

From EEG Dependency Multichannel Matching Pursuit to Sparse Topographic Decomposition

Daniel Studer, * Ulrich Hoffmann,[†] Thomas Koenig *

** University Hospital of Clinical Psychiatry (Director: Prof. W. Strik, M.D.),
Department of Psychiatric Neurophysiology,
Bolligenstrasse 111, CH-3000 Berne, Switzerland.*

*[†] Ecole Polytechnique Fédérale de Lausanne (EPFL),
Signal Processing Institute,
CH-1015 Lausanne, Switzerland.*

Corresponding author: Daniel Studer. Email: daniel.studer@puk.unibe.ch,
tel: +41(31)930 9930 fax: +41(31)930 9961

Abstract

In this work we present a multichannel EEG decomposition model based on an adaptive topographic time-frequency approximation technique. It is an extension of the Matching Pursuit algorithm and called Dependency Multichannel Matching Pursuit (DMMP). It takes the physiologically explainable and statistically observable topographic dependencies between the channels into account, namely the spatial smoothness of neighboring electrodes that is implied by the electric leadfield. DMMP decomposes a multichannel signal as a weighted sum of atoms from a given dictionary where the single channels are represented from exactly the same subset of a complete dictionary. The decomposition is illustrated on topographical EEG data during different physiological conditions using a complete Gabor dictionary. Further the extension of the single-channel time-frequency distribution to a multichannel time-frequency distribution is given. This can be used for the visualization of the decomposition structure of multichannel EEG. A clustering procedure applied to the topographies, the vectors of the corresponding contribution of an atom to the signal in each channel produced by DMMP, leads to an extremely sparse topographic decomposition of the EEG.

Key words: adaptive topographic time-frequency approximation, matching pursuit, dependency multichannel matching pursuit, energy distribution, multichannel energy distribution, sparse topographic decomposition, clustering, EEG, BCI.

1 Introduction

Scalp recorded EEG signals result directly from the synchronous firing of large neural populations in the brain. The recorded signals are well organized in space and time: Both EEG and event-related potentials (ERPs) have typical scalp distributions (topographies) that reproduce well across subjects and that are closely associated with the activity of specific brain regions and functions.

As function of time, EEG signals have long proven to provide sensitive probes into the brain's current functional state (i.e. the level of arousal, sleep stage, drug induced states, event-related synchronization and desynchronization). They also reflect well many pathological processes like epilepsy, neurodegeneration or tumors. Despite the EEGs sensitivity for many relevant brain functional states, the isolation and characterization of its constituting elementary patterns has remained difficult. This is at least partly due to the fact that the scalp EEG is additive: signals recorded at a single electrode can be composed of a mixture of activities that originate from different locations (mixture in space, e.g. resulting from bilateral activity) and that may have different temporal patterns (mixture in time, e.g. sleep spindles overlapping with slow waves). In order to attempt the isolation of basic EEG patterns in time and space, the effects of this mixing have to be taken into account. Thus, the decomposition of the EEG requires the identification of both spatial and temporal elements that can account for an EEG in an additive way.

For a physiologically plausible and comprehensive decomposition, it is necessary that these spatial and temporal elements have some elementary properties. In the following part, we will thus introduce and discuss the properties of 'atomic' spatial and temporal elements that can serve as a useful basis for EEG decomposition. In space, the properties of atomic elements are physically well defined. As starting point that cannot be further subdivided, we use the activity of a single, point-like dipole localized anywhere in the brain. From the location, orientation and strength of a given dipole, one can compute the scalp potential differences generated by this dipole, the so called EEG topography. The physics that determine the topography produced by a single dipole are well known and are essentially determined by the volume conduction properties of the brain, skull, and

scalp. They are summarized in the so-called lead-field (Lopes da Silva, 2004), a matrix that translates the intracerebral activity to scalp electric topographies. In general, it has been shown that the lead-field is a spatial smoothing operator that strongly attenuates intracerebral high spatial frequencies. This implies that even single, point-like sources produce smooth scalp topographies that generally cover the entire scalp. Spatial smoothness implies that signals recorded at neighboring electrodes can be expected to be similar, see Pascual et al. (1988) and Jimenez et al. (1995). Furthermore, since volume conduction is instantaneous, the signal from any single source will affect all electrodes simultaneously, and time delays between electrodes cannot be accounted for by a single source. Finally, since we deal with sum of dipoles, the resulting field will in general have areas with opposite (positive and negative) polarity, although these areas may not always be covered well by the electrode array.

For the characterization of atomic elements in time, the knowledge about the physiological mechanisms of the generation of EEG oscillations (Lopes da Silva, 1991) is still insufficient to provide models that are general enough to account for the full spectrum of EEG events. Therefore, the current, more pragmatic practice is to employ user-defined elements that are assumed to be 'EEG like' and suit the specific research objectives. In the following section, we will briefly review the most typical approaches: In quantitative EEG, the most widely used atomic elements in time are continuous sine- and cosine functions. This is the classical FFT that has provided many important, well-established and clinically relevant findings (John et al., 1977, 1994; Valdes et al., 1992; Dumermuth et al., 1987 for a review) and helped to attribute specific brain functions to specific, narrowly defined EEG frequency bands (Harmony et al., 2004). Sine / cosine functions are especially useful for investigating brain states that are assumed to fluctuate only little or very slowly (e.g. sleep stages: Achermann et al., 2003, drug states: Saletu et al., 1987, maturation: John et al., 1980, or neuropathology: Prichep et al., 1994). However, the variance of FFT transformed EEG across channels and analysis periods has been used to obtain a robust factorization of EEG spectral power into components that are defined in time, frequency, and space (Miwakeichi et al., 2004).

It is noteworthy that in Fourier-transformed mul-

tichannel EEG data, one typically observes that more than 75% of the variance across electrodes is compatible with sources oscillating at a common phase (Lehmann et al., 1990). The amount of common phase across the electrodes has been shown to depend on the subject’s general functional state (schizophrenia: Koenig et al., 2001; dementia: Koenig et al., 2005). However, many changes of the brain’s state occur at a timescale way below a second. These events are ill-represented when using the FFT.

Time domain analysis uses instantaneous delta-functions as atomic events. In time domain EEG analysis, it has been shown that multichannel EEG can be parsed into subsecond epochs that are stable in spatial configuration. These so called microstates last about 100 ms and are separated by rapid topographical changes (Lehmann et al., 1987). It is worth noting that the observation of microstates implies that the EEG is composed of a continuous sequence of short time periods where most of the brain activity visible in the EEG appears highly synchronized. Furthermore, the observed microstate configurations seem to concentrate in a limited number of prototypical maps and can be efficiently classified (Wackermann et al., 1993; Pascual et al., 1995). Therefore quantitative descriptions of EEG data in terms of i.e. durations of microstates of some typical configuration became feasible. In normal subjects, EEG microstate topography has been shown to indicate different cognitive modes (Lehmann et al., 1998; Mueller et al., 2004), maturational states (Koenig et al., 2002) and abnormal conditions including dementia (Strik et al., 1997; Dierks et al., 1997) or schizophrenia (Koenig et al., 1999, Lehmann et al., 2005).

While microstate analysis provides a high time resolution, there is little information about frequency. In order to perform analyses that provide information both in time and frequency, i.e. in order to identify transient, frequency-locked oscillatory states, the atomic events of choice are Gabor-functions. Gabor-functions have been used to identify sleep spindles (Zygierevicz et al., 1999; Durka et al., in press), epileptic seizures (Durka, 2004), movement related synchronization and desynchronization (Durka et al., 2001; Alegre et al., 2003) and cognitive paradigms (Braeutigam et al., 2001; Gurtubay et al., 2001). Koenig et al. (2001) have proposed a multichannel Topographic Time-Frequency analysis based on Gabor-functions that explains the EEG

as transient oscillations of a small set of topographies. This is again similar to the concept employed in microstate analysis and in the FFT approximation: Within the events that are identified, there are no phase differences across electrodes. A similar approach that is based on the same assumption of transient oscillations with a common phase across electrodes has been proposed for the improvement of source localization procedures (Gonzalez Andino et al. 2001). A recent study analyzing simultaneously recorded EEG and fMRI could demonstrate that such transient, spatially synchronous, oscillatory events do correlate with the hemodynamic activation of localized cortical networks (Koenig et al., 2004).

Furthermore, in order to obtain an optimal time and frequency resolution of these Gabor-functions, Matching-Pursuit has been proposed by Durka et al. (1995). Matching Pursuit initially introduced by Mallat et al. (1993) is an adaptive approximation technique, that decomposes the signal into a linear expansion of waveforms, for example Gabor-functions. The approximation is greedy in the sense that the waveforms are iteratively chosen to correspond optimally to the most important signal structure. The Matching Pursuit technique is adaptive because at the begin of the analysis the precise number of iteration steps to produce a decomposition with a given approximation error is unpredictable. As the algorithm tries to use a small number of waveforms it generates a sparse decomposition. Matching pursuit has previously been applied to EEG (Durka et al., 1995, 2001, 2004; Zygierevicz et al., 1999). However, the application of Matching Pursuit to EEG has been on a channel-wise basis. Durka et al. (in press) have proposed a multichannel EEG matching pursuit that is based on the decomposition of the average signal across channels. More general multichannel Matching Pursuit decompositions have been proposed by Gribonval (2003), Leviatan et al. (2003), Lutoborski et al. (2003) and Bonnet et al. (2004). Alternatively to greedy algorithms like Matching Pursuit Marti-Lopez et al. (2003) have proposed a topographic time-frequency decomposition based on the method of frames. Also basis pursuit, see Chen et al. (1996), is a non-greedy decomposition technique that can serve for topographic time-frequency decompositions.

The aim of the current paper was the implementation of a multichannel matching pursuit algorithm that represents the EEG as a set of spatial topogra-

phies that are modulated by Gabor-functions. Specific to the needs of EEG analysis is the objective that the spatial distribution of the weights of an atom (i.e. the atom's topography) should be compatible with physiologically plausible intracerebral sources. This implies namely that a) there is smoothness of the topographies and b) an atomic event found by our algorithm does not have phase differences across electrodes. After a formal introduction of the method, some examples of normal and abnormal EEG are given and both the visualization and statistical quantification are demonstrated. The method is then discussed in the context of quantitative EEG analysis methods.

2 Methods

2.1 Multichannel Matching Pursuit

The single-channel Matching Pursuit algorithm (MP) is an adaptive approximation technique that iteratively decomposes a signal f in the Hilbertspace \mathcal{H} to a linear combination of atoms drawn from a complete dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$, a family of unit vectors in \mathcal{H} , such that finite linear combinations of the atoms are dense in \mathcal{H} . The principle of the MP can easily be generalized to a linear decomposition of a multichannel signal $\mathbf{f} = (f^1, \dots, f^r) \in \mathcal{H}^r$ of r single channels into atoms of the form $\mathbf{g}_\gamma = (g_\gamma, \dots, g_\gamma) \in \mathcal{D}^r$, called Multichannel Matching Pursuit (MMP). We note explicitly, that contrary to independent channel decomposition, MMP represents every component f^l of \mathbf{f} as a weighted sum of the same elements from the dictionary

$$f^l = \sum_{n=0}^{\infty} c_{n,l} g_{\gamma_n}, \quad g_{\gamma_n} \in \mathcal{D}, \quad l = 1, \dots, r.$$

First define for every channel l the 0^{th} -order residual

$$(1) \quad R^0 f^l = f^l,$$

and then apply the following iteration

1. For a given residual $R^n f^l$ of order $n \geq 0$ select the best atom $g_{\gamma_n} \in \mathcal{D}$ in the sense that

$$(2) \quad \sum_{l=1}^r |C(R^n f^l, g_{\gamma_n})|^2 \geq \alpha \sup_{\gamma \in \Gamma} \sum_{l=1}^r |C(R^n f^l, g_\gamma)|^2,$$

where C is a correlation function that measures the similarity $C(f, g_\gamma)$ between f and g_γ

and $\alpha \in (0, 1]$ is an optimality factor (see also Gribonval (2003)).

2. Compute for all $l = 1, \dots, r$ the residual $R^{n+1} f^l$ of order $n + 1$:

$$(3) \quad R^{n+1} f^l = R^n f^l - C(R^n f^l, g_{\gamma_n}) g_{\gamma_n}.$$

M iterations based on our new multichannel selection criterion will finally lead to the following expansion formula for a component f^l

$$(4) \quad f^l = \sum_{n=0}^{M-1} C(R^n f^l, g_{\gamma_n}) g_{\gamma_n} + R^M f^l.$$

In the MP, initially introduced by Mallat et al. (1993), the correlation function is the inner product of \mathcal{H} : $C(f, g_\gamma) = \langle f, g_\gamma \rangle$. When using this correlation function it follows from theorem 1 in appendix A, that the error $\sum_{l=1}^r \|R^M f^l\|^2$ tends to zero, where $\|\cdot\|$ is the norm in \mathcal{H} , and if \mathcal{H} has a finite dimension N , from lemma 3 in appendix A, that $\sum_{l=1}^r \|R^M f^l\|^2$ decays exponentially:

$$(5) \quad \sum_{l=1}^r \|R^M f^l\|^2 \leq \sum_{l=1}^r \|f^l\|^2 (1 - \alpha I_r)^{M+1},$$

where

$$(6) \quad I_r = \inf_{f \in \mathcal{H}} \sup_{\gamma \in \Gamma} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle|^2 / \sum_{l=1}^r \|f^l\|^2 > 0$$

is the smallest correlation ratio of a function $\mathbf{f} \in \mathcal{H}^r$ with respect to \mathcal{D} .

2.2 Dependency Multichannel Matching Pursuit

Our decomposition algorithm should also take into account possible statistical dependencies between the channels. For this we replace step 1 of the iteration from the previous section by the following rule:

$R^n f^l$ is the given n^{th} -order residual where $n \geq 0$. Let $\mathbf{f}_{n+1, \gamma} = (f_{n+1, \gamma}^1, \dots, f_{n+1, \gamma}^r)^\top$, where

$$(7) \quad f_{n+1, \gamma}^l = \sum_{m=0}^{n-1} C(R^m f^l, g_{\gamma_m}) g_{\gamma_m} + C(R^n f^l, g_\gamma) g_\gamma$$

then we want to select an atom $g_\gamma \in \mathcal{D}$, $\gamma \in \Gamma$, such that $\mathbf{f}_{n+1,\gamma}$ or the contribution $(C(R^n f^l, g_\gamma)g_\gamma)_{l=1,\dots,r}$ to $\mathbf{f}_{n+1,\gamma}$ has a similar property as the signal \mathbf{f} .

Let $D^r : \mathcal{H}^r \rightarrow \mathbb{R}_0^+$ be a cost function that assesses the power of the desired property of $\mathbf{f}_{n+1,\gamma}$ and write

$$(8) \quad D_n(\mathbf{f}, \gamma) = D^r(\mathbf{f}_{n+1,\gamma}).$$

Let $S_n : \mathbb{R}_0^+ \rightarrow [0, W]$ be a function bounded by $W \in \mathbb{R}^+$ with $S_n(0) = 0$ and define the weighting function

$$(9) \quad W_n = S_n(D_n).$$

The new selection criterion to get a Dependency Multichannel Matching Pursuit (DMMP) for the best atom $g_{\gamma_n} \in \mathcal{D}$ at iteration n is based on the maximization of the ratio of the correlation objective and the weighted cost objective:

$$(10) \quad \frac{\sum_{l=1}^r |C(R^n f^l, g_{\gamma_n})|^2}{1 + W_n(\mathbf{f}, \gamma_n)} \geq \alpha \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |C(R^n f^l, g_\gamma)|^2}{1 + W_n(\mathbf{f}, \gamma)}$$

where $\alpha \in (0, 1]$ is an optimality factor.

When applying M times our new iteration rule using the correlation function $C(f^l, g_\gamma) = \langle f^l, g_\gamma \rangle$, again it follows from theorem 1 that $\sum_{l=1}^r \|R^M f^l\|^2$ goes to zero and in the finite dimensional case, it follows from lemma 3 in appendix A that $\sum_{l=1}^r \|R^M f^l\|^2$ decays exponentially:

$$(11) \quad \sum_{l=1}^r \|R^M f^l\|^2 \leq \sum_{l=1}^r \|f^l\|^2 (1 - \alpha I_r \frac{1}{1+W})^{M+1}.$$

From this formula, it is seen, that the upper bound W of the cost W_n is essential for a reasonable convergence rate. An interesting candidate for S_n , where D_n is based on a linear function, is presented in section 2.5.

2.3 Multichannel Time-Frequency Energy Distribution

Let \mathbf{f} be a multichannel signal with components $f^l \in L(\mathbb{R})^2$ and let $g(t) = 2^{1/4} e^{-\pi t^2}$ be a Gaussian function and for any scale $s > 0$, translation u and frequency v define the Gabor function

$$(12) \quad g_{(s,u,v)}(t) = K_{(s,u,v)} g\left(\frac{t-u}{s}\right) e^{i v t},$$

depending on the parameters s, u, v , where $K_{(s,u,v)}$ normalizes the norm of $g_{(s,u,v)}(t)$. Denote $\gamma = (s, u, v)$ and write $K_\gamma = K_{(s,u,v)}$ and $g_\gamma(t) = g_{(s,u,v)}(t)$. The set of all these atoms g_γ builds a well known time-frequency dictionary, the so called Gabor dictionary. Other time-frequency dictionaries can be derived when replacing the Gaussian function $g(t)$ with an appropriate function. When we use a time-frequency dictionary to decompose \mathbf{f} it is our goal to visualize the decomposition structure of \mathbf{f} with a color plot in the time-frequency plane to answer the question how the present frequencies in the decomposition of the signal are distributed over time. For this we first have a look at the time-frequency energy distribution introduced by Mallat et al. (1993) of a function $f \in L(\mathbb{R})^2$ that was decomposed into a linear combination of atoms drawn from an arbitrary time-frequency dictionary \mathcal{D}

$$(13) \quad f(t) = \sum_{n=0}^{\infty} c_n g_{\gamma_n}(t), g_{\gamma_n} \in \mathcal{D}.$$

We use the Wigner distribution of two functions $h_1(t)$ and $h_2(t)$ defined by

$$(14) \quad W[h_1, h_2](t, \omega) = \int_{-\infty}^{\infty} h_1\left(t + \frac{\tau}{2}\right) \bar{h}_2\left(t - \frac{\tau}{2}\right) e^{-i\omega\tau} d\tau,$$

depending on the time variable t and the frequency variable ω . If we have $h_1 = h_2$ then write $W h_1(t, \omega)$ instead of $W[h_1, h_2](t, \omega)$. It follows that

$$(15) \quad \begin{aligned} W f(t, \omega) &= \sum_{n=0}^{\infty} |c_n|^2 W g_{\gamma_n}(t, \omega) \\ &+ \sum_{m \neq n} c_n \bar{c}_m W [g_{\gamma_n}, g_{\gamma_m}](t, \omega). \end{aligned}$$

Its time-frequency energy distribution is defined as the first sum of the Wigner distribution that doesn't contain the so called cross terms that build the double sum, thus we have

$$(16) \quad E f(t, \omega) = \sum_{n=0}^{\infty} |c_n|^2 W g_n(t, \omega).$$

If we apply formula 16 to a component f^l of the multichannel signal \mathbf{f} , that was decomposed by DMMP

into a linear combination of atoms drawn from a Gabor dictionary, we get

$$(17) \quad E f^l(t, \omega) = \sum_{n=0}^{\infty} |C(R^n f^l, g_{\gamma_n})|^2 W g_{\gamma_n}(t, \omega),$$

where the Wigner distribution $W g_{\gamma_n}$ is now a two-dimensional Gaussian bell-shaped curve and is given by

$$(18) \quad W g_{\gamma_n}(t, \omega) = 2e^{-2\pi((\frac{t-u_n}{s_n})^2 + (\frac{s_n(\omega-v_n)}{2\pi})^2)}$$

depending on the parameter $\gamma_n = (s_n, u_n, v_n)$. The point (u_n, v_n) defines the center of the bell shaped curve in the time-frequency plane and the parameter s_n regulates the expansion along the time and the frequency axes. Because $E f^l(t, \omega)$ satisfies the equation

$$(19) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E f^l(t, \omega) dt d\omega = \|f^l\|^2,$$

it can be interpreted as an energy density of f^l in the time-frequency plane. We can now draw an intensity plot of $E f^l(t, \omega)$ in the time-frequency plane, where we assign each point (t, ω) a color corresponding to the value $E f^l(t, \omega)$ that is interpreted as height with respect to the time-frequency plane. The range of colors is chosen automatically based on the minimum and maximum values of $E f^l(t, \omega)$. This plot then reveals information on which frequencies are present in the decomposition of f^l and how those frequencies change over time. This principle can easily be generalized to visualize the decomposition structure of the multichannel signal \mathbf{f} in the time-frequency plane. By adding the equality 19 for $l = 1, \dots, r$ we get

$$(20) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{l=1}^r E f^l(t, \omega) dt d\omega = \sum_{l=1}^r \|f^l\|^2.$$

We can now define the time-frequency distribution of the multichannel signal \mathbf{f} by

$$(21) \quad E \mathbf{f}(t, \omega) = \sum_{l=1}^r E f^l(t, \omega).$$

As in the single-channel case drawing an intensity plot of $E \mathbf{f}(t, \omega)$ in the time-frequency plane shows how the frequencies of the decomposition of \mathbf{f} are localized in time. In practice we only know the M -term decomposition \mathbf{f}_M of the multichannel signal

\mathbf{f} produced by DMMP and therefore we normally draw the intensity plot of $E \mathbf{f}_M(t, \omega)$. From experimenting it turns out, that the intensity plot reveals more information about the decomposition structure of \mathbf{f}_M , if we draw an intensity plot of a scaled multichannel energy distribution defined by

$$(22) \quad E_q \mathbf{f}_M(t, \omega) = (E \mathbf{f}_M(t, \omega))^q,$$

where $0 < q \leq 1$.

2.4 Discrete Real Gabor Dictionary

To decompose a real multichannel signal \mathbf{f} with components $f^l \in L(\mathbb{R})^2$ such that the expansion coefficients also become real we use a time-frequency dictionary originally introduced by Mallat et al. (1993). Let $g(t)$ be the Gaussian function and for any scale $s > 0$, translation u , frequency modulation v , phase $w \in [0, 2\pi)$ denote $\gamma = (s, u, v)$ and define the cosine-modulated Gabor function

$$(23) \quad g_{(\gamma, w)}(t) = K_{(\gamma, w)} g\left(\frac{t-u}{s}\right) \cos(vt + w),$$

where $K_{(\gamma, w)}$ normalizes the norm of $g_{(\gamma, w)}(t)$. The set of all these atoms $g_{(\gamma, w)}$ builds the real Gabor dictionary. To build the multichannel energy distribution $E \mathbf{f}$ of \mathbf{f} we first have to rewrite the decomposition

$$(24) \quad f^l(t) = \sum_{n=0}^{\infty} \langle R^n f^l, g_{(\gamma_n, w_n)} \rangle g_{(\gamma_n, w_n)}(t),$$

produced by DMMP using a real Gabor dictionary such that it is written as a linear combination of atoms drawn from the Gabor dictionary introduced in section 2.3. For this let $\gamma_n = (s_n, u_n, v_n)$ and $\gamma_n^- = (s_n, u_n, -v_n)$ and then apply the relation

$$(25) \quad g_{(\gamma_n, w_n)}(t) = \frac{K_{(\gamma_n, w_n)}}{2} (e^{iw_n} g_{\gamma_n}(t) + e^{-iw_n} g_{\gamma_n^-}(t)).$$

between the complex version 12 and the real version 24 of the gabor function to get

$$(26) \quad f^l(t) = \sum_{n=0}^{\infty} \langle R^n f^l, g_{(\gamma_n, w_n)} \rangle \cdot \frac{K_{(\gamma_n, w_n)}}{2} (e^{iw_n} g_{\gamma_n}(t) + e^{-iw_n} g_{\gamma_n^-}(t))$$

Then Mallat and Zhang define the energy distribution as

$$(27) \quad E f^l(t, \omega) = \sum_{n=0}^{\infty} |\langle R^n f^l, g_{(\gamma_n, w_n)} \rangle|^2 \cdot \frac{1}{2} (W g_{\gamma_n}(t, \omega) + W g_{\gamma_n^-}(t, \omega)).$$

This is justified as $E f^l$ of equation 27 also satisfies equation 19. This version of $E f^l$ serves then as an input to calculate $E f$. To decompose discrete channels of N samples, we define \mathcal{H} as the set of finite discrete single channels of length N and then the discrete real Gabor dictionary is the set of real Gabor atoms for $s \in (1, N)$, integers $u \in (1, N]$, $v \in (2\pi k/N)_{k=0, \dots, N-1}$, $w \in [0, 2\pi)$, where $K_{(\gamma, w)}$ normalizes the discrete norm of $g_{(\gamma, w)}$. As suggested by Mallat and Zhang, it is enough to compute the correlation functions of the signal residuals with a subset $\mathcal{D}_\alpha = (g_{(\gamma, w)})_{(\gamma, w) \in \Gamma_\alpha \times [0, 2\pi)}$ of the discrete real Gabor dictionary to select the best atom $g_{(\gamma_n, w_n)}$ at iteration step n . The parameters of the atoms are chosen from a -adic sequences of integers: The index set $\Gamma_\alpha \times [0, 2\pi)$ is given by all $(\gamma, w) = ((a^j, p a^j \Delta u, k a^{-j} \Delta v), w)$, with $a > 1$, $\Delta u = 1/2$, $\Delta v = \pi$, $0 < j < \log_a(N)$, $0 \leq p < N a^{-j} / \Delta u$, $0 \leq k < a^j 2$ and $w \in [0, 2\pi)$. The exact multichannel and DMMP version of Mallat's theorem from which it follows that it is possible to replace the discrete real Gabor dictionary with the subdictionary \mathcal{D}_α are stated in appendix A as theorem 4 and corollary 5. Further simplifications can be done when using a discrete real Gabor dictionary. If it appears to be computationally intensive to find in each iteration step the best atom that satisfies the selection criterion of inequality 45, then it follows from corollary 6 that it is permissible to select the best atom in the following sense. For $\gamma \in \Gamma$ define

$$(28) \quad w_\gamma = \arg \sup_{w \in [0, 2\pi)} \sum_{l=0}^r |\langle R^n f^l, g_{(\gamma, w)} \rangle|^2,$$

then select the best atom $g_{(\gamma, w_\gamma)}$ in the discrete real Gabor dictionary in the sense that

$$(29) \quad \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w_\gamma)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w_\gamma))} \geq \alpha^* \sup_{(\gamma, w) \in \Gamma_\alpha \times [0, 2\pi)} \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w))},$$

where $\alpha^* \in (0, 1]$ is an optimality factor.

2.5 Spatial Multichannel Matching Pursuit of the EEG

In EEG applications the dependency objective is a spatial objective: It follows from the electric leadfield, which behaves as smoothing operator, that measurements of neighboring channels should have similar values. So we select at iteration step n that atom g_γ whose coefficients $C(R^n f^1, g_\gamma), \dots, C(R^n f^r, g_\gamma)$ are smoothest in space. That means if L is the $r \times r$ -graph Laplacian operator matrix of spatial smoothness of the electrode surface, we look for coefficients, that minimize

$$(30) \quad c \mapsto |Lc| \quad \text{with } c \in \mathbb{R}^r,$$

thus we have

$$(31) \quad D_n(\mathbf{f}, \gamma) = |L(C(R^n f^1, g_\gamma), \dots, C(R^n f^r, g_\gamma))^\top|,$$

where $|\cdot|$ denotes the euclidian norm. (For a discussion of the graph Laplacian operator see Mohar (1991)). Now define $C_n^2(\mathbf{f}, \gamma) = \sum_{l=1}^r |C(R^n f^l, g_\gamma)|^2$ for all $n \geq 0$, $\gamma \in \Gamma$, then a first candidate for a weighting function can be defined as

$$(32) \quad D_n^2(\mathbf{f}, \gamma) / C_n^2(\mathbf{f}, \gamma),$$

the cost $D_n^2(\mathbf{f}, \gamma)$ with respect to $C_n^2(\mathbf{f}, \gamma)$. We introduce additional parameters $0 \leq p_1, 0 \leq p_2 \leq p_3$ and $p_4 \in [0, 1]$ that permit to influence this ratio and define the weighting function

$$(33) \quad \tilde{W}_n(\mathbf{f}, \gamma) = p_1 \frac{(D_n^2(\mathbf{f}, \gamma))^{p_3}}{(C_n^2(\mathbf{f}, \gamma))^{p_2}} p_4^{\ln(n+1)}.$$

To get a suitable signal approximation we have to focus more on the global convergence of the MP as on the local fit of the statistical dependencies between the channels, thus we design the weighting function \tilde{W}_n to prefer an atom among those atoms with small cost $\tilde{W}_n(\mathbf{f}, \gamma)$ which guarantees a reasonable global convergence of the MP. For this we assume that \mathcal{H} has a finite dimension N and, as suggested by Mallat et al. (1993), we limit the search for atoms that match best the signal residuals to the sub-dictionary $\mathcal{D}_\alpha = (g_\gamma)_{\gamma \in \Gamma_\alpha}$ introduced by theorem 4 in appendix A. So let $\mathcal{S}^n \subset \Gamma_\alpha$ be an index set of s atoms,

$s > 0$, which match the signal best in the sense, that for each $\gamma \in \Gamma_\alpha$ exists a $\gamma' \in \mathcal{S}^n$, such that

$$(34) \quad C_n^2(\mathbf{f}, g_{\gamma'}) \geq C_n^2(\mathbf{f}, g_\gamma).$$

Then the final weighting function

$$(35) \quad W_n(\mathbf{f}, \gamma) = \begin{cases} \tilde{W}_n(\mathbf{f}, \gamma), & \gamma \in \mathcal{S}^n \\ W, & \gamma \in \Gamma_\alpha \setminus \mathcal{S}^n \end{cases}$$

with $W > p_1 \lambda^{p_3} (C_0^2(\mathbf{f}, \gamma))^{p_3 - p_2}$, where λ is the largest eigenvalue of $L^\top L$, lets the ratio

$$(36) \quad \frac{C_n^2(\mathbf{f}, \gamma)}{1 + W_n(\mathbf{f}, \gamma)}$$

become large, if $\gamma \in \mathcal{S}^n$, i.e. by Lemma 7 atoms g_γ with $\gamma \in \mathcal{S}^n$ will be preferred. As DMMP is very general model, it is not restricted to the proposed graph laplacian operator constraint (30) for the function D_n . A more sophisticated constraint that can be used with DMMP can be found in Pascual-Marqui et al. (1988). They model the relationship between the electrodes with the surface laplacian that is based on spherical harmonic Fourier expansion.

2.6 From topographic clustering to sparse topographic decomposition

Let $\mathbf{c}_{\gamma_n} = (C(R^n f^1, g_{\gamma_n}), \dots, C(R^n f^r, g_{\gamma_n})) \in \mathbb{R}^r$ for all $n = 0, \dots, M-1$ be the topography of g_{γ_n} , the vector of coefficients corresponding to the contributions of the atom g_{γ_n} to the signal in the channels 1 to r . The number M of topographies and the corresponding atoms can now be reduced to get a more sparse decomposition of the signal (Pascual et al., 1995). Using the method of Pascual et al. (1995), we partition the set of topographies $\{\mathbf{c}_{\gamma_0}, \dots, \mathbf{c}_{\gamma_{M-1}}\}$ into p clusters $\mathcal{T}_1, \dots, \mathcal{T}_p$ with representative topographies $\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_p$, the most central objects of the clusters. The optimal value of p can be determined using cross-validation (Pascual et al., 1995). The scoring of the cluster structure is validated by the silhouette measure (Rousseeuw, 1987). (We partition also the M atoms $\{g_{\gamma_0}, \dots, g_{\gamma_{M-1}}\}$ into p clusters $\mathcal{A}_1, \dots, \mathcal{A}_p$, where an atom g_{γ_n} is assigned to cluster \mathcal{A}_i if the topography \mathbf{c}_{γ_n} matches best the representative topography $\boldsymbol{\tau}_i$ in the sense of covariance:

$$(37) \quad g_{\gamma_n} \in \mathcal{A}_i : \iff i = \arg \max_{m=1, \dots, p} |\boldsymbol{\tau}_m^\top \mathbf{c}_{\gamma_n}|$$

Let

$$(38) \quad \rho_{\gamma_n} = \boldsymbol{\tau}_{m^*}^\top \mathbf{c}_{\gamma_n}$$

where

$$(39) \quad m^* = \arg \max_{m=1, \dots, p} |\boldsymbol{\tau}_m^\top \mathbf{c}_{\gamma_n}|$$

then the new representative waveform a_i building the most central object of the cluster \mathcal{A}_i is defined by

$$(40) \quad a_i(t) = \sum_{g_{\gamma_n} \in \mathcal{A}_i} g_{\gamma_n}(t) \rho_{\gamma_n}.$$

This leads to the sparse topographic multichannel decomposition

$$(41) \quad \mathbf{h}_p(t) = \sum_{i=1}^p a_i(t) \boldsymbol{\tau}_i$$

of the EEG signal $\mathbf{f} \in \mathcal{H}^r$ (See also Koenig et al. (2000) for spatial decompositions). To analyze the timecourse of activity and the frequencies of a representative waveform a_i it is helpful to create an intensity plot of a scaled version of the energy distribution $E a_i(t, \omega)$ in the time-frequency plane. It follows from equation 16 that

$$(42) \quad E a_i(t, \omega) = \sum_{g_{\gamma_n} \in \mathcal{A}_i} |\rho_{\gamma_n}|^2 W g_{\gamma_n}(t, \omega),$$

where $W g_{\gamma_n}(t, \omega)$ is given by equation 18. Our scaled version is then defined by

$$(43) \quad E_q a_i(t, \omega) = (E a_i(t, \omega))^q,$$

where $0 < q \leq 1$.

2.7 Practical Pathfinder

It follows a brief summary about the basic structure of DMMP that should help the reader to apply successfully the method.

Define for every channel l the 0^{th} -order residual

$$(44) \quad R^0 f^l = f^l,$$

then apply the following iteration

1. For a given residuals $R^n f^l$, $n \geq 0$, select the best atom $g_{\gamma_n} \in \mathcal{D}$ in the sense that

$$(45) \quad \frac{\sum_{l=1}^r |C(R^n f^l, g_{\gamma_n})|^2}{1 + W_n(\mathbf{f}, \gamma_n)} \geq \alpha \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |C(R^n f^l, g_{\gamma})|^2}{1 + W_n(\mathbf{f}, \gamma)}$$

where $\alpha \in (0, 1]$.

2. Compute for all $l = 1, \dots, r$ the residual $R^{n+1} f^l$:

$$(46) \quad R^{n+1} f^l = R^n f^l - C(R^n f^l, g_{\gamma_n}) g_{\gamma_n}.$$

M iteration steps lead to the decomposition

$$(47) \quad f^l = \sum_{n=0}^{M-1} C(R^n f^l, g_{\gamma_n}) g_{\gamma_n} + R^M f^l.$$

In practice let f^l be a discrete channel of N samples and the correlation function C be the scalar product $C(f, h) = \sum_{t=0}^{N-1} f(t)h(t)$ of two vectors f and h . As a dictionary \mathcal{D} choose the discrete real Gabor introduced in section 2.4. In part one of the iteration step it is permissible to find an atom of the discrete real Gabor dictionary that is only better than the atoms of the sub-dictionary \mathcal{D}_α of the discrete real Gabor dictionary. As the sub-dictionary \mathcal{D}_α contains only a finite number of atoms there is no need to know the optimality factor α . Formula 35 defines a simple weighting function W_n based on the dependency criterion 31 that can easily be used with the discrete real Gabor dictionary. Finally the M -term decompositions with $l = 1, \dots, r$

$$(48) \quad f_M^l = \sum_{n=0}^{M-1} \langle R^n f^l, g_{(\gamma_n, w_n)} \rangle g_{(\gamma_n, w_n)}.$$

define the topographies

$$(49) \quad \mathbf{c}_{(\gamma_n, w_n)} = (\langle R^n f^1, g_{(\gamma_n, w_n)} \rangle, \dots, \langle R^n f^r, g_{(\gamma_n, w_n)} \rangle)$$

which serve as an input for the sparse topographic decomposition of section 2.6.

3 Examples

We present one type of 74 channel and two types of 21 channel EEG signals: Eyes-closed resting state, stage 2 sleep and spike/waves. All data was sampled at the rate of $\lambda = 250\text{Hz}$, the duration of the analyzed signals was 5s (1250 samples). DMMP was

applied to the EEG against average reference after bandpass filtering between 2 and 20 Hz with a Butterworth zero phase filter. Let the example slices of the filtered rest, sleep and spike/waves signals (denoted by a superscript index in brackets) be $R, S2, SW$.

As an example we find in figure 1 the graph which the laplacian operator matrix L for the 21 channel EEG signals was derived from. The laplacian operator matrix for the 74 channel EEG signal can be derived from a similiar graph describing the neighboring relationship between the electrodes. Since we are only interested in atoms with physiologically interpretable frequencies, we discard all atoms with frequencies above the Nyquist-frequency, e.g. the frequency parameter v has at the most the value π . The parameter a of the discrete Gabor dictionary was set to 1.5. All three decompositions returned by the DMMP consist of $M = 100$ atoms. For each example the square root of the remaining overall relative signal energy is given by

$$(50) \quad e_M = \sum_{l=1}^r \|R^M f^l\| / \sum_{l=1}^r \|f^l\|$$

with $r = 74$ or $r = 21$. In the first example, the resting EEG in figure 2, the signal is dominated by alpha activity, occasionally we find also slower waves and overlapping beta spindles. The sparse topographic decomposition consists of 8 topographic clusters and their new waveforms. The time-frequency planes in figure 3 shows that all topographies have large portions of alpha activity that is separated by phase, but not well localized in time. All topographies show additional activity in a lower or higher frequency band. Almost all topographies (except cluster 1, 6, 8) show higher activity not well localized in frequency. (DMMP parameters: $p_1 = 1.2, p_2 = 0.7, p_3 = 1, p_4 = 0.9$ and $s = 400$, that leads to $e_{100} = 0.278$.)

The second example, a sleep-stage 2 EEG, is shown in figure 4. The signal shows activity in the frequency band of 11 to 15Hz which is the typical range of sleep spindles. It is interrupted by very slow cycles with frequencies below 3Hz. The sparse topographic decomposition is made up of 7 topographic clusters and their waveforms. The time-frequency planes in figure 5 display that the topographies in principle are made up of slow waves differently localized in

time, superimposed by waves with higher frequencies. (DMMP parameters: $p_1 = 1.2, p_2 = 0.7, p_3 = 1, p_4 = 0.9$ and $s = 400$, that gives $e_{100} = 0.263$.) In the third example, the spike/waves EEG in figure 6, the signal shows a dominant 3 per second spike and wave pattern. The sparse topographic decomposition resulted in 5 topographic clusters with their new waveforms. The time-frequency planes in figure 7 show that cluster 1 has in time the most extended pattern of slow waves. All clusters show spikes as vertical lines not localized in frequency, occurring at different moments in time. (DMMP parameters: $p_1 = 1.2, p_2 = 0.2, p_3 = 1, p_4 = 0.9, s = 400$ and get $e_{100} = 0.355$.)

4 Discussion

The paper presents a matching pursuit based multichannel decomposition technique tailored to EEG data. It is an improvement of the method proposed by Koenig et al. (2001) in the sense that the result space does not cover all the atoms that were initially in the dictionary, but only the most relevant ones. It therefore also simplifies subsequent statistics considerably. The decomposition technique stepwise selects an atom from the dictionary such that the vector of coefficients formed by the multichannel decomposition explains a maximal amount of the signal while it is smooth in space. The decomposition thus makes two assumptions, namely that an 'atomic' event is seen simultaneously in all channels, and that the topographic distribution of the event is smooth. Both assumptions are well justified by the volume conduction properties of the brain: Volume conduction is instantaneous, which implies that latency differences in EEG signals cannot result from the same event. Spatial smoothness is justified by the fact that the vector of coefficients of a selected atom corresponds to the scalp field produced by the intracerebral sources that assumingly produced that atom. Since the leadfield operator that relates intracerebral sources to scalp electric potential differences implies smooth topographies, we expect the vector of coefficients of physiologically meaningful atoms to be smooth as well. Imposing temporal simultaneity and spatial smoothness in multichannel EEG decompositions thus biases the solution towards physiologically plausible results. Contrary to the method proposed by Durka et al. (in press), the method is well capable to account for EEG deflec-

tions with opposite polarity that result from, single dipolar events.

The choice of Gabor elements as atoms of the dictionary matches well with many needs of EEG analysis: They are well defined in frequency and thus cover a domain where most of the analyses of continuous EEG data have been made and where there is a vast body of literature that can be referred to (see introduction). On the other side, they reduce the assumption of stationarity to small, subsecond time-windows, which is probably more appropriate to ongoing brain activity than assuming stationarity over the entire analysis period (Ferber, 1987; Gonzalez Andino et al. 2001). Furthermore they correspond well to the so-called spindles that are often used in visual descriptions of EEG.

The clustering of the coefficient vectors extends previous methods on the identification of common phase oscillations (Koenig et al., 2001; Gonzalez Andino et al., 2001) by grouping these oscillations into topographically similar classes. This is efficient in reducing the spatial redundancy of the results and produces a result space that is low-dimensional, which is convenient for later statistics. The observation that EEG data can be explained well by a small number of spatial topographies has been made previously (Koenig et al., in press) and indicates that there seems to be a set of meta-states that are preferentially visited during given experimental conditions. These meta-states assumingly correspond to different modes of information processing (Lehmann et al., 1987) and usually cover large areas of the scalp. The specific content of the single events may thus be represented in the variance of the different topographic classes.

The combination of Gabor atoms with a given frequency, the total energy of the coefficient vectors and the assignment of the coefficient vectors to a low number of classes makes it also possible to assess the energy vs. frequency distribution of spatially defined EEG events. With a sufficient number of events, this distribution can serve to estimate the frequency spectrum of the set of events that belongs to a given class. This may then be used for the systematic comparison of EEG recorded under different conditions, i.e. during different states of vigilance, during different tasks, under the influence of substances or when different groups of subjects are studied. Furthermore, the analysis may be extended by using i.e. the count of events above a certain energy level or the duration of the events.

For visualization, the waveforms of all atoms that belong to a class of topographies can be combined to yield a single representative waveform that accounts for the timecourse of the activity assigned to that class. The duration (scale), latency (translation), frequency and total energy of the atoms belonging to class of topographies can furthermore be used to visualize the energy distribution of the activity of that class on the time-frequency plane. The visualization of sample EEG data under different conditions has shown that the expected main features of the EEG (e.g. sleep spindles or spikes) are well represented in the obtained decomposition.

This new topographic multichannel decomposition approach may be refined in many directions. For a further understanding of the events, the obtained topographies may be submitted to a source localization analysis. Also, one may choose different weighting functions for the spatial objective or include additional objectives. For example one may construct weighting functions based on none logarithmic functions or based on a more adaptive scheme. For further improvement of the method, the identified events may be used for EEG signal classification based on time-frequency and ambiguity function signal classifiers or discriminant pursuit (Buckheit et al. 1995). Furthermore, as described by Pon-Varma et al. (2001) one can identify standard dictionaries of waveshapes for given classes of EEG signals that can serve for their discrimination, e.g. when developing brain-computer interfaces.

Acknowledgement

This work was funded in part by the Swiss National Science Foundation (Grant no. 2153-067852).

A Appendix: Theorems and proofs

In the following let $\mathbf{f} \in \mathcal{H}^r$ and let the residue $R^n f^l$, where $n \geq 0$, $l = 1, \dots, r$ always be defined by equation (1) and induction equation (3), where the selection criterion is always based on equation (45).

Theorem 1

For all $l = 1, \dots, r$ we have

$$(51) \quad f^l = \sum_{n=0}^{\infty} \langle R^n f^l, g_{\gamma_n} \rangle g_{\gamma_n}, \|f^l\|^2 = \sum_{n=0}^{\infty} |\langle R^n f^l, g_{\gamma_n} \rangle|^2.$$

The proof of this theorem is based on the proof of theorem 1 given by Mallat et al. (1993), that goes back to Jones's (1987) proof for the convergence of projection pursuit regression and mainly on the new and more general proof given by Bonnet et al. (2004).

Lemma 2

Let $l = 1, \dots, r$, $m > 0$ and $h_{m,l} = \langle R^m f^l, g_{\gamma_m} \rangle g_{\gamma_m}$, then for all $k > 0$

$$(52) \quad |\langle h_{m,l}, R^k f^l \rangle|^2 \leq \frac{1+W}{\alpha} \|h_{m,l}\|^2 \sum_{j=1}^r |\langle R^k f^j, g_{\gamma_k} \rangle|^2.$$

Proof: We have $|\langle \frac{h_{m,l}}{\|h_{m,l}\|}, R^k f^l \rangle|^2 = |\langle g_{\gamma_m}, R^k f^l \rangle|^2 \leq \sum_{j=1}^r |\langle g_{\gamma_m}, R^k f^j \rangle|^2$. Then with

$$(53) \quad \alpha \frac{\sum_{j=1}^r |\langle g_{\gamma_m}, R^k f^j \rangle|^2}{1+W_k(\mathbf{f}, \gamma_m)} \leq \frac{\sum_{j=1}^r |\langle g_{\gamma_k}, R^k f^j \rangle|^2}{1+W_k(\mathbf{f}, \gamma_k)}$$

and $0 \leq W_k(\mathbf{f}, \gamma_k) \leq W$ the lemma follows. \square

Proof of theorem 1: First we have to show that the sequence $(R^n f^l)_{n \in \mathbb{N}}$ is a Cauchy sequence for all $l = 1, \dots, r$. Let $M > N > 0$, then

$$(54) \quad \begin{aligned} & \|R^N f^l - R^M f^l\|^2 \\ &= \|R^N f^l\|^2 + \|R^M f^l\|^2 - 2\text{Re}(\langle R^N f^l, R^M f^l \rangle) \\ &= \|R^N f^l\|^2 + \|R^M f^l\|^2 - 2\text{Re}(\langle R^M f^l \\ &\quad + \sum_{m=N}^{M-1} h_{m,l}, R^M f^l \rangle) \\ &\leq \|R^N f^l\|^2 + \|R^M f^l\|^2 + 2 \sum_{m=N}^{M-1} |\langle h_{m,l}, R^M f^l \rangle| \end{aligned}$$

By lemma 2 we get

$$(55) \quad \begin{aligned} & \|R^N f^l - R^M f^l\|^2 \\ &\leq \|R^N f^l\|^2 - \|R^M f^l\|^2 + 2\left(\frac{1+W}{\alpha}\right)^{1/2} \\ &\quad \cdot \sum_{m=N}^{M-1} \|h_{m,l}\| \left(\sum_{j=1}^r |\langle R^M f^j, g_{\gamma_M} \rangle|^2\right)^{1/2} \end{aligned}$$

We have

$$\left(\sum_{j=1}^r |\langle R^M f^j, g_{\gamma_M} \rangle|^2\right)^{1/2} \leq r^{1/2} \sup_{j=1, \dots, r} \|h_{M,j}\|$$

and $\sum_{m=N}^{M-1} \|h_{m,l}\| \leq \sum_{m=N}^{M-1} \sup_{j=1,\dots,r} \|h_{m,j}\|$, so we get

$$(56) \quad \begin{aligned} & \|R^N f^l - R^M f^l\|^2 \\ & \leq \|R^N f^l\|^2 - \|R^M f^l\|^2 + 2\left(\frac{1+W}{\alpha}\right)^{1/2} \\ & \quad \cdot \sup_{j=1,\dots,r} \|h_{M,j}\| \left(\sum_{m=N}^{M-1} \sup_{j=1,\dots,r} \|h_{m,j}\|\right)^{1/2} \end{aligned}$$

From $R^N f^l = \langle R^N f^l, g_{\gamma_N} \rangle g_{\gamma_N} + R^{N+1} f^l$ follows the well known energy conservation equation $\|R^N f^l\|^2 = |\langle R^N f^l, g_{\gamma_N} \rangle|^2 + \|R^{N+1} f^l\|^2$, that means $(\|R^n f^l\|^2)_{n \in \mathbb{N}}$ is monotonically decreasing and thus convergent, so for large values of N and M the difference $\|R^N f^l\|^2 - \|R^M f^l\|^2$ goes to zero.

For all $j = 1, \dots, r$ we have $\sum_{m=0}^{\infty} \|h_{m,j}\|^2 < \infty$ and thus $H^2 = \sum_{m=0}^{\infty} \sup_{j=1,\dots,r} \|h_{m,j}\|^2 < \infty$. Let $\epsilon > 0$ then for almost all M we have $\sup_{j=1,\dots,r} \|h_{M,j}\| < \frac{\epsilon}{\sqrt{M}}$, otherwise $H^2 = \infty$.

Define $s(m) = \sup_{j=1,\dots,r} \|h_{m,j}\|$ then by the Cauchy-Schwarz inequality we get $\sum_{m=N}^{M-1} |s(m)| \leq \sqrt{M-N} (\sum_{m=N}^{M-1} |s(m)|^2)^{1/2} \leq \sqrt{M-N} H \leq \sqrt{M} H$ and have now

$$(57) \quad \sup_{j=1,\dots,r} \|h_{M,j}\| \sum_{m=N}^{M-1} \sup_{j=1,\dots,r} \|h_{m,j}\| \leq \frac{\epsilon}{M} \sqrt{M} H \leq \epsilon H$$

and it follows, that $(R^n f^l)_{n \in \mathbb{N}}$ is Cauchy sequence. Let $R^\infty f^l = \lim_{n \rightarrow \infty} R^n f^l$ and show that $\|R^\infty f^l\|^2 = 0$. Since for all $\gamma \in \Gamma$

$$(58) \quad \begin{aligned} & |\langle R^n f^l, g_\gamma \rangle|^2 \\ & \leq \sup_{\gamma \in \Gamma} \sum_{j=1,\dots,r} |\langle R^n f^j, g_\gamma \rangle|^2 \\ & \leq \sum_{j=1,\dots,r} |\langle R^n f^j, g_{\gamma_n} \rangle|^2 \frac{1+W}{\alpha} \\ & \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

we have $|\langle R^\infty f^l, g_\gamma \rangle| = 0$ for all $\gamma \in \Gamma$, thus $R^\infty f^l$ lies in the orthogonal complement of the closed linear span of the dictionary \mathcal{D} , but \mathcal{D} is complete, so $R^\infty f^l = 0$ and $\|R^\infty f^l\| = 0$.

Lemma 3

Let \mathcal{H} be finite dimensional. For all $n > 0$,

$$(59) \quad \sum_{l=1}^r \|R^n f^l\|^2 \leq \sum_{l=1}^r \|f^l\|^2 (1 - \alpha I_r \frac{1}{1+W})^n.$$

Proof: Show first, that $I_r > 0$. For this assume, that f is an element of \mathcal{H} and apply a single-channel Matching Pursuit to f to get a M -term decomposition of f . In this case one can define the single-channel version $I_1 = \inf_{f \in \mathcal{H}} \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|^2 / \|f\|^2$. Mallat et al. (1993) have shown that $I_1 > 0$. The proof of the multichannel version $I_r > 0$ is a simple carry-over of this proof and is not reproduced here. From $\|R^{n+1} f^l\|^2 = \|R^n f^l\|^2 - |\langle R^n f^l, g_{\gamma_n} \rangle|^2$ for all $l = 1, \dots, r$ and all $n > 0$ follows

$$(60) \quad \sum_{l=1}^r \|R^{n+1} f^l\|^2 = \sum_{l=1}^r \|R^n f^l\|^2 - \sum_{l=1}^r |\langle R^n f^l, g_{\gamma_n} \rangle|^2.$$

For all $\mathbf{h} = (h^1, \dots, h^r)^\top \in \mathcal{H}^r$ define

$$(61) \quad \lambda^n(\mathbf{h}) = \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |\langle h^l, g_\gamma \rangle|^2}{\sum_{l=1}^r \|h^l\|^2} \frac{1}{1 + W_n(\mathbf{h}, \gamma)},$$

then $\frac{\sum_{l=1}^r |\langle R^n f^l, g_{\gamma_n} \rangle|^2}{\sum_{l=1}^r \|R^n f^l\|^2} \frac{1}{(1 + W_n(\mathbf{R}^n \mathbf{f}, \gamma_n))} \geq \alpha \lambda^n(\mathbf{R}^n \mathbf{f})$ and we get

$$(62) \quad \sum_{l=1}^r |\langle R^n f^l, g_{\gamma_n} \rangle|^2 \geq \alpha \lambda^n(\mathbf{R}^n \mathbf{f}) \sum_{l=1}^r \|R^n f^l\|^2,$$

where $\mathbf{R}^n \mathbf{f} = (R^n f^1, \dots, R^n f^r)^\top$, thus we have

$$(63) \quad \sum_{l=1}^r \|R^{n+1} f^l\|^2 \leq \sum_{l=1}^r \|R^n f^l\|^2 (1 - \alpha \lambda^n(\mathbf{R}^n \mathbf{f})).$$

By

$$(64) \quad \begin{aligned} \lambda^n(\mathbf{R}^n \mathbf{f}) & \geq \inf_{\mathbf{h} \in \mathcal{H}^r} \lambda^n(\mathbf{h}) \\ & \geq \inf_{\mathbf{h} \in \mathcal{H}^r} \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |\langle h^l, g_\gamma \rangle|}{\sum_{l=1}^r \|h^l\|^2} \frac{1}{1+W} \\ & = I_r \frac{1}{1+W} \end{aligned}$$

we get

$$(65) \quad \sum_{l=1}^r \|R^{n+1} f^l\|^2 \leq \sum_{l=1}^r \|R^n f^l\|^2 (1 - \alpha I_r \frac{1}{1+W}),$$

and finally $\sum_{l=1}^r \|R^n f^l\|^2 \leq \sum_{l=1}^r \|f^l\|^2 (1 - \alpha I_r \frac{1}{1+W})^n$. \square

Now follows the multichannel version of the theorem from Mallat et al. (1993) that the fast numerical implementation of a multichannel matching pursuit in a real Gabor dictionary is based on.

Theorem 4

Let Δu and Δv be a time and a frequency discretization interval that satisfy $\Delta u = \frac{\Delta v}{2\pi} < 1$. Let $a > 1$ be a dilation factor. Define the discrete set $\Gamma_\alpha = \{(a^j, pa^j \Delta u, ka^{-j} \Delta v) | (j, p, k) \in \mathbb{Z}^3\} \subset \mathbb{R}^+ \times \mathbb{R}^2$. There exists a $\alpha \in \mathbb{R}^+$, such that for all $f^1, \dots, f^r \in \mathbf{L}^2(\mathbb{R})$

$$(66) \quad \sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle|^2 \geq \alpha \sup_{\gamma \in \Gamma} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle|^2.$$

Proof: We only have to show, that there is a constant $\alpha > 0$ such that $\sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle| \geq \sqrt{\alpha} \sup_{\gamma \in \Gamma} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle|$, then the theorem follows by induction over r . Set $\gamma = (a^j, pa^j \Delta u, ka^{-j} \Delta v) \in \Gamma_\alpha$ and write $g_{j,p,k} = g_\gamma$. From Daubechies (1990) we know that for the gaussian function $g, (g_{0,p,k})_{(p,k) \in \mathbb{Z}^r}$ is a frame of $\mathbf{L}^2(\mathbb{R})$ with the dual frame $(\tilde{g}_{0,p,k})_{(p,k) \in \mathbb{Z}^r}$, where $\tilde{g} \in \mathbf{L}^2(\mathbb{R})$ and $\tilde{g}_{j,p,k}(t) = \frac{1}{\sqrt{a^j}} \tilde{g}(\frac{t - pa^j \Delta u}{a}) e^{ika^{-j} \Delta v t}$. Let $\gamma_0 = (s, u, v) \in \Gamma, j \in \mathbb{Z}$ be such that $a^{j-1/2} < s \leq a^{j+1/2}$. For all $f^l \in \mathbf{L}^2(\mathbb{R})$, where $l = 1, \dots, r$, we know from Mallat et al. (1993) that

$$(67) \quad \langle f^l, g_{\gamma_0} \rangle = \sum_{p=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle f^l, g_{j,p,k} \rangle \langle \tilde{g}_{j,p,k}, g_{\gamma_0} \rangle.$$

Thus it follows, that

$$(68) \quad \begin{aligned} & \sum_{l=1}^r |\langle f^l, g_{\gamma_0} \rangle| \\ & \leq \sum_{p=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=1}^r |\langle f^l, g_{j,p,k} \rangle| |\langle \tilde{g}_{j,p,k}, g_{\gamma_0} \rangle|. \end{aligned}$$

Because of $g_{j,p,k} = g_\gamma$ with $\gamma \in \Gamma_\alpha$ we have

$$(69) \quad \begin{aligned} & \sum_{l=1}^r |\langle f^l, g_{\gamma_0} \rangle| \\ & \leq \sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle f^l, g_\gamma \rangle| \sum_{p=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\langle \tilde{g}_{j,p,k}, g_{\gamma_0} \rangle|. \end{aligned}$$

Then Mallat et al. (1993) have shown that there is a constant $\alpha > 0$, such that

$$(70) \quad \sum_{p=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\langle \tilde{g}_{j,p,k}, g_{\gamma_0} \rangle| \leq \sqrt{\alpha},$$

and then the theorem follows. \square

From theorem 4 follows the DMMP version

Corollary 5

Let the condition of theorem 4 be satisfied. Then there exists a $\beta \in \mathbb{R}^+$, such that for all n and l

$$(71) \quad \sup_{\gamma \in \Gamma_\alpha} \frac{\sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2}{1 + W_n(\mathbf{f}, \gamma)} \geq \beta \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2}{1 + W_n(\mathbf{f}, \gamma)}.$$

Proof: By theorem 4 there exists a constant $\alpha \in \mathbb{R}^+$ such that

$$(72) \quad \sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2 \geq \alpha \sup_{\gamma \in \Gamma} \sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2,$$

then it follows that

$$(73) \quad \begin{aligned} & \frac{\alpha}{1 + W} \sup_{\gamma \in \Gamma} \frac{\sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2}{1 + W_n(\mathbf{f}, \gamma)} \\ & \leq \frac{\alpha}{1 + W} \sup_{\gamma \in \Gamma} \sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2 \\ & \leq \frac{1}{1 + W} \sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2 \\ & \leq \sup_{\gamma \in \Gamma_\alpha} \frac{\sum_{l=1}^r |\langle R^n f^l, g_\gamma \rangle|^2}{1 + W_n(\mathbf{f}, \gamma)}. \end{aligned}$$

So we can set $\beta = \alpha/(1 + W)$. \square

In case of using a discrete real Gabor dictionary one can obviously use the suboptimal selection criterion that is implied by the following

Corollary 6

Let the condition of corollary 5 be satisfied and let w_γ be defined by equation 28. Then there exists a $\beta^* \in \mathbb{R}^+$, such that for all n and l

$$(74) \quad \begin{aligned} & \sup_{\gamma \in \Gamma_\alpha} \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w_\gamma)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w_\gamma))} \\ & \geq \beta^* \sup_{(\gamma, w) \in \Gamma_\alpha \times [0, 2\pi)} \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w))}. \end{aligned}$$

Proof: Select an appropriate $\alpha^* \in \mathbb{R}^+$ such that

$$\begin{aligned}
(75) \quad & \sup_{\gamma \in \Gamma_\alpha} \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w_\gamma)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w_\gamma))} \\
& \geq \frac{1}{1 + W} \sup_{\gamma \in \Gamma_\alpha} \sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w_\gamma)} \rangle|^2 \\
& \geq \frac{\alpha^*}{1 + W} \sup_{(\gamma, w) \in \Gamma_\alpha \times [0, 2\pi]} \sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w)} \rangle|^2 \\
& \geq \frac{\alpha^*}{1 + W} \sup_{(\gamma, w) \in \Gamma_\alpha \times [0, 2\pi]} \frac{\sum_{l=1}^r |\langle R^n f^l, g_{(\gamma, w)} \rangle|^2}{1 + W_n(\mathbf{f}, (\gamma, w))}
\end{aligned}$$

and set now $\beta^* = \alpha^*/(1 + W)$. \square

Lemma 7

For all $n \geq 0$ let W_n be the weighting function defined by equation (35) then with $\gamma \in \mathcal{S}^n$ and $\tilde{\gamma} \in \Gamma_\alpha \setminus \mathcal{S}^n$

$$(76) \quad \frac{C_n^2(\mathbf{f}, \gamma)}{1 + W_n(\mathbf{f}, \gamma)} > \frac{C_n^2(\mathbf{f}, \tilde{\gamma})}{1 + W_n(\mathbf{f}, \tilde{\gamma})}.$$

Proof: If λ is the largest eigenvalue of $L^\top L$, then of course $|Lc|^2 \leq \lambda|c|^2$ for all $c \in \mathbb{R}^r$ and then

$$\begin{aligned}
(77) \quad W_n(\mathbf{f}, \gamma) &= p_1 \frac{(D_n^2(\mathbf{f}, \gamma))^{p_3}}{(C_n^2(\mathbf{f}, \gamma))^{p_2}} p_4^{\ln(n+1)} \\
&\leq p_1 \frac{(\lambda C_n^2(\mathbf{f}, \gamma))^{p_3}}{(C_n^2(\mathbf{f}, \gamma))^{p_2}} \\
&\leq p_1 \lambda^{p_3} (C_0^2(\mathbf{f}, \gamma))^{p_3 - p_2} \\
&< W \\
&= W_n(\mathbf{f}, \tilde{\gamma}).
\end{aligned}$$

Because of $C_n^2(\mathbf{f}, \gamma) \geq C_n^2(\mathbf{f}, \tilde{\gamma})$ the lemma follows. \square

References

- Achermann, P Borbely, A A. Mathematical models of sleep regulation. *Front Biosci.*, 2003: May, 8, 683-93.
- Alegre, M Labarga, A Gurtubay, I G Iriarte, J Malanda, A Artieda, J. Movement-related changes in cortical oscillatory activity in ballistic, sustained and negative movements. *Brain Res.*, 2003: Jan, 148(1), 17-25.
- Bonnet, C Partington, J R Sorine, M. A modified matching pursuit algorithm applied to the approximation of vectorcardiogram data. *Internat. J. Pure Appl. Math.*, 2004: 15, 25-35.

Braeutigam, S Bailey, A J Swithenby, S J. Phase-locked gamma band responses to semantic violation stimuli. *Brain Res. Cogn. Brain Res.*, 2001: Jan, 10(3), 365-77.

Buckheit, J Donoho, D L. Improved linear discrimination using time-frequency dictionaries. *Proc SPIE*, 1995: 2569, 540-51.

Chen, S Donoho, D L Saunders, M A. Atomic Decomposition by Basis Pursuit. Technical report, Department of Statistics, Stanford University 1996.

Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. on Information Theory*, 1990: Sept, 36, 961-1005.

Dierks, T Jelic, V Julin, P Maurer, K Wahlund, L O Almkvist, O Strik, W K Winblad, B. EEG-microstates in mild memory impairment and Alzheimer's disease: possible association with disturbed information processing. *J. Neural. Transm.*, 1997: 104(4-5), 483-95.

Dumermuth, G Molinari, L. Spectral analysis of the EEG. Some fundamentals revisited and some open problems. *Neuropsychobiology*, 1987: 17(1-2), 85-99.

Durka, P J Blinowska, K J. Analysis of EEG transients by means of matching pursuit. *Ann. Biomed. Eng.*, 1995: Sep-Oct, 23(5), 608-11.

Durka, P J Ircha, D Neuper, C Pfurtscheller, G. Time-frequency microstructure of event-related electroencephalogram desynchronisation and synchronisation. *Med. Biol. Eng. Comput.*, 2001: May, 39(3), 315-21.

Durka, P J. Adaptive time-frequency parametrization of epileptic EEG spikes. *Physical Review E*, 2004: 60, 051914.

Durka, P J Matysiak, A Martnez Montes, E Valds Sosa, P Blinowska, K J. Multichannel matching pursuit and EEG inverse solutions. *Journal of Neuroscience Methods*, in Press, available online 23 May 2005.

Ferber, G. Treatment of some nonstationarities in the EEG. *Neuropsychobiology*, 1987: 17(1-2), 100-4, Review.

Gonzalez Andino, S L Grave de Peralta, R Thut, G Michel, C M. Non stationary distributed source approximation: An alternative to improve localization procedures. *Human Brain Mapping*, 2001: 14, 81-95.

Gribonval R. Piecewise linear source separation. *Wavelets: Applications in Signal and Image Processing X*. Edited by Unser, Michael A.; Aldroubi, Akram; Laine, Andrew F. *Proceedings of the SPIE*, Volume 5207, pp. 297-310 (2003). 2003: Nov, 297-310.

Gurtubay, I G Alegre, M Labarga, A Malanda, A Iriarte, J Artieda, J. Gamma band activity in an auditory oddball paradigm studied with the wavelet transform. *Clin. Neurophysiol.*, 2001: Jul, 112(7), 1219-28.

- Harmony, T Fernandez, T Gersenowies, J Galan, L Fernandez-Bouzas, A Aubert, E Diaz-Comas, L. Specific EEG frequencies signal general common cognitive processes as well as specific task processes in man. *Int. J. Psychophysiol.*, 2004: Aug, 53(3), 207-16.
- Haueisen, J Schack, B Meier, T Curio, G Okada, Y. Multiplicity in the high-frequency signals during the short-latency somatosensory evoked cortical activity in humans. *Clin. Neurophysiol.*, 2001: Jul, 112(7), 1316-25.
- Jimenez, C Biscay, R Montoto, O. Modeling the electroencephalogram by means of spatial spline smoothing and temporal autoregression. *Biol. Cybern.*, 1995: 72(3), 249-59.
- John, E R John, B Z Corning, W C Easton, P Brown, D Ahn, H John, M Harmony, T Prichep, L S Thatcher, R Kaye, H Valdes, P Schwartz, E. Neurometrics: Numerical taxonomy identifies different profiles of brain functions within groups of behaviorally similar people. *Science*, 1977: Dec, 196, 1383-1410.
- John, E R Ahn, H Prichep, L Trepetin, M Brown, D Kaye, H. Neurometrics: Developmental equations for the electroencephalogram. *Science*, 1980: Dec, 210(4475), 1255-8.
- John, E R Prichep, L S Alper, K R Mas, F G Cancro, R Easton, P Sverdllov, L. Quantitative electrophysiological characteristics and subtyping of schizophrenia. *Biol. Psychiatry*, 1994: Dec, 36(12), 801-26.
- Jones, L K. On a conjecture of Huber concerning the convergence of projection pursuit regression. *The Annals of Statistics*, 1987: 15(2), 800-82.
- Koenig, T Lehmann, D Merlo, M C Kochi, K Hell, D Koukkou, M. A Deviant EEG Brain Microstate in Acute, Neuroleptic-Naive Schizophrenics at Rest. *European Archives of Psychiatry and Clinical Neuroscience*, 1999: 249(4), 205-11.
- Koenig, T Lehmann, D Saito, N Kuginuki, T Kinoshita, T Koukkou, M.. Decreased functional connectivity of EEG theta-frequency activity in first-episode, neuroleptic-naive patients with schizophrenia: preliminary results. *Schizophr. Res.*, 2001: May, 50(1-2), 55-60.
- Koenig, T Marti-Lopez, F Valdes-Sosa, P. Topographic time-frequency decomposition of the EEG. *Neuroimage*, 2001: Aug, 14(2), 383-90.
- Koenig, T Prichep, L Lehmann, D Valdes-Sosa, P Braeker, E Kleinlogel, H Isenhardt, R John, E R. Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *Neuroimage*, 2002: May, 16(1), 41-8.
- Koenig, T Melie-Garcia, L Hubl, D Valdes-Sosa, P Dierks, T. Multichannel time frequency analysis of spontaneous EEG and fMRI correlates. *Proceedings of the Mediterranean Conference on Medical and Biological Engineering (Medicon)*, 2004.
- Koenig, T Prichep, L Dierks, T Hubl, D Wahlund, L O John, E R Jelic, V. Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment. *Neurobiol. Aging*, 2005: Feb, 26(2), 165-71.
- Koenig, T Studer, D Hubl, D Melie, L Strik, W . Brain connectivity at different time scales measured in EEG. In *press in the Philosophical Transactions of the Royal Society, Series B*.
- Lehmann, D Ozaki, H Pal, I. EG alpha map series: Brain microstates by space-oriented adaptive segmentation. *Electroencephalogr. Clin. Neurophysiol.*, 1987: 67, 271-88.
- Lehmann, D Michel, C M. Intracerebral dipole source localization for FFT power maps. *Electroencephalogr. Clin. Neurophysiol.*, 1990: Sep, 76(3), 271-6.
- Lehmann, D Strik, W K Henggeler, B Koenig, T Koukkou, M. Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *Int. J. Psychophysiol.*, 1998: June, 29(1), 1-11,
- Lehmann, D Faber, P L Galderisi, S Herrmann, W M Kinoshita, T Koukkou, M Mucci, A Pascual-Marqui, R D Saito, N Wackermann, J Winterer, G Koenig, T. EG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. *Psychiatry Res.*, 2005: Feb, 138(2), 141-56.
- Leviatan, D Temlyakov, V. Simultaneous approximation by greedy algorithms. *Tech. Rep. 0302, IMI, Dept of Mathematics, University of South Carolina*, 2003.
- Lopes da Silva, F. Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol*, 1991: Aug, 79(2), 81-93.
- Lopes da Silva, F. Functional localization of brain sources using EEG and/or MEG data: volume conductor and source models. *Magn. Reson. Imaging*, 2004: Dec, 22(10), 1533-8.
- Lutoborski, A Temlyakov, A. Vector Greedy Algorithms. *J. Complexity*, 2003: Aug, 19(4), 458-73.
- Mallat, S Zhang, Z. Matching pursuit with time frequency dictionaries. *IEEE Trans. Signal Process*, 1993: 41(12), 3397-415.
- Marti-Lopez, F Koenig, T. Approximating method of frames. *Digital Signal Processing*, 2003: 13, 519-29.
- Miwakeichi, F Martinez-Montes, E, Valdes-Sosa, P A Nishiyama, N Mizuhara, H Yamaguchi, Y. Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *Neuroimage*, 2004: Jul, 22(3), 1035-45.
- Mohar, B. *The Laplacian Spectrum of Graphs. Graph Theory, Combinatorics, and Applications Wiley*, 1991: 2, 871-98, Alavi, Y and Chartrand, Y and Oellermann, O R and Schwenk, A J, editors.
- Mueller, T J, Koenig, T Wackermann, J Kalus, P Fallgatter, A Strik, W Lehmann, D. Subsecond changes of global brain state in illusory multistable motion perception. *J. Neural Transm.*, 2004: Aug.
- Pascual-Marqui, R D Gonzalez-Andino, S L Valdes-Sosa, P A Biscay-Lirio, R. Current source density estimation and interpolation based on spherical harmonic fourier expansion. *Int. J. Neurosci.*, 1988: Dec, 43(3-4), 237-49.

- Pascual-Marqui, R D Michel, C M Lehmann, D. Segmentation of brain electrical activity into microstates: Model estimation and validation. *IEEE Trans. Biomed. Eng.*, 1995: Jul, 42(7), 658-65.
- Prichep, L S John, E R Ferris, S H Reisberg, B Almas, M Alper, K Cancro, R. Quantitative EEG correlates of cognitive deterioration in the elderly. *Neurobiol. Aging*, 1994: Jan-Feb, 15(1), 85-90.
- Valdes, P Bosch, J Grave, R Hernández, J Riera, J Pascual, R Biscay, R. Frequency domain models of the EEG. *Brain Topogr.*, 1992, 4(4), 309-19.
- Pon Varma, S Papandreou-Suppappola, A Suppappola, S B. Detecting faults in structures using time-frequency techniques. *International Conference on Acoustics, Speech, and Signal Processing*, (Salt Lake City, Utah), 2001: May, 6, 3593-96.
- Saletu, B Anderer, P Kinsperger, K Grunberger, J. Topographic brain mapping of EEG in neuropsychopharmacology—Part II. Clinical applications (pharmaco EEG imaging). *Methods Find. Exp. Clin. Pharmacol. Int. J. Psychophysiol.*, 1987: Jun, 9(6), 385-408.
- Rousseeuw, P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 1987: 20, 53-65.
- Strik, W K Chiaramonti, R Muscas, G C Paganini, M Mueller, T J Fallgatter, A J Versari, A Zappoli, R. Decreased EEG microstate duration and anteriorisation of the brain electrical fields in mild and moderate dementia of the Alzheimer type. *Methods Find. Exp. Clin. Pharmacol. Psychiatry Res.*, 1997: Oct, 75(3), 183-191.
- Wackermann, J Lehmann, D Michel, C M Strik, W K. Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *Int. J. Psychophysiol.*, 1993: May, 14(3), 269-83.
- Zygierewicz, J Blinowska, K J Durka, P J Szelenberger, W Niemcewicz, S Androsiuk, W. High resolution study of sleep spindles. *Clin. Neurophysiol.*, 1999: Dec, 110(12), 2136-47.

Figures

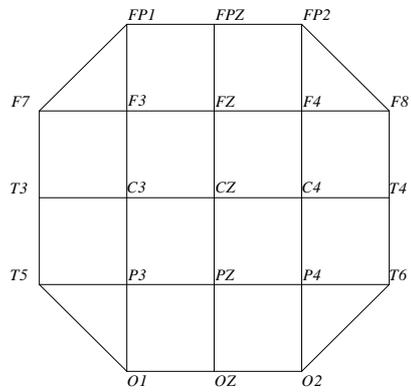


Figure 1: Vertices of the graph are given by the 21 electrodes distributed over the scalp, where the neighboring ones are connected by edges.

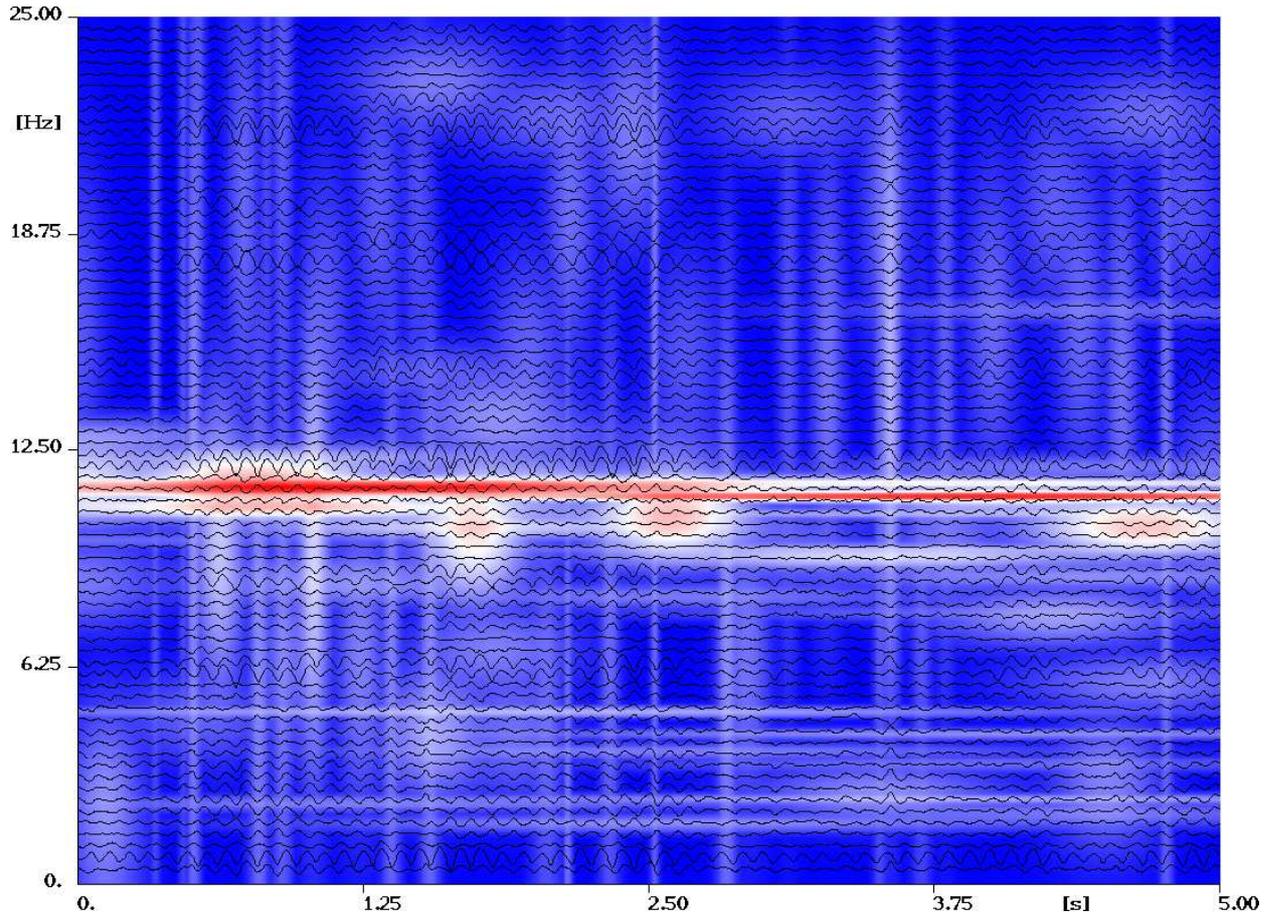


Figure 2: Rest(R): Intensity plot in the time-frequency plane of the scaled multichannel energy distribution $E_q \mathbf{f}_M^{(R)}(t, w)$ of the DMMP decomposition $\mathbf{f}_M^{(R)}$ of the resting EEG $\mathbf{f}^{(R)}$. Blue areas indicate low, red areas indicate high activity. The plot is overlaid by the original EEG traces $\mathbf{f}^{(R)}$. The multichannel energy distribution is scaled with $q = 0.3$. Explained energy of the DMMP decomposition: 72.2%.

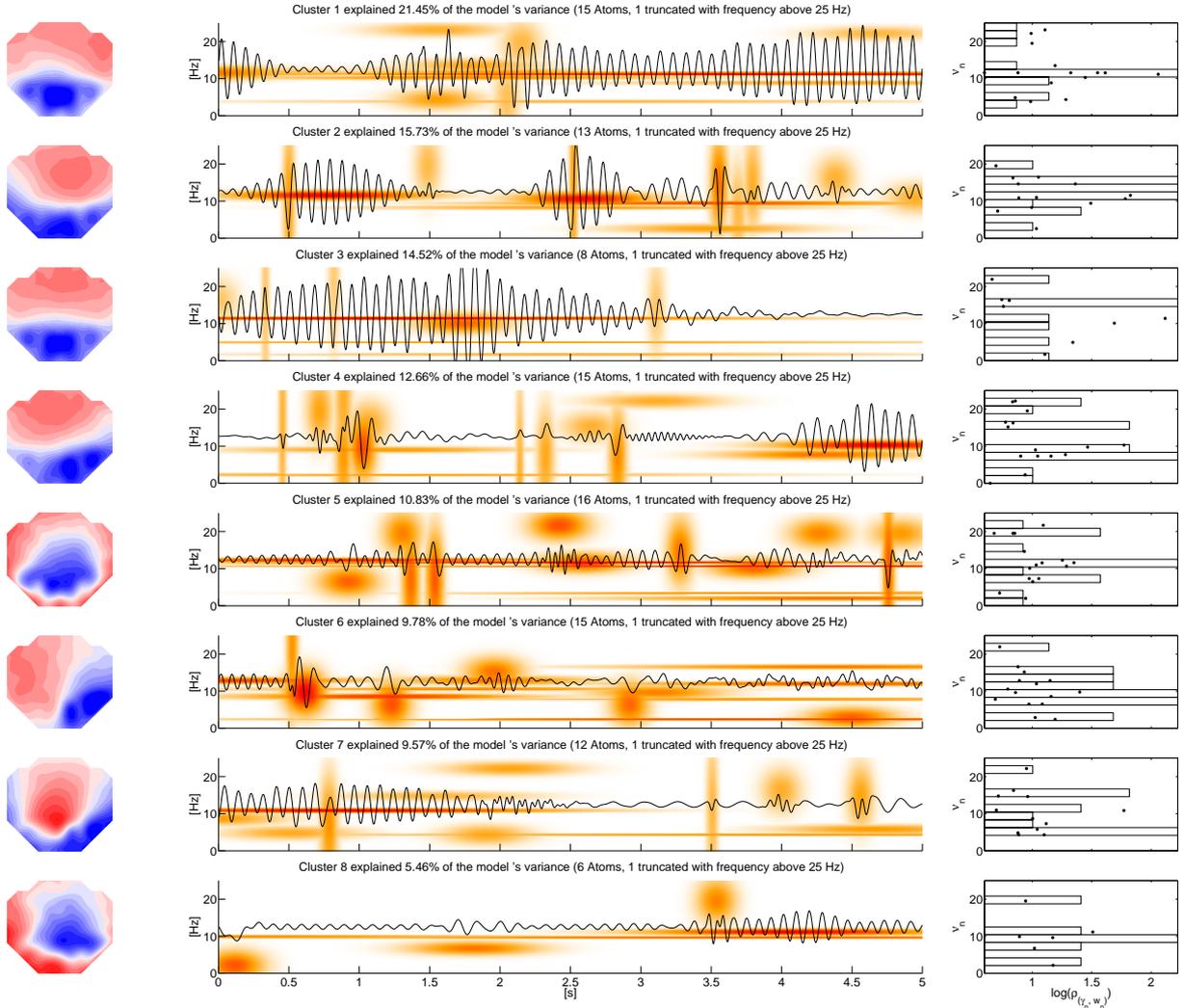


Figure 3: Rest(R): The middle panels show the intensity plots in the time-frequency plane of the scaled energy distributions $E_q a_i^{(R)}(t, w)$ of the representative waveforms $a_i^{(R)}(t) = \sum_{g_{\gamma_n} \in \mathcal{A}_i} g_{(\gamma_n, w_n)}(t) \rho_{(\gamma_n, w_n)}$ of the sparse topographic decomposition $\mathbf{h}_p^{(R)}$ of the resting EEG $\mathbf{f}^{(R)}$ with $i = 1, \dots, 4$. White areas indicate low, red values indicate high values of $E_q a_i^{(R)}(t, w)$. The energy distributions are scaled with $q = 0.2$. The plots are overlaid by the corresponding representative waveforms $a_i^{(R)}$. The left panels show the corresponding representative topographies $\tau_i^{(R)}$ shown as isocontour potential maps (left ear is left, nose is up), red and blue areas indicate opposite polarities. The right panels show the amplitude-frequency spectra of the corresponding representative waveforms $a_i^{(R)}(t)$, that indicate the log-scaled absolute amplitude $\log(|\rho_{(\gamma_n, w_n)}|)$ against the frequency $\nu_n = \lambda v_n / (2\pi)$ of each frequency component $g_{(\gamma_n, w_n)}$ of $a_i^{(R)}$ with $\gamma_n = (s_n, u_n, v_n)$ and sampling rate $\lambda = 250\text{Hz}$. These plots are overlaid by the corresponding frequency distributions histogram showing the number of log-scaled absolute amplitudes falling into each of several ranges of frequency values ν_n . The sparse topographic decomposition build of those waveforms and topographies explains 85.4% of the models variance. Overall silhouette value of the cluster structure: 0.4429.

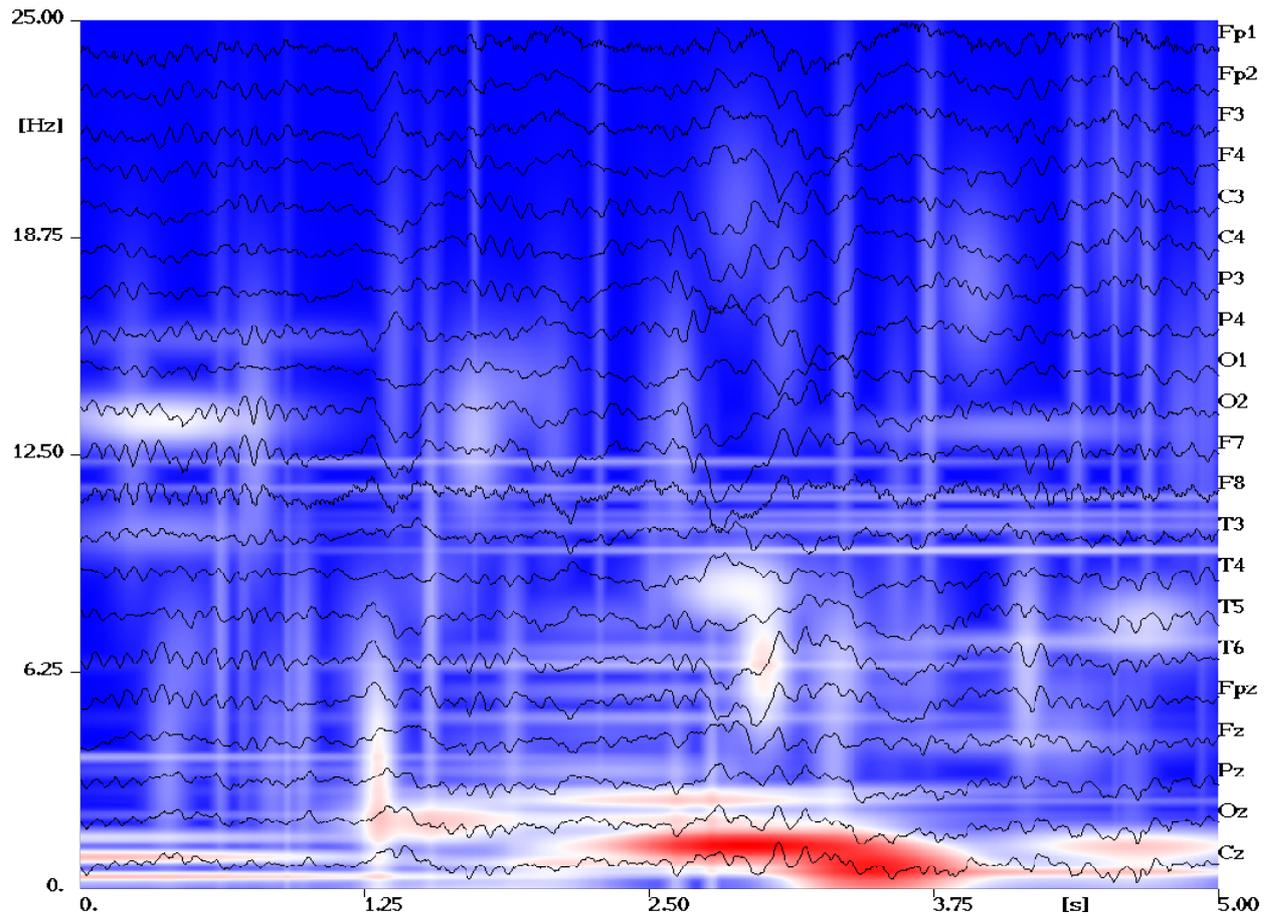


Figure 4: Sleep(S2): Intensity plot in the time-frequency plane of the scaled multichannel energy distribution of the DMMP decomposition of the sleep stage 2 EEG $f^{(S2)}$. Format identical to fig. 2. Explained energy of the DMMP decomposition: 73.7%.

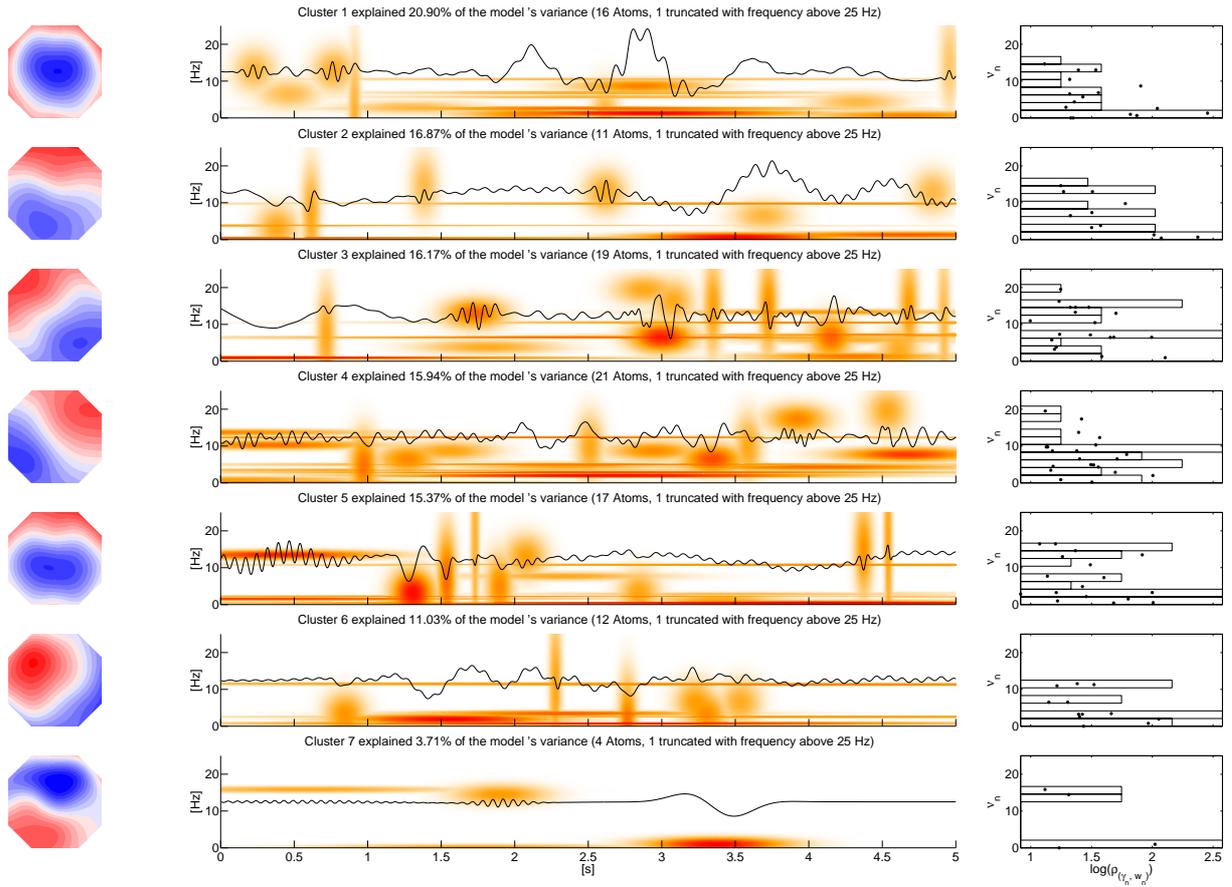


Figure 5: Sleep(S2): Sparse topographic decomposition of sleep stage 2 EEG $f^{(S2)}$. Layout as described in the legend of fig. 3. The sparse topographic decomposition explains 82.0% of the models variance. Overall silhouette value of the cluster structure: 0.5087.

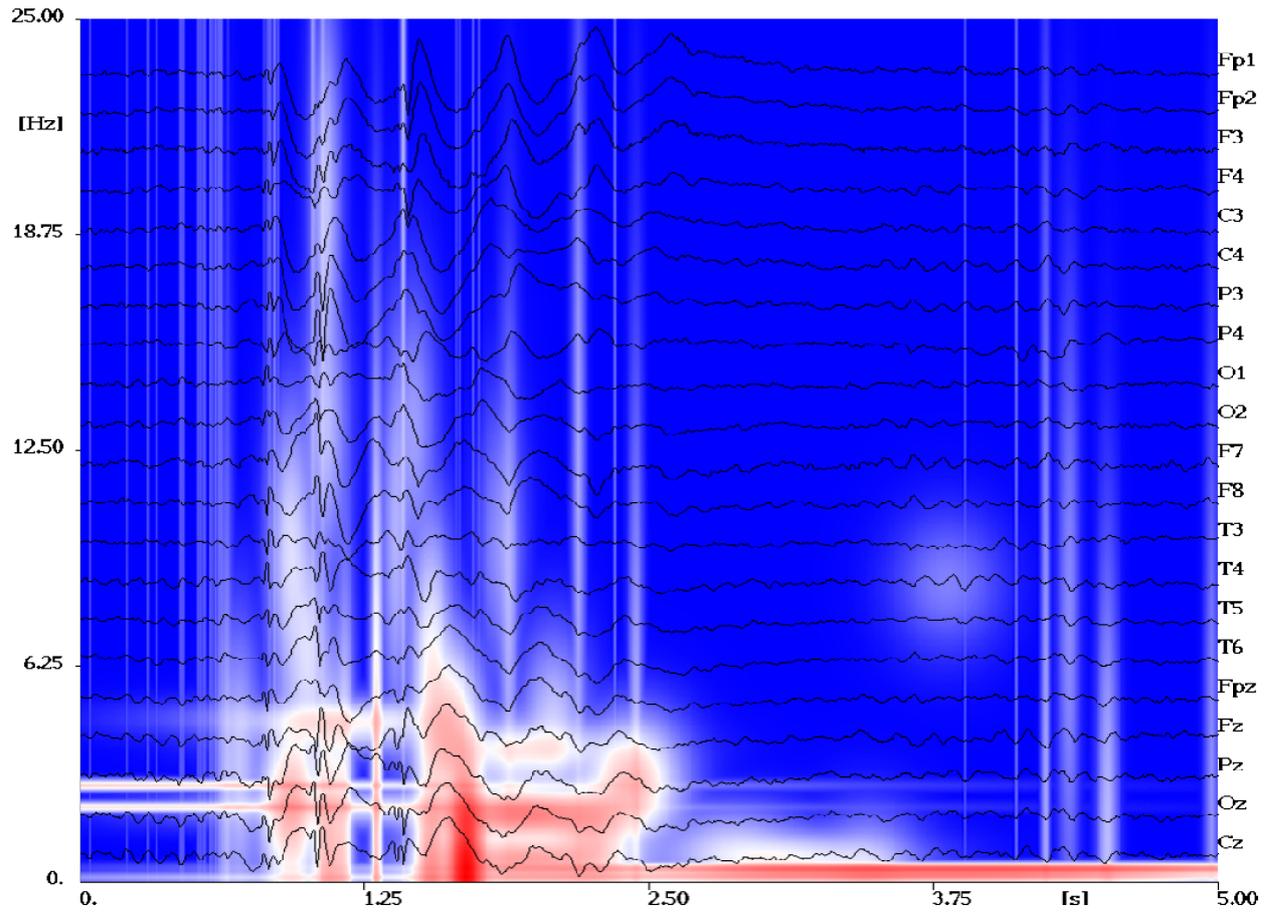


Figure 6: Spike/Waves(SW): Intensity plot in the time-frequency plane of the scaled multichannel energy distribution of the DMMP decomposition of the spike and wave EEG $f^{(SW)}$. Layout as in fig. 2. Explained energy of the DMMP decomposition: 64.5%.

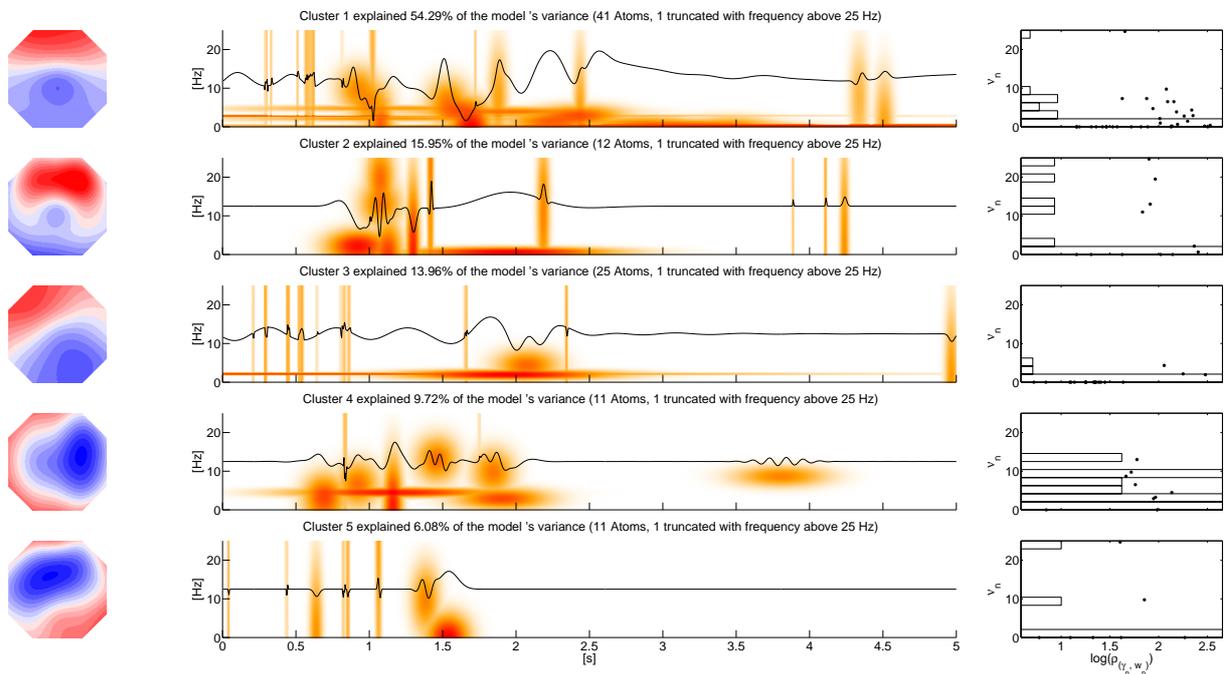


Figure 7: Spike/Waves(SW): Sparse topographic decomposition of the spike and wave EEG $f^{(SW)}$. Layout as described in the legend of fig. 3. The sparse topographic decomposition explains 77.1% of the models variance. Overall silhouette value of the cluster structure: 0.5937.