Time-Frequency analysis for multi-channel and/or multi-trial signals

B. Torrésani

Aix-Marseille Univ.
Laboratoire d'Analyse, Topologie et Probabilités

ESI, December 2012

Outline

Motivation :

- Multi-sensor biosignals, such as EEG, MEG,... contain information that shows up differently in various channels, and may be difficult to extract from single channel.
- In this context, one often looks for features that are localized in some joint space-time-frequency domain.
- To detect weak signals, experiments are often repeated several times : multi-trial signals
- Problem : tackle inter-trial variability... which may sometimes be modelled as time-frequency jitter and amplitude variability...

Outline

Time-frequency analysis :

- Time-frequency transforms are inherently single channel techniques.
- Can be trivially extended to multi-channel signals, by individually transforming each channel ; multi-channel cooperation is enforced by post-processing.
- Synthesis-based frameworks allow one to enforce multi-channel information sharing already in the first stage.
- The multi-trial situation is much more complex... need of time-frequency registration techniques prior to trial averaging.

Notations : Gabor atoms

Modulated and translated copies of a reference window

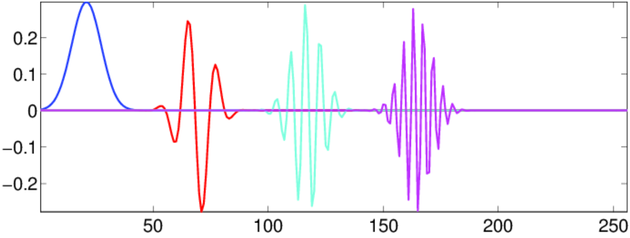
$$g_{kn}[t] = e^{2i\pi n\nu_0(t-kb_0)/L} g[t-kb_0]\ ,\quad k\in\mathbb{Z}_K\ ,\ n\in\mathbb{Z}_N$$

where ν_0 and b_0 are divisors of L, $K = L/b_0$ et $N = L/\nu_0$.
Given $f\in\mathbb{C}^L$, the family of coefficients

$$\mathcal{V}_g f[k,n] = \langle f, g_{kn}\rangle = \sum_{t=0}^{L-1} f[t]\overline{g}[t-kb_0]e^{-2i\pi n\nu_0(t-kb_0)/L}$$

form a **short time Fourier transform** (if $b_0=\nu_0=1$) or a **Gabor transform** of f.

Examples of Gabor atoms :



Notations : MDCT atoms

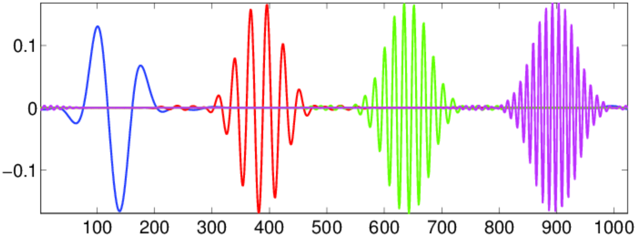
In $\mathbb{C}^L$, let $M \in \mathbb{Z}^+$ be a un divisor of L.

- $\mathbb{Z}_L$ is segmented into $K = L/N$ intervals of length N
- For all $k = 0, \dots K-1$, let $w_k \in \mathbb{C}^L$ be such that
  - $w_k[t] = 0$ for $t < (k - 1/2)N$ and $t > (k + 3/2)N$.
  - $w_k[kN + \tau] = w_{k+1}[kN - \tau]$ for all $\tau = 1 - N/2, \dots N/2 - 1$
  - $w_k[kN + \tau]^2 + w_{k+1}[kN + \tau]^2 = 1$ for all $\tau = 1 - N/2, \dots N/2 - 1$
- Denote by $u_{kn} \in \mathbb{C}^L$ the vectors defined by

$$u_{kn}[t] = \sqrt{\frac{2}{N}}\, w_k[t] \cos\left(\pi\left(n + \frac{1}{2}\right)(t - kN)\right)$$

- The collection $\{u_{kn}\}$ is an orthonormal basis of $\mathbb{C}^L$.

Examples of MDCT atoms :



Being a basis has a price : the time-frequency localization of MDCT atoms is more difficult to control.

Frames and bases

- Given a time-frequency basis $\Psi = \{\psi_{tf}\}$: the transform $x \in \mathbb{C}^L \rightarrow \{\langle x, \psi_{tf}\rangle\}$ is unitary. Any x has a unique expansion

$$x = \sum_{tf} \alpha_{tf} \psi_{tf} \ .$$

- Given a time-frequency frame $\Psi = \{\psi_{tf}\}$ (which is not a basis). Any x has infinitely many expansions of the form

$$x = \sum_{tf} \alpha_{tf} \psi_{tf} \ ,$$

finding the most relevant one requires extra information,... and is application dependent.

Outline

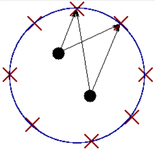
Multichannel signals :

$$\underline{x} = \{x^c, c = 1, \dots N_c\}$$

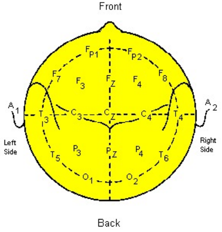
- signals from different channels are often dependent
- the dependence structure is often complex, and not necessarily known in advance

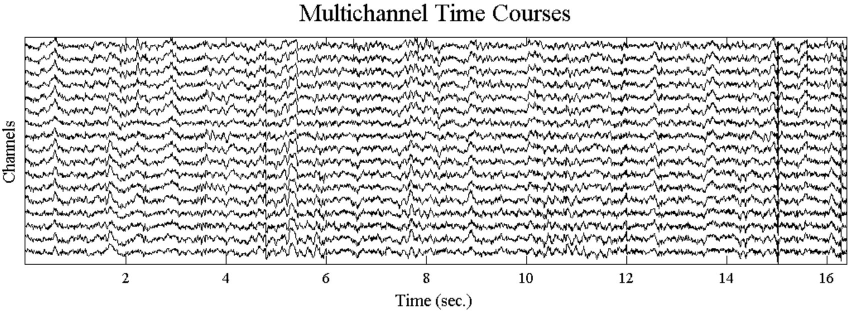
Example (Propagation from two sources)

- Signals, originating from two inner "sources", propagating to the boundary of some region where they are measured.
- Quasi-static approximation : time-locked signals



Example : EEG signals





Multichannel time-frequency expansions

In the framework of quasi-static type approximations (no time delay) : use the same time-frequency dictionary for all channels : $\Psi = \{\psi_{tf}\}$.

Transform + post-processing : example

- Compute time-frequency transform coefficients

$$\underline{\alpha} = \{\alpha_{tf}^c\}\ , \qquad \alpha_{tf}^c = \langle x^c, \psi_{tf}\rangle$$

- Describe the data cube $\underline{\alpha}$ via space-time-frequency modes, using factor decomposition (PARAFAC, Kruskal,...)

$$\underline{\alpha} = \sum_k C_k \otimes T_k \otimes F_k\ +\ \text{res.}\ , \qquad \alpha_{tf}^c = \sum_k C_k^c T_{kt} F_{kf}\ +\ \text{res.}$$

Outline

Synthesis models

estimate multichannel time-frequency expansions of the form

$$x^c = \sum_{t,f} \alpha_{tf}^c \psi_{tf} + \text{noise}$$

- Elementary models : channel, time and frequency are independent variables ; e.g. *bridge regression* type approaches : for $1 \leq p \leq 2$, solve

$$\min_{\underline{\alpha}} \left[\frac{1}{2} \sum_c \left\| x^c - \sum_{t,f} \alpha_{tf}^c \psi_{tf} \right\|^2 + \frac{\mu}{p} \|\underline{\alpha}\|_p^p \right]$$

- More complex models : introduce correlation structures, via either function space models (involving sophisticated mixed norms) or probabilistic models.

Gaussian and Gaussian mixture models

- Gaussian prior model :
$$p(\underline{\alpha}) \sim \mathcal{N}(0, \mathbf{\Sigma})$$
- Gaussian mixture prior model : for example
$$p(\underline{\alpha}) \sim \sum_{k=1}^{K} p_k \mathcal{N}(0, \mathbf{\Sigma}_k)$$
- Generalizations...

In most cases, $\mathbf{\Sigma}$ and/or $\mathbf{\Sigma}_k$ are large matrices : difficult to estimate... and to exploit.

Multi-channel/multi-trial time-frequency analysis

Gaussian mixture prior

$$\min \left[\frac{1}{2\sigma_0^2} \left\| \underline{x} - \sum_{t,f} \underline{\alpha}_{tf} \psi_{tf} \right\|^2 - \ln(p(\underline{\alpha})) \right]$$

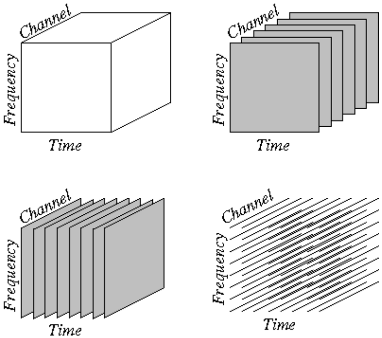
with p a Gaussian or Gaussian mixture prior.

- Gaussian prior : the (explicit) solution requires the inversion of a large matrix involving the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ and the Gram matrix of the frame.
- Gaussian mixture priors : MM numerical strategies require at each iteration the inversion of matrices of the same size.

Typical size : $N_c \approx 20$ channels, time-frequency blocks of dimension $MN \approx 1000$... yields matrices of size $\approx 20000 \times 20000$.

Extra structure has to be assumed for the covariance model : coefficient cube α with independent slices

Extra structure has to be assumed for the covariance model : coefficient cube α with independent slices



If the time-frequency frame and the covariance structure are compatible, corresponding estimation algorithms can be designed.

Definition (Translation invariant TF frame)

A time-frequency frame Ψ is invariant by (circular) translations if the columns of the corresponding matrix $\mathbf{\Psi}$ satisfy

$$\psi_\lambda[k] = \psi_{m,n}[k] = \psi_{0,n}[k-m]\ ,\ m = 0,\dots M-1,\ n = 0,\dots N-1\ .$$

The corresponding Gram matrix $\mathbf{\Psi}^*\mathbf{\Psi}$ is block circulant.

$$\mathbf{G} = \mathbf{\Psi}^*\mathbf{\Psi} = \begin{pmatrix} G_0 & G_1 & G_2 & \dots & G_{N-1} \\ G_{N-1} & G_0 & G_1 & \dots & G_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_1 & G_2 & \dots & G_{N-1} & G_0 \end{pmatrix},$$

Examples

- Gabor frames (time locked version)
- MDCT bases
- Translation invariant wavelet frames
- Arbitrary subband frames can be made translation invariant
- ...

Theorem (MM approach convergence)

Consider the Gaussian mixture prior model. Set

$$\mathbf{A} = \sum_k p_k \mathbf{\Sigma}_k^{-1}\ , \qquad C(\underline{\alpha}) = \ln p(\underline{\alpha}),$$

and let ε *be a positive integer.*

1. *The iteration* $\underline{\alpha}_n \longmapsto \underline{\alpha}_{n+1}$ *defined by*

$$\left[\frac{1}{\sigma_0^2}\mathbf{\Psi}^*\mathbf{\Psi} + 2(\mathbf{A}+\varepsilon\mathbf{I})\right]\underline{\alpha}_{n+1} = \left[\frac{1}{\sigma_0^2}\mathbf{\Psi}^*\underline{x} - \nabla C(\underline{\alpha}_n) - 2(\mathbf{A}+\varepsilon\mathbf{I})\underline{\alpha}_n\right]$$

   converges to a local minimum of the objective function.

2. *If* Ψ *is translation invariant, and the coefficient cube* $\underline{\alpha}$ *has independent fixed-time slices, the matrix* $\mathbf{M}$ *below is block circulant*

$$\mathbf{M} = \frac{1}{\sigma_0^2}\mathbf{\Psi}^*\mathbf{\Psi} + 2(\mathbf{A}+\varepsilon\mathbf{I})$$

Kronecker product : $\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}B & \dots & a_{1,N'_a}B \\ a_{2,1}B & \dots & a_{2,N'_a}B \\ \vdots & \ddots & \vdots \\ a_{N_a,1}B & \dots & a_{N_a,N'_a}B \end{pmatrix}$.

Proposition (De Mazancourt)

Block-circulant matrices $\mathbf{M}$ *can be diagonalized using the block Fourier transform* $\mathbf{F}$ *(Kronecker product of the standard Fourier transform and the identity), yielding* $\mathbf{M} = \mathbf{F}^*\mathbf{PF}$ *with* $\mathbf{P}$ *invertible block-diagonal.*

- Hence, the size of the matrices to be inverted is reduced.
- If further dimension reduction is needed : frequency-channel matrices can be seeked in the form of Kronecker products :

$$\mathbf{\Sigma}_{(cf)} = \mathbf{\Sigma}_{(c)} \otimes \mathbf{\Sigma}_{(f)} , \qquad \mathbf{\Sigma}_{(cf)}^{-1} = \mathbf{\Sigma}_{(c)}^{-1} \otimes \mathbf{\Sigma}_{(f)}^{-1} .$$

Related problem : estimation of the model parameters :

- Covariance matrices $\boldsymbol{\Sigma}_k$ (or Kronecker factors),
- Membership probabilities p_k.

Current solution : (ad hoc) re-estimation at each iteration of the algorithm. No convergence proof for the combined approach.

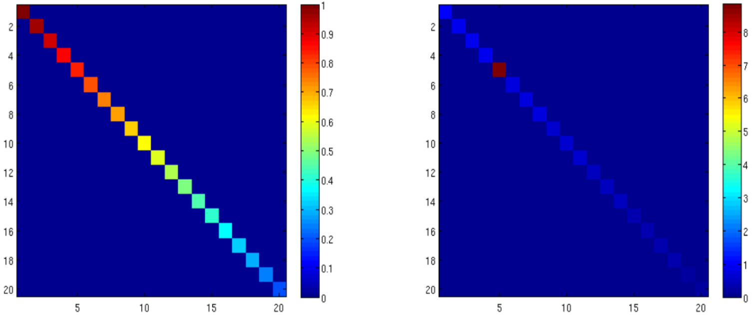
Related problem : estimation of the model parameters :

- Covariance matrices $\mathbf{\Sigma}_k$ (or Kronecker factors),
- Membership probabilities p_k.

Current solution : (ad hoc) re-estimation at each iteration of the algorithm. No convergence proof for the combined approach.

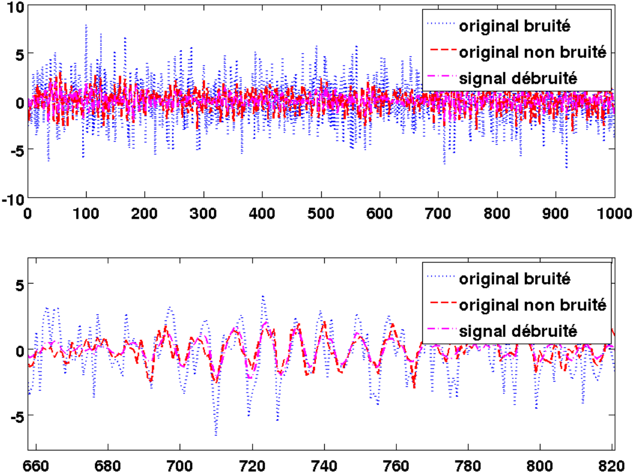
Numerical simulation : Preliminary : single sensor, Gaussian mixture ($N = 2$) with known covariance matrices.

Simulation



Frequency covariance matrices (state 2 : alpha waves)

Simulation



Original, noisy and reconstructed signals

Outline

So far : fixed-time coefficient vectors were assumed independent.

Multichannel harmonic hidden Markov model

- A hidden state $t \to X_t \in \{1, 2, \ldots N_s\}$ controls the distribution of corresponding coefficients α.
- Fixed time coefficients $\alpha_{t\cdot}^{\cdot}$ are modeled as before as a Gaussian random vector $\mathcal{N}(0, \mathbf{\Sigma}_s)$, whose covariance depends on the state X_s.
- Conditional to the hidden states, fixed time coefficient vectors $\alpha_{t\cdot}^{\cdot}$ are statistically independent.
- The dynamics of hidden states is governed by a Markov chain : transition $X_t = s$ to $X_{t+1} = s'$ with fixed probabilities.

Problems to solve

- Estimate the model parameters :
  - Covariance matrices : $\boldsymbol{\Sigma}_{fc}$ or $\boldsymbol{\Sigma}_f \otimes \boldsymbol{\Sigma}_c$
  - Characteristics of the chain : transition probabilities $\mathbb{P}\{X_{t+1} = s' | X_t = s\}$, initial probabilities $\mathbb{P}\{X_0 = s\}$.
- Estimate the hidden states sequences

Answers

- MDCT or Wilson basis : standard procedure
  - Computation of TF coefficients
  - Parameter estimation : Baum Welch algorithm (provable convergence even for Kronecker covariance matrices)
  - Hidden states estimation : Viterbi algorithm (low complexity)
- For Gabor frames : ad hoc procedures... not really satisfactory

Details : Forward and backward variables

$$a_t^s = \mathbb{P}\left\{X_t = s | \underline{\alpha}_{0:t}\right\} \times L_t$$

with L_t the likelihood of the observations until time t,

$$b_t^s = \mathbb{P}\left\{\underline{y}_{(t+1):(N_t-1)} | X_t = s\right\} .$$

are computed recursively using the forward-backward equations.

$$a_{t+1}^s = f_s(\underline{\alpha}_{t+1}) \sum_{s'=1}^{N_s} \pi_{s's} a_t^{s'} , \qquad b_t^s = \sum_{s'=1}^{N_s} \pi_{ss'} f_{s'}(\underline{\alpha}_{t+1}) b_{t+1}^{s'} .$$

with π the transition matrix of the chain, and f_s the pdf of fixed-time coefficient vectors $\underline{\alpha}_t$ in state s.

Details : Initial probabilities and transition matrix re-estimation

$$\widehat{\nu_s} = \frac{a_0^s b_0^s}{\mathscr{L}}$$

$$\widehat{\pi_{s,s'}} = \pi_{s,s'} \frac{\frac{1}{\mathscr{L}} \sum_{t=0}^{N_t-2} a_t^s b_{t+1}^{s'} f_{s'}(\underline{\alpha}_{t+1})}{\frac{1}{\mathscr{L}} \sum_{t=0}^{N_t-2} a_t^s b_t^s} ,$$

with

$$\mathscr{L} = L_{N_t-1} = \sum_{s=1}^{N_s} a_t^s b_t^s$$

Estimation of $\Sigma_s^{(c)}$ given $\Sigma_s^{(f)}$: define $M_t^s(c,c') = \langle (\boldsymbol{\Sigma}_s^{(f)})^{-1}\underline{\alpha}_t^c\,, \underline{\alpha}_t^{c'}\rangle$ and set

$$\widetilde{\boldsymbol{\Sigma}_s^{(c)}} = \frac{1}{N_f}\frac{\sum_{t=0}^{N_t-1}\mathbb{P}\{X_t=s\}M_t^s}{\sum_{t=0}^{N_t-1}\mathbb{P}\{X_t=s\}}$$

Normalization : set

$$\widehat{\boldsymbol{\Sigma}_s^{(c)}} = \widetilde{\boldsymbol{\Sigma}_s^{(c)}}\Big/\left\|\widetilde{\Sigma_s^{(c)}}\right\|_F ,$$

Estimation of $\boldsymbol{\Sigma}_s^{(f)}$ given $\widehat{\boldsymbol{\Sigma}_s^{(c)}}$: define $P_t^s(f,f') = \langle (\boldsymbol{\Sigma}^{(c)})^{-1}\underline{\alpha}_{tf}, \underline{\alpha}_{tf'}\rangle$ and set

$$\widehat{\boldsymbol{\Sigma}_s^{(f)}} = \frac{1}{N_c}\frac{\sum_{t=0}^{N_t-1}\mathbb{P}\{X_t=s\}P_t^s}{\sum_{t=0}^{N_t-1}\mathbb{P}\{X_t=s\}}$$

Application to rest EEG

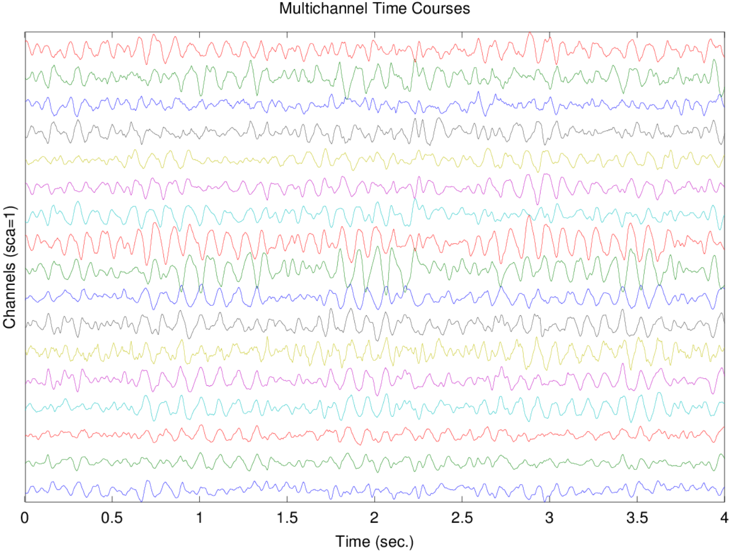
Rest EEG basically features Alpha waves : short duration time-localized oscillations (frequencies around 10 Hz) which appear in specific situations ; topographically localized in specific sensors located in posterior regions of the head.

Alpha wave occurrence may be considered a departure from a stationary background signal. This motivates the use of hidden Markov models as described above.

Remark (Time-frequency resolution)

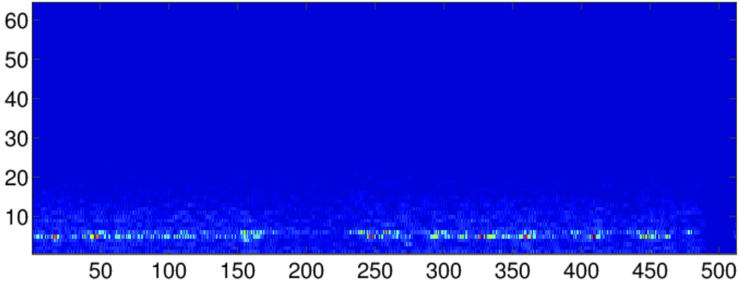
alpha waves are actually close to the Heisenberg limit. One needs frequency resolution of approximately 4Hz, and time resolution of approximately 250 msec....

Multi-channel/multi-trial time-frequency analysis



Alpha waves in rest EEG

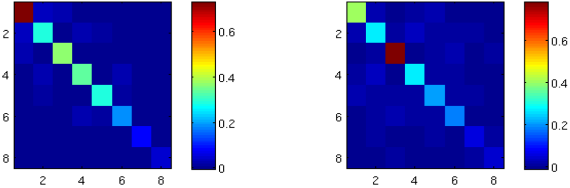
Application to rest EEG : real data



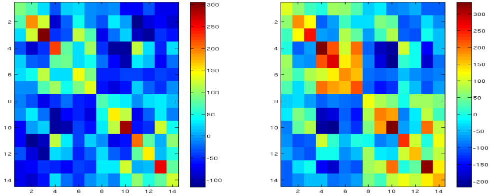
MDCT coefficients of a 30 sec. long EEG recording (rest EEG)

Aix*Marseille université

Application to rest EEG

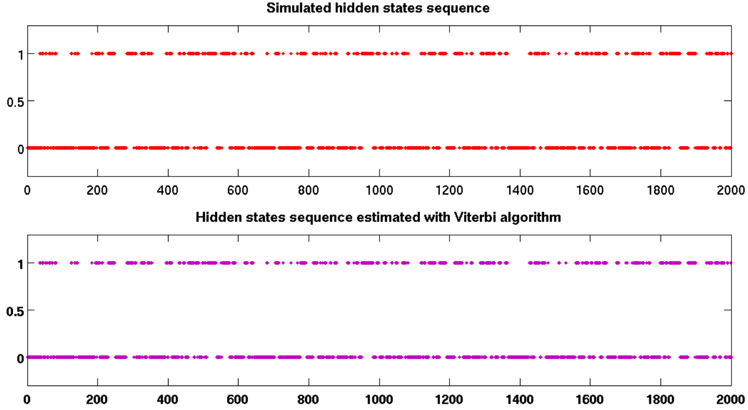


Frequency covariance matrices estimates for the two classes



Channel covariance matrices estimates for the two classes

Hidden states estimation : simulated data



Outline

Multiple sclerosis

Multiple sclerosis has been reported to affect the left-right synchronization in the alpha band. This assumption can be tested using the model.

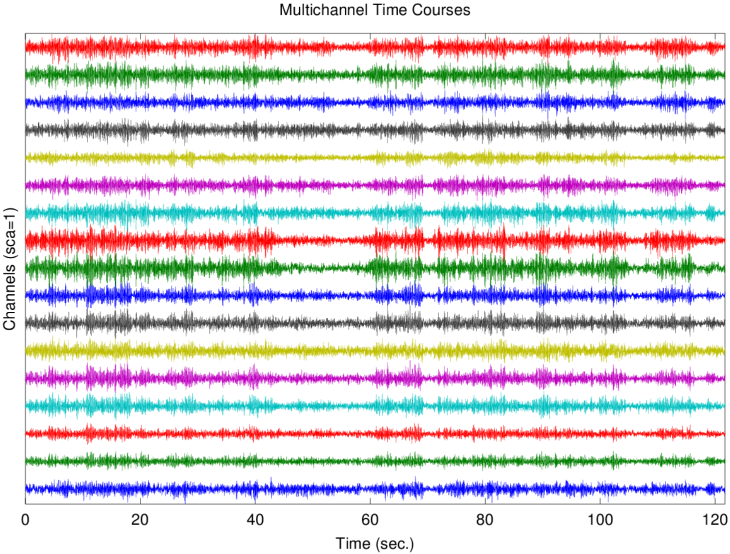
Dataset

EEG data originating from the CODYSEP dataset, designed to study the impact of multiple sclerosis in inter-hemispherical transfer.

The dataset consists in 31 patients and 20 controls ; 17 channels EEG signals were collected at a 256 Hz sampling rate.

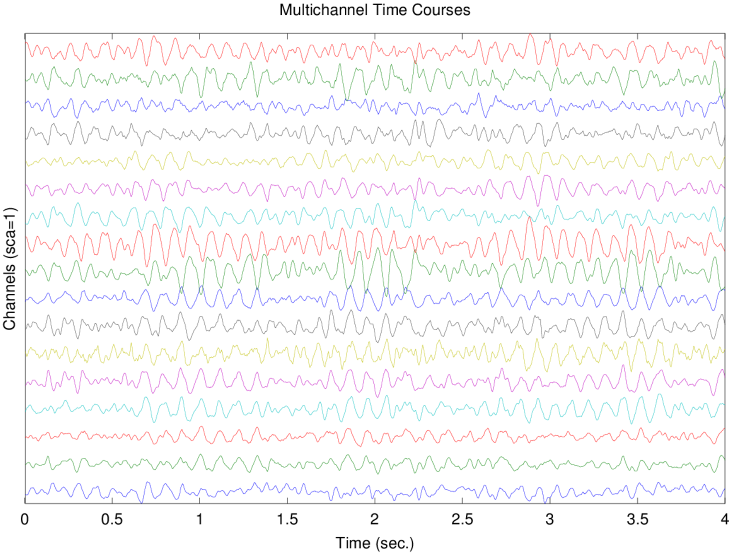
EEG data essentially contain alpha waves bursts.

Multi-channel/multi-trial time-frequency analysis



2 minutes of recording

Multi-channel/multi-trial time-frequency analysis



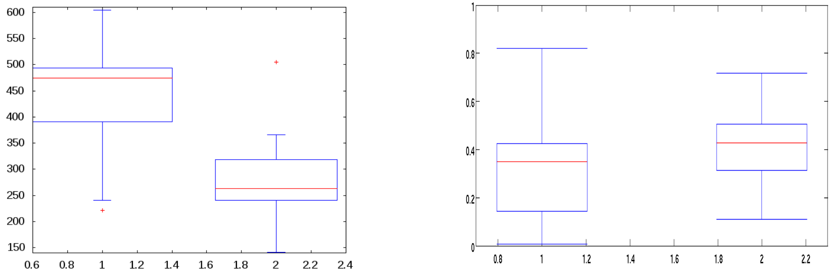
4 seconds of recording

Testing protocole :

1. For both classes (patient and control)
    - Select relevant left and right subsets of the set of sensors
    - For each subset :
        - Estimate corresponding model parameters
        - Estimate left and right hidden states sequence $X^{(L)}$ and $X^{(R)}$
    - Compute the Hamming distance between left and right hidden states sequences : $d_H = \|X^{(L)} - X^{(R)}\|_1$.
2. Compare estimated Hamming distances of controls and patients : boxplots, p-values,...
3. Compare with the results obtained using inter-coherence : left-right cross-correlation after band pass filtering.

Aix*Marseille université

Results



Left : boxplots of Hamming distances d_H between hidden states
Right : boxplots of inter-coherences d_C between signals

Mann-Whitney test : P-value$\approx$ 0.0384 : confirms quantitatively the hypothesis of two distinct distributions.

Faculté des Sciences Aix-Marseille Université

Conclusions

- When going multi-channel, one has to fight the curse of dimensionality.
- Factorized models can help in this respect
- Two approaches were presented, tackling two different problems. Next question : how to keep the best of the two ?
- Multi-trial : matching pursuit type approach (Consensus matching pursuit)

Thanks

- Joint works with E. Villaron and S. Anthoine
- Pleasant collaborations and discussions with the LATP signal processing group
- NuHaG and partners for organizing this nice event
- ...
- The audience for your attention !