



HAL
open science

Detection and analysis of oscillatory patterns in multisensor data by time-frequency hidden Markov models: an application to EEG

Emilie Villaron, Bruno Torr sani

► To cite this version:

Emilie Villaron, Bruno Torr sani. Detection and analysis of oscillatory patterns in multisensor data by time-frequency hidden Markov models: an application to EEG. 2012. hal-01737437

HAL Id: hal-01737437

<https://hal.science/hal-01737437>

Preprint submitted on 19 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.

Detection and analysis of oscillatory patterns in multisensor data by time-frequency hidden Markov models: an application to EEG

Emilie Villaron, Bruno Torr sani

LATP, Aix-Marseille Univ./CNRS, UMR 7353,
CMI, 39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
email: Emilie.Villaron@cmi.univ-mrs.fr Bruno.Torresani@univ-amu.fr
web: www.latp.univ-mrs.fr

Abstract

A model combining time-frequency representation, hidden Markov models and multichannel signal processing is developed, for the characterization of locally oscillatory phenomena in multisensor signals, such as in particular EEG and MEG signals). The model is based upon a multichannel MDCT transform of the signal; the multichannel MDCT coefficients are modelled by a hidden Markov chain, that is common to all sensors. The coefficients are distributed according to a multivariate Gaussian distribution (that depends on the hidden state), with a Kronecker-structured covariance matrix.

Corresponding estimation algorithms are described and demonstrated, on both simulated and real data. The relevance of the proposed approach is demonstrated on a case study on rest EEG signals, for which it is shown that the model parameters can discriminate between control subjects and patients with multiple sclerosis.

Keywords: EEG signals, time-frequency representation, MDCT basis, hidden Markov model, Kronecker product

1. Introduction

This paper is concerned with a new approach for the description, modeling and detection of oscillatory patterns in multi-trial, multichannel signals. Prototypes of such signals are electro-physiological signals (such as EEG, MEG...), which often feature components that are localized simultaneously in time, frequency and spatial domains. In other words, they often involve short oscillatory rhythms, that are most visible on specific sensors or scalp areas.

The joint localization in time and frequency domains is often described in terms of time-frequency or time-scale representations (such as short time Fourier or wavelet transforms). The latter are generally used as alternative representations of the signals, from which relevant features are extracted, and validated using statistical techniques. The *synthesis* aspect of time-frequency/scale transforms, i.e. the fact that signals can be synthesized from suitably chosen time-frequency or time-scale coefficients is more rarely exploited. The approach we develop here is somewhat different, in the sense that we introduce an explicit modeling of the signals under consideration, that involves time-frequency-space localization, together with a latent (hidden) time dependent variable, aiming at describing some underlying (time varying) state that generate different signal shapes. The model is a probabilistic model, which allows us to use the machinery of maximum likelihood estimation techniques.

More precisely, the model we propose is written directly in the time-frequency domain. Starting from a multichannel signal, the coefficients of a corresponding time-frequency-channel transform are modelled as multivariate Gaussian vectors, whose

characteristics (means, covariance matrices) depend upon some unobserved latent state (in EEG/MEG signals, the latent state aims at accounting for some underlying cerebral state). The latter state is time-dependent, and its time evolution is governed by a Markov chain, characterized by (unknown) transition probabilities (i.e. probabilities for jumping from a state to another at a given time). Such model architectures are known under the name of HMM (Hidden Markov Models), and methods and algorithms for solving the corresponding estimation problems have been thoroughly studied in various application domains. The contribution of the current paper is to extend such models and estimation algorithms to specific multichannel time-frequency signal representations. The corresponding covariance matrices involve frequency and channel degrees of freedom, which generally result in high dimensional problems. To cope with the curse of dimensionality, we also introduce covariance matrices of a specific type, namely Kronecker products of frequency and channel covariance matrices. This requires an adaptation of the estimation methods and algorithms.

This paper is organized as follows. Mathematical tools (time-frequency, Markov models, algorithms) are described in Section 2, where the main aspects of the estimation problem and corresponding algorithms are explained. Results of numerical simulations are also given there. Numerical results on real data are given in Section 3, where we describe in some details a case study on rest EEG signals originating from both control subjects and multiple sclerosis patients. We show in that section that the proposed approach is able first to detect automatically alpha rhythms bursts and estimate their statistical characteristics,

and second, to discriminate between patients and controls, using a measure of desynchronization between left and right sensors. More technical aspects are given in the appendix.

Preliminary results of this work have been presented at the Eusipco'10 conference [14].

2. Mathematical tools and methods

Throughout this paper, we shall be concerned with finite-dimensional real vectors. We denote by $u, v \in \mathbb{R}^N \rightarrow \langle u, v \rangle = u^T v$ the inner product of two (column) vectors (' T ' denoting matrix transposition). The Frobenius norm $\|M\|_F$ of a matrix is defined as the square root of the sum of its squared elements. Given a vector v labelled by a time index t , we shall make use of the notation $v_{t_1:t_2}$ to denote the sequence $\{v_{t_1}, \dots, v_{t_2}\}$.

Multichannel signals will be denoted generically using underlined symbols, for example $\underline{x} = \{x^c, c = 1, \dots, N_c\}$. We shall use the letters c, t, f to label respectively channels, time and frequency indices. When several realizations (trials) of the same signal are available, we shall label them using the letter r .

We shall also use the symbols \mathbb{P} and \mathbb{E} for probabilities and expectations, and denote by $\mathbb{P}\{X|Y\}$ conditional probabilities.

2.1. Time-frequency transforms

Time-frequency and time-scale transforms (see for example [3, 7, 10]) have been used for quite a long time for analyzing signals featuring non-stationary oscillatory behaviors. Such transforms also belong to the standard biosignal analysis toolkits (such as for example the EEGLAB toolbox [5]). For tasks such as signal analysis, pattern detection/recognition, parameter estimation, redundant transforms such as continuous wavelet transform or short time Fourier transform are often preferred, since they yield (at least in principle) better resolution. Among the alternatives, let us also mention various versions of the sparse regression techniques such as matching pursuit and variants (multichannel, multi-trial) or multichannel basis pursuit denoising [9].

We shall rather concentrate here on discretized versions of time-frequency representations, namely MDCT transforms, which are very popular in the domain of signal coding (in particular audio coding). MDCT is actually an expansion with respect to an orthonormal basis of modulated windows, which can be implemented through fast algorithms. The time-frequency resolution that can be achieved using MDCT is weaker than the time-frequency resolution of redundant transforms (because of a coarser sampling of the time-frequency plane, and of limitations on the analysis window). Nevertheless, the MDCT transform of a signal is in one to one correspondence with the signal, which will be important in our approach.

For the sake of simplicity, we limit the current discussion to the practically realistic case of discrete, finite duration signals, of length N . It has been shown (see [16, 10] for a review) that for suitable choices of a window function $w \in \mathbb{R}^N$, the family of waveforms $\mathcal{B} = \{u_{tf}, t = 0, \dots, N_t - 1, f = 0, \dots, N_f - 1\}$ defined by

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

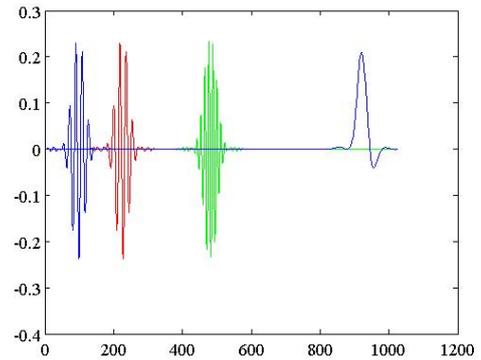


Figure 1: A few examples of MDCT atoms, with various locations and frequencies.

is an orthonormal basis of \mathbb{R}^N . Here $t = 0, \dots, N_t - 1$ (resp. $f = 0, \dots, N_f - 1$) denotes time (resp. frequency) index, and N_t and N_f are such that $N_t N_f = N$. Therefore, any signal $x \in \mathbb{R}^N$ is characterized by its MDCT coefficients

$$y[t, f] = \langle x, u_{tf} \rangle = \sum_{n=0}^{N-1} x[n] u_{tf}[n],$$

through the inversion formula

$$x = \sum_{t=0}^{N_t-1} \sum_{f=0}^{N_f-1} y[t, f] u_{tf}.$$

Examples of MDCT atoms with various values of time and frequency indices are displayed in Figure 1

We are actually interested in vector-valued (i.e. multichannel) discrete signals $\underline{x} = \{x^c, c = 0, \dots, N_c - 1\}$ of fixed duration N , which we shall represent by a family of multichannel MDCT coefficients $\underline{y} = \{y^c, c = 0, \dots, N_c - 1\}$

$$\underline{x} = \sum_{t=0}^{N_t-1} \sum_{f=0}^{N_f-1} \underline{y}[t, f] u_{tf}, \quad y^c[t, f] = \langle x^c, u_{tf} \rangle. \quad (2)$$

In other words, the same basis is used for all signal channels. The coefficients are therefore organized in the form of a three-way array. In the model described below, this three-way array is reshaped as a matrix with $N_f \times N_c$ rows and N_t columns. A signal model of the form (2) will be called *harmonic signal model*. A random version in which the coefficients are modeled by a hidden Markov chain (HMC) was developed in [11], under the name *harmonic hidden Markov model* (HHMM). In the multichannel case considered here, we shall talk about *multichannel harmonic hidden Markov model* (MHHMM).

2.2. Multichannel harmonic hidden Markov model

In order to model the piecewise stationarity and local oscillatory properties of the signals, we now introduce a probabilistic model for the MDCT coefficients in the expansion (2). We denote by $\underline{y}_t = \{y^c[t, f], f = 0, \dots, N_f - 1, c = 1, \dots, N_c\}$ the vector

of fixed time MDCT coefficients (i.e. a column of the reshaped \underline{y} matrix). The model assumes that the vectors \underline{y}_t are dependent random vectors, whose distribution is governed by a latent (hidden) state $X_t \in \{1, \dots, N_s\}$, which models the subject's cerebral state at a given time. The dynamics of cerebral states is itself modeled by a Markov chain (see [12] for a review).

More precisely, we assume N_s possible cerebral states. The details of the model are as follows:

1. When the latent state X_t is set to state s , the vector \underline{y}_t is a multivariate Gaussian random vector with zero mean and covariance matrix Σ_s . In more mathematically precise terms, conditional to $\underline{X} = \{s_t, t = 0, \dots, N_t - 1\}$ the vectors \underline{y}_t are mutually independent Gaussian vectors, with density

$$\Phi_s(\underline{y}) = \frac{\exp\left\{-\frac{1}{2}\langle \underline{y}, \Sigma_s^{-1} \underline{y} \rangle\right\}}{\sqrt{(2\pi)^{N_c N_f} \det(\Sigma_s)}}, \quad \underline{y} \in \mathbb{R}^{N_c N_f}. \quad (3)$$

2. The latent states (which depend on the time index only, and are common to all channels) are random variables, that form a Markov chain, characterized by its transition matrix π and the vector of initial probabilities ν . More precisely, the transition matrix is the matrix of conditional probabilities

$$\pi_{ss'} = \mathbb{P}\{X_{t+1} = s' | X_t = s\}, \quad s, s' = 1, \dots, N_s,$$

and the initial probabilities are defined by

$$\nu_s = \mathbb{P}\{X_0 = s\}, \quad s = 1, \dots, N_s.$$

In order to control the curse of dimensionality when the number of channels N_c is large (for example, when channels coincide with sensors -from 32 to 64 sensors are usually used in EEG), we shall be particularly concerned with models in which covariance matrices take the form of Kronecker products of frequency and channel matrices.

$$\Sigma_s = \Sigma_s^{(c)} \otimes \Sigma_s^{(f)}, \quad s = 1, \dots, N_s, \quad (4)$$

in other words, denoting by Y_{tf}^c the random coefficients, and dropping the time index for the sake of simplicity,

$$\mathbb{E}\{Y_f^c Y_{f'}^{c'}\} = \Sigma_{cc'}^{(c)} \Sigma_{ff'}^{(f)}.$$

This is very much in the spirit of [1, 2], where Kronecker products and sums of Kronecker products were considered for time-scale covariance matrices. Our approach assumes (conditional) independance of fixed-time coefficient vectors, and does not assume stationarity, therefore correlations between fixed-frequencies coefficient vectors are introduced. Notice that such a Kronecker product form is not unique, i.e. given Σ_s , $\Sigma_s^{(c)}$ and $\Sigma_s^{(f)}$ are defined up to a multiplicative factor. This fact has to be taken care of in the estimation procedure, by enforcing a normalization constraint on one of the covariance matrices.

The classical hidden Markov model inference techniques have to be modified in order to account for such a particular structure. More details are given below.

2.3. Estimation and algorithms

We are interested in the following situations. We assume we are given N_r realizations (trials) $\underline{x}^{(r)}$, $r = 1, \dots, N_r$ of a MHHMM signal as above, each having its own sequence of hidden states $\underline{X}^{(r)}$. The problem at hand is the following:

- Estimate from the observations the model parameters: the channel and frequency covariance matrices Σ^c and Σ^f , as well as the transition matrix π and the vector of initial probabilities ν .
- Estimate the sequence of latent states $X_t^{(r)}$, $t = 1, \dots, N_t$ for each trial r .

The parameter estimation problem does not require the knowledge of the latent states, and can be solved using a single trial, or several trials. The hidden states estimation requires the knowledge of the model parameters (either from the estimation procedure, or from a prior training stage), and is to be performed for each trial independently.

We outline below the main stages of the estimation procedure, details are given in the appendix.

2.3.1. Parameter estimation

The parameter estimation problem is solved using fairly classical tools, that are adapted to the present situation. First, starting from the observed signals (trials) $\underline{x}^{(r)}$, $r = 1, \dots, N_r$, corresponding MDCT coefficients $y^{(r):c}[t, f] = \langle x^{(r):c}, u_{tf} \rangle$ are computed. This is done using the freely available matlab toolbox LTFAT [13].

We use the EM algorithm to perform the parameter estimation. EM is an efficient iterative procedure to compute the Maximum Likelihood (ML) parameter estimate in the presence of latent data. We outline here the main points of the algorithm. The so-called forward and backward variables are instrumental in the algorithm. They represent respectively the normalized distribution of the latent state X_t conditional to the observed coefficients $\underline{y}_{0:t}$:

$$\alpha_t^s = \mathbb{P}\{X_t = s | \underline{y}_{0:t}\} \times L_t \quad (5)$$

where L_t is the likelihood of the observations until time t , and the backward variables β_t^s are the likelihoods of the observations $\underline{y}_{(t+1):(N_t-1)}$ conditional to $X_t = s$ for $s = 1, \dots, N_s$ and $t = 0, \dots, N_t - 1$.

$$\beta_t^s = \mathbb{P}\{\underline{y}_{(t+1):(N_t-1)} | X_t = s\}. \quad (6)$$

Given the model parameters, the forward and backward variables can be computed recursively (see the Appendix), and yield the likelihood of the parameters, which in turn can be optimized.

More precisely, each iteration of EM involves two steps.

- In the E-step (Expectation), the log-likelihood of the observed data (for each trial) is estimated given the current estimate of the model parameters, using the so-called forward and backward equations described in the appendix.

- In the M-step, the global likelihood function (involving all trials) is maximized leading to the so-called Baum-Welch re-estimation formulas. This yields new estimates for the parameters of the multivariate Gaussian distributions (i.e. the covariance matrices), and the parameters of the Markov chain (transition and initial probabilities).

Convergence is ensured since it may be shown that the algorithm is guaranteed to increase the likelihood at each iteration.

2.3.2. Estimation of the hidden states sequence

Given the estimated MAP parameters, we use the Viterbi dynamic programming algorithm to find the most likely sequence of hidden states

$$X_{0:(N_t-1)}^{MAP} = \arg \max_{\ell_{0:(N_t-1)}} \mathbb{P}\{X_{0:(N_t-1)} = \ell_{0:(N_t-1)} \mid \underline{y}_{0:(N_t-1)}\}. \quad (7)$$

This task is classically achieved using dynamic programming techniques (see [12] for a detailed account). This procedure is also linear in t but requires numerical evaluation of probability densities, i.e. inversion of covariance matrices. Therefore, when N_f and N_c are large, using tensor product decompositions yields more stability in the numerical evaluation.

2.4. Numerical simulations

EM and Viterbi algorithms are standard and well established tools for statistical inference in hidden Markov models. The situations we want to face in electro-physiological signals however correspond to models with large dimensional observations, which raise important identifiability questions. The goal of the present section is to validate the modification of the algorithms induced by the Kronecker type matrices, and show that they yield more precise estimates than the classical algorithms.

To this end, we simulated signals following the Kronecker model and tested both the standard and Kronecker estimation procedures, for several parameter choices. Given the alpha waves case study we shall be interested in below, we limited ourselves to $N_s = 2$ hidden states and generated frequency covariance matrices similar to the estimated matrices for the real EEG signals studied in section 3.2 (see Figure 6), and artificial channel covariance matrices that exhibit topographically distinct sources for the two states (see Figure 2).

We display here the results obtained using 8 channels, single trial ($N_r = 1$) 40 seconds long signals, sampled at 256 Hz. The MDCT transform was tuned so as to generate 32 frequency bins (of bandwidth 4 Hz) for each channel, with a Gauss like window. Out of these, we selected $N_f = 8$ frequency bins so as to cover a sufficient spectrum to deal with the oscillatory patterns of interest for us.

A reference transition probability $\pi_{11} = \mathbb{P}\{X_t = 1 \mid X_{t-1} = 1\} = 0.85$ was chosen (corresponding to a typical value for alpha waves duration, i.e. 700ms), simulations were run for various values of $\pi_{00} = \mathbb{P}\{X_t = 0 \mid X_{t-1} = 0\}$ in the interval $[0.75; 0.95]$ (100 runs per value of π_{00}), and relative errors were computed, defined as

$$\text{Err} = \frac{\|\hat{\Sigma} - \Sigma\|_F}{\|\Sigma\|_F}.$$

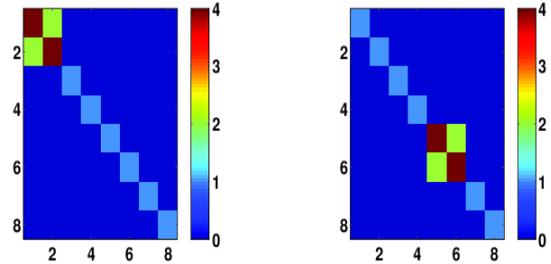


Figure 2: Channel covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

The results, displayed in Figure 3, clearly show that the Kronecker product based estimate outperforms the classical one in terms of accuracy (for comparable running time). As expected, when π_{00} grows, the average length of zero-state regions increases, and the accuracy of the estimate of Σ_0 (resp. Σ_1) grows (resp. decays). The hidden state identification error rates are displayed in Figure 4. The error rates are very comparable for the two methods and always remain at an extremely low level (less than 0.5%). The accuracy of the identification seems to decay when π_{00} increases. One can also see that the error rate is higher when Viterbi algorithm is based on the Kronecker form parameters. We have no simple explanation for these phenomena, nevertheless the error rates remain very low in all considered cases.

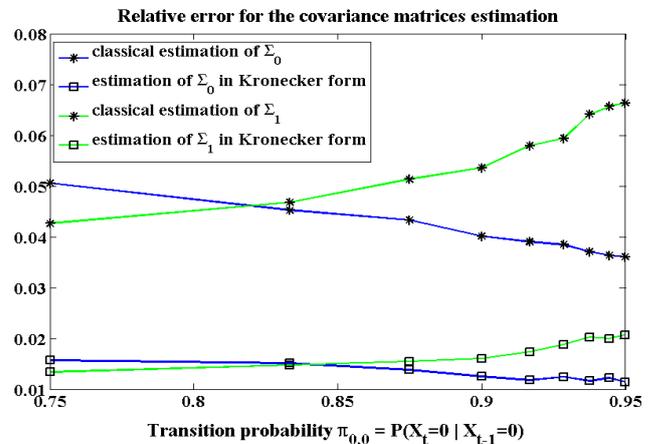


Figure 3: Evolution of the estimation relative errors as a function of the transition probability π_{00}

Numerical simulations using several trials lead to similar results, increasing the number of trials leading to better precision estimates.

3. A case study: alpha waves based characterization of multiple sclerosis

We now turn to results obtained on real data, more particularly to the problem of detection of alpha waves in rest EEG signal. Notice however that the proposed approach can be adapted

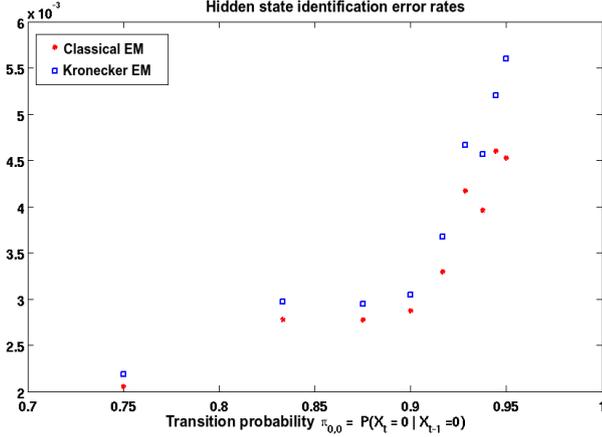


Figure 4: Evolution of the hidden states identification error rates as a function of the transition probability π_{00}

to much more general situations, namely whenever local oscillations show up (apparently randomly) in multichannel signals.

Alpha waves are short duration time-localized oscillations (with frequency around 10 Hz) which appear in EEG signals in some specific situations, and can naturally be accounted for by time-frequency representations. They predominantly originate from the occipital lobe, and are therefore topographically localized in specific sensors located in posterior regions of head. The alpha wave occurrence may be considered a non-stationary effect, i.e. a departure from a stationary background signal. This therefore motivates the use of hidden Markov models as described above.

3.1. Problem statement and data

Alpha waves detection. Our first goal is the automatic detection of alpha waves, more precisely the segmentation of the signal into time epochs where alpha waves are present (state 1) or absent (state 0). Alpha waves detection is generally considered easy, and can be achieved using many signal processing techniques. Among these, a simple one would amount to band pass filter signals to the appropriate frequency band, and make a decision based upon the energy within specified time frames. In any case, such strategies would require choices for time frame length, thresholds, channels,... The MHHMM we have described here provides a more systematic approach.

The main goal of the current study is to use the model in an unsupervised manner, in order to test its ability to blindly detect alpha waves. The case study below provides a positive answer. In other situations though (i.e. with more complex signals), supervised training may be necessary in order to obtain accurate detection.

For the problem under consideration, we use the model with two instances for the latent states: alpha ($s = 1$)/non-alpha ($s = 0$) whose parameters are

- the initial hidden state distribution ν , with components $\nu_s = \mathbb{P}\{X_0 = s\}$

- the transition matrix of the Markov Chain π where $\pi_{s,s'} = \mathbb{P}\{X_{t+1} = s' | X_t = s\}$,
- the channel and frequency covariance matrices corresponding to the two states $\Sigma_0^{(f)}, \Sigma_1^{(f)}, \Sigma_0^{(c)}, \Sigma_1^{(c)}$.

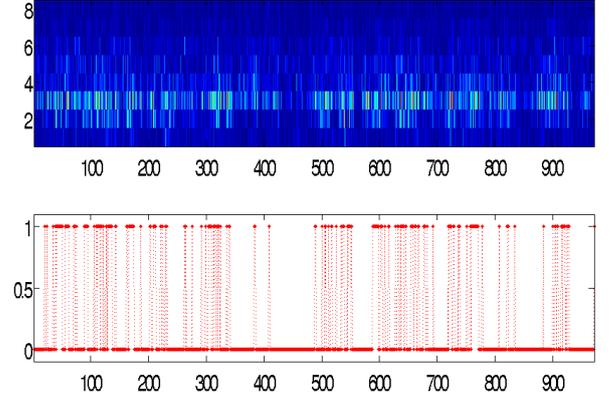


Figure 5: MDCT coefficients for a fixed sensor (top) and hidden states sequence estimated (from all sensors) using Viterbi algorithm (bottom)

We shall consider two different situations. In the *subject by subject* case, parameters will be estimated independently for each subject (in this case, a single trial will be used for the estimation), and a sequence of hidden states will be inferred for each subject too. In the *multi-subject* case, we shall assume that the model parameters are common to all subjects, a hidden states sequence being estimated for each individual subject.

Given the model parameters, the hidden states sequences are estimated for all subjects. These sequences will be the main ingredients for the case study described below, as they represent the bursts in the time-frequency images which corresponds to alpha waves. As an example, we can clearly see in Figure 5 that values equal to one correspond to time domains whose corresponding MDCT coefficients take significantly larger values in the alpha band.

Data and calibration of the time-frequency parameters. The approach was applied to EEG data originating from the CODYSEP dataset, designed to study the impact of sclerosis in inter-hemispherical transfer. The dataset consists in 31 patients (hereafter termed SEP) and 20 controls (TEM); 17 channels EEG signals were collected at a 256 Hz sampling rate.

A subset of 14 subjects and 16 controls was selected, namely those exhibiting sufficiently similar time-frequency contents, in particular in the 8-12 Hz range. The calibration of the time-frequency analysis (for us, the length of the analysis window, also called *hop size*) is a crucial issue. Indeed, if the window is too large, the time resolution can be insufficient to properly describe alpha bursts, while short windows may not yield sufficient frequency resolution to distinguish alpha waves. To set a reasonable trade-off between time and frequency resolution, we set the hop size to $N_f = 32$ samples (i.e. a time resolution of 125 ms), resulting in a moderate frequency resolution

(each MDCT basis function having a bandwidth of approximately 4Hz).

This choice, which is obviously questionable, was motivated by the need of focusing on the same phenomenon (i.e. with the same time-frequency patterns) for all retained subjects. Other choices were of course tested, the corresponding results were of lower precision. It is worth noticing that using Gabor frames rather than MDCT bases would yield more freedom in the calibration, without eliminating the need of a time-frequency trade-off, which results from Heisenberg's uncertainty.

A subset of sensors was also selected that correspond to relevant scalp locations for observing alpha waves.

3.2. Estimation results

Corresponding MDCT transforms were computed, and the model parameters (for a two-state model: alpha and non alpha states) were estimated using the above algorithms. Examples of covariance matrices for the two states are displayed in Figure 6 and 7. The frequency covariance matrices shown in Figure 6 are almost diagonal, which indicates a decorrelation of the frequency bands. The main difference between the two matrices appears for the frequency band 8-12Hz (matrix element (3,3)), which is significantly bigger in 'alpha' state. This state therefore succeeds to capture alpha waves. The channel covariance matrices in Fig 7 are also significantly different for some channels: in particular, channels O1 and O2 (matrix elements (4,4) and (11,11)) are overactivated in 'alpha' state. Again this is coherent, since those sensors correspond to regions where alpha wave signals are most visible. This phenomenon is even more obvious on the graphical representation of topographies (Figure 8), actually representing the diagonals of matrices of Fig 7 on which the relevant sensors O1 and O2 are clearly more "energetic" in 'alpha' state.

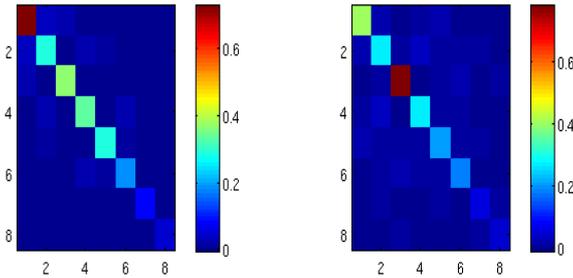


Figure 6: Frequency covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

After completion of parameter estimation, a maximum likelihood estimate for the sequence of latent states is obtained via the Viterbi algorithm. An example is provided in Figure 5.

3.3. A study of alpha wave desynchronization

Let us now turn to a more specific case study in the CODYSEP dataset, namely the study of alpha wave desynchronization between left hand and right hand sensors. We will test the assumption that multiple sclerosis can affect the left-right

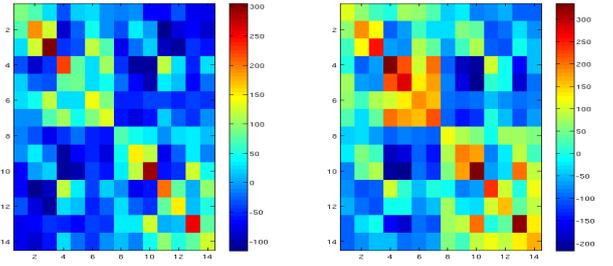


Figure 7: Channel covariance matrices for state 'non-alpha' (left) and 'alpha' (right)

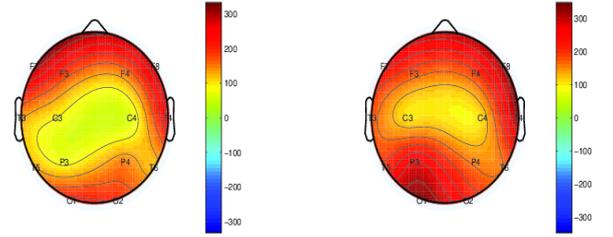


Figure 8: Representations of the channels variances for state 'non-alpha' (left) and 'alpha' (right)

synchronization in the alpha band, by estimating, for each subject, the hidden states sequences in left and right sensors separately, and comparing them. The assumption is supported by known physiological results, and losses in synchronization between EEG signals in left and right hemispheres have been reported in the literature. For instance, [4] states : "in multiple sclerosis, both axonal damage and demyelination occur. Therefore it can be expected that the connectivity between the different regions in the brains of MS patients will be impaired compared to healthy controls." In the literature, desynchronization between signals is often detected and assessed using correlation type measures, such as coherence. The MHHMM method described here is an alternative natural candidate for testing and evaluating such desynchronization effects.

To test the desynchronization assumption, hidden states sequences were estimated for all subjects in the dataset, separately for left sensors and right sensors. The latter sequences take the form of sequences of zeros (non-alpha state) and ones (alpha state). For each subject, the Hamming distance between the left sensors $X_t^{(L)}$ and right sensors $X_t^{(R)}$ sequences was computed

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

and the empirical distributions of the so-obtained distances were compared using a non parametric test. The figure 9 shows the boxplots of the two families (SEP and TEM) of distances, and seems to indicate a significant difference: SEP data exhibit a larger left-right asymmetry, in accordance with the above mentioned desynchronization hypothesis. To assess statistically the significance of the result, we performed a Mann-Whitney test. The corresponding P-value was found to equal $P = 0.0384$,

which confirms quantitatively the hypothesis of two distinct distributions.

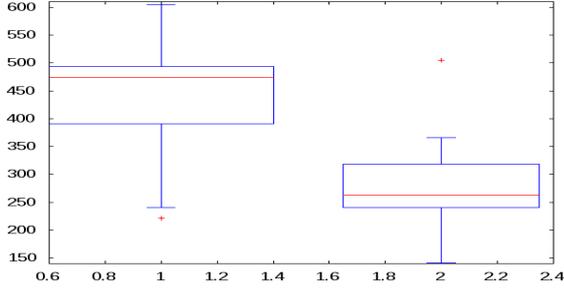


Figure 9: Boxplots of the Hamming distances between left and right hidden states for SEP (left) and TEM (right)

For the sake of comparison, we also computed a corresponding version of the coherence between left and right sensors. More precisely, we compute an MDCT version of the coherence function, for a specific frequency band: given the MDCT coefficients $y^1[t, f_3]$ and $y^2[t, f_3]$ (for the most significant sensors O1 and O2) for the frequency range $f_3 \sim 8 - 12\text{Hz}$, we computed the quantity

$$C = \frac{|y^1[t, f_3]y^2[t, f_3]|}{\|y^1[t, f_3]\| \|y^2[t, f_3]\|},$$

for all subjects (SEP and TEM) under consideration. The results are summarised by a boxplot in Figure 10, where it can be seen that even though TEM subjects tend to have a larger “left-right” coherence, the difference is absolutely not significant. We tested several similar coherence measures, with essentially the same results. The sequences of hidden states provided by our model yields the most significant decision between SEP and TEM datasets.

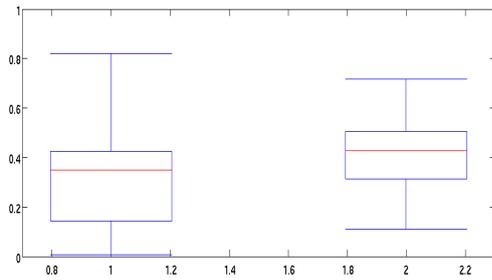


Figure 10: Boxplots of the coherences between left and right hidden states for SEP (left) and TEM (right)

Remark: a multi-subject study. As stressed before, the approach developed here can also cope with multitrial signals, assuming that all trials share the same parameter set (covariance matrices and transition probabilities), but not necessarily the same latent variable sequences. Numerical simulations produce results similar to the results in the previous section in terms of accuracy. Such extensions can be of interest in some situations,

for instance when several epochs are available (for example in BCI protocols).

Although the model presented in section 2.2 has been defined as a multitrial model, it is also interesting to test it in a *multi-subject* situation, for example as a test of the inter-subject variability of the model parameters. Indeed, if the intersubject variability is weak enough, applying our approach to multi-subject data can yield improved parameter estimations. In addition, one may think of using the model parameters learnt from a training set of subjects to estimate hidden states sequences for a test set.

We attempted such an approach on the Codysep dataset, processing independently the two (SEP and TEM) groups. For each group, we considered the whole signals as independent and identically distributed realisations of the same Multichannel Harmonic Hidden Markov Model. This boiled down to choose $N_r = 16$ trials for the first group (SEP) and $N_r = 14$ trials for the second one (TEM). Unfortunately, numerical tests were not conclusive as they yielded irrelevant parameters and hidden states sequences for the problem of alpha waves detection.

To interpret such a negative conclusion, we display in Figure 11 two time-frequency images associated to two patients from the SEP group. These images clearly exemplify the inter-subject variability which dominates the intrinsic variability of the alpha and no-alpha regions. Indeed, the two subjects present very different time-frequency profiles. For the first one, the prevailing activity deals with the alpha band (8-12 Hz) whereas the second subject exhibits some more scattered activity with large coefficients associated with very low frequencies (from 0 to 4 Hz). Such differences explain the fact that the estimation algorithms (EM + Viterbi) do not lead to automatic alpha waves detection.

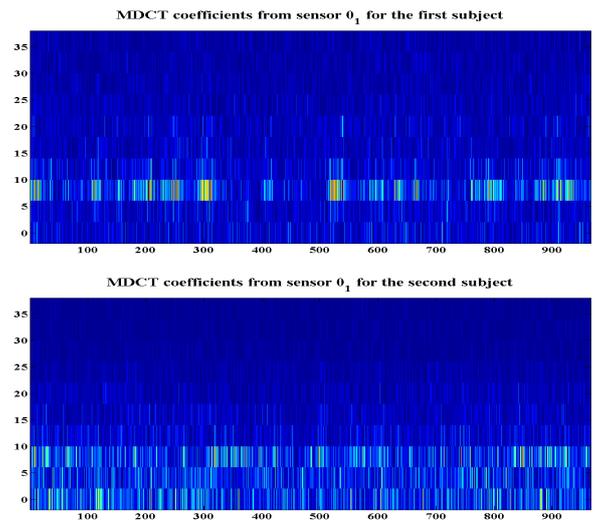


Figure 11: MDCT coefficients computed from sensor O1 for two subjects from group SEP

4. Conclusions and perspectives

We have described in this paper a new model for the study of multi-trial, multi-channel signals exhibiting oscillatory patterns. Time-frequency representations are clearly relevant to account for such features, and we defined models for the time-frequency coefficients rather than the signals themselves. The use of MDCT basis allows us to extend the structure of hidden Markov models and enable us to model explicitly the non-stationarity of such signals. In order to get tractable algorithms on realistic data (in terms of length, time-frequency resolution and number of channels...), we proposed an adaptation to the case where covariance matrices take the form of Kronecker products of smaller matrices gathering respectively topographic and frequency correlations. We extend the classical algorithms (EM, Viterbi) to tackle this new problem of parameters estimation and test them on simulated data. We validated the estimation procedure on numerical simulations, and tested the algorithms developed on a case study devoted to alpha rhythms characterization and detection from EEG signals. While the algorithms described here are not new, the model itself and its adaptation to EEG type data is original. Furthermore, the case study clearly shows the relevance of the model for real data processing, since it yields a biologically relevant result.

In this work, fixed-time vectors of MDCT coefficients were modeled as multivariate Gaussian vectors whose distribution is governed by a hidden state representing some mental state of the subject. Such Gaussian mixtures can also be used to account for different cerebral activities. Thus, this model can potentially be used to deal with various problems of similar nature (i.e. detection and characterization of non-stationary features) as long as the signals of interest exhibit oscillatory features that are different enough in terms of channels and frequency localizations. Further work will concern the validation of this approach to different types of EEG detection and characterization problems. Applications to BCI (brain computer interfaces) problems (namely, beta rebounds detection) are currently under study.

From a more fundamental point of view, the approach developed here strongly relies on the fact that the time-frequency transform corresponds to an orthonormal basis (extensions to other bases than the MDCT can also be considered). However, MDCT bases are known for having important limitations in terms of time-frequency resolution. It would therefore be desirable to turn to better quality time-frequency expansions, such as Gabor frames (see [6, 8] for a detailed account of theory and algorithms). Gabor frames yield expansions of multi-channel signals in the form $x^c = \sum_{t,f} y_{t,f}^c \phi_{t,f}$, where the $\phi_{t,f}$ are the Gabor atoms (time-frequency shifted copies of a reference window function ϕ), and the coefficients $(y_{t,f}^c)$ are *synthesis coefficients*, to be modelled. This raises two major difficulties: first, a choice has to be made for synthesis coefficients (which are generally not unique), which also have to be estimated from the input signals. In [15] we propose an algorithm in the spirit of Majorization-Minimization methods to solve this problem when the vectors $y_{t,f}$ are supposed to be generated by a multivariate Gaussian mixture. Second, the non-orthogonality of

Gabor atoms complicates the use of Markov models and the task of inference since standard algorithms won't be applicable without complex structural refinements. Such an issue will be sought in a forthcoming study.

Another important problem that has to be dealt with when such approaches are applied to multichannel signals is the curse of dimensionality. Using large dimensional Gaussian vectors often leads to numerical instabilities, which can be better controlled when dimension is reduced. A first way to do that is to use Kronecker-structured covariance matrices, as we did in this work. Other complementary approaches could also be investigated. For example, dimension reduction techniques (principal component analysis, or more sophisticated techniques) could be applied in the channel space. One may also think of constructing the channel dimension reduction so as to favour discrimination between states, or develop a state dependent dimension reduction scheme. The latter problem, which is probably very relevant (in EEG signals, different states generally correspond to different sources in the cortex and appear most significantly in different sensors) would however require major modifications in the model and the estimation procedure.

A last extension that could also be envisioned from our current work would be the coupling with inverse problem approaches, to introduce the modelling directly in the source domain. Again, that would require important modifications in the estimation algorithms.

Acknowledgements

This work was partly supported by CNRS, through the *NeuroInformatique* program, which allowed the recording of the **CODYSEP** data used in this work, and by the ANR project **CO-ADAPT** (ANR-09-EMER-002-05).

Appendix A. EM algorithm for Kronecker structured covariance matrices

We provide here details on the estimation algorithms and for implementations. EM is a fairly classical tool in computational probability and statistics, we focus on the modifications necessary to adapt the algorithm to Kronecker structured covariance matrices.

Appendix A.1. Expectation:

Within this subsection, we consider any realization (trial) $\underline{x} = \underline{x}^{(r)}$ of the model. Each trial gives rise to a family of forward and backward variables $\alpha_t^{(r)}$ and $\beta_t^{(r)}$. Throughout the current subsection Appendix A.1 we drop the trial label “r” for the sake of simplicity. We recall that for $t \in 0, \dots, N_t - 1$, the vectors of MDCT coefficients between time t_1 and time t_2 is denoted by $\underline{y}_{t_1:t_2} = \{y[t, f], t = t_1, \dots, t_2, f = 0, \dots, N_f - 1\}$.

The forward variables defined in (5) are computed recursively for $t = 1, \dots, N_t - 1$, while the backward variables defined

in (6) are obtained by recursion too, operating on decreasing indices $t = N_t - 1$ down to 0:

$$\alpha_{t+1}^s = \Phi_s(\underline{y}_{t+1}) \sum_{s'=1}^{N_s} \pi_{s's} \alpha_t^{s'}, \quad (\text{A.1})$$

$$\beta_t^s = \sum_{s'=1}^{N_s} \pi_{ss'} \Phi_{s'}(\underline{y}_{t+1}) \beta_{t+1}^{s'}. \quad (\text{A.2})$$

Thanks to these recursive equations, the complexity of the computation is linear in t , which makes the algorithm extremely efficient.

Given the forward and backward variables, the distribution of the transition (X_t, X_{t+1}) for $t = 0, \dots, N_t - 2$ conditional to the observations up to final time $N_t - 1$ reads

$$\mathbb{P}\{X_t = s, X_{t+1} = s' | \underline{y}\} = \frac{1}{\mathcal{L}} \alpha_t^s \pi_{s,s'} \Phi_{s'}(\underline{y}_{t+1}) \beta_{t+1}^{s'} \quad (\text{A.3})$$

where the constant \mathcal{L} is the likelihood of the observations until final time $\mathcal{L} = L_{N_t-1}$, obtained as

$$\mathcal{L} = L_{N_t-1} = \sum_{s=1}^{N_s} \alpha_t^s \beta_t^s \quad (\text{A.4})$$

for every considered time t .

As a consequence, the distribution of any hidden state X_t , $0 \leq t \leq N_t - 2$ conditional to the observations up to final time $N_t - 1$ satisfies

$$\mathbb{P}\{X_t = s | \underline{y}_{0:N_t-1}\} = \frac{1}{\mathcal{L}} \alpha_t^s \beta_t^s \quad (\text{A.5})$$

Remark 1. In order to avoid underflow potentially caused by very small values of the probabilities, normalized versions of the forward and backward variables α and β have to be used. In this work we used the normalization proposed by Rabiner in [12], to which we refer for more details.

Appendix A.2. Maximization:

Given the forward and backward variables $\alpha_t^{(r);s}$ and $\beta_t^{(r);s}$ associated with any trial $\underline{x}^{(r)}$ and state s , the maximum likelihood estimates for the parameters of the Markov chain are obtained as follows. We denote generically \hat{u} the new estimate, obtained from a previous value u (in the iterative algorithm, the result of the previous iteration). Denoting by $\mathcal{L}^{(r)}$ the final likelihood of trial r ,

$$\hat{v}_s = \frac{1}{N_r} \sum_{r=1}^{N_r} \frac{\alpha_0^{(r);s} \beta_0^s}{\mathcal{L}^{(r)}} \quad (\text{A.6})$$

$$\widehat{\pi}_{s,s'} = \pi_{s,s'} \frac{\sum_{r=1}^{N_r} \frac{1}{\mathcal{L}^{(r)}} \sum_{t=0}^{N_t-2} \alpha_t^{(r);s} \beta_{t+1}^{(r);s'} \Phi_{s'}(\underline{y}_{t+1})}{\sum_{r=1}^{N_r} \frac{1}{\mathcal{L}^{(r)}} \sum_{t=0}^{N_t-2} \alpha_t^{(r);s} \beta_t^{(r);s}}, \quad (\text{A.7})$$

where equations (A.3) and (A.5) are used for numerator and denominator, respectively.

Given the Kronecker structure of the covariance matrices, the corresponding re-estimation procedure is an alternating minimization procedure: consider $\underline{y}_t^c = (y_{t,f}^c, f = 0, \dots, N_f - 1)$ and $\underline{y}_{t,f}^c = (y_{t,f}^c, c = 1, \dots, N_c)$ and iterate the following steps

- *Estimation of $\Sigma_s^{(c)}$ given $\Sigma_s^{(f)}$:* define

$$M_t^{(r),s}(c, c') = \langle (\Sigma_s^{(f)})^{-1} \underline{y}_t^{(r),c}, \underline{y}_t^{(r),c'} \rangle$$

and set

$$\widehat{\Sigma}_s^{(c)} = \frac{1}{N_f} \frac{\sum_{r=1}^{N_r} \sum_{t=0}^{N_t-1} \mathbb{P}\{X_t^{(r)} = s\} M_t^{(r),s}}{\sum_{r=1}^{N_r} \sum_{t=0}^{N_t-1} \mathbb{P}\{X_t^{(r)} = s\}} \quad (\text{A.8})$$

- *Normalization:* set

$$\widehat{\Sigma}_s^{(c)} = \widehat{\Sigma}_s^{(c)} / \left\| \widehat{\Sigma}_s^{(c)} \right\|_F, \quad (\text{A.9})$$

$\| \cdot \|_F$ denoting the Frobenius norm.

- *Estimation of $\Sigma_s^{(f)}$ given $\widehat{\Sigma}_s^{(c)}$:* define

$$P_t^{(r),s}(f, f') = \langle (\widehat{\Sigma}_s^{(c)})^{-1} \underline{y}_{t,f}^{(r)}, \underline{y}_{t,f'}^{(r)} \rangle$$

and set

$$\widehat{\Sigma}_s^{(f)} = \frac{1}{N_c} \frac{\sum_{r=1}^{N_r} \sum_{t=0}^{N_t-1} \mathbb{P}\{X_t^{(r)} = s\} P_t^{(r),s}}{\sum_{r=1}^{N_r} \sum_{t=0}^{N_t-1} \mathbb{P}\{X_t^{(r)} = s\}} \quad (\text{A.10})$$

These estimators are obtained by alternate optimization of the log-likelihood with respect to $\Sigma_k^{(c)}$ and $\Sigma_k^{(f)}$ respectively. The normalization step described here above enables us to solve the indetermination since $\Sigma_k^{(f)}$ and $\Sigma_k^{(c)}$ are defined up to a constant. Strictly speaking, this corresponds to a GEM algorithm for which the convergence to local minimum is still ensured by the fact that each step increases the likelihood of the observations with respect to the model.

Remark 2. As an alternative, the full covariance matrix Σ can also be estimated classically, the channel and frequency covariances being estimated afterwards by factorization, i.e. by minimizing a mean square error $\left\| \Sigma_k - \Sigma_k^{(c)} \otimes \Sigma_k^{(f)} \right\|_F^2$ under the constraint $\left\| \Sigma_k^{(c)} \right\|_F = 1$.

References

- [1] Bijma, F., De Munck, J., Huizenga, H., Heethaar, R., 2003. A mathematical approach to the temporal stationarity of background noise in MEG/EEG measurements. *NeuroImage* 20, 233–243.
- [2] Bijma, F., De Munck, J. C., 2008. A space-frequency analysis of MEG background processes. *Neuroimage* 43, 478–488.
URL <http://www.few.vu.nl/~fetsje/BijmaSumKPFreq.pdf>
- [3] Carmona, R., Hwang, W., Torr sani, B., 1998. *Practical Time-Frequency Analysis: continuous wavelet and Gabor transforms, with an implementation in S*. Vol. 9 of *Wavelet Analysis and its Applications*. Academic Press, San Diego.
- [4] Cover, K. S., Vrenken, H., Geurts, J. J., van Oosten, B. W., Jelles, B., Polman, C. H., Stam, C. J., van Dijk, B. W., February 2006. Multiple sclerosis patients show a highly significant decrease in alpha band inter-hemispheric synchronization measured using MEG. *NeuroImage* 29 (3), 783–788.
- [5] Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134, 9–21.
URL <http://www.sccn.ucsd.edu/eeglab/>

- [6] Feichtinger, H. G., Strohmer, T., 1997. Gabor Analysis and Algorithms : Theory and Applications.
- [7] Flandrin, P., 1999. Time-frequency/time-scale analysis. Vol. 10 of Wavelet Analysis and its Applications. Academic Press Inc., San Diego, CA, with a preface by Yves Meyer, Translated from the French by Joachim Stöckler.
- [8] Gröchenig, K., 2001. Foundations of Time-Frequency Analysis. Birkhäuser.
- [9] Kowalski, M., Torrèsani, B., May 2008. Décompositions parcimonieuses et persistantes de signaux multicanaux. applications aux signaux MEEG. In: d'Alché Buc, F. (Ed.), Proceedings of CAP'08, 10eme Conference d'Apprentissage Porquerolles. Cepadues Editions, Porquerolles, France, pp. 105–120.
URL <http://hal.archives-ouvertes.fr/hal-00347436/fr/>
- [10] Mallat, S., 2007. A wavelet tour of signal processing, the sparse way. Academic Press, second edition.
- [11] Molla, S., Torrèsani, B., 2005. An hybrid audio scheme using hidden Markov models of waveforms. Applied and Computational Harmonic Analysis 18 (2), 137–166.
- [12] Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 257–286.
- [13] Soendergaard, P., 2010. LTFAT, the linear time-frequency analysis toolbox. URL <http://sourceforge.net/projects/ltfat>
- [14] Torrèsani, B., Villaron, E., 2010. Harmonic hidden Markov models for the study of EEG signals. In: EUSIPCO 2010, European Signal Processing Conference, Aalborg, Denmark. Session L-BSP-1.
URL <http://hal.archives-ouvertes.fr/hal-00492800>
- [15] Villaron, E., Anthoine, S., Torrèsani, B., 2011. Modèles de mélange de gaussiennes temps-fréquence pour le débruitage des signaux. In: Proceedings of GRETSI'11.
- [16] Wickerhauser, M. V., 1994. Adapted Wavelet Analysis from Theory to Software. AK Peters, Boston, MA, USA.