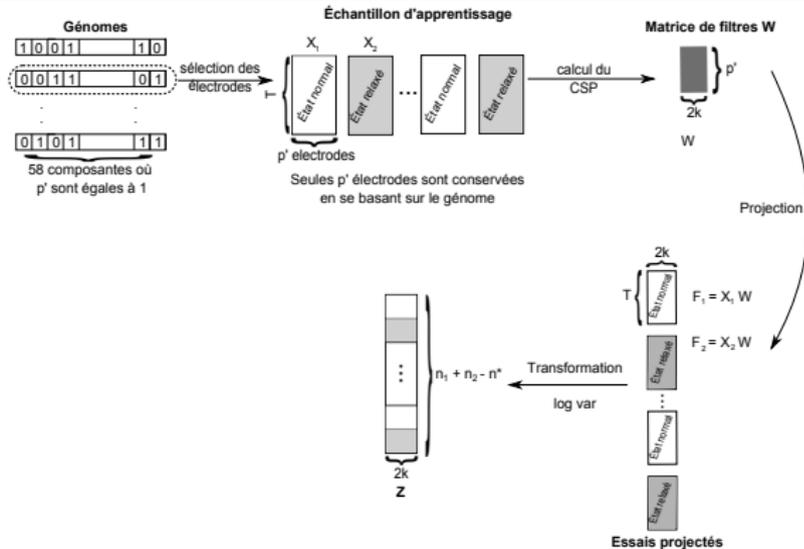
0 1 0 1 1 1

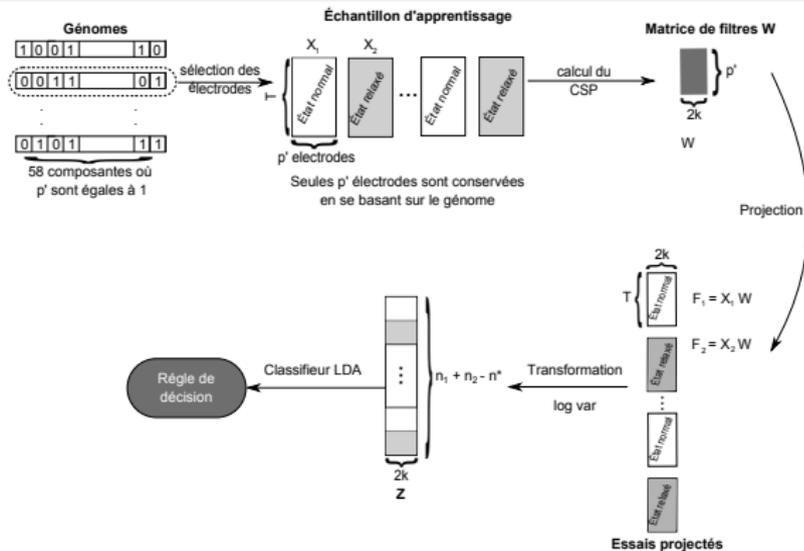
58 composantes où  
p' sont égales à 1

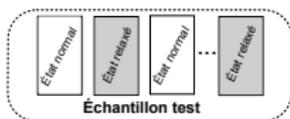
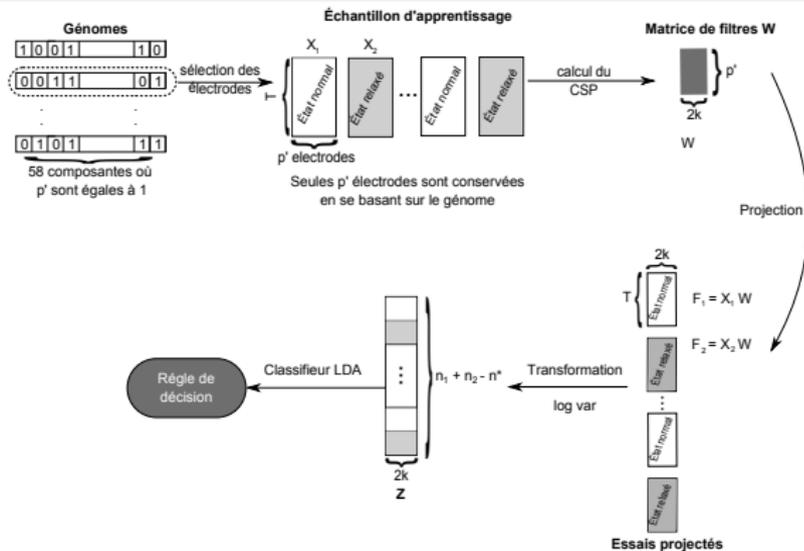


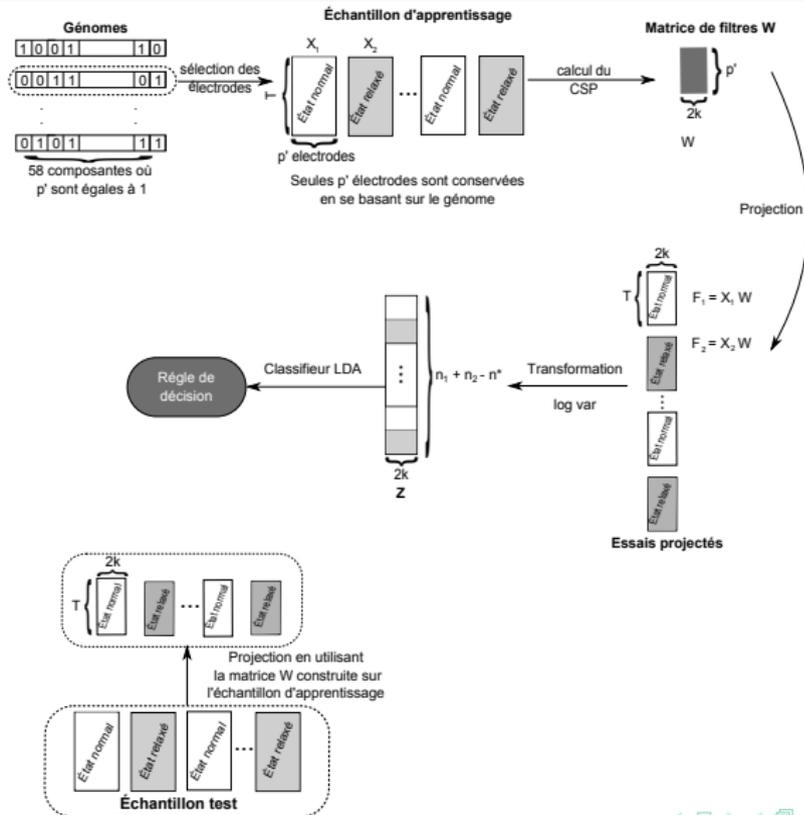


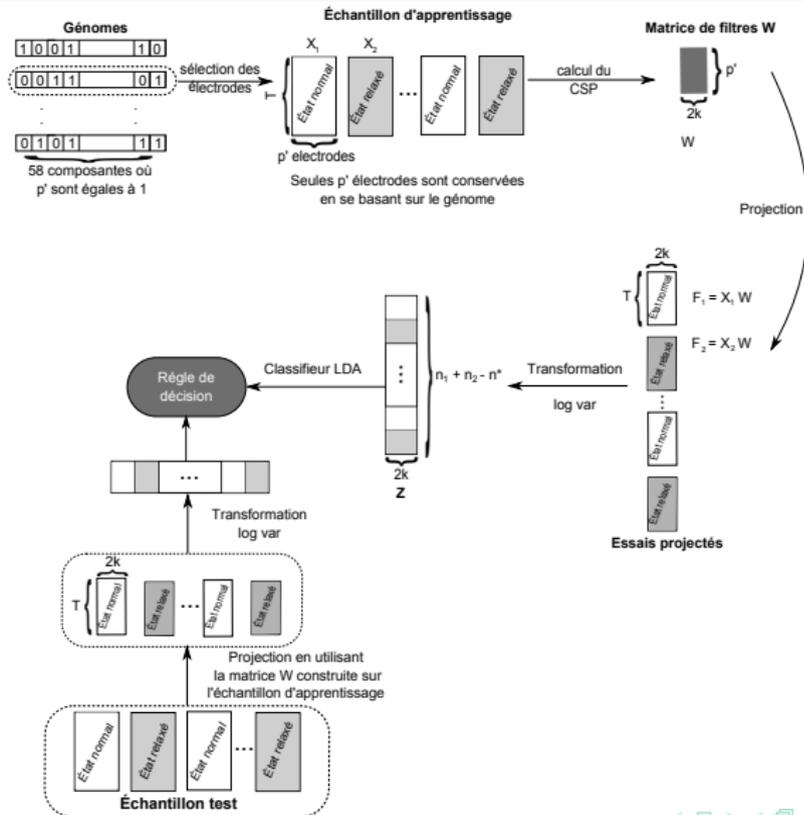


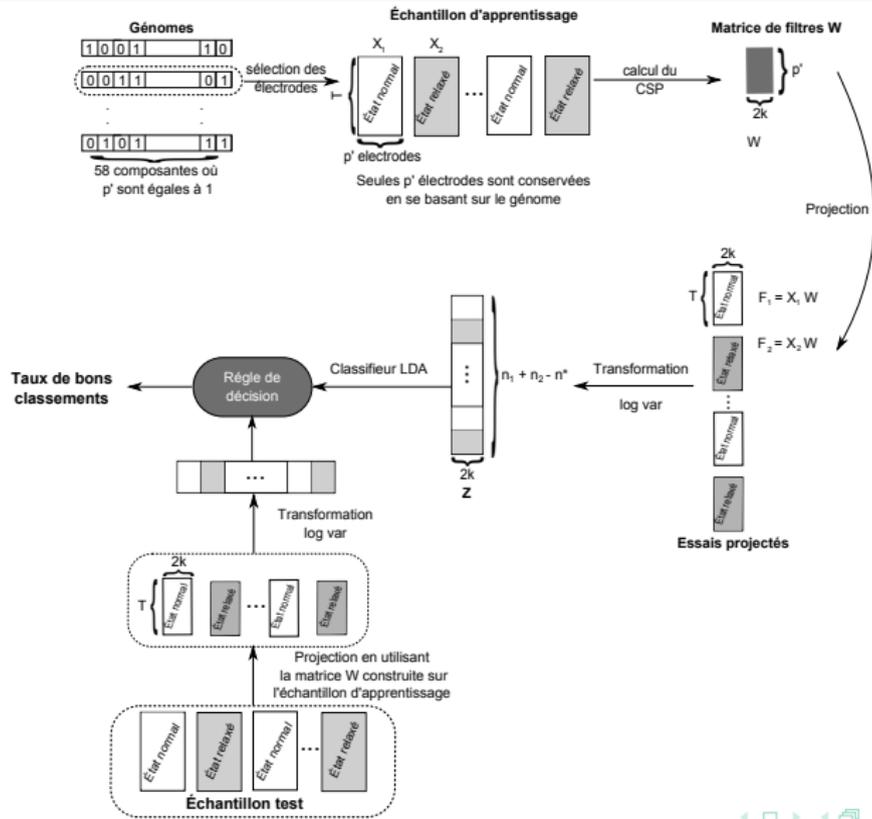




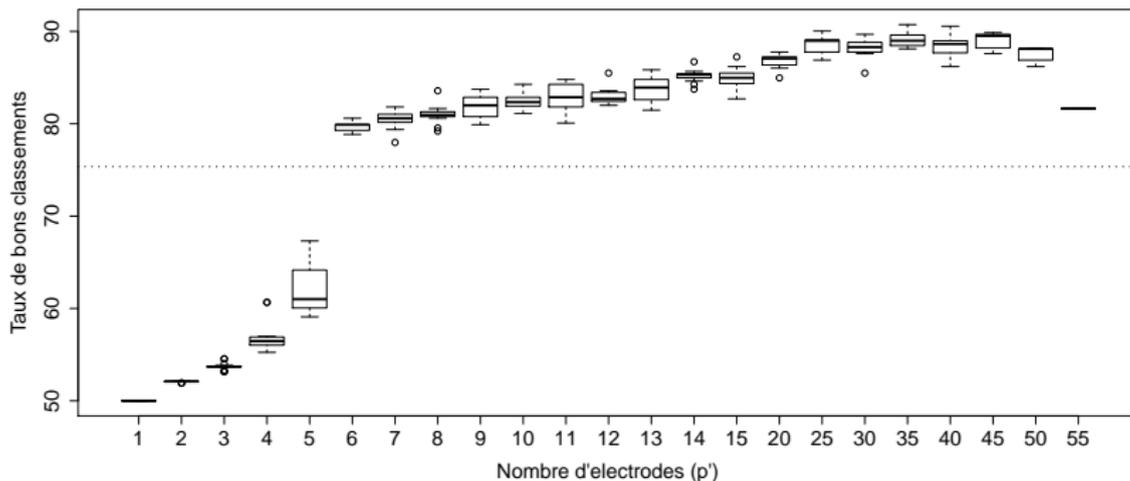




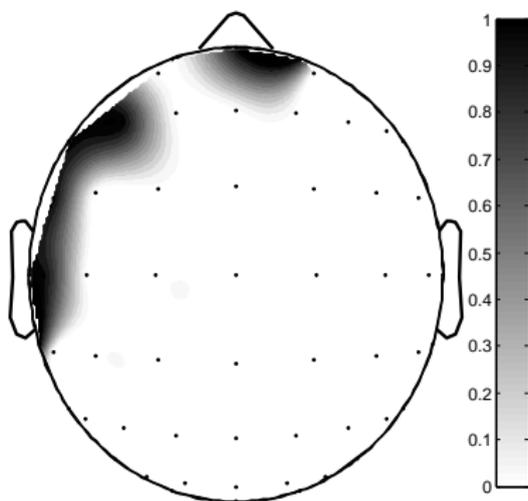




- 15 lancements de l'algorithme génétique pour chaque valeur de  $p'$  (où  $p' = 1, 2, \dots, 15, 20, 25, 30, 35, 40, 45, 50, 55$ ).
- Paramètres : 300 parents et 150 enfants.



À l'intérieur de l'algorithme génétique : estimation du TBC par LOO sur les sujets de la première campagne ⇒ APPRENTISSAGE			
$p'$	Moyenne du nombre d'itérations	Moyenne des TBC des meilleurs génomés	Écart types des TBC des meilleurs génomés
1	12	50 (50)	0 (0)
2	55.93	52.06 (53.85)	0.07 (0)
3	71.07	53.73 (53.85)	0.41 (0)
4	78.6	56.85 (58.72)	1.66 (0.023)
5	87.07	62.05 (65.13)	2.68 (0.057)
6	111.6	79.72 (86.41)	0.59 (0.038)
7	108.2	80.43 (86.67)	0.96 (0.025)
8	130.53	81 (86.41)	0.97 (0.035)
9	121.2	81.89 (90.26)	1.28 (0.029)
10	141.47	82.47 (89.74)	0.93 (0.028)
11	137.87	82.84 (91.03)	1.59 (0.035)
12	141.67	82.96 (90.77)	0.87 (0.032)
13	146	83.8 (90.51)	1.36 (0.032)
14	169.07	85.19 (92.05)	0.68 (0.027)
15	163.67	84.93 (92.56)	1.19 (0.031)
20	165.73	86.75 (94.1)	0.72 (0.025)
25	185.53	88.5 (95.13)	0.95 (0.018)
30	168.47	88.22 (95.13)	0.99 (0.027)
35	180.6	89.16 (94.87)	0.8 (0.024)
40	168	88.41 (96.67)	1.29 (0.029)
45	181.13	89.06 (96.15)	0.8 (0.025)
50	145.8	87.49 (92.56)	0.81 (0.01)
55	72.53	81.64 (80.77)	0 (0)
58	-	75.35 (73.08)	-



Taux de bons classements de 71.59% (75%) sur les essais de l'échantillon **TEST**.

[Retour](#)

# Démonstration

## Preuve

On recherche le vecteur  $\mathbf{w} \in \mathbb{R}^p$  tel que  $\max_{\mathbf{w} \in \mathbb{R}^p} f(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}}$ . Si  $\left. \frac{\partial f}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_1} = 0$ , alors  $\mathbf{w}_1$  est un optimum local de  $f$ . On a :

$$\begin{aligned} \left. \frac{\partial f}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_1} = 0 &\Leftrightarrow \frac{\left( \frac{\partial}{\partial \mathbf{w}_1} (\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1) \right) (\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1) - (\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1) \left( \frac{\partial}{\partial \mathbf{w}_1} (\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1) \right)}{(\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1)^2} = 0 \\ &\Leftrightarrow 2\mathbf{C}_1 \mathbf{w}_1 (\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1) - (\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1) 2\mathbf{C}_2 \mathbf{w}_1 = 0 \\ &\Leftrightarrow \mathbf{C}_1 \mathbf{w}_1 = \frac{\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1}{\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1} \mathbf{C}_2 \mathbf{w}_1 \\ &\Leftrightarrow \mathbf{C}_2^{-1} \mathbf{C}_1 \mathbf{w}_1 = \frac{\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1}{\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1} \mathbf{w}_1 \end{aligned}$$

Retour

# L'algorithme itératif de Zou, Hastie et Tibshirani

Pour un  $\alpha$  fixé

On se ramène donc à  $k$  problèmes de régressions LASSO où l'on cherche, pour  $j = 1, \dots, k$ ,

$$\min_{\beta_j} \|\mathbf{f}_j - X\beta_j\|_2^2 + \lambda^j \|\beta_j\|_1$$

Chaque colonne  $\beta_j$  de  $\beta$  est obtenue par régression LASSO de  $\mathbf{f}_j = X\alpha_j$  sur  $X$ .

Démonstration

Pour un  $\beta$  fixé

$$\hat{\alpha} = \arg \max_{\alpha} \text{Tr}(\alpha^T X^T X \beta)$$

s. c.  $\alpha^T \alpha = I_k$

Démonstration

D'après [Zou et al., 2006], il vient que  $\hat{\alpha} = UV^T$  où  $X^T X \beta = UDV^T$ .

# Démonstration

## Preuve

Dans un premier temps, on va prouver la relation 5 en partant du terme de gauche.

$$\begin{aligned}\|X - X\beta\alpha^T\|^2 &= \text{Tr} \left[ (X^T - \alpha\beta^T X^T)(X - X\beta\alpha^T) \right] \\ &= \text{Tr} \left[ X^T X - \alpha\beta^T X^T X - X^T X\beta\alpha^T + \alpha\beta^T X^T X\beta\alpha^T \right]\end{aligned}$$

Comme  $\text{Tr}(AB) = \text{Tr}(BA)$ , il vient que :

- $\text{Tr}(\alpha\beta^T X^T X) = \text{Tr}(\beta^T X^T X\alpha)$
- $\text{Tr}(X^T X\beta\alpha^T) = \text{Tr}(\alpha^T X^T X\beta)$
- $\text{Tr}(\alpha\beta^T X^T X\beta\alpha^T) = \text{Tr}(\beta^T X^T X\beta\alpha^T\alpha) = \text{Tr}(\beta^T X^T X\beta)$

De plus, comme  $\text{Tr}(A^T) = \text{Tr}(A)$ , il vient que  $\text{Tr}(\beta^T X^T X\alpha) = \text{Tr}(\alpha^T X^T X\beta)$ .  
Ainsi,  $\|X - X\beta\alpha^T\|^2 = \text{Tr}(X^T X) - 2\text{Tr}(\alpha^T X^T X\beta) + \text{Tr}(\beta^T X^T X\beta)$ , ce qui démontre bien la relation (5).

# Démonstration

## Preuve

Dans un second temps, on va prouver la relation 6 en partant de la relation 5.

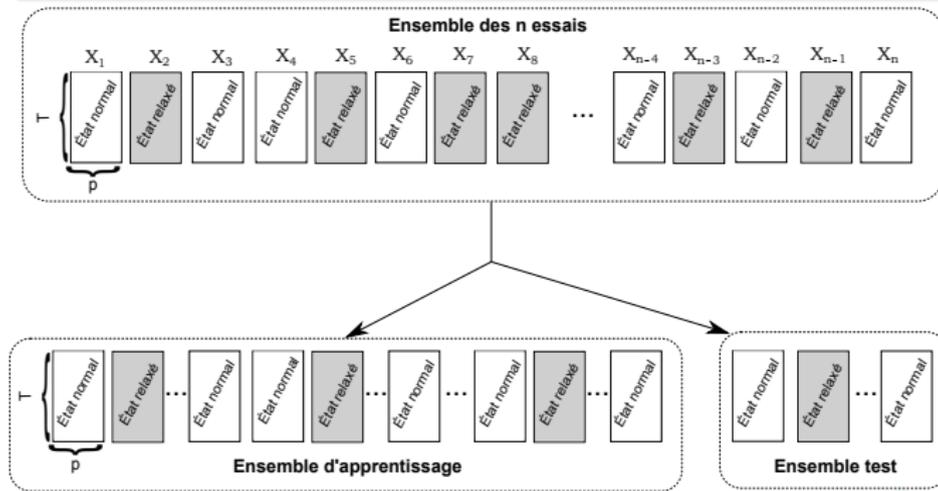
$$\text{Ainsi, } \text{Tr}(\alpha^T X^T X \beta) = \sum_{j=1}^k \alpha_j^T X^T X \beta_j \text{ et } \text{Tr}(\beta^T X^T X \beta) = \sum_{j=1}^k \beta_j^T X^T X \beta_j.$$

$$\text{D'où, } \|X - X\beta\alpha^T\|^2 = \text{Tr}(X^T X) - 2 \sum_{j=1}^k \alpha_j^T X^T X \beta_j + \sum_{j=1}^k \beta_j^T X^T X \beta_j$$

Retour

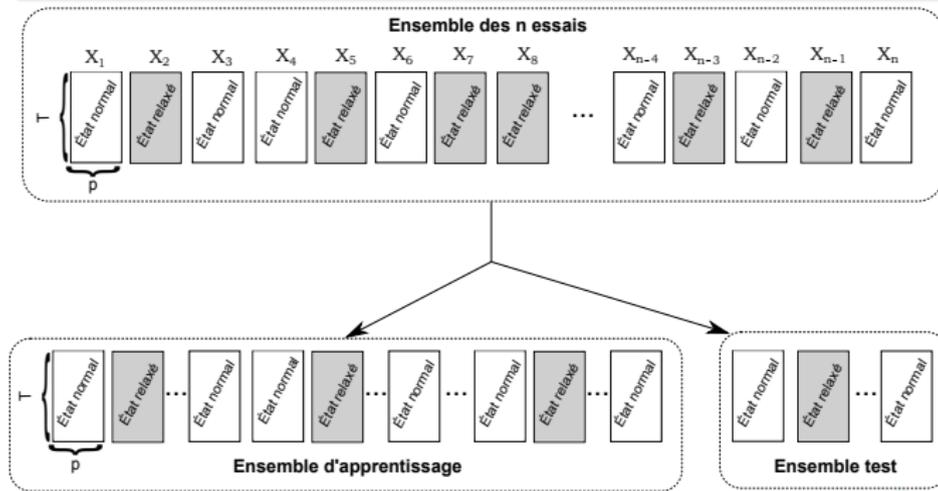
# Prédiction

- Découpage en échantillon d'apprentissage et échantillon test.
- La règle de décision (filtres CSP + classifier) est formée grâce aux essais de l'échantillon d'apprentissage.
- Taux de bons classements estimé sur l'échantillon test.



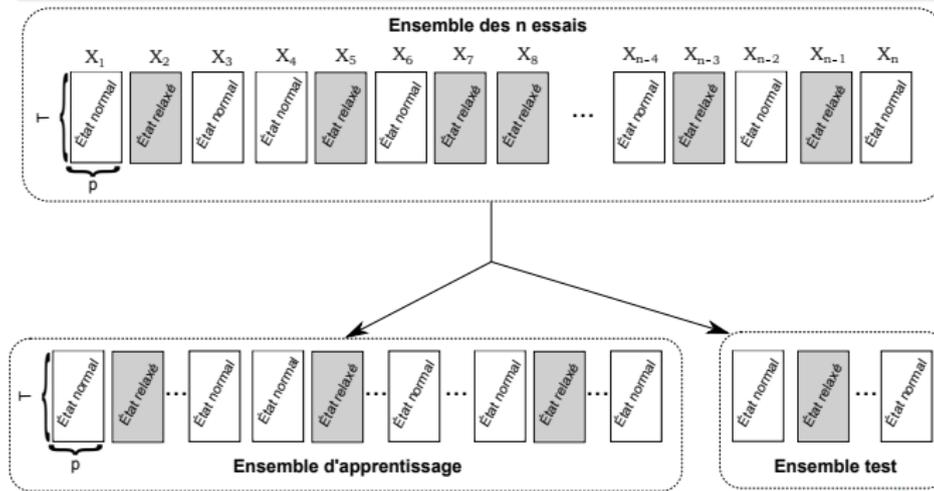
# Prédiction

- Découpage en échantillon d'apprentissage et échantillon test.
- La règle de décision (filtres CSP + classifier) est formée grâce aux essais de l'échantillon d'apprentissage.
- Taux de bons classements estimé sur l'échantillon test.

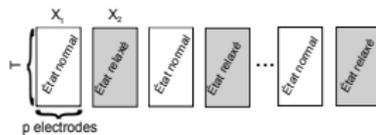


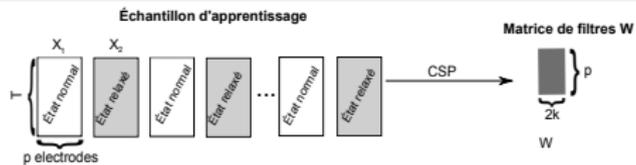
# Prédiction

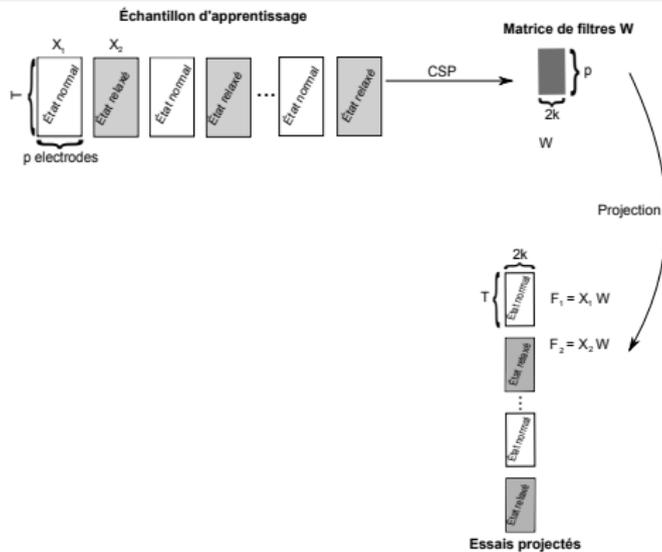
- Découpage en échantillon d'apprentissage et échantillon test.
- La règle de décision (filtres CSP + classifieur) est formée grâce aux essais de l'échantillon d'apprentissage.
- Taux de bons classements estimé sur l'échantillon test.

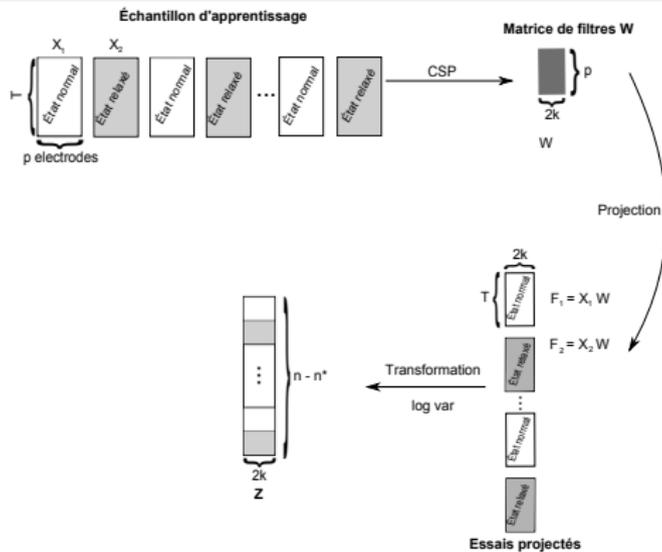


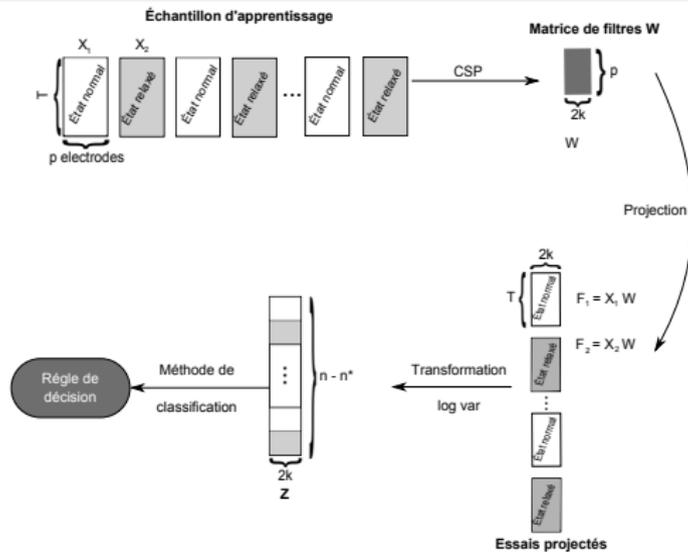
### Échantillon d'apprentissage

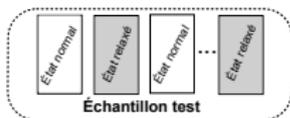
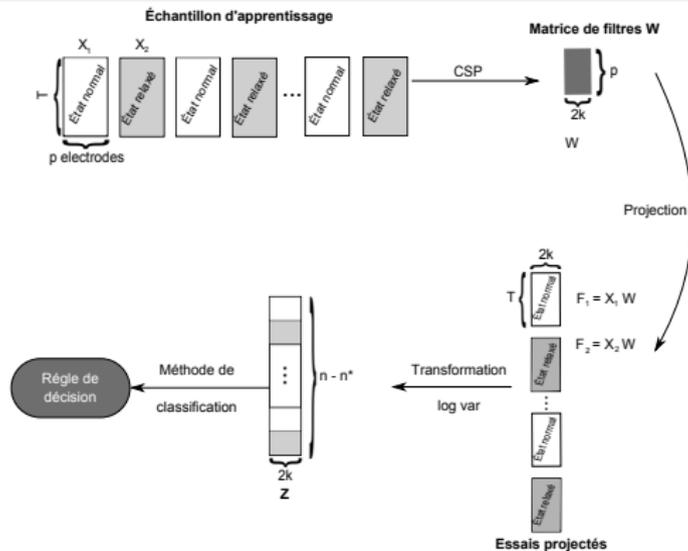


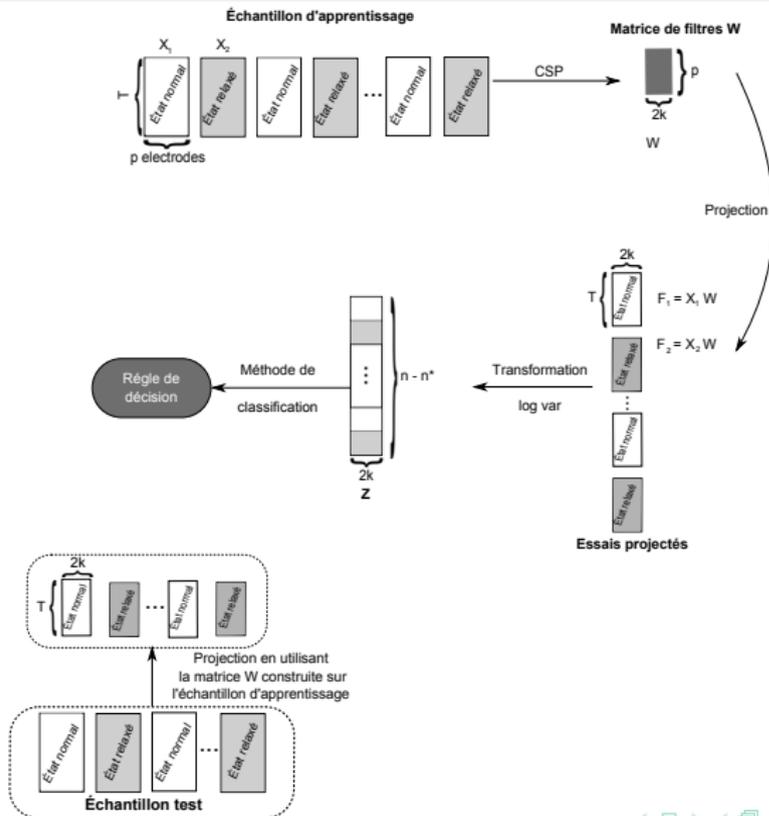




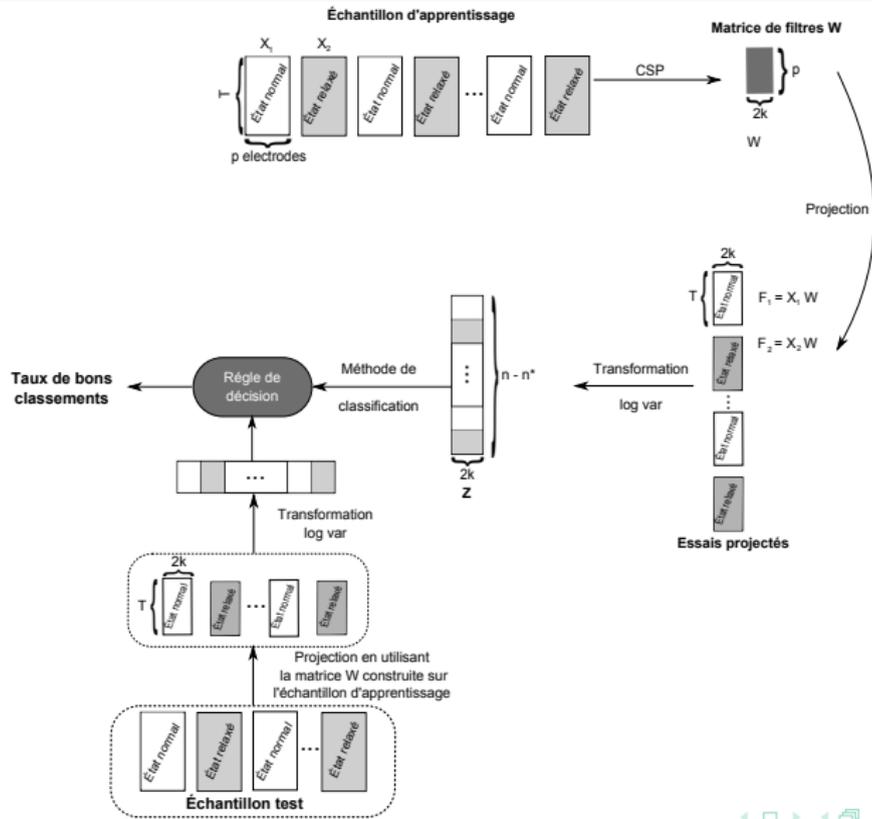












# Stratégie d'optimisation

## Proposition

Soit  $X$  une matrice de dimensions  $n \times p$  et soient  $\beta$  et  $\alpha$  deux matrices de dimensions  $p \times k$ . Alors,

$$\|X - X\beta\alpha^T\|^2 = \text{Tr}(X^T X) - 2\text{Tr}(\alpha^T X^T X \beta) + \text{Tr}(\beta^T X^T X \beta) \quad (5)$$

$$= \text{Tr}(X^T X) - 2 \sum_{j=1}^k \alpha_j^T X^T X \beta_j + \sum_{j=1}^k \beta_j^T X^T X \beta_j \quad (6)$$

Démonstration

## Pour un $\alpha$ fixé

$$\begin{aligned}\|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 &= \text{Tr}(X^T X) + \sum_{j=1}^k [\beta_j^T X^T X \beta_j - 2\mathbf{f}_j^T X \beta_j] + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \\ &= \text{Tr}(X^T X) + \sum_{j=1}^k [\|\mathbf{f}_j - X\beta_j\|_2^2 - \mathbf{f}_j^T \mathbf{f}_j] + \sum_{j=1}^k \lambda^j \|\beta_j\|_1\end{aligned}$$

d'où

$$\min_{\beta} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \Leftrightarrow \min_{\beta} \sum_{j=1}^k [\|\mathbf{f}_j - X\beta_j\|_2^2 + \lambda^j \|\beta_j\|_1]$$

On se ramène donc à  $k$  problèmes de régressions LASSO où l'on cherche, pour  $j = 1, \dots, k$ ,

$$\min_{\beta_j} \|\mathbf{f}_j - X\beta_j\|_2^2 + \lambda^j \|\beta_j\|_1$$

Chaque colonne  $\beta_j$  de  $\beta$  est obtenue par régression LASSO de  $\mathbf{f}_j = X\alpha_j$  sur  $X$ .

[Retour](#)

## Pour un $\beta$ fixé

On cherche  $\hat{\alpha}$  tel que :

$$\hat{\alpha} = \arg \min_{\alpha} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \quad (7)$$

s. c.  $\alpha^T \alpha = I_k$

D'après (5),  $\|X - X\beta\alpha^T\|^2 = \text{Tr}(X^T X) - 2\text{Tr}(\alpha^T X^T X \beta) + \text{Tr}(\beta^T X^T X \beta)$ . D'où (7) équivaut à :

$$\hat{\alpha} = \arg \max_{\alpha} \text{Tr}(\alpha^T X^T X \beta)$$

s. c.  $\alpha^T \alpha = I_k$

où

- $X\alpha$  sont les composantes principales orthogonales,
- $X\beta$  sont les composantes principales parcimonieuses.

D'après [Zou et al., 2006], il vient que  $\hat{\alpha} = UV^T$  où  $X^T X \beta = UDV^T$ .

Retour

**Entrées**  $\mathbf{X}_1$  et  $\mathbf{X}_2$  : matrices obtenues par concaténation des essais recueillis dans la condition 1 et 2,

$k$  : nombre de filtres CSP à construire,

$[\lambda^1, \dots, \lambda^k, \phi^1, \dots, \phi^k]$  : valeurs des paramètres de régularisation des filtres  $\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k$ .

**Sortie(s)** Obtention de la liste des  $p'$  électrodes sélectionnées et de la matrice  $\tilde{W}$  des filtres CSP obtenue avec ces  $p'$  électrodes.

**Initialisation**

**pour**  $c = 1, 2$  faire

$$\mathbf{C}_c = \frac{1}{T_c} \mathbf{X}_c^T \mathbf{X}_c$$

$$[U_c, \Sigma_c, V_c] = \text{SVD}(\mathbf{C}_c)$$

$$\mathbf{C}_c^{-\frac{1}{2}} = U_c \Sigma_c^{-\frac{1}{2}} V_c^T$$

**fin du pour**

$$\mathbf{Z}_1 = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}} \quad \text{et} \quad \mathbf{Z}_2 = \mathbf{X}_2 \mathbf{C}_1^{-\frac{1}{2}}$$

**Sélection de variables pour les filtres  $\mathbf{w}_1, \dots, \mathbf{w}_k$**

$$[\beta_1] = \text{sPCA}(\mathbf{Z}_1^T \mathbf{Z}_1, k, [\lambda^1, \dots, \lambda^k])$$

$\beta_1$  est la matrice de dimensions  $p \times k$  des loadings parcimonieux des  $k$  premières composantes de l'ACP de  $\mathbf{Z}_1$ .

**Sélection de variables pour les filtres  $\mathbf{v}_1, \dots, \mathbf{v}_k$**

$$[\beta_2] = \text{sPCA}(\mathbf{Z}_2^T \mathbf{Z}_2, k, [\phi^1, \dots, \phi^k])$$

$\beta_2$  est la matrice de dimensions  $p \times k$  des loadings parcimonieux des  $k$  premières composantes de l'ACP de  $\mathbf{Z}_2$ .

$$[\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \text{Liste\_Var}] = \text{Select\_Variables}(\beta_1, \beta_2, \mathbf{X}_1, \mathbf{X}_2)$$

Sélection des  $p'$  variables ayant des coefficients non nuls dans  $\beta_1$  ou  $\beta_2$ . Ainsi,  $\tilde{\mathbf{X}}_1$  et  $\tilde{\mathbf{X}}_2$  sont des matrices de dimensions  $T_1 \times p'$  et  $T_2 \times p'$ .

---

**Obtention des filtres CSP**

$$[\tilde{W}] = \text{CSP}(\tilde{X}_1, \tilde{X}_2, k)$$

Obtention de la matrice des filtres  $\tilde{W}$ , de dimensions  $p' \times 2k$  en utilisant un CSP uniquement sur les variables sélectionnées par CSP parcimonieux.

**Retour**  $\tilde{W}$ , Liste\_Var

---

Retour

---

**Entrées**  $X$  : matrice des données centrées,  
 $k$  : nombre de composantes d'ACP parcimonieuse à considérer.  
**Sortie(s)** Obtention de la grille matricielle  $\Lambda$  des paramètres de régularisation.

**Initialisation**

$$(\mathbf{y}_1, \dots, \mathbf{y}_k) = \text{ACP}(X, k)$$

Les  $k$  composantes principales de l'ACP classique de  $X$  sont obtenues.

**Calcul de la grille  $\Lambda$**

**pour**  $j = 1, \dots, k$  **faire**

$$\lambda_1^j, \dots, \lambda_{r_j}^j = \text{glmnet}(X, \mathbf{y}_j)$$

Obtention de la grille des paramètres du chemin de régularisation pour la régression LASSO de  $\mathbf{y}_j$  sur  $X$ .

**fin du pour**

$$r = \min_{j=1, \dots, k} r_j$$

$$\Lambda = (\lambda_i^j)_{r \times k}$$

**Retour**  $\Lambda$

---