An exemplar-based statistical model for the dynamics of neural synchrony

Justin Dauwels*†, François Vialatte*, Theophane Weber†, and Andrzej Cichocki*

*RIKEN Brain Science Institute, Saitama, Japan
†Massachusetts Institute of Technology, Cambridge, MA.
{jdauwels,theo_w}@mit.edu
{fvialatte,cia}@brain.riken.jp

Abstract. A method is proposed to determine the similarity of a collection of time series. As a first step, one extracts events from the time series, in other words, one converts each time series into a point process ("event sequence"); next one tries to align the events from those different point processes. The better the events can be aligned, the more similar the original time series are considered to be. The proposed method is applied to predict mild cognitive impairment (MCI) from EEG and to investigate the dynamics of oscillatory-event synchrony of steady-state visually evoked potentials (SSVEP).

1 Introduction

The problem of detecting correlations between neural signals ("neural synchrony") has recently attracted much attention in the neuroscience community (e.g., [1]). For instance, it has frequently been reported that abnormalities in neural synchrony lie at the heart of brain disorders such as Alzheimer's disease (e.g., [2]).

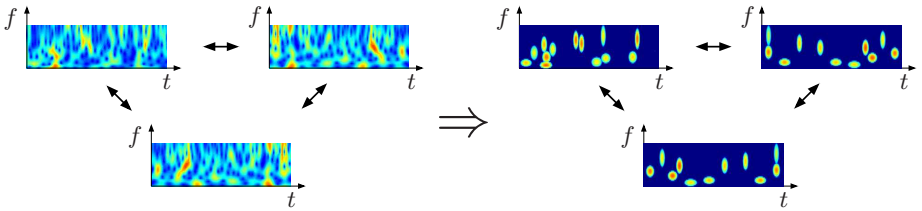
In this paper, we present a method to quantify dynamical interdependencies between a collection of time series (e.g., spike data, EEG or MRI signals; see Fig. 1). As a first step, we extract "events" from each time series, in other words, we transform each time series into a point process, i.e., a sequence of events. Next we try to align the events of each of those point processes. The better the point processes can be aligned, the more similar the point processes and hence the original time series are considered to be. In our method, the similarity may vary over time.

Our approach is inspired by the "stochastic event synchrony" measures (SES) of [3], which are also based on event alignment. However, those measures are only applicable to *pairs of signals*, in addition, they are not time-dependent. As a result, they cannot be applied to study the *dynamics* of neural synchrony. The proposed technique can deal with *collections* of signals, moreover, it treats the similarity parameters as stochastic processes; therefore, it may be used to investigate how the synchrony of a collection of neural signals *evolves* over time.

The proposed method makes use of so-called "exemplars", which are events that serve as representatives of each cluster; it is related to exemplar-based approaches for clustering such as affinity propagation [4] and the convex clustering algorithm of [5]. The exemplar-based formulation allows us to extend the pairwise similarity measures of [3] to multivariate similarity measures.

We will use our method to predict mild cognitive impairment (MCI) from EEG. This application was also investigated in [3], however, the method of [3] only considers pairwise synchrony; we will quantify the synchrony of multiple brain areas *simultaneously*, which improves the sensitivity of EEG to diagnose MCI and leads to a more detailed understanding of the abnormalities in EEG synchrony in MCI patients. As a second application, we study the dynamics of oscillatory-event synchrony of steady-state visually evoked potentials (SSVEP).

This paper is organized as follows. In the following section, we outline the exemplar-based statistical model for synchrony; in Section 3 we describe how to perform inference in that model, and we characterize the underlying combinatorial problem. As an illustration, we apply our method to detect MCI induced perturbations in EEG synchrony (Section 4), and to analyze the neural synchrony of steady-state visually evoked potentials (Section 5).



(a) Time-frequency maps. (b) Bump models.

Fig. 1. Similarity of 3 EEG signals ($N = 3$); from their time-frequency transforms (left), one extracts bump models (right), which are then aligned by the proposed algorithm.

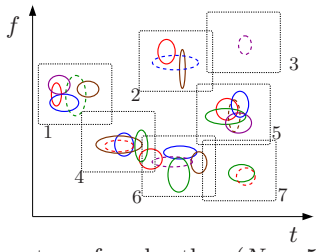


Fig. 2. Five bump models on top of each other ($N = 5$; each model has a different color); the dashed boxes indicate clusters, the dashed ellipses correspond to exemplars.

2 Exemplar-Based Statistical Model

We consider N signals $S_1, \ldots, S_N$ from which we extract point processes $X_1, \ldots, X_N$ by some method. Each point process X_i is a list of n_i points ("events") in a given multi-dimensional set $\mathcal{S} \subseteq \mathbb{R}^M$, i.e., $X_i = \{X_{i,1}, X_{i,2}, \ldots, X_{i,n_i}\}$ with $X_{i,k} \in \mathcal{S}$ for $k = 1, \ldots, n_i$ and $i = 1 \ldots N$. Let us consider the example of bump models [6] extracted from the time-frequency maps of EEG signals (see Fig. 1). The time-frequency ("wavelet") transform of each EEG signal is approximated as a sum of half-ellipsoid basis functions, referred to as "bumps" [6]; each bump is described by five parameters: time T, frequency F, width ΔT, height ΔF, and amplitude W. (In this paper, we use precisely the same procedure and same parameter settings as in [6] to extract bumps.) How similar are the N resulting bump models $X_i = ((T_{i,1}, F_{i,1}, \Delta T_{i,1}, \Delta F_{i,1}, W_{i,1}), \ldots, (T_{i,n_i}, F_{i,n_i}, \Delta T_{i,n_i}, \Delta F_{i,n_i}, W_{i,n_i}))$

(with $i = 1, 2, \ldots, N$)? Intuitively speaking, N bump models X_i may be considered well-synchronized if bumps appear in all models (or almost all) simultaneously, potentially with some slowly varying offset in time and frequency. In other words, if one overlays N more or less resembling bump models (cf. Fig. 2 with $N = 5$), and removes the potential offsets in time and frequency, bumps naturally appear in clusters that contain precisely one bump from all (or almost all) bump models. In the example of Fig. 2, cluster 1, 5 and 6 contain bumps from all 5 models X_i, cluster 2, 4 and 7 contains bumps from 3, 4, and 2 models respectively, and cluster 3 consists of a single bump.

This intuitive concept of similarity may readily be translated into a generative stochastic model. In that model, the N point processes X_i are treated as independent noisy observations of a hidden "mother" process $\tilde{X}$. An observed sequence X_i is obtained from $\tilde{X}$ by the following three-step procedure:

1. COPY: generate a copy of the mother bump model $\tilde{X}$,
2. DELETION: delete some of the copied mother bumps,
3. PERTURBATION: alter the position and shape of the remaining mother bump copies, amounting to the bump model X_i.

As a result, each sequence X_i consists of "noisy" copies of a non-empty subset of mother bumps. The point processes X_i may be considered well-synchronized if there are only few deletions (cf. Step 2) and if the bumps of X_i are well aligned with the corresponding mother bumps (cf. Step 3), apart from some slowly varying offset in time and frequency. One way to determine the synchrony of given point processes X_i is to first reconstruct the hidden mother process $\tilde{X}$, and next to align the point processes X_i with the mother process $\tilde{X}$. Inferring the mother process is a high-dimensional estimation problem, the underlying probability distribution typically has a large number of local extrema. Therefore, we will use an alternative procedure: as in [4], we will assume that each cluster contains one *identical* copy of a mother bump, the other bumps in that cluster are *noisy* copies of that mother bump. The identical copy, referred to as "exemplar", plays the role of "center" or "representative" of each cluster (see Fig. 2).

The exemplar-based formulation amounts to the following inference problem: given the point processes X_i, we need to identify the bumps that are exemplars and the ones that are noisy copies of some exemplar. Obviously, this inference problem also has potentially many locally optimal solutions, however, in contrast to the original (continuous) inference problem, we can in practice find the global optimum by integer programming (see Section 3).

We now discuss the above exemplar-based statistical model in more detail. The number M of mother events $\tilde{X}_m$ is geometrically distributed with parameter $\lambda \operatorname{vol}(S)$. Each mother event $\tilde{X}_m$ for $m = 1, \ldots, M$ is uniformly distributed in S.

The noisy copies are modeled as follows. The number C_m of copies is modeled by a prior $p(c_m|\theta^c)$, parameterized by θ^c, which in turn has a prior $p(\theta^c)$. We consider as prior for C_m a binomial distribution $\mathrm{Bi}(p_s)$ with $N - 1$ trials and probability of success p_s. We adopt a conjugate prior for the parameters p_s, i.e., the beta distribution $\mathrm{B}(\kappa, \lambda)$. Conditional on the number C_m of copies, the copies are attributed uniformly at random to other signals X_j, with the

constraints of at most one copy per signal and $j \neq i(m)$; since there are $\binom{N-1}{c_m}$ possible attributions $\mathcal{A}_m \subseteq \{1, \ldots, i(m)-1, i(m)+1, \ldots, N\}$ with $|\mathcal{A}_m| = c_m$, the probability mass of an attribution $\mathcal{A}_m$ is $p(\mathcal{A}_m|c_m) = \binom{N-1}{c_m}^{-1}$.

The process of generating a noisy copy $X_{i,r}$ from a mother bump $\tilde{X}_m$ is described by a conditional distribution $p_x(x_{i,r}|\tilde{x}_m; \theta_i^x)$. The vectors θ_i^x may be treated as random vectors with non-trivial priors $p(\theta_i^x)$. In the case of bump models (cf. Fig. 2), we generate copies by shifting the mother bump center while the other parameters (width, height, and amplitude) are drawn from some prior distribution, independently for each copy. The center offset may be modeled as a Gaussian random variable with mean vector $(\delta_{t,i}, \delta_{f,i})$ and diagonal non-isotropic covariance matrix $V_i = \mathrm{diag}(s_{t,i}, s_{f,i})$, and hence, $\theta_i^x = (\delta_{t,i}, \delta_{f,i}, s_{t,i}, s_{f,i})$. The model allows an *average offset* $(\delta_{t,i}, \delta_{f,i})$ between bump models; even if the average offset is zero, there may still be random offsets between the exemplars and their copies (see Fig. 2). We will assume that $s_{t,i} = s_t$ and $s_{f,i} = s_f$ for all i. We adopt improper priors $p(\delta_{t,i}) = 1 = p(\delta_{f,i})$ for $\delta_{t,i}$ and $\delta_{f,i}$ respectively, and conjugate priors for s_t and s_f, i.e. scaled inverse chi-square distributions.

The parameters $\theta = (\theta^c, \theta^x)$ might be constant or time-varying. In the latter case, we make the reasonable assumption that the parameters θ vary smoothly over time. The prior on $\theta(t)$ can then be chosen as:

$$p(\theta(t)) = \frac{1}{Z(\beta)} \prod_j \exp\left[\beta_j \int_0^T \left(\frac{d^2\theta_j}{dt^2}\right)^2 dt\right], \tag{1}$$

where $Z(\beta)$ is a normalization constant and β_j (for j=1,2,...) is a real positive number. For later convenience, we will introduce some more notation. The exemplar associated to mother event $\tilde{X}_m$ is denoted by $X_{i(m),k(m)}$, it is the event $k(m)$ in point process $X_{i(m)}$. We denote the set of pairs $(i(m), k(m))$ by $\mathcal{I}^{\mathrm{ex}}$. A noisy copy of $\tilde{X}_m$ is denoted by $X_{j(m),\ell(m)}$, it is the event $\ell(m)$ in point process $X_{j(m)}$ with $j(m) \in \mathcal{A}_m$. We denote the set of all pairs $(j(m), \ell(m))$ associated to $\tilde{X}_m$ by $\mathcal{I}_m^{\mathrm{copy}}$, and furthermore define $\mathcal{I}^{\mathrm{copy}} \triangleq \mathcal{I}_1^{\mathrm{copy}} \cup \cdots \cup \mathcal{I}_M^{\mathrm{copy}}$ and $\mathcal{I} = \mathcal{I}^{\mathrm{ex}} \cup \mathcal{I}^{\mathrm{copy}}$; the latter contains the indices of all exemplars and their copies. In this notation, the exemplar-based probabilistic model may be written as:

$$\begin{aligned} p(\tilde{X}, X, \mathcal{I}, \theta) = p(\theta^c)p(\theta^x)(1 - \lambda\,\mathrm{vol}(S))\lambda^M N^{-M} \prod_{m=1}^{M} \delta\big(x_{i(m),k(m)} - \tilde{x}_m\big) \\ \cdot\, p(c_m|\theta^c)\binom{N-1}{c_m}^{-1} \prod_{(i,j)\in\mathcal{I}_m^{\mathrm{copy}}} p_x(x_{i,j}|\tilde{x}_m, \theta^x). \end{aligned} \tag{2}$$

If the point processes $X = (X_1, \ldots, X_N)$ are well-synchronized, almost all processes X_i contain a copy of each mother bump $\tilde{X}_m$; the sets $\mathcal{I}_m^{\mathrm{copy}}$ are either of size $N-1$ or are slightly smaller. In the case of bump models, the variances s_t and s_f are then small. Note that $\mathcal{I}$ specifies the exemplars and their copies, and as a result, from $\mathcal{I}$ one can deduce various properties of the clusters, e.g., average number of bumps per cluster; the parameters s_t and s_f are part of the parameter vector θ. Therefore, given point processes $X = (X_1, \ldots, X_N)$, we wish to infer $\mathcal{I}$ and θ, since those variables contain information about similarity.

3 Statistical Inference

A reasonable approach to infer $(\mathcal{I}, \theta)$ is maximum a posteriori (MAP) estimation:

$$(\hat{\mathcal{I}}, \hat{\theta}) = \underset{(\mathcal{I},\theta)}{\operatorname{argmax}} \log p(\tilde{X}, X, \mathcal{I}, \theta). \tag{3}$$

There is no closed form expression for (3), therefore, we need to resort to numerical methods. A simple technique to try to find (3) is coordinate descent: We first choose initial values $\hat{\theta}^{(0)}$, and then perform the following updates for $r \geq 1$ until convergence:

$$\hat{\mathcal{I}}^{(r)} = \underset{\mathcal{I}}{\operatorname{argmax}} \log p(\tilde{X}, X, \mathcal{I}, \hat{\theta}^{(r-1)}) \tag{4}$$

$$\hat{\theta}^{(r)} = \underset{\theta}{\operatorname{argmax}} \log p(\tilde{X}, X, \hat{\mathcal{I}}^{(r)}, \theta). \tag{5}$$

3.1 Integer Program

We can write the update (4) as an integer program, i.e., a discrete optimization problem with linear objective function and linear (equality and inequality) constraints; we will omit the details here due to space constraints, and will merely discuss some general observations. The update (4) is for $N = 2$ equivalent to *bipartite maximum weighted matching* optimization, a problem that can be solved in polynomial time, for instance by using the LP relaxation of the corresponding IP formulation, or by the max-product message-passing algorithm detailed in [3]. We have shown that for $N > 2$, the combinatorial problem (4) is equivalent to *weighted k-dimensional matching* optimization, an NP-hard problem in the general case. Therefore, the extension from 2 time series to more than 2 is far from trivial. In practice (see the applications of Section 4 and 5), we were often able to solve the corresponding integer program very efficiently (using CPLEX). For integer programs with more than 10'000 variables and 5'000 constraints, the solution of a given was obtained in less than 1 second on a fast processor (3GHz). The total running time of the algorithm (iterations of equations (14) and (15) until convergence) was under 7 seconds on average.

3.2 Parameter Estimation

We now consider the update (5), i.e., estimation of the parameters $\theta = (\theta^x, \theta^c)$. First we treat constant parameters θ. The estimate $\hat{\theta}^{(r+1)} = (\hat{\theta}^{x^{(r+1)}}, \hat{\theta}^{c^{(r+1)}})$ (5) is often available in closed-form. This is in particular the case for the parametrization $\theta_i^x = (\delta_{t,i}, \delta_{f,i}, s_t, s_f)$. The point estimates $\hat{\delta}_{t,i}^{(r+1)}$ and $\hat{\delta}_{f,i}^{(r+1)}$ are the (sample) mean of the timing and frequency offset respectively, computed between all noisy copies in X_i and their associated exemplars. The estimates $\hat{s}_t^{(r)}$ and $\hat{s}_f^{(r)}$ are obtained similarly. The expression for the parameter p_s of the binomial prior for the number of copies C_m is also straightforward. Now we treat time-varying parameters θ (cf. (1)). The estimate $\hat{\theta}^{(r+1)}(t) = \left(\hat{\theta}^{x^{(r+1)}}(t), \hat{\theta}^{c^{(r+1)}}(t)\right)$ (5) are usually not available in closed-form. Let us again consider the parametrization $\theta_i^x = (\delta_{t,i}, \delta_{f,i}, s_t, s_f)$. The point estimates $\hat{\delta}_{t,i}^{(r+1)}$ are given by:

$$\hat{\delta}_{t,i}^{(r)} = \underset{\delta_{t,i}}{\operatorname{argmin}} \sum_{k,i',k'} \hat{b}_{i,k,i',k'}^{(r)} \frac{\left(T_{i,k} - T_{i',k'} - \delta_{t,i}(T_{i',k'})\right)^2}{\hat{s}_t^{(r)}(T_{i',k'})} + \beta_{\delta_t} \int_0^T \left(\frac{d^2\delta_{t,i}}{dt^2}\right)^2 dt, \tag{6}$$

where $\hat{b}_{i,k,i',k'}^{(r)}$ equals one if, according to $\hat{\mathcal{I}}^{(r)}$ (4), bump $X_{i,k}$ is considered as "noisy" copy of the exemplar $X_{i',k'}$ and zero otherwise. Similarly we have:

$$\hat{s}_{t,i}^{(r)} = \underset{s_{t,i}}{\operatorname{argmin}} \sum_{k,i',k'} \hat{b}_{i,k,i',k'}^{(r)} \left[\log s_t(T_{i',k'}) + \frac{\left(T_{i,k} - T_{i',k'} - \hat{\delta}_{t,i}^{(r)}(T_{i',k'})\right)^2}{s_t(T_{i',k'})}\right] + \beta_{s_t} \int_0^T \left(\frac{d^2 s_{t,i}}{dt^2}\right)^2 dt. \tag{7}$$

Note that the updates (6) and (7) are coupled. The processes $\hat{\delta}_{t,i}^{(r)}$ and $\hat{s}_{t,i}^{(r)}$ may be determined jointly by coordinate descent: with an initial guess for $\hat{s}_{t,i}^{(r)}$ one determines $\hat{\delta}_{t,i}^{(r)}$ (6), with the resulting estimate $\hat{\delta}_{t,i}^{(r)}$ one updates $\hat{s}_{t,i}^{(r)}$ through (7), and so on, until convergence. A similar procedure can be applied to jointly determine $\hat{\delta}_{f,i}^{(r)}$ and $\hat{s}_{f,i}^{(r)}$. The updates (6)–(7) may be solved by a variational approach, in particular, by solving the associated Euler-Lagrange equations numerically. However, it is well known that the unique solution of optimization problems of the form (6) over the space of twice differentiable functions are smoothing cubic splines [8]. As a result, the expression (6) can be found by cubic-spline smoothing. This does not apply, however, to the updates (7), since it involves non-quadratic terms. One option is to solve this expression by numerically integrating the associated Euler-Lagrange equations, as we pointed out earlier. Alternatively, one may determine the second-order Taylor expansion of the first term in (7). This results in quadratic approximations, known as saddle-point or Laplace approximation, of the non-quadratic terms. The solutions of those relaxed variational problems are smoothing cubic spline. In other words, the expressions $\hat{s}_{t,i}^{(r)}$ and $\hat{s}_{f,i}^{(r)}$ can be found *approximately* by cubic-spline smoothing. The latter is much faster than numerically solving the associated Euler-Lagrange equations, which is a major advantage. One can derive similar variational problems for the parameter p_s of the binomial prior, which in turn can also be solved practically by cubic-spline smoothing.

4 Diagnosis of MCI from EEG

As a first application, we consider the problem of diagnosing mild cognitive impairment (MCI) from EEG. For the sake of comparison, we used the same EEG data as in [3], i.e., rest eyes-closed EEG data recorded from 21 sites on the scalp based on the 10–20 system [9] with sampling frequency of 200Hz and band-pass filter between 4 and 30Hz. The subjects consist of two groups: (i) 22 patients suffering from mild cognitive impairment (MCI), who subsequently developed mild AD; (ii) a control set of 38 age-matched, healthy subjects who had no memory or other cognitive impairments. Pre-selection was conducted to

ensure that the data were of a high quality, as determined by the presence of at least 20s of artifact free data. We aggregated the 21 bump models in five regions (frontal, temporal left and right, central, occipital) by means of the aggregation algorithm described in [6], resulting in a bump model for each of those five regions ($N = 5$).

A large variety of *classical* synchrony measures (more than 30 in total) have in [3] been applied to both data sets with the aim of detecting MCI induced perturbations in EEG synchrony; none of those classical measures except full frequency Directed Transfer Function (ffDTF) [7], which is a Granger causality measure, was able to detect significant loss of EEG synchrony in MCI patients; note that since we considered many methods simultaneously, we need to apply Bonferroni postcorrection: the p-values need to be multiplied by the number of measures. On the other hand, the stochastic-event-synchrony measure ρ_{spur}, proposed in [3], resulted in significant differences between both subject groups ($p = 2.1{\cdot}10^{-4}$). This seems to indicate that there is an increase of unsynchronized activity in MCI patients.

The results from our exemplar-based approach are summarized in Table 1; we adopted constant parameters, because time-varying parameters for spontaneous EEG. We studied the following statistics: posterior distribution $p(c_m = i|X) = p_i^c$ of the number of copies of each exemplar c_m, parameterized by $(p_0^c, p_1^c, \ldots, p_4^c)$; $\bar{c}_m$: average number of copies per cluster; s_t: variance in time domain ("time jitter"); s_f: variance in frequency domain ("frequency jitter"); $\Delta\bar{T}$: average width of bumps; $\Delta\bar{F}$: average height of bumps; $\bar{F}$: average frequency of bumps. We also consider the linear combination h^c of all parameters p_i^c that optimally separates both subject groups. Interestingly, the latter statistic amounts to about the same p-value as the index ρ_{spur} of SES [3]. The posterior $p(c_m|X)$ mostly differs in p_1^c, p_2^c and p_4^c: in MCI patients, the number of clusters of size five (p_4^c) significantly decreases; on the other hand, the number of clusters of size one (p_1^c) and two (p_2^c) significantly increases. This explains and confirms the observed increase of ρ_{spur} in MCI patients [3]. Combining h^c with ffDTF and $\Delta\bar{T}$ allows to separate the two groups quite well (more than 90% correctly classified); this is far better than what can be achieved by means of classical similarity measures (about 75% correctly classified). Classification rates between 80 and 85% can be obtained by combining two features. (The classification rates were obtained through crossvalidation, i.e., the leave-one-out procedure.)

We also verified that the measures p_i^c, $\bar{c}_m$, h^c, and s_t are not correlated with other synchrony measures, e.g., Pearson correlation coefficient, magnitude and phase coherence, phase synchrony etc. (Pearson r, $p > 0.10$). In contrast to the classical measures, they quantify the synchrony of oscillatory events instead of more conventional amplitude or phase synchrony, therefore, they provide *complementary* information about EEG synchrony.

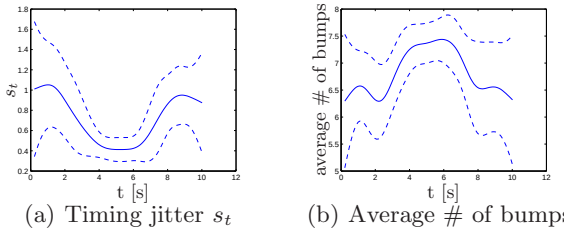
Stat.	p_0^c	p_1^c	p_2^c	p_3^c	p_4^c	$\bar{c}_m$	h^c	s_t	s_f	$\Delta\bar{T}$	$\Delta\bar{F}$	$\bar{F}$
p-value	0.016	$2.9.10^{-4**}$	0.089	0.59	0.0054^{*}	1.10^{-3**}	1.10^{-4**}	0.46	0.28	$2.3.10^{-4**}$	0.023^{*}	2.10^{-3**}

Table 1. Sensitivity of multivariate SES for diagnosing MCI (p-values for Mann-Whitney test; * and ** indicate $p < 0.05$ and $p < 0.005$ respectively).

5 Dynamics of Oscillatory Event Synchrony in SSVEP

As an application of the exemplar-based statistical model with *time-varying* parameters, we investigate how EEG synchrony evolves in response to a visual stimulus. In particular, human scalp EEG was recorded in a dark room while a subject was exposed to flickering light. The stimulus was a single flashing white square that flickers at 16Hz during 4 sec, and produces a steady-state response of the human visual system at the same frequency, referred to as "steady-state visually evoked potentials" (SSVEP). In total 50 trials were recorded with a baseline period of 3 sec, stimulation during 4 sec, and a post-stimulation period of 3 sec during which EEG resumed to the baseline activity. EEG data was recorded from 64 sites on the scalp, based on the extended 10-20 standard system, with a sampling frequency of 2048Hz; the EEG was high-pass filtered off-line with cut-off frequency of 3Hz (Butterworth forward and reverse), next it was downsampled by a factor of 8. A Biosemi system with average reference was used. As in Section 4, we aggregated the electrode sites into separate zones (here 9 instead of 5, since we used 64 instead of 21 channels), and extracted a bump model from each zone by means of the algorithms of [6]. The time-frequency maps were restricted to the frequency range of SSVEP, i.e., 15–17Hz.

Fig. 3 shows that the timing jitter significantly decreases during stimulation ($t \in [3s, 7s]$); also the average number of bumps per cluster clearly increases. The parameters β_{δ_t}, β_{δ_f}, β_{s_t}, and β_{s_f} (cf. (6)(7)) were determined by cross-validation.



(a) Timing jitter s_t (b) Average # of bumps per cluster

Fig. 3. Time-varying parameters (solid line: average; dashed lines: ± one standard deviation over the 50 trials).

References

1. E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophsyiological signals," *Progress in Neurobiology,* 77 (2005) 1–37.
2. P. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology," *Neuron,* 52:155–168, 2006.
3. J. Dauwels, F. Vialatte, T. Rutkowski, and A. Cichocki, "Measuring neural synchrony by message passing," *Advances in Neural Information Processing Systems 20,* MIT Press.
4. B. Frey and D. Dueck, "Clustering by passing messages between data points," *Science,* vol. 315, No. 5814, pp. 972–976, 2007.
5. D. Lashkari and P. Golland, "Convex clustering with exemplar-based models," In *Advances in Neural Information Processing Systems 20,* MIT Press.
6. F. Vialatte, C. Martin, R. Dubois, J. Haddad, B. Quenet, R. Gervais, and G. Dreyfus, "A Machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics," *Neural Networks,* 2007, 20:194–209.
7. M. Kamiński and Hualou Liang, "Causal Influence: Advances in Neurosignal Analysis," *Critical Review in Biomedical Engineering,* 33(4):347–430 (2005).
8. I. J. Schoenberg, "Spline functions and the problem of graduation," *Mathematics* 52: 974–50.
9. P. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG,* Oxford University Press, 2006.