Biol. Cybern. (2003)
DOI 10.1007/s00422-003-0433-7

Biological Cybernetics

Estimation of single-trial multicomponent ERPs: Differentially variable component analysis (dVCA)

Wilson Truccolo[1], **Kevin H. Knuth**[2], **Ankoor Shah**[3,4], **Steven L. Bressler**[5], **Charles E. Schroeder**[3,4], **Mingzhou Ding**[5]

[1] Department of Neuroscience, Brown University, 190 Thayer Street, Providence, RI 02912, USA
[2] Computational Sciences Division, Code IC, NASA Ames Research Center, Moffett Field, CA 94035, USA
[3] Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[4] Cognitive Neuroscience and Schizophrenia Dept., Nathan S. Kline Institute, Orangeburg, NY 10962, USA
[5] Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA

Received: 27 January 2003 / Accepted: 4 September 2003 / Published online: 4 December 2003

Abstract. A Bayesian inference framework for estimating the parameters of single-trial, multicomponent, event-related potentials is presented. Single-trial recordings are modeled as the linear combination of ongoing activity and multicomponent waveforms that are relatively phase-locked to certain sensory or motor events. Each component is assumed to have a trial-invariant waveform with trial-dependent amplitude scaling factors and latency shifts. A *Maximum a Posteriori* solution of this model is implemented via an iterative algorithm from which the component's waveform, single-trial amplitude scaling factors and latency shifts are estimated. Multiple components can be derived from a single-channel recording based on their differential variability, an aspect in contrast with other component analysis techniques (e.g., independent component analysis) where the number of components estimated is equal to or smaller than the number of recording channels. Furthermore, we show that, by subtracting out the estimated single-trial components from each of the single-trial recordings, one can estimate the ongoing activity, thus providing additional information concerning task-related brain dynamics. We test this approach, which we name differentially variable component analysis (dVCA), on simulated data and apply it to an experimental dataset consisting of intracortically recorded local field potentials from monkeys performing a visuomotor pattern discrimination task.

1 Introduction

Relations between brain and behavior are often studied by recording event-related potentials (ERPs) over repeated presentations of a sensory stimulus or task performance. The obtained data are commonly modeled as the linear combination of ongoing activity and event-related components that are relatively phase-locked to the onset of the event. The ongoing activity includes signals that are not related to the event and possibly also signals that are induced by the event but that are not phase-locked to its onset. Based on the classic signal-plus-noise model, the event-related signals are extracted by averaging across the ensemble of trials, resulting in the average event-related potential (AERP). The main assumption of this approach is that an event-related signal exists that is invariant across trials. However, evidence amassed thus far indicates that this assumption is no longer tenable. More realistic models have been advanced where trial-to-trial amplitude and latency variability are taken into account (Woody 1967; Coppola et al. 1978; Truccolo et al. 2002a).

The existence of trial-to-trial variability in event-related activity implies that the simple ensemble mean is not the optimal estimator for event-related components. Woody (1967) proposed a method for estimating the latency of single-trial event-related activity by use of a matched filter. Following Woody's technique, averages can then be computed after adjusting for the estimated latency variability over trials. Subsequent work has extended Woody's idea by including amplitude variability and by placing the estimation of single-trial parameters in a maximum likelihood framework (Jaskowski and Verleger 1999; Lange et al. 1997; McGillem et al. 1985; Pham et al. 1987). Other techniques based on wavelet (Bartnik et al. 1992; Quiroga 2000; Quiroga and Garcia 2003) and clustering analysis (Haig et al. 1995) have also been introduced in the literature.

The goal of the present paper is threefold. First, starting from a Bayesian perspective, we propose a comprehensive framework that can serve as a theoretical guideline for modeling and estimating parameters of single-trial multicomponent ERPs. The ERPs are modeled as the linear combination of multiple components whose waveforms, single-trial latency shifts, and amplitude scaling factors are to be estimated based on the components' differential variability from trial to trial, a technique we entitle differentially variable component analysis (dVCA). In this regard, Woody's

Coreespondece to: Mingzhou Ding (e-mail: ding@fau.edu)

(Woody 1967) and related algorithms are seen as special cases implementing the *maximum a posteriori* solution of the posterior probability. Second, we emphasize that single-trial parameters such as amplitude scaling factors and latency shifts are themselves interesting physiological variables, providing essential information for understanding stages of information processing in the cortex (Nowak and Bullier 1997; Schmolesky et al. 1998; Schroeder et al. 1998). Especially interesting is the finding of differential variability for estimated components in the same or different channels. Third, recent work shows that the ongoing activity, which contains non-phase-locked signals, is the main source for the study of functional interdependences between neuronal populations (Truccolo et al. 2002a). Traditionally, ongoing activities have been obtained via subtraction of the AERP from single-trial recording. We demonstrate that a more appropriate way to extract the ongoing process is to perform single-trial estimation of event-related potential components and then subtract them from the single-trial time series.

The entire dVCA approach is first tested on simulated data and then applied to an experimental dataset consisting of intracortically recorded single-trial local field potentials (LFPs) in monkeys performing a GO/NO-GO visuomotor pattern discrimination task.

2 Theory

2.1 Model of single-trial LFPs

Our starting point is a generative model for the recorded single-trial LFPs. Such a model should capture at least four main properties of LFP recordings in the event-related paradigm: (a) the existence of signals that are relatively phase-locked to a specific event onset, (b) the trial-to-trial variability in amplitude and latency of the event-related signals, (c) the possibility that the event-related response may be the superposition of multiple components with differential variability in their single-trial amplitude scaling factors and latency shifts, and (d) the existence of signals that are not phased-locked to the event, including activity that is unrelated to the experiment (e.g., measurement noise) and activity that is induced by the event. As a first approximation, the waveforms of the event-related components are assumed to be invariant across trials. A model that incorporates all these features can be written as

$$x_r(t) = \sum_{n=1}^{N} a_{nr} s_n(t - \tau_{nr}) + \eta_r(t) \quad , \tag{2.1}$$

where $x_r(t)$ is the LFP recording from the r-th trial, $s_n(t)$ is the n-th event-related component waveform with a trial-to-trial variable amplitude scaling factor and latency shift given by a_{nr} and τ_{nr}, respectively, and N is the total number of components. The process $\eta_r(t)$, henceforth referred to as the ongoing activity, includes all the non-phase-locked signals and is assumed to have a zero mean. From a physiological point of view we note that each component could correspond either to the activity of a distinct neural source or generator, or to the activity of the same source but at a different neural processing stage.

2.2 Bayesian estimation framework

The problem of estimating the single-trial parameters $s_n(t)$, a_{nr}, and τ_{nr} is formulated from a Bayesian perspective. According to Bayes theorem, the posterior probability of model parameters M given data D and prior information I can be written as

$$p(M|D,I) = \frac{p(D|M,I)p(M|I)}{p(D|I)} \quad . \tag{2.2}$$

For the LFP model in Eq. 2.1, the posterior probability becomes

$$\begin{aligned} &p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\theta_\eta(t)|\{x_r(t)\},I) \\ &\quad = \frac{p(\{x_r(t)\}|\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\theta_\eta(t),I)}{p(\{x_r(t)\}|I)} \\ &\qquad \cdot p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\theta_\eta(t)|I) \quad , \end{aligned} \tag{2.3}$$

where $\{\cdot\}$ refers to the set of parameters for all the components and the whole ensemble of trials, $\theta_\eta(t)$ denotes the parameters for the ongoing process, and $p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\theta_\eta(t)|I)$ is the prior probability for the model parameters. For this additive model, the likelihood $p(\{x_r(t)\}|\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\theta_\eta(t),I)$ turns out to be simply given by the probability model of the ongoing activity, i.e., $p(\eta(t)|I)$. In the absence of precise knowledge about the temporal structure of the ongoing activity, we assign $\eta(t)$ to be independent identically distributed with an (unknown) time-independent variance σ_η^2 and zero mean. In this way, Eq. 2.3 is rewritten as

$$\begin{aligned} &p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\sigma_\eta|\{x_r(t)\},I) \\ &\quad = \frac{p(\{\eta(t)\}|\sigma_\eta,I)p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\sigma_\eta|I)}{p(\{x_r(t)\}|I)} \quad . \end{aligned} \tag{2.4}$$

Under the constraint of a given mean and σ_η^2 and following the principle of maximum entropy (Sivia 1996), a Gaussian density is assigned to the likelihood function. After dropping the normalization term $1/p(\{x_r(t)\}|I)$, the posterior can be rewritten as

$$\begin{aligned} &p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\sigma_\eta|\{x_r(t)\},I) \\ &\propto p(\{s_n(t)\},\{a_{nr}\},\{\tau_{nr}\},\sigma_\eta|I)(2\pi\sigma_\eta^2)^{\frac{-RT}{2}} \\ &\quad \cdot \exp\left[\frac{-\sum_{r=1}^{R}\sum_{t=1}^{T}\left[x_r(t) - \sum_{n=1}^{N} a_{nr}s_n(t-\tau_{nr})\right]^2}{2\sigma_\eta^2}\right], \end{aligned} \tag{2.5}$$

where R is the total number of trials and T is the total number of sampled data points in a given trial. For notational simplicity we have assumed the sampling

interval to be unity in the above expression. In practice, the real time can be recovered by multiplying the integer time index t by the sampling interval.

In the absence of detailed knowledge about the parameters a_{nr}, τ_{nr}, and $s_n(t)$, we take their prior distributions to be uniform, with appropriate cutoffs reflecting physiologically reasonable ranges of values. That is,

$$p(s_n(t)|I) = \text{const.}, \forall n \ , \tag{2.6}$$

$$p(a_n|I) = \text{const.}, \ \text{for } 0 < a_n \leq a_{\max}, \forall n \ , \tag{2.7}$$

$$p(\tau_n|I) = \text{const.}, \ \text{for } \tau_{\min} < \tau_n \leq \tau_{\max}, \forall n \ . \tag{2.8}$$

Treating the variance of the ongoing activity as a nuisance parameter and assigning the Jeffreys prior $p(\sigma_\eta) = \sigma_\eta^{-1}$ (Sivia 1996), we marginalize the posterior over σ_η:

$$\begin{aligned} &p(\{s_n(t)\}, \{a_{nr}\}, \{\tau_{nr}\}|\{x_r(t)\}, I) \\ &\quad \propto p(\{s_n(t)\}, \{a_{nr}\}, \{\tau_{nr}\}|I) \int_{-\infty}^{\infty} (2\pi\sigma_\eta^2)^{\frac{-RT}{2}} \sigma_\eta^{-1} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{r=1}^{R} \sum_{t=1}^{T} \left[x_r(t) - \sum_{n=1}^{N} a_{nr} s_n(t - \tau_{nr}) \right]^2 \right) d\sigma_\eta \end{aligned} \tag{2.9}$$

and obtain

$$\begin{aligned} &p(\{s_n(t)\}, \{a_{nr}\}, \{\tau_{nr}\}|\{x_r(t)\}, I) \\ &\quad \propto p(\{s_n(t)\}, \{a_{nr}\}, \{\tau_{nr}\}|I) (2\pi)^{\frac{-RT}{2}} \Gamma\left(\frac{RT}{2}\right) \\ &\quad \cdot \left(\sum_{r=1}^{R} \sum_{t=1}^{T} \left[x_r(t) - \sum_{n=1}^{N} a_{nr} s_n(t - \tau_{nr}) \right]^2 \right)^{\frac{-RT}{2}} , \end{aligned} \tag{2.10}$$

where $\Gamma(\cdot)$ is the gamma function.

The evaluation of the posterior probability and computation of its moments can be obtained via Markov Chain Monte Carlo (MCMC) methods. Although highly informative, these methods carry the disadvantage of being computationally intensive. Here, instead, we summarize the posterior density by seeking the *maximum a posteriori* (MAP) solution, i.e., a set of parameters that maximize the posterior probability. In the context of Eq. 2.2, the MAP solution for the model parameters M is

$$\begin{aligned} \hat{M} &= \arg\max_M [p(D|M, I) p(M|I)] \\ &= \arg\max_M [\ln p(M|I) + \ln p(D|M, I)] \ . \end{aligned} \tag{2.11}$$

Because waveforms, amplitude scaling factors, and latency shifts are being estimated simultaneously, the model has degeneracy. We solve this problem by constraining the ensemble mean of the amplitude scaling factors and latency shifts of each component to equal one and zero, respectively.

Intuition about the characteristics of the MAP solution can be gained by examining the partial derivatives of the logarithm of the posterior probability with respect to each of the model parameters. This leads to a practical and simple estimation algorithm. In what follows, the time t assumes discrete values corresponding to digital sampling. Let

$$Q = \sum_{r=1}^{R} \sum_{t=1}^{T} \left[x_r(t) - \sum_{n=1}^{N} a_{nr} s_n(t - \tau_{nr}) \right]^2 \ . \tag{2.12}$$

Let P represent the posterior probability in Eq. 2.10. Then its logarithm can be simply written as

$$\ln P = -\frac{RT}{2} \ln Q + \text{const} \ . \tag{2.13}$$

For the partial derivatives we use j, p, and q to denote specific index values for the generic running indices n, r, and t, respectively. The first partial derivative with respect to $s_j(q)$ is

$$\frac{\partial \ln P}{\partial s_j(q)} = -\frac{RT}{2} Q^{-1} \frac{\partial Q}{\partial s_j(q)} \ , \tag{2.14}$$

where

$$\frac{\partial Q}{\partial s_j(q)} = -2 \sum_{r=1}^{R} \left[W a_{jr} - (a_{jr})^2 s_j(q) \right] \tag{2.15}$$

and

$$W = x_r(q + \tau_{jr}) - \sum_{\substack{n=1 \\ n \neq j}}^{N} a_{nr} s_n(q - \tau_{nr} + \tau_{jr}) \ . \tag{2.16}$$

Setting $\frac{\partial Q}{\partial s_j(q)} = 0$ gives

$$\hat{s}_j(q) = \frac{\sum_{r=1}^{R} W a_{jr}}{\sum_{r=1}^{R} (a_{jr})^2} \ , \tag{2.17}$$

with $\hat{s}_j(q)$ denoting the estimated parameter. The above equation does not have a closed-form solution since the right-hand side depends on the other estimated parameters. However, intuition about the type of solution can be obtained by examining the term W. Basically, this term involves the following two elements: (a) the data are shifted according to the latency shift of the estimated component, i.e., $x_r(q + \tau_{jr})$; and (b) the other scaled and time-shifted components, i.e., $a_{nr} s_n(q - \tau_{nr} + \tau_{jr})$, for $n \neq j$ are subtracted from the data. The properly scaled residuals, where the scaling is given by the term a_{jr}, are then averaged across trials.

Similarly, we obtain the estimate for the amplitude scaling factors a_{nr}:

$$\hat{a}_{jp} = \frac{\sum_{t=1}^{T} UV}{\sum_{t=1}^{T} V^2} \ , \tag{2.18}$$

where $U = x_p(t) - \sum_{n=1_{n\neq j}}^{N} a_{np} s_n(t-\tau_{np})$ and $V = s_j(t-\tau_{jp})$. Notice that the formula derived for $\hat{a}_{jp}$ is related to a matched filter solution. That is, $\hat{a}_{jp}$ is given by projecting U, which is the data after removing the contribution from the other scaled and time-shifted components, onto V, which is the current component under estimation.

For the latency shift parameter, setting $\frac{\partial Q}{\partial \tau_{jp}} = 0$ leads to the following equation:

$$2\sum_{t=1}^{T}\left[\left[x_p(t) - \sum_{\substack{n=1\\n\neq j}}^{N} a_{np} s_n(t-\tau_{np})\right] a_{jp} s_j'(t-\tau_{jp}) - a_{jp}^2 s_j'(t-\tau_{jp}) s_j(t-\tau_{jp})\right] = 0 \ , \quad (2.19)$$

where $s_j'(t-\tau_{jp})$ is the time derivative of $s_j(t-\tau_{jp})$. The solution for $\hat{\tau}_{jp}$ is more difficult since τ appears in the argument of the waveform function. Again, intuition can be gained by directly examining the condition for the maximization of the logarithm of the posterior, which is equivalent to the minimization of the term Q in Eq. 2.12. Expansion of this term results in

$$\sum_{r=1}^{R}\sum_{t=1}^{T}\left[x_r^2(t) + \left[\sum_{n=1}^{N} a_{nr} s_n(t-\tau_{nr})\right]^2 - 2x_r(t)\sum_{n=1}^{N} a_{nr} s_n(t-\tau_{nr})\right] \ . \quad (2.20)$$

As τ_{jp} is varied, only the cross terms in $x_p(t)\cdot \sum_{n=1}^{N} a_{np} s_n(t-\tau_{np})$ for $n = j$ are relevant for the minimization of Eq. 2.12 (as long as the event-related components $s_n(t)$ can be considered zero outside some time interval (t_0, t_f)). Thus the optimal parameter $\hat{\tau}_{jp}$ is found by maximizing

$$\rho(\tau) = \sum_{t=1}^{T}\left[a_{jp} s_j(t-\tau)\left[x_p(t) - \sum_{\substack{n=1\\n\neq j}}^{N} a_{np} s_n(t-\tau)\right]\right] \ , \quad (2.21)$$

which, if properly normalized, is just the cross correlation between the estimated component and the data after the contributions from the other components have been subtracted out. Thus

$$\hat{\tau}_{jp} = \arg\max_{\tau} \rho(\tau) \ . \quad (2.22)$$

This result corresponds to Woody's matched filter algorithm for latency estimation (Woody 1967).

2.3 Algorithm implementation

Given that the partial derivatives and the Hessian matrix of the posterior probability are readily available, techniques such as conjugate gradients and quasi-Newton trust region methods can be employed in the search for the MAP solution. At each iteration step, the optimization can be carried out in stages, i.e., first with respect to the latency shifts, then the waveforms, and finally the amplitude scaling factors. These techniques are much less computationally intensive than MCMC but still require multiple searches starting at different initial conditions or stochastic annealing to avoid local maxima.

However, the analysis in the previous section suggests a simpler heuristic algorithm. After an initial guess, at each iteration step the parameters for all components are updated in sequence as mentioned before: first the latency shifts, then the waveforms, and finally the amplitude scaling factors. Specifically, let $s_j^m(t)$, a_{jr}^m, τ_{jr}^m denote the estimated values of the parameters in the m-th iteration. Also, let the latency shifts assume only discrete integer values, with unit corresponding to the sampling interval. To avoid degeneracy in the model, the averages of the amplitude scaling factors and latency shift values in each iteration are constrained to $\langle a_{jr}^m \rangle_r \equiv 1$ and $\langle \tau_{jr}^m \rangle_r \equiv 0$. In this way, if there is no trial-to-trial variability both in amplitude and latency, the superposition of the estimated component waveforms should equal the AERP. For a single channel dataset $\{x_r(t)\}$, the algorithm consists of the following steps:

(0) At $m = 0$, the initial guess for the amplitude scaling factors and latency shifts are set to $a_{jr}^0 = 1, \tau_{jr}^0 = 0, \forall j, r$. For simplicity, the decision on

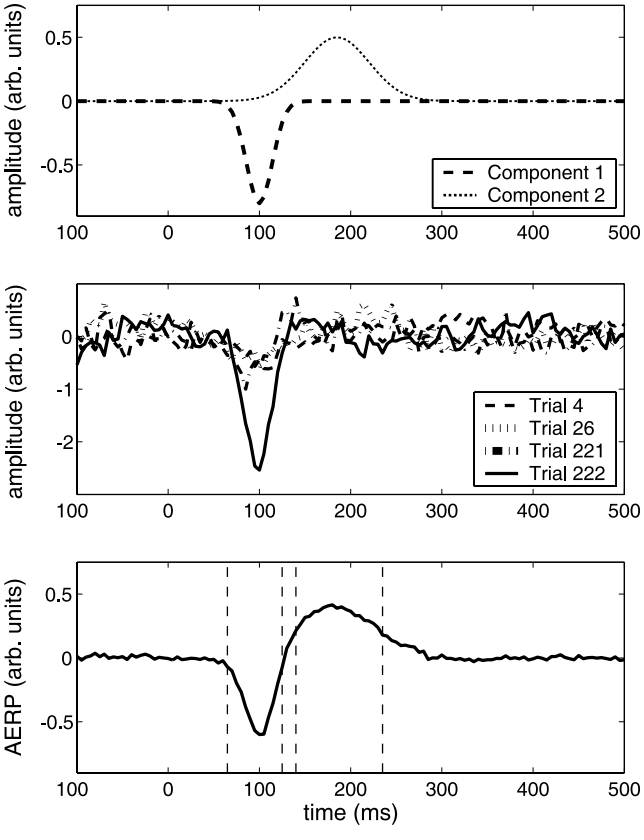

Fig. 1. Synthetic data generated to evaluate the performance of the dVCA algorithm (see text for details)

the number of components N is based on the inspection of the AERP (see Figs. 1 and 4). Similarly, N non overlapping segments of the AERP are taken as the initial guesses for the N components' waveforms $s_j^0(t)$. After this initialization, each iteration consists of four steps:

(1) For all trials, estimate the single-trial latency shifts for one component at a time, starting with the first and proceeding up to the N-th component, according to $\tau_{jr}^{m+1} = \arg\max \rho^m(\tau)$. Reexpressed in time units, the estimated latency shift is simply $\tau_{jr}\Delta t$, where Δt is the sampling interval. Given an approximate knowledge of where in time the component is expected to happen, an interval for the search of the optimal latency shift τ_{jr} can be stipulated. In this way, the possibility that the component matches by chance the waveform of unrelated ongoing activity is diminished. In the application to experimental data in Sect. 4, the following intervals for the search of the optimal latency shift were employed: [−30 ms, 30 ms] for early components and [−60 ms, 60 ms] for later components.

(2) Estimate the waveforms according to

$$s_j^{m+1}(t) = \frac{\sum_{r=1}^{R} W a_{jr}^m}{\sum_{r=1}^{R} (a_{jr}^m)^2} \quad ,$$

with

$$W = x_r(t + \tau_{jr}^{m+1}) - \sum_{\substack{n=1 \\ n \neq j}}^{N} a_{nr}^m s_n^m (t - \tau_{nr}^{m+1} + \tau_{jr}^{m+1}) \quad .$$

(3) For all the trials and components, estimate the amplitude scaling factors according to

$$a_{jr}^{m+1} = \frac{\sum_{t=1}^{T} UV}{\sum_{t=1}^{T} V^2} \quad ,$$

with $U = x_r(t) - \sum_{\binom{n=1}{n \neq j}}^{N} a_{nr}^m s_n^{m+1}(t - \tau_{nr}^{m+1})$ and $V = s_j^{m+1}(t - \tau_{jr}^{m+1})$.

(4) Increment the iteration index: $m = m + 1$; repeat (1) through (3) for M iterations.

We note that the differential variability of the components on a trial-by-trial basis is the foundation of the estimation technique. We thus term our algorithm differentially variable component analysis (dVCA). For the experimental data employed here, where the overlap between the components is not large and the signal-to-noise ratio (SNR) is high, two iterations already give reasonable results. The above algorithm is especially well suited for single-trial analysis of LFPs, e.g., intracortically recorded potentials, where the spatial overlap of different and distant sources is not severe as in noninvasive recordings like EEG and MEG. However, if we suspect that the components might be highly overlapped or that a more automatic way to specify the number of components and initialize their waveforms is called for, the algorithm can be improved with the following consideration.

The algorithm is initialized with a single component whose waveform can be taken to be a short segment around the largest fluctuation of the AERP. Single-trial amplitude scaling factors and latency shifts for this component are initialized as before, and the same sequence of steps (1–4) is performed until the estimated parameters converge (given a tolerance criterion) or a maximum number of iterations is reached. Then the estimated single-trial scaled and translated component is subtracted from the corresponding single-trial data, resulting in residual data. The ensemble average of these residual data is computed, leading to a new AERP. A new component is introduced into the estimation if the largest fluctuation of this new AERP happens to be outside of a confidence interval (e.g., $\pm 2\sigma(t)/\sqrt{R}$, where $\sigma(t)$ is the ensemble standard deviation of the residual). The initial value of this new component is taken to be the segment around this largest fluctuation. The whole procedure is repeated until the ensemble average of the residual data no longer shows any significant structure, i.e., fluctuations outside the confidence interval. Besides the use of a confidence interval as mentioned above, a model selection approach using information criteria (e.g., Akaike's information criterion) for the selection of the number of components could also be explored.

3 Application to simulated data

Simulations were used to evaluate the robustness of the dVCA in the presence of noise. The synthetic data mimicked local electric field potentials containing two independent, event-related components as shown in Fig. 1 (top panel). The components have Gaussian waveforms motivated by those found in the experimental data presented in the next section. A single-trial time series was produced by a linear summation of component 1, component 2, and the ongoing noise process that was assumed to have a 1/f spectrum with standard deviation denoted by σ_{noise}. (For white noise ongoing processes we obtained similar results.) Trial-to-trial variations in each component were introduced by multiplying the component with an amplitude scaling factor and by shifting it in time with a specific latency shift. Single-trial amplitude scaling factors for component 1 were varied as a lognormal distribution with sample mean $\mu_{\text{amp}} = 1.0$ and sample standard deviation $\sigma_{\text{amp}} = 1.0$. For component 2, the amplitude scaling factors had a sample mean $\mu_{\text{amp}} = 1.0$ and a sample standard deviation of $\sigma_{\text{amp}} = 1.5$. Single-trial latency shifts for component 1 were varied as a normal distribution with sample mean $\mu_{\text{lat}} = 0.0$ and sample standard deviation $\sigma_{\text{lat}} = 10.0$ ms. For component 2, the latency shifts had a sample mean $\mu_{\text{lat}} = 0.0$ and sample standard deviation $\sigma_{\text{lat}} = 25.0$ ms.

Twelve synthetic datasets were generated, each consisting of 222 simulated trials sampled at 200 Hz and

corresponding to a given ongoing noise level σ_{noise}. Specifically, the signal-to-noise ratios (SNR) of the datasets were {–34.6574, –27.7259, –20.7944, –13.8629, –9.8083, –6.9315, –2.8768, 0, 6.9315, 13.8629, 20.7944, 27.7259 dB} with respect to component 2, where the SNR was defined as:

$$SNR_{\text{component}} = 20 \log_{10} \frac{\sigma_{\text{component}}}{\sigma_{\text{noise}}} \text{dB} \ . \tag{3.1}$$

Here $\sigma_{\text{component}}$ was calculated as the square root of the variance of the component across time. It was critical to assign SNRs separately for each component in each experiment since the standard deviations of the two components differ.

Examples of single-trial time series are shown in the middle panel of Fig. 1 where the SNR with respect to component 2 is approximately −7 dB. It is clear that component 2 is not readily visible in these single-trial time series and that component 1 is not immediately obvious except in trial 222.

The dVCA algorithm was applied to each of the synthetic datasets as explained above in Sect. 2.3. The initial estimates for the component waveforms were taken as the AERP between nonoverlapping time periods and as zero elsewhere. Inspection of the AERP in each of the synthetic datasets revealed that the first peak, defined as component 1, occurred between 65 and 125 ms while the second peak, defined as component 2, occurred between 140 and 235 ms (see Fig. 1 bottom panel). Initial estimates of the single-trial amplitude scaling factors were initialized to 1.0, and those for the single-trial latency shifts were set to 0 ms. After initialization, dVCA was applied to each synthetic dataset until the estimated component waveforms varied less than 1% between iterations or until 15 iterations of the algorithm were performed. Performance of the algorithm was evaluated by comparing the estimated waveforms, amplitude scaling factors, and latency shifts with their original counterparts. The accuracies of estimated waveforms were measured by calculating the fractional root mean square (RMS) error for the j-th component as follows:

$$E^j_{\text{wave}} = \frac{\sqrt{\sum_{t=1}^{T} \left(s_j(t) - \hat{s}_j(t)\right)^2}}{\sqrt{\sum_{t=1}^{T} \left(s_j(t)\right)^2}} \ , \tag{3.2}$$

where T is the total number of time points, $s_j(t)$ is the original component waveform, and $\hat{s}_j(t)$ is the estimated component waveform. As shown in the top panel of Fig. 2, the error in the estimated waveforms decreases as the SNR increases for both components. A waveform error of less than 25% is achieved at around −10 dB and is below 10% at a SNR of 0 dB for each component.

The accuracy of estimated amplitude scaling factors and latency shifts was evaluated by examining the distribution of the differences between estimated and original values. Figure 2 (middle and lower panels) shows the results of comparing the estimated values and the

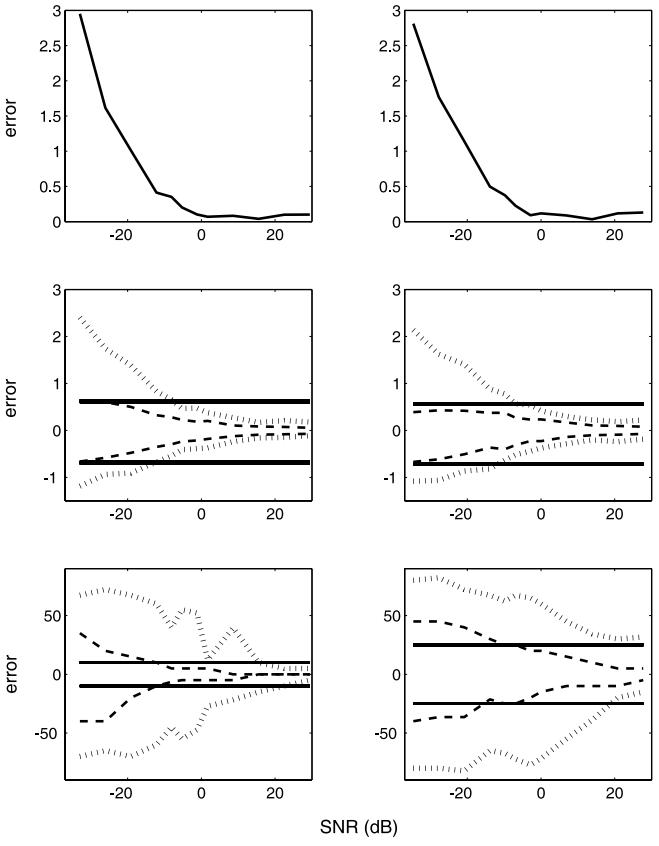

Fig. 2. Accuracy of the dVCA estimates of component waveforms, amplitude scaling factors, and latency shifts across various signal-to-noise ratios (see text for details)

original values for the two components across various SNR levels. The horizontal black lines denote the one standard deviation of the original latency shift and amplitude scaling factor values with respect to their means. The broken lines represent the one standard deviation of the differences between the estimated values and the original values. We note that, when these broken lines fall within the black horizontal lines, it means that the trial by trial estimated latency shifts and amplitude scaling factors are better representations of the true values than the ensemble or grand average. For amplitude estimation, trial-by-trial estimations produced better results than grand average for all the SNR levels examined (middle panels). For latency estimation, we achieved this result for SNR above −11 dB (lower panels). The dashed lines in the middle and lower panels are two standard deviations of the differences between the estimated values and the original values.

For two datasets with very different SNRs we examined directly the correlation between the estimated and the original values for waveform, amplitude scaling factor, and latency shift. The results are shown in Fig. 3. When the SNR is approximately 14 dB (high SNR, left column) with respect to component 2, the estimated waveforms of components 1 and 2 are an almost perfect replication of the original waveforms, as shown in the top left panel of Fig. 3. Similarly, the estimated amplitude scaling factors are nearly perfectly correlated with

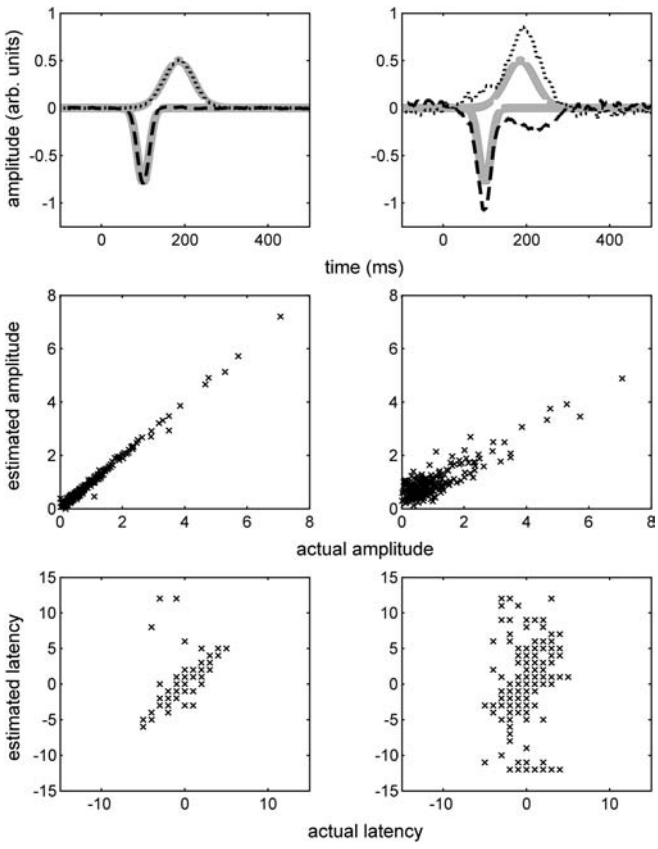

Fig. 3. Cortical sites (from *top* to *bottom*: striate, parietal, and somatosensory)

the original values ($r^2_{\text{component1}} = 0.99$; $r^2_{\text{component2}} = 0.99$) (middle left panel). Good results are also achieved for latency shifts ($r^2_{\text{component1}} = 0.52$; $r^2_{\text{component2}} = 0.70$) (lower left panel). For the second dataset, the SNR of the data with respect to component 2 was approximately -21 dB (low SNR, right column). The estimated waveforms capture the major features of the original components but also contain some error. The estimated amplitude scaling factors are still well correlated with the original values ($r^2_{\text{component1}} = 0.77$; $r^2_{\text{component2}} = 0.85$), but the latency shifts are no longer well correlated with the original values ($r^2_{\text{component1}} = 0.04$; $r^2_{\text{component2}} = 0.10$). We note that the SNR in this case is extremely small. It is rather remarkable that the waveform and amplitude scaling factor estimations are still able to produce reasonable results.

Application of the dVCA algorithm to synthetic data demonstrates that the technique is robust in estimating the parameters of the two components from a single channel of data. Our results suggest that the method works effectively for all the component parameters for SNRs down to approximately -11 dB. For waveforms of the components and their amplitude scaling factors, the method works effectively for SNRs down to -21 dB. As the SNR decreases, the dVCA estimates of the latency shifts are most affected. It is worth noting that the simulation here used two components that have a temporal overlap of about 35 ms. This is rather substantial given that each component has a length of about 60 to 70 ms. For situations with less component overlap, the dVCA technique is expected to work even more effectively. In summary, the results illustrated here demonstrate that the dVCA can be a valuable and accurate tool in identifying multiple components from a single channel of data and that its application to real data shown below, where the components do not appear to overlap temporally, should yield reasonable results.

4 Application to experimental LFP data

The experimental dataset was collected by Dr. Richard Nakamura in the Laboratory of Neuropsychology at the U.S. National Institute of Mental Health. Visual evoked responses in two macaque monkeys were sampled at 200 Hz from chronically implanted surface-to-depth bipolar electrodes at 15 and 14 cortical sites in the hemisphere contralateral to the hand executing the response. Prior to estimation of the single-trial parameters, the time series were resampled (linear interpolation) to 1 kHz. The monkeys performed a GO/NO-GO visual pattern discrimination task. The prestimulus stage began when the monkey, while viewing a computer screen, depressed a hand lever with the preferred hand. Following a random interval from 0.5 to 1.2 s, a visual stimulus appeared for 100 ms on the screen. Two types of visual patterns were presented: four dots arranged as a DIAMOND or as a LINE. The monkeys were rewarded for lifting the hand in response (GO) to one pattern type or for maintaining pressure (NO-GO) to the other pattern type. The contingency between stimulus pattern and response type was reversed on successive sessions. For the applications in this paper, two ensemble datasets were employed: one with 222 GO response trials to the DIAMOND pattern and another consisting of the 100 GO response trials to the LINE pattern that also had the fastest response times. A detailed description of the experiment and data preprocessing have been presented elsewhere (Bressler et al. 1993; Ding et al. 2000).

Examples of the estimation framework covering two main applications to the analysis of LFP data are presented. First, the algorithm is applied to LFPs from three diverse cortical sites: striate, parietal, and somatosensory (arm area), illustrating the capability and versatility of the dVCA algorithm to recover single-trial parameters. Second, by subtracting out the event-related phase-locked potentials from the single-trial time series we estimate the ongoing activities. From the ongoing activity we study the event-related modulation of power spectra as a function of time.

4.1 Differential variability of event-related phase-locked components

Figure 4 shows the AERPs from three very diverse cortical sites: striate, parietal, and somatosensory (arm area). Inspection of the AERPs suggests the presence of two main components. The time intervals for the

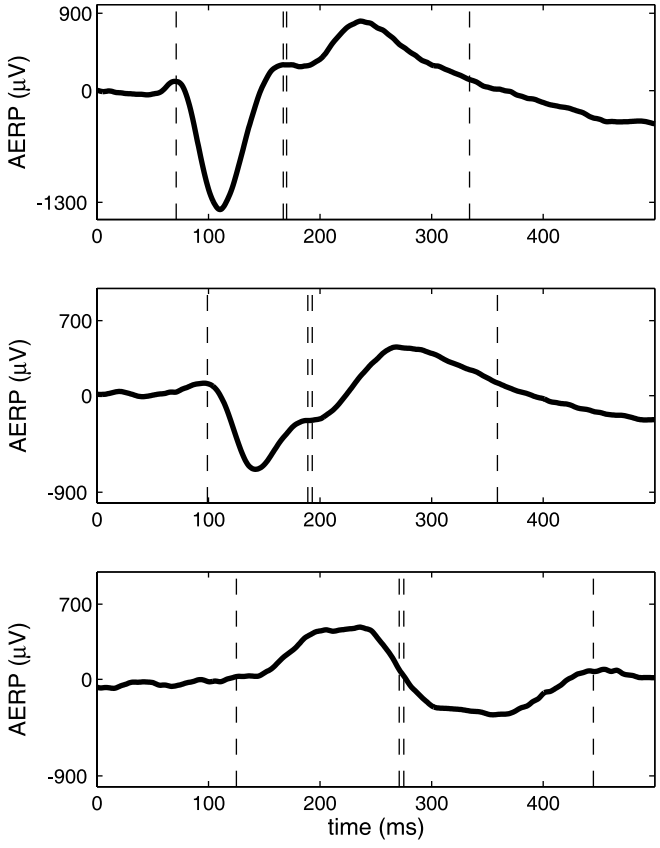

Fig. 4. Event-related phase-locked components and AERPs (see text for details)

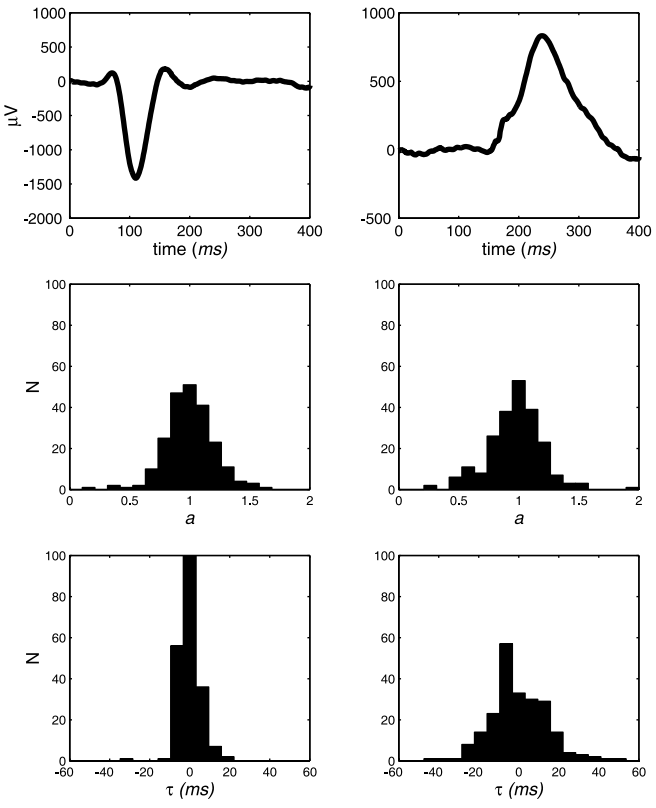

Fig. 5. Estimated components and estimated parameter distributions for the striate channel. Ensemble size: 222 trials. Two iterations of the algorithm were performed

components are delineated by dashed vertical lines. Figures 5, 6, and 7 show the results of applying the dVCA estimation algorithm to the two components from each of the three cortical areas. The left (right) columns of the three figures are the reconstructed waveforms (top), histograms of the estimated amplitude scaling factors (middle), and histograms of the estimated latency shifts (bottom) of the first (second) component. We sometimes also refer to the first and second components as the early and later components. It is interesting to note that the latency shift histograms and amplitude scaling factor histograms (bottom plots) all show a single peaked distribution, suggesting that the estimation is capturing the trial-to-trial variations of the underlying events that are relatively phase-locked to the stimulus onset.

Two observations are in order. First, for all the cortical sites, the variability in both amplitude scaling factor and latency shift is larger for the later component (right column in Figs. 5, 6, and 7) in comparison with the earlier one (left column in Figs. 5, 6, and 7), as indicated by the broader latency shift distributions. This by itself argues for the existence of differential variability of the event-related phase-locked components. Although the AERP may suggest a monolithic structure, the differential variability entails that the event-related components should be treated individually. Second, the variability both in amplitude scaling factor and latency shift increases as the cortical site becomes more distant from the primary visual cortex, suggesting that other intervening processes contribute to the increased variability.

Given the trial-to-trial variability of latency shifts and amplitude scaling factors, one is naturally led to investigate their relations to the variability of behavioral variables (e.g., relationships between response time and latency shift or amplitude scaling) and also to examine the relations between the components' parameters (i.e., between amplitude scaling factors and latency shifts). Here we focus on the relationships between the single-trial response time (RT) and component latency shifts. Scatter plots in Fig. 8 show that significant correlation appears for both components at the parietal site (middle plots), although it is much higher for the second one, indicating that this component is more related to the motor phase of the task than the sensory phase [r^2 values of 0.09 and 0.25 for the first and second components ($p < 0.001$); the two coefficients were significantly different ($p < 0.02$)]. Indeed, the mean response time is around 260 ms after stimulus onset, which is about the same time as the peak of the second parietal component. A much stronger correlation with RT is observed for the components in the somatosensory site (top plots) (r^2 values of 0.60 and 0.62 ($p < 0.001$) for the first and second component, respectively). The first component actually precedes the single-trial RT, indicating that this component arises possibly as an early activation from frontal cortex (motor or premotor). No significant

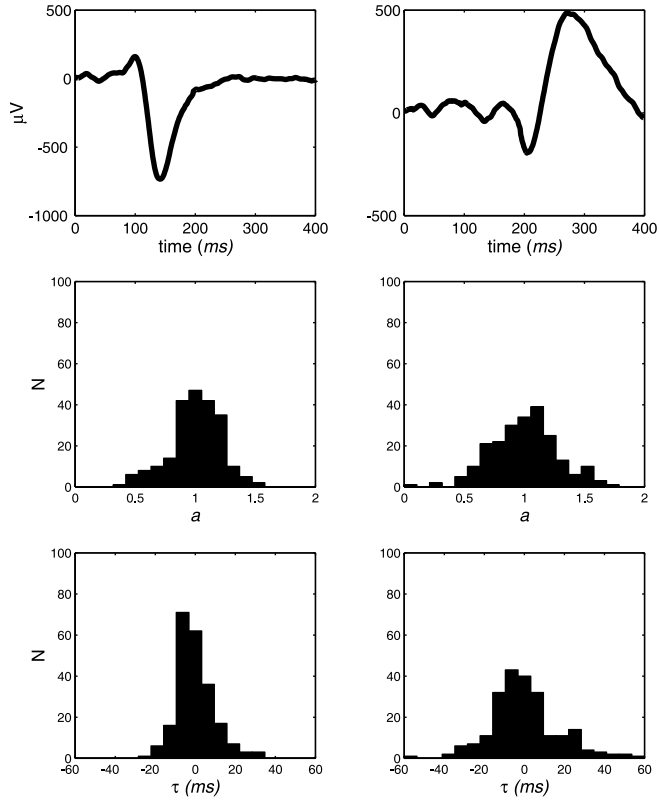

Fig. 6. Estimated components and estimated parameter distributions for the parietal site. The same conventions as in Fig. 5 are employed. Ensemble size: 222 trials (see text for details)

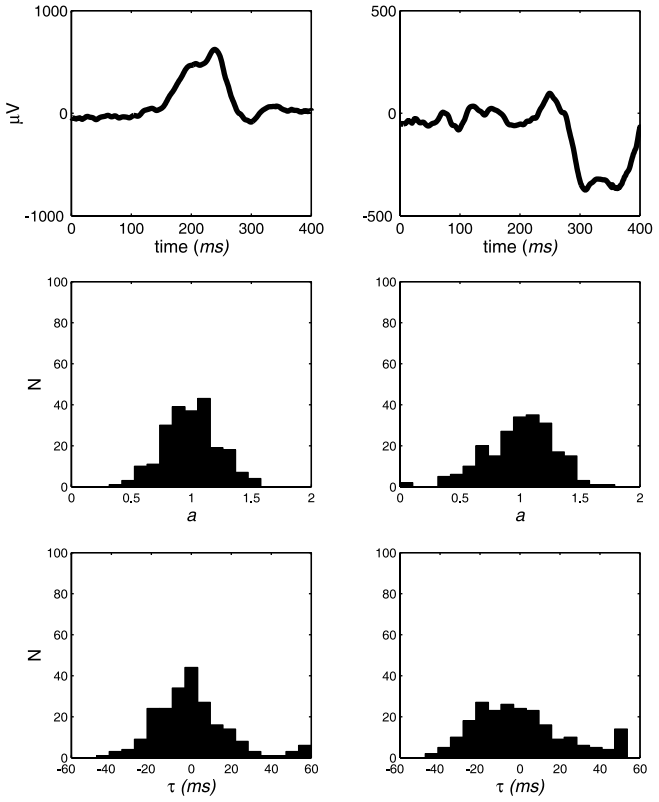

Fig. 7. Estimated components and estimated parameter distributions for the somatosensory site. The same conventions as in Fig. 5 are employed. Ensemble size: 222 trials (see text for details)

correlations are observed for the striate site ($p < 0.05$) (bottom plots).

To provide supporting evidence for the above correlation results, we sorted the trials according to the level of RT and plotted the single-trial time series as a gray-scale-coded "raster plot" in Fig. 9. The RT is superimposed on the raster plots as the black curve. It is quite apparent that for the striate site (top plot) little relation exists between RT and the component latency. For the somatosensory site (bottom plot), the latency for both components clearly tracks the RT curve.

Another useful application of single-trial parameter estimation is in the investigation of cortical processes at different cortical sites that appear to be related by examination of their AERPs. The AERPs alone, however, cannot reveal whether transient phase-locked processes are related. An illustrative case is presented for the striate and the somatosensory channel. As seen in Fig. 4, the second component at the striate site peaks at about the same time as the first component at the somatosensory site, just before the mean response time (260 ms). However, the single-trial latency shifts for the somatosensory component are highly correlated with RT, while there is no significant correlation for the striate site (Fig. 8). Further corroboration of this result is obtained by visual inspection of the single-trial waveforms sorted according to the respective single-trial RTs (Fig. 9). It is clear that, although the striate and somatosensory AERPs exhibit components at similar times, indicated by the temporal coincidence of peaks or large fluctuations, the activation in the somatosensory cortex systematically tracks the RT, whereas the striate activation seems to behave independently. This indicates that the activations measured by these potentials are related to different and possibly largely independent processes.

4.2 Separation of ongoing activity and event-related phase-locked components

Single-trial LFP time series can stem from many distinct cortical processes. In early sensory cortices, for example, phase-locked transient signals are thought to arise mainly from the response to thalamic inputs, while modulations observed in ongoing activity signals are thought to represent changes in effective connectivity or other intrinsic parameters of the cortical networks (Pfurtscheller and Lopes da Silva 1999). An important step in the analysis of LFPs recorded in the event-related paradigm is then to decompose the time series into these two main signals. Once the decomposition is achieved, the analysis proceeds by the investigation of the temporal dynamics in each of the decomposed signals separately.

The model of event-related recordings presented in Sect. 2 implies that simple subtraction of the AERP from each single-trial time series does not result in the desired decomposition due to the trial-to-trial variability

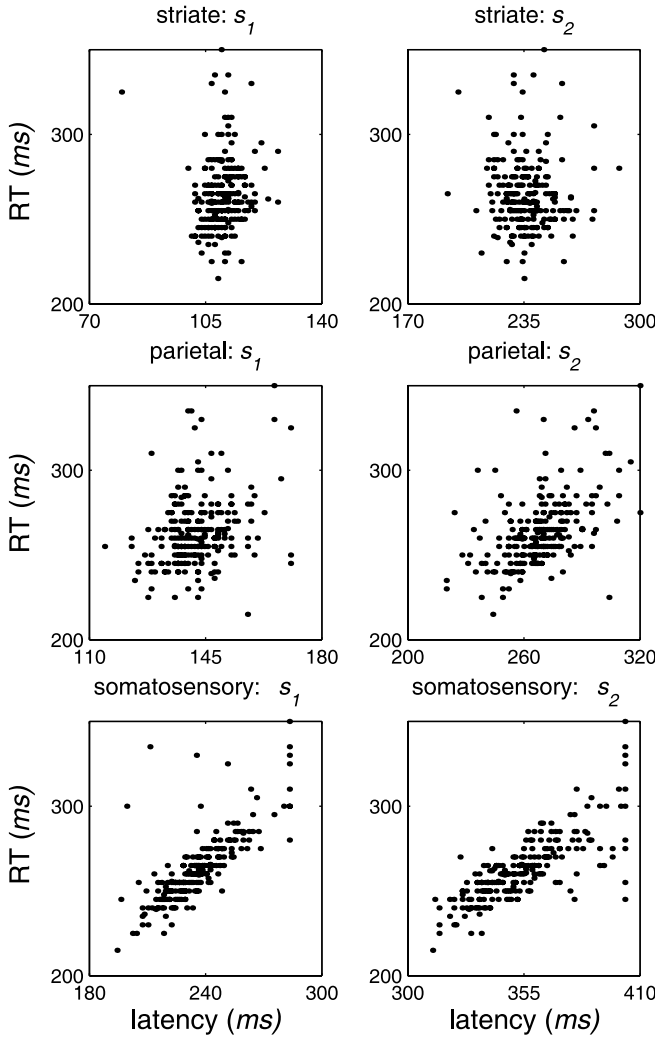

Fig. 8. Correlation between response time (RT) and single-trial latency shift for each estimated component in three cortical sites (see text for details)

of event-related activity. The problems resulting from this simple-minded decomposition have been analyzed in Truccolo et al. (2002a). We reproduce some basic arguments here for completeness. For simplicity, consider the case of a single component with trial-to-trial variability in amplitude scaling factors only:

$$x_r(t) = a_r s(t) + \eta_r(t) \ . \tag{4.1}$$

The ensemble average (AERP) based on this model becomes $\langle x_r(t)\rangle_r = \langle a_r\rangle s(t)$. After the AERP is subtracted from each single trial, the residual time series is given by

$$\begin{aligned}\xi_r(t) &= a_r s(t) - \langle x_r(t)\rangle_r + \eta_r(t) \\ &= (a_r - \langle a_r\rangle)s(t) + \eta_r(t) \\ &= S_r(t) + \eta_r(t) \ ,\end{aligned} \tag{4.2}$$

which contains two components – the ongoing activity $\eta_r(t)$ and a component related to the AERP waveform

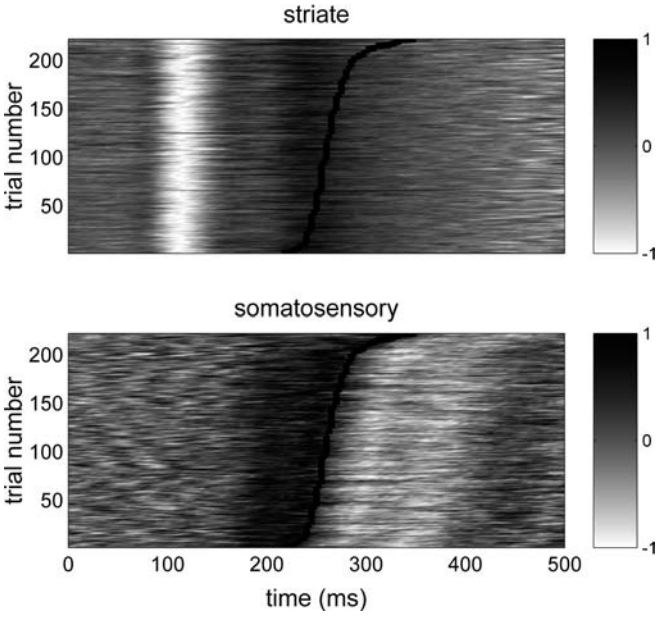

Fig. 9. Single-trial latencies and RT. The amplitudes are denoted by the grayscale and have been normalized for illustration purposes

$$S_r(t) = (a_r - \langle a_r\rangle)s(t) \ . \tag{4.3}$$

Thus, a phase-locked component still remains in the residual LFP time series even after the AERP is subtracted from each trial. Consequently, the variance of $x_r(t)$ over the ensemble of trials at a given time t becomes

$$\begin{aligned}\sigma_x^2(t) &= \langle [\xi_r(t)]^2\rangle_r \\ &= \langle [S_r(t)]^2\rangle_r + \langle [\eta_r(t)]^2\rangle_r \\ &= \sigma_S^2(t) + \sigma_\eta^2(t) \ ,\end{aligned} \tag{4.4}$$

which is also clearly modulated by the time dependence of $\sigma_S^2(t)$. Similarly, the power spectral density time function of the residual, $\langle |\xi_r(f,t)|^2\rangle$, computed in a sliding time window centered at time t will also exhibit temporal modulations at frequencies characteristic of the AERP waveform. Specifically, since $s(t)$ is often oscillatory with a very distinct main frequency, the quantity $\langle |\xi_r(f,t)|^2\rangle$ at this frequency will be modulated and should also exhibit a significant increase during the event-related response time period. In other words, even if the ongoing activity is not changing significantly in time, a significant transient change in the ensemble variance and power time functions will be observed, with the time course of this transient being related to $s(t)$ and to the shape of the AERP. Another important possibility considered below appears in the case where large transients resulting from the remnant phase-locked component can mask important changes in the ongoing activity.

Therefore, rather than simply subtracting the AERP, a better solution for the decomposition is obtained when the single-trial phase-locked components are estimated and subtracted from each single-trial time series. The ongoing activity, $\hat{\eta}_r(t)$, is then obtained as

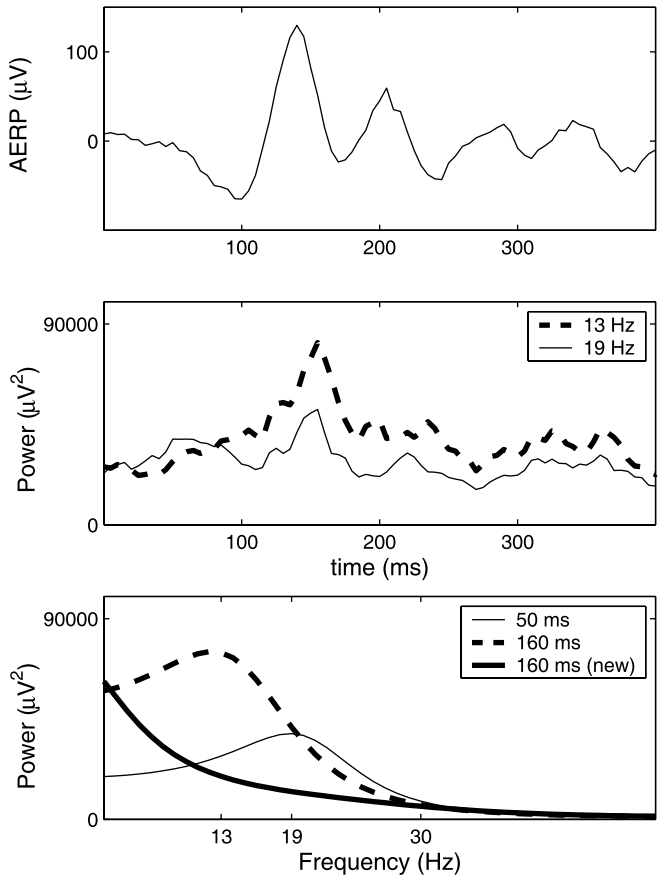

Fig. 10. Discriminating event-related transients in ongoing activity from transients in phase-locked components. The *middle plot* shows the power at 13 and 19 Hz as a function of time computed from the AERP-subtracted residual time series. *Bottom plot*: power spectra computed from a short window centered at 50 and 160 ms (*dashed curve*: spectrum from AERP subtracted time series; *thick curve*: spectrum from the time series after the single-trial phase-locked component had been subtracted

$$\hat{\eta}_r(t) = x_r(t) - \sum_{n=1}^{N} \hat{a}_{nr}\hat{s}_n(t - \hat{\tau}_{nr}) \ . \tag{4.5}$$

We now illustrate this approach by looking at the time course of neural oscillations observed from a frontal site (superior principle) for one of the monkeys. The 100 trials corresponding to the fastest response times to the GO response (stimulus LINE) were employed. This choice was done primarily to highlight the modulation of ongoing activity between the pre- and postsensory activation periods.

Figure 10 (top plot) shows the AERP's waveform with a characteristic frequency around 13 Hz and a main peak at $\sim$ 130 ms following stimulus onset. To contrast the performance of the ongoing process estimation in Eq. 4.5 with that of the simple AERP subtraction method, we first compute the power spectrum time function on the residual time series that is obtained by subtracting out the AERP from each single-trial time series. The spectrum is derived by fitting a fifth-order autoregressive model to the residual time series using a 50-ms-long analysis window that is shifted one data point at a time (Ding et al. 2000) through the trial. As shown by the thin solid line in the bottom plot of Fig. 10, there is a clear peak in the spectrum at 19 Hz for the time window centered at 50 ms. This oscillatory phenomenon is also characteristic of the whole prestimulus period. Analysis done in Liang et al. (2002) implicated the 19-Hz oscillation corresponding to this peak as possibly related to anticipatory attention. To rule out the possibility that this oscillation was linked to reward anticipation or motor preparation, it was necessary to study the time course of the oscillation. The solid line in the middle plot of Fig. 10 is the power at 19 Hz as a function of time. It shows no decline as the trial progresses. But the small peak seen around 130 ms poststimulus (coinciding with the main peak of the AERP) and the time course of the power at 13 Hz (dashed line in the middle plot of Fig. 10), peaking also at 130 ms, suggest that the persistence of the power at 19 Hz is an artifact due to the interference from the evoked activity that is not cleanly removed by the simple subtraction of the AERP.

To address this possibility, we estimate the evoked activity on a single-trial basis. A single component is modeled with the initial guess for its waveform obtained from the AERP around 130 ms. The final estimated component's waveform preserves the main shape of the AERP, including its characteristic oscillation around 13 Hz. The power spectrum performed on the ongoing activity obtained from (3.5) is then computed. A single time snapshot for the power computed from a window at 160 ms is shown in Fig. 10 (bottom plot). Interestingly, the power peak at 13 Hz is no longer present, and the 19 Hz ongoing activity is reduced to one third of the presensory activation level. Based on these findings it is concluded that the prestimulus oscillation in the beta frequency range is not related to motor preparation or reward but to anticipatory attention.

5 Discussion

The main contribution of this paper is to present, from a Bayesian perspective, a unified framework for the estimation of single-trial ERP parameters where all the assumptions in the inference process are made explicit. The resulting algorithm is referred to as differentially variable component analysis (dVCA). The framework makes it easy to analyze and relate different algorithms. For example, Woody's matched filter method is seen as a way of implementing the MAP solution for the posterior probability under specific assumptions. The dVCA approach is thoroughly tested on simulated data and then applied to the analysis of LFP data from behaving monkeys, where we demonstrate the benefits of single-trial parameters and the proper extraction of the ongoing process in delineating the physiological meaning of the various aspects of LFP recordings. Our experience shows that the dVCA implemented here works well for the estimation of relatively low-frequency and large-amplitude event-related components. High-frequency components (e.g., evoked gamma bursts) may

represent a challenge to this heuristic algorithm because of their low SNR and the larger impact of trial-to-trial latency variability when the recordings are obtained from broad field potentials.

The framework deals mainly with the analysis of recordings from a single channel. This channel-by-channel strategy is reasonable for multichannel recordings where the separation between electrodes is sufficiently large that different channels are not measuring the same generators. However, for recordings made with densely packed sensors such as in high-density scalp EEG recordings, the incorporation of multiple channels into the component estimation process becomes important. A multichannel implementation within this same framework has been briefly described by Truccolo et al. (2002b) and Knuth et al. (2001) and used to perform a detailed investigation of intralaminar cortical interactions (Shah et al. 2001). It is worth noting that many currently used component separation techniques, including cluster analysis, factor analysis, PCA, and ICA (Bell and Sejnowski 1995; Jung et al. 1999), can be seen as special cases of the multichannel multicomponent Bayesian approach depending on variations of the underlying model, the choice of the prior probabilities, and the parameters to be estimated (Knuth 1999; Hinton and Ghahramani 1997; Knuth et al. 2002; Knuth et al., manusript in preparation).

Improvements in the proposed framework are expected to come especially in three main aspects. First, the current LFP model assumes the linear superposition of event-related phase-locked components and ongoing activities. Future experimental evidence may result in better models, allowing for nonlinear effects as well as for more informative priors for the estimated parameters. Second, the current approach models the ongoing activity as white noise. This is clearly an oversimplification. There are some important considerations regarding improvements in this respect, though. To begin with, we note that the inclusion of temporal correlations of ongoing activity does not change the location of the modes of the posterior probability. (This partly explains why in Sect. 3 we are able to get good estimation results even though the ongoing process is modeled as 1/f noise.) It only affects their width, i.e., the spread of the probability, and by consequence the variance of the estimators. In this way the MAP solution remains the same regardless of whether the ongoing activity is assumed to be white or temporally correlated. Furthermore, the inclusion of the autocorrelation function or any other higher-order statistics in the time domain is very cumbersome. A better solution would be to rewrite the generative model in the spectral domain (Brillinger 1975) and use the power spectrum of the preevent LFP as an estimate for the ongoing activity. However, based on our experience and the considerations of Sect. 4.2, the ongoing activity may be highly nonstationary during the transition from pre- to postevent segments. Future research could consider the alternative of including the power spectrum of the ongoing activity as a parameter to be simultaneously estimated. That would also reduce the risk of overfitting the data, in other words spurious "locking" of the estimated components' waveforms to the ongoing activity.

Finally, a third aspect for improvement refers to the algorithmic implementation itself. The choice of the number of event-related components and how to model each component is a key aspect of the algorithm. The solution implemented in this work depends largely on inspection of the AERP waveform and is somewhat ad hoc. An improvement in this aspect may be obtained by taking explicit advantage of differential variability for amplitude scaling factors and latency shifts between different components (Sect. 4.1). The alternative for an automatic and more principled way of selecting the number of components suggested in Sect. 2.3 makes use of such differential variability as a source of separability for the component analysis. This approach is currently under investigation in the context of a multichannel multicomponent framework (Knuth et al., manusript in preparation).

Acknowledgements. This work was supported by NIMH (MH64204 and MH42900), NSF (IBN0090717), ONR (N000149910062), T32M07288, NARSAD Young Investigator Award (KHK), NASA IDU/IS/CICT Program, and the NASA Aerospace Technology Enterprise.

References

Bartnik EA, Blinowska KJ, Durka PJ (1992) Single evoked-potential reconstruction by means of wavelet transform. Biol Cybern 67: 175–181

Bell AJ, Sejnowski TJ (1995) An information maximization approach to blind separation and blind deconvolution. Neural Comput 7(6): 1129–1159

Bressler SL, Coppola R, Nakamura R (1993) Episodic multiregional cortical coherence at multiple frequencies during visual task performance. Nature 366: 153–156

Brillinger DR (1975) Time series analysis: data analysis and theory. Holt, New York

Coppola R, Tabor R, Buchsbaum MS (1978) Signal to noise ratio and response variability measurements in single trial evoked potentials. Electroencephalogr Clin Neurophysiol 44: 214–222

Ding M, Bressler SL, Yang W, Liang H (2000) Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol Cybern 83: 35–45

Haig AR, Gordon E, Rogers G, Anderson J (1995) Classification of single-trial ERP sub-types: application of globally optimal vector quantization using simulated annealing. Clin Neurophysiol 94: 288–297

Hinton GE, Ghahramani Z (1997) Generative models for discovering sparse distributed representations. Philosoph Trans R Soc Lond Ser B Biol Sci 352(1358): 1177–1190

Jaskowski P, Verleger R (1999) Amplitude and latencies of single-trial ERP's estimated by a maximum-likelihood method. IEEE Trans Biomed Eng 46(8): 987–993

Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (1999) Analyzing and visualizing single-trial event-related potentials. In: Proceedings of the conference on advances in neural information processing systems (NIPS1999), Denver, 30 November–2 December 1999, pp 118–124

Knuth KH (1999) A Bayesian approach to source separation. In: Cardoso JF, Jutten C, Loubaton P (eds) Proceedings of the 1st

international workshop on independent component analysis and signal separation (ICA'99), Aussios, France, 11–15 January 1999, pp 283–288

Knuth KH, Truccolo WA, Bressler SL, Ding M (2001) Separation of multiple evoked components using differential amplitude and latency variability. In: Proceedings of the 3rd international conference on independent component analysis and blind signal separation, San Diego, 9–12 December 2001, pp 463–468

Knuth KH, Clanton ST, Shah AS, Truccolo WA, Ding M, Bressler SL, Trejo LJ, Schroeder CE (2002) Multiple component event-related potential (mcERP) estimation. In: Abstracts of the Society for Neuroscience annual meeting, San Diego, 2–7 November 2002, vol 28, Prog no 506.4.

Liang H, Bressler SL, Ding M, Truccolo WA, Nakamura R (2002) Synchronized activity in prefrontal cortex during anticipation of visuomotor processing. Neuroreport 13(6): 2011–2015

Lange DH, Pratt H, Inbar GF (1997) Modeling and estimation of single evoked brain potential components. IEEE Trans Biomed Eng 44(9): 791–799

McGillem CD, Aunon JI, Pomalaza CA (1985) Improved waveform estimation procedures for event-related potentials. IEEE Trans Biomed Eng 32(6): 371–379

Nowak LG, Bullier J (1997) The timing of information transfer in the visual system. In: Rockland KS (ed) Cerebral cortex, vol 12. Plenum Press, New York, pp 205–241

Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110: 1842–1857

Pham DT, Mocks J, Kohler W, Gasser T (1987) Variable latencies of noisy signals: estimation and testing in brain potential data. Biometrika 74(3): 525–533

Quiroga RQ (2000) Obtaining single stimulus evoked potentials with wavelet denoising. Physics D 145(3–4): 278–292

Quiroga RQ, Garcia H (2003) Single-trial event-related potentials with wavelet denoising. Clin Neurophysiol 114(2): 376–390

Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. J Neurophysiol 79: 3272–3278

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cerebral Cortex 8: 575–592

Shah AS, Knuth KH, Mehta AD, Fu KG, Johnston TA, Dias EC, Truccolo WA, Ding M, Bressler SL, Schroeder CE (2001) Functional connectivity between visual structures in behaving monkeys. In: Abstracts of the Society for Neuroscience annual meeting, San Diego, 10–15 November 2001, vol 27, Prog no 721.3

Sivia DS (1996) Data analysis: a Bayesian tutorial. Oxford University Press, New York

Truccolo WA, Ding M, Knuth KH, Nakamura R, Bressler SL (2002a) Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. Clin Neurophysiol 113: 206–226

Truccolo WA, Knuth KH, Bressler SL, Ding M (2002b) Bayesian analysis of single trial cortical event-related components. In: Fry RL (ed) Bayesian inference and Maximum entropy and Bayesian methods in science and engineering. In: Proceedings of AIP conference, Baltimore, MD, 4–9 August 2001, 617: 64–73. American Institute of Physics, Melville, NY

Woody CD (1967) Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med Biol Eng 5: 539–553