Topography-Time-Frequency Atomic Decomposition for Event-Related M/EEG Signals.

Christian Bénar, Théodore Papadopoulo, Maureen Clerc

Adaptive Time-Frequency Models for Single-Trial M/EEG Analysis

Christian Bénar[1], Maureen Clerc[2], and Théodore Papadopoulo[2]

[1] INSERM U751, La Timone, Marseille, 13006 France
[2] INRIA/ENPC/ENS Odyssée Laboratory, INRIA Sophia-Antipolis, 06902 France

Abstract. A new method is introduced for estimating single-trial magneto- or electro-encephalography (M/EEG), based on a non-linear fit of time-frequency atoms. The method can be applied for transient activity (e.g. event-related potentials) as well as for oscillatory activity (e.g. gamma bursts), and for both evoked or induced activity. In order to benefit from all the structure present in the data, the method accounts for (i) spatial structure of the data via multivariate decomposition, (ii) time-frequency structure via atomic decomposition and (iii) reproducibility across trials via a constraint on parameter dispersion. Moreover, a novel iterative method is introduced for estimating the initial time-frequency atoms used in the non-linear fit. Numerical experiments show that the method is robust to low signal-to-noise conditions, and that the introduction of the constraint on parameter dispersion significantly improves the quality of the fit.

1 Introduction

A classical method for analyzing brain electric and magnetic waves in humans consists in averaging many EEG or MEG trials, obtained in similar conditions, in order to improve the signal to noise ratio (SNR).

When the waves are sufficiently time-locked with respect to the reference time (usually the time of stimulation), the average can be performed directly in the time-domain ("event-related potentials", ERPs, in EEG and "event-related fields", ERFs, in MEG). For activity with higher time dispersion with respect to one wave period, the resulting variation of phase across trials can cause the waves to cancel out in the average signal. This is particularly relevant for high-frequency activity (above 20 Hz), where a small time delay can cause a large phase difference. This cancellation can be circumvented by averaging the power of the signal in the time-frequency or time-scale plane. Several methods have been introduced for evaluating the average increase of energy in given frequency band, whether time-locked ("evoked" energy) or not ("induced" energy) (e.g. [1]).

For both ERPs and oscillations, the averaging procedure relies on the assumption of similarity of the activity of interest across trials. However, there is often significant variability in shape and latency from one trial to another, even

when measuring responses to repetitions of the exact same stimulus. This variability can arise from habituation effects, fluctuations in the level of attention and arousal, or different response strategies.

Inter-trial variations are problematic when averaging. In particular, fluctuations in latency result in a blurring of the average signal, producing a false impression of events lasting longer than their actual duration. Moreover, comparing the amplitude of averages between conditions (for example, response to rare events versus response to frequent events), as done routinely in event-related potentials, does not permit to distinguish between an actual amplitude effect across conditions or a higher dispersion in latencies within one condition.

Nevertheless, variability can also be a source of information. For example, Jung and colleagues have demonstrated different trial-to-trial behaviors of event-related waves measured on ICA components, some being phased-locked to the stimulus and others to the subjects' responses [2]. Recently, studies have demonstrated correlations between the fluctuations of energy in the gamma band and the phase of theta oscillations, indicating an interaction between activities at different frequencies [3]. Variability between trials can also be used to measure relations between brain regions, or relations between different modalities recorded simultaneously, such as EEG and functional MRI [4].

Many different classes of methods have been introduced for detecting trial-to-trial variations in event-related potentials (ERPs) or event-related fields (ERFs). These include spline models, autoregressive models, template matching, neural networks and other multivariate classification methods, Bayesian analysis, non-linear analysis.

A promising class of methods is based on time-scale (i.e. wavelet) analysis and on time-frequency analysis. These methods permit to adapt the analyzing functions to the signals of interest, based on prior shape information. Quian Quiroga and colleagues have proposed to estimate single-trial ERPs by building fixed wavelet filters based on the average signal [5]. Estimating the wavelet filters from the average signal is subject to the difficulties mentioned above, i.e. that the average may not be fully representative of the single trials. The wavelet basis used is orthogonal, which allows fast decomposition and reconstruction of filtered signals, but is not translation-invariant and therefore ill-suited to represent latency jitters. Finally, the inter-trial fluctuations in latency or scale may necessitate the selection of a large number of coefficients in the time-scale plane, independently of the trial under consideration, which can be sensitive to noise. Some of these drawbacks have been addressed by Benkherrat and colleagues [6], who consider the average energy of the wavelet transforms of single trials instead of the transform of the average, and by [7] who use translation-invariant wavelet transforms. Statistical issues in the detection of activity have been studied in the work of Durka and colleagues, who use the matching pursuit approach and bootstrap resampling [8].

Up to now, most effort in single-trial analysis has been directed to slow-varying ERPs (in the range of 1-20Hz). Less attention has been given to the estimation of single-trial oscillatory activity, for example gamma activity around 40 Hz.

Such oscillatory activity has been hypothesized to play a major role in the communication between different brain areas, for example in feature binding or matching stimuli to a target stored in memory (reviews in [9], [10]).

Another important aspect is the use of the full range of structure contained in the data, i.e. not only according to its the temporal structure, but also to its spatial structure (obtained directly across sensors or via estimated brain sources) [11], [2] and its time-frequency signature.

The new method presented here is designed to track fluctuations in brain electromagnetic activity, for any given set of frequency bands. Our method has several original features. First, we introduce a methodology for defining a reference set of Gabor time-frequency atoms, or "template", that is capable of modelling both low frequency event-related potentials and high frequency oscillations. Second, the template is deformed across trials with nonlinear optimization, which permits to follow accurately the fluctuations of the actual signal and obtain a sparse final representation of the data. Third, the deformations for each trial are constrained using information arising from all the trials, which increases the robustness of the fit even for low SNRs.

Our method is closely related to Bayesian modelling, and can be in fact seen as a Maximum a Posteriori (MAP) approach, but we are concentrating here on the nonlinear minimization aspect. Our model could be be extended in order to incorporate richer *a priori* information, or trial-to-trial variations in spatial patterns (i.e., topographies) of the signals.

The principles of the method and the mathematical framework are introduced first, followed by a description of the validation procedure and its results.

2 General Principles and Mathematical Framework

2.1 Decomposition into realizations of deformable templates

We suppose that the signal contains classes of EEG or MEG activity that have a similar spatio-temporal structure. Such a structure will be approximated by a parametric template with parameters par. The signal $S(t,\theta)$, as a function of time t and spatial position θ, is modelled as the sum of a model signal $\hat{S}$ and a noise term:

$$S(t,\theta) = \hat{S}(t,\theta|par) + E(t,\theta)\ . \tag{1}$$

The model signal $\hat{S}$ is in turn composed of K repetitions (or trials) of N parametric templates

$$\hat{S}(t,\theta|par) = \sum_{n=1}^{N}\sum_{k=1}^{K} \beta_{n,k} T_n(t - t_{n,k}, \theta - \theta_{n,k}|par)\ . \tag{2}$$

In the above model, inter-trial variability is apparent in the form of a spatio-temporal shift by $(t_{n,k}, \theta_{n,k})$, and an amplitude modulation by $\beta_{n,k}$. Additional sources of inter-trial variability can be incorporated in the additional parameters par. The expected value of the noise $E(t,\theta)$ is assumed to be zero.

We further assume that a given template T_n, $n \in \{1...N\}$, can be decomposed as the product of a temporal pattern A_n and a spatial pattern (or topography) M_n (see Section 3.1 for more details).

$$T_n(t,\theta) = A_n(t)M_n(\theta)\ . \tag{3}$$

In the sequel, a simplified version of this model will be used, with $N = 1$, i.e. only one template, making the template index n no longer necessary. Moreover, the topography will be assumed stable across trials ($\theta_k = 0$), simplifying the model to

$$\hat{S}(t,\theta|par) = \sum_{k=1}^{K} \beta_k A(t - t_k|par)\, M(\theta)\ . \tag{4}$$

2.2 Time-Frequency Analysis

As we are interested in modelling signals across a wide range of frequencies (e.g., event-related potentials or induced oscillations), we propose to use time-frequency descriptors. With this in view, the Gabor atoms offer the best trade-off between time and frequency compactness. The temporal template is modeled as the linear combination of P Gabor atoms, in the real domain:

$$A(t) = \sum_{p=1}^{P} \beta_{k,p} e^{-\frac{1}{2\sigma_{k,p}^2}(t-d_{k,p})^2} cos(\omega_{k,p}(t - d_{k,p})). \tag{5}$$

The number of atoms P is fixed (see section 3.1); the template temporal shift $d_{k,p}$ is equal to the onset of the stimulation (i.e. the initial time for each trial). The trial-specific parameters to be estimated are, for each atom, its latency, amplitude, width and modulating frequency: $par = \{d_{k,p}, \beta_{k,p}, \sigma_{k,p}, \omega_{k,p}\}$.

The present study aims to model evoked potentials, hence the ratio between the width and the frequency is kept fixed (similar to Morlet wavelets). For bursts of oscillation, this constraint can be relaxed, in order to also adapt the number of oscillations to each trial.

2.3 Cost Function

The model defined by (1), (4) and (5) can be fitted to the data by the non-linear minimization of a cost function $C(par)$ composed of two parts. The first part $C_1(par)$ maximizes the fit to the data. The second part $C_2(par)$ minimizes the dispersion of the parameters around their mean and avoids over-fitting the noise. This can be formulated as:

$$C_1(par) = \sum_t \left[S(t) - \hat{S}(t|par)\right]^2, C_2(par) = \sum_{k,j,p} \left[\frac{par_{j,k}^p - \overline{par_j^p}}{std(par_j^p)}\right]^2\ , \tag{6}$$

with : t index of time samples, j index on parameters, p index on time-frequency atoms; $\overline{par_j^p}$ and $std(par_j^p)$ the sought values for the parameter mean and standard deviation across trials, respectively.

The complete cost function is

$$C(par) = C_1(par)/var(E) + \lambda\, C_2(par)\ , \tag{7}$$

where $var(E)$ is the variance of the background noise, and λ a regularization hyperparameter which balances the dispersion constraint.

The minimization is performed using the Levenberg-Marquardt algorithm, which is suited for minimizing cost functions consisting of sum of squares. We use the implementation provided in the immoptibox toolbox[3] [12]. In a first step, some quantities need to be estimated:

- the projection of the spatio-temporal data $S(t, \theta)$ into a single time-topography template,
- the initial values of the parameters $par^{init} = \{\beta_{k,p}, \sigma_{k,p}, d_{k,p}, \omega_{k,p}\}^{init}$,
- the hyperparameters $\lambda, var(E), \overline{par_j^p}$ and $std(par_j^p)$.

The next section presents strategies for estimating the time-topography signal projection, the hyperparameters and the initial values of the parameters.

3 Initialization of the Parameters

3.1 Time-Topography Decomposition

Many methods are available for decomposing spatio-temporal data (sensors $\times$ time) into fixed spatial topographies and corresponding time courses, as in (3). Examples of such methods include dipolar source localization with fixed dipoles, principal component analysis (PCA), independent component analysis (ICA), and PARAFAC (e.g. when including the frequency dimension). Such decompositions rely on the assumptions that (i) the spatial and temporal aspects of the activity of interest are independent of one another within a trial and (ii) that there is no spatial variability from one trial to the other. Albeit these are strong assumptions, they have however proven useful in a variety of situations. Moreover, they can be seen as a way of initializing spatially adaptive models such as (2).

The Singular Value Decomposition allows to project the data into a lower dimensional space that is tractable. In particular, if the data can be assumed to originate from a single source, the SVD allows to recover its spatial topography on the sensors. In a more realistic situation with several sources, the SVD can be performed as a dimension reduction preprocessing, prior to ICA. In order to select automatically the component that captures the temporal activity of the source, we assess the reproducibility of the component across trials. A score for each component is computed at each time point by dividing the mean energy by the standard deviation of the component; we select the component with highest score.

[3] available at www2.imm.dtu.dk/ hbn/immoptibox/, in the Matlab (Mathworks, Natick, MA, USA) environment.

3.2 Initial Time-Frequency Atoms

The parameter initialization for the time-frequency atoms is performed via an iterative wavelet analysis. For this, Morlet wavelets are introduced, which are Gabor atoms with a fixed number of oscillations, defined in the frequency domain by:

$$\Psi_s^\xi(\omega) = \frac{\sqrt{s}}{(4\pi)^{\frac{1}{4}}} e^{-\frac{1}{2}(s\omega-\xi)^2} \tag{8}$$

The shifting frequency ξ determines the number of oscillations of the wavelet over its time-support. For a given shifting frequency ξ, the scale parameter s stretches or compresses the wavelet, controlling simultaneously the time-support (proportional to s) and the frequency (equal to $\xi/(2\pi s)$) at which the wavelet oscillates. The optimal number of oscillations depends on the type of activity under examination, smaller for evoked potentials/fields and larger for oscillations. Therefore a set of shifting frequencies is considered : $\xi \in \{1, 1.5, 3, 5, 9\}$.

An initial time-frequency decomposition $W_k^0(t,\omega)$ is obtained by projecting each trial $S_k(t)$ on a set of analyzing functions:

$$W_k^0(t,s,\xi) = \int S_k(t')\overline{\psi}_s^\xi(t-t')dt' \ . \tag{9}$$

In the above relation, the time index t belongs to a time window $[t_k, t_k + L]$ containing the kth trial (t_k defined in Section 2.2), and $\overline{\psi}$ denotes the complex conjugate of ψ. The energy for each ξ is then averaged across trials

$$TF^0(t,s,\xi) = \frac{1}{K}\sum_k |W_k^0(t-t_k,s,\xi)|^2. \tag{10}$$

The average energy $TF^0(t,s,\xi)$ could be normalized by the average power value in a baseline part of the window [1], [13]. One can also consider dividing by the standard deviation of the noise power estimated at the point of interest [6]. Such normalizations allow to be more robust to noise and to detect high-frequency activity that is hidden in the original signal due to the 1/f behavior of EEG and MEG.

The iterative parameter selection proceeds à la matching pursuit [14], iteratively subtracting from the single-trial data its projection on the atom $\psi_{s^0}^{\xi^0}$ whose parameters maximize the average energy:

$$(t^0, s^0, \xi^0) = argmax(TF^0(t,s,\xi))$$

This leads to a new set of trials $S_k^1(t)$ on which the above-described procedure can again be applied. The number of iterations P is determined by hand; however, one could consider using a quantitative criterion, based for example on the energy of the residuals after subtraction. The main difference with matching pursuit resides in the use of the data over a collection of trials, thereby taking advantage of the information arising from reproducibility across trials.

3.3 Constraint on Parameters Dispersion

The hyperparameters $\overline{par_j^p}$ and $std(par_j^p)$ help to reduce the parameter dispersion across trials, in order to increase robustness to noise. We choose to estimate these parameters from the data itself, in a first step that consists in minimizing only the first part $C_1(par)$ of the cost function (6) (i.e., without the control of the parameter dispersion). We make the assumption that the fit will be successful for a majority of trials, failing only in a small proportion of outliers due to the noise. The mean and standard deviation of each parameter are then estimated with a robust MM-estimator [15].

3.4 Regularization Hyperparameters

The hyperparameters $var(E)$ and λ appearing in (7) control the relative weights given on the one hand to the quality of the fit (6), and on the other hand to the constraint on dispersion $C2$ in (6).

The variance of the noise $var(E)$ is estimated on the residuals after the first fit (i.e., using only the first part of the cost function (6)), using robust estimators as in 3.3.

The regularization parameter λ is estimated with the L-curve method. The fit is performed using (7) for a series of λ ranging from 10^{-4} to 10^4, with a higher sampling around $\lambda = 1$. The minimization procedure results in optimal values C_1^{opt} and C_2^{opt}, for each λ. The L-curve method consists in finding the inflection point in the curve $C_2^{opt}(\lambda) = f(C_1^{opt}(\lambda))$. This point is estimated as the minimum of the cost function

$$c(\lambda) = \frac{C_1^{opt}(\lambda)}{std(C_1^{opt})} + \frac{C_2^{opt}(\lambda)}{std(C_2^{opt})} \tag{11}$$

with std standard deviation across values of λ.

4 Methodology of Evaluation

First of all, simulations were performed to evaluate the robustness of the method to different levels of noise, with a particular emphasis on the contribution of the regularization term C2 in (6). Secondly, the method was applied to real data originating from an oddball experiment [4].

4.1 Simulated data

A series of 50 event-related potentials (ERPs) were computed, corresponding to a central radial source within a sphere, with 83 electrodes (10/10 system), at a sampling frequency of 1 kHz. The potentials consisted of a sum of Gaussian-shaped waves (N100-like and P300-like) and of high frequency oscillations (range 38-42 Hz) with a Gaussian envelope. The latency, amplitude, width and frequency parameters of the waves and oscillations were pseudo-random numbers with a

truncated distribution, with values lying three standard deviations around the mean. The latency parameter was assigned a skewed distribution, similar to that of realistic reaction times. See figure 1 for an overview of the simulated signals.

Stationary background noise was simulated, with spatial and temporal correlations ressembling real EEG data. This noise was obtained by distributing dipoles uniformly within a sphere, with a random amplitude following a normal distribution [16]. Each channel was then filtered and scaled with an autoregressive filter, whose parameters had been fitted on real EEG data. The resulting correlated noise was added to the simulated ERPs, with a multiplicative factor corresponding to different values of the signal-to-noise ratio: SNR $\in \{0, 0.1, 0.5, 1, 3, 5, 10, 100, 10^6\}$. The SNR was computed as the ratio of the sums of squares across all channels. For the high frequency oscillations, the data and noise were first filtered with a 35-45 Hz bandpass filter before computing the scaling factor.

For each SNR, a singular value decomposition (SVD) was applied to the simulated data, and the time-course with the highest reproducibility across trials was retained. Reproducibility was assessed by computing the average energy across trials. The method was then applied on the time-course, across trials, as described in the previous sections.

4.2 Real data

We used an EEG dataset recorded within the MRI scanner, from an simultaneous EEG-fMRI oddball protocol presented in detail in [4]. In summary, the stimulation consisted in pure tones, with low frequency tones presented frequently, and high frequency tones presented rarely. The subject was asked to respond to the rare tones by a button press. This protocol is known to provoke a stereotyped positive response on the EEG around 300 ms, called P300.

5 Results

The data obtained after spatial filtering by SVD are presented in figure 1, along with the iterative time-frequency decomposition, for a SNR of 100 and a SNR of 0.5. This demonstrates that the iterative decomposition is robust even for low SNR. Figure 2 presents raster plots of the data across trials, for a SNR of 1, for the raw data (left plot) and for the estimated atoms (right plot). For display purposes, the data is sorted with respect to the simulated latency of the second wave, an information that was not used during the fit. The algorithm was able to recover the simulated waves. For each SNR, we have computed the correlation at each trial between the fitted data on a SNR of 10^6 and the fitted data at the SNR under consideration, with and without the constraint on parameter dispersion (figure 3). Without the constraint on the parameters, the quality of the fit deteriorates rapidly with decreasing SNR. The constraint allows the fit to maintain a good quality up to a SNR of 0.5.

The outcome of the fit on real data (auditory oddball in simultaneous EEG-fMRI) is presented in figure 4. The application of the proposed method results in a very significant denoising of this data, which had extremely low SNR due to the simultaneous fMRI scanning.

6 Discussion

A new method has been presented for estimating single-trial EEG or MEG activity, based on a non-linear fit. The method can be used for both transient (e.g., event-related potentials) and oscillatory (e.g., gamma bursts) activity, and for both evoked or induced activity.

The method aims to use all the structure present in the data. First, it takes advantage of the spatial structure via multivariate decomposition. Second, it uses the time-frequency structure via atomic decomposition. Third, reproducibility across trials is imposed via a constraint on parameter dispersion.

A novel method was introduced for estimating the initial time-frequency atoms used in the non-linear fit, which runs in an iterative manner. This method is similar to matching pursuit [14], with the originality of operating across trials. We use Gabor atoms, which have good time-frequency properties, but it could be interesting to learn the atom structure directly from the data. We also present a novel non-linear minimization framework, contrary to previous approaches using wavelet transforms [6], [7]. This leads to a more parcimonious (sparse) description of the data, in particular for large jitters in the parameters (for example latency jitters), which could result in higher robustness to noise.

The same data was used to estimate the hyper-parameters (mean and standard deviation of the parameters, used in the dispersion constraint) and the actual parameters of the time-frequency atoms. Such an approach must be performed with caution, since estimating (the hyper-parameters) and detecting (the parameters of the atoms) on the exact same dataset could lead to a bias of the results. However the risk of bias is low in our case, as the parameters for only *one* trial are being estimated, based on hyper-parameters estimated on *all* trials; i.e. the influence of the trial under investigation on the definition of the hyper-parameters is low. Moreover, the robust estimation of the hyper-parameters further lowers the influence of a given trial: if the parameters of a trial are outliers, they are not taken into account in the computation of the hyper-parameters, and therefore do not influence the estimation.

As shown by the tests on simulated and real data, the proposed method is robust to low signal-to-noise conditions. The introduction of a constraint on parameter dispersion, estimated from the data itself, significantly improves the quality of the fit. Future work will concentrate on the estimation of high-frequency (gamma) activity and on taking into account the structure of the noise.

Acknowledgment

The authors thank the team of the fMRI center of Marseille (Jean-Luc Anton, Bruno Nazarian and Muriel Roth) for the EEG-fMRI data.

References

1. Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J.: Stimulus specificity of phase-locked and non-phase-locked 40 hz visual responses in human. J. Neurosci. **16**(13) (July 1996) 4240–4249
2. Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.: Analysis and visualization of single-trial event-related potentials. Human Brain Mapping **14** (2001) 166–185
3. Canolty, R., Edwards, E., Dalal, S., Soltani, M., Nagarajan, S., Kirsch, H., Berger, M., Barbaro, N., Knight, R.: High gamma power is phase-locked to theta oscillations in human neocortex. **313**(5793) (September 2006) 1626–1628
4. Bénar, C., Schön, D., Grimault, S., Nazarian, B., Burle, B., Roth, M., Badier, J., Marquis, P., Liegeois-Chauvel, C., Anton, J.: Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI. Human Brain Mapping **in press**
5. Quian Quiroga, R., Garcia, H.: Single-trial event-related potentials with wavelet denoising. Clin. Neurophysiol. **114**(2) (2003) 376–390
6. Benkherrat, M., Burle, B., Allain, S., Hasbroucq, T., Vidal, F.: Individual evoked potential extraction by multiresolution wavelets decomposition. In: Proceedings EUROCON2005. (2005)
7. Wang, Z., Maier, A., Leopold, D., Logothetis, N., Liang, H.: Single-trial evoked potential estimation using wavelets. Comput. Biol. Med. **in press**
8. Durka, P., Zygierewicz, J., Klekowicz, H., Ginter, J., Blinowska, K.: On the statistical significance of event-related eeg desynchronization and synchronization in the time-frequency plane. IEEE Trans. Biomed. Eng. **51**(7) (July 2004) 1167–1175
9. Kahana, M.: The cognitive correlates of human brain oscillations. J. Neurosci. **26**(6) (Feb 2006) 1669–1672
10. Fell, J., Fernandez, G., Klaver, P., Elger, C., Fries, P.: Is synchronized neuronal gamma activity relevant for selective attention? Brain Res. Rev. **42** (2003) 265–272
11. Ranta-aho, P., Koistinen, A., Ollikainen, J., Kaipio, J., Partanen, J., Karjalainen, P.: Single-trial estimation of multichannel evoked-potential measurements. IEEE Trans. Biomed. Eng. **50**(2) (2003) 189–196
12. Madsen, K., Nielsen, H., Tingleff, O.: Methods for Non-Linear Least Squares Problems, 2nd edition. IMM DTU (2004)
13. Makeig, S.: Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones. Electroencephalogr. Clin. Neurophys. **86**(4) (April 1993) 283–293
14. Mallat, S., Zhang, Z.: Matching pursuit with time-frequency dictionaries. IEEE Trans. Signal Processing **41**(12) (December 1993) 3397–3415
15. Croux, C., Dhaene, G., Hoorelbeke, D.: Robust standard errors for robust estimators. Research Report DTEW 0367, K.U.Leuven (2003)
16. de Munck, J., Vijn, P., Lopes da Silva, F.: A random dipole model for spontaneous brain activity. IEEE Trans. Biomed. Eng. **39**(8) (1992) 791 – 804

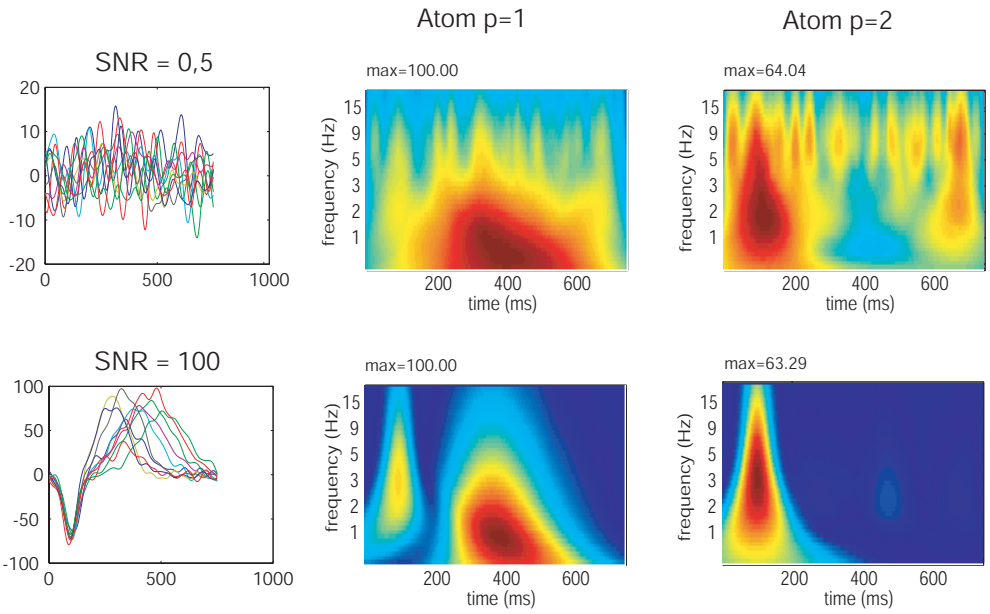

Fig. 1. Left column: Simulated data after spatial filtering, for a SNR of 0.5 (upper part) and 100 (lower part) (the SNR corresponds to the data before filtering). Middle and right column: time-frequency analysis used for the iterative definition of time-frequency atoms, for p=1 (first atom) and p=2 (second atom). The maximum value indicated is the maximum energy in the time-frequency plane with respect to the maximum energy of the first iteration (i.e, first atom).

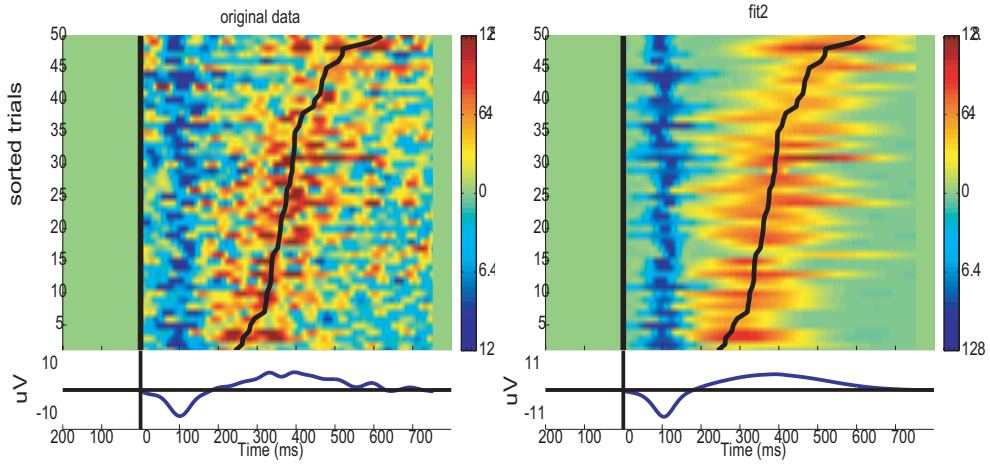

Fig. 2. Left: original data for a SNR of 1. For display purposes, the data is sorted with respect to the simulated latency of the second wave, an information that was not used during the fit. Right: fitted atoms, with constraint on parameter dispersion.

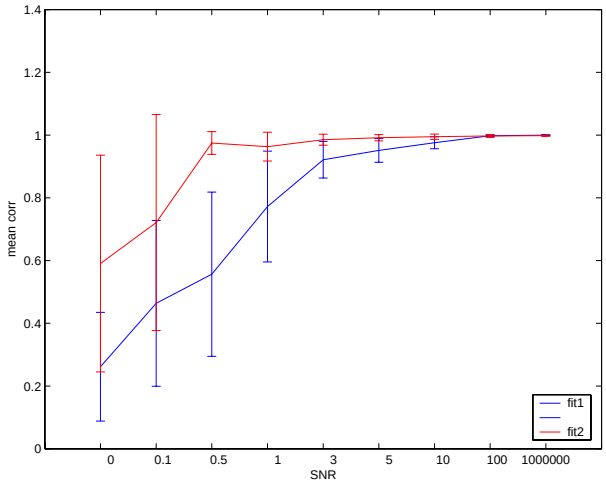

Fig. 3. Correlation between the fitted time-course at a given SNR and the fitted time-course for a SNR of 10^6, considered as the reference. The use of a constraint on parameter dispersion (fit2, in red, upper curve) improves the results compared to a fit without this constraint (fit1, in blue, lower curve).

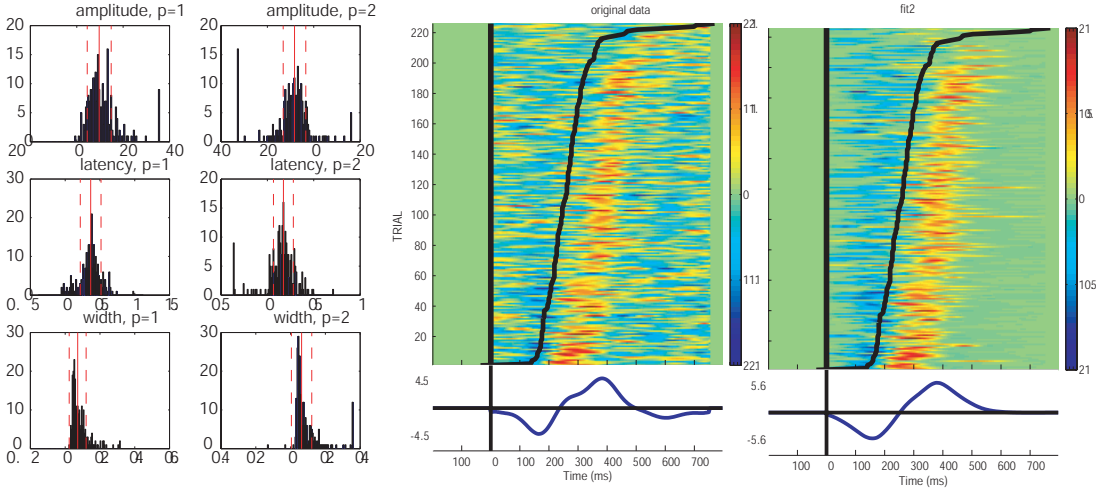

Fig. 4. Results for the oddball data. Left: histograms of parameters fitted in fit1 (no constraint on dispersion). The red (light grey) vertical bars show the mean and standard deviation of the parameters, which were not influenced by outliers (robust estimation). Middle: raster plot original data (y-axis: trials corresponding to rare events), sorted by reaction time (dark line). Right: result of the fitting procedure (fit2, with constraint on the parameters).