

Decomposing Stimulus and Response Components in ERP Recordings

Jun Zhang^{1,CA}, Gang Yin²

(1 Department of Psychology, The University of Michigan)

(2 Sichuan Cancer Hospital, Radiotherapy Center, Chengdu, China)

Abstract: Event-related potentials (ERPs) reflect the brain activities related to specific behavioral events, and are obtained by averaging across many trial repetitions to improve signal-to-noise ratio. Individual trials are aligned either to the occurrence of a specific event, such as the onset of stimulus (s-aligned), or to the onset of the behavioral response (r-aligned). However, the s-aligned and r-aligned ERP waveforms do not purely reflect, respectively, underlying stimulus (S-) or response (R-) component waveforms, since they cross-contaminate each other in the s-aligned or r-aligned ERP waveforms due to variability in reaction times (measured as the time lapse between the stimulus onset and response onset) across these trials. Here, we provide a method for decomposing of ERP according to two markers (stimulus and the behavioral response). The method was first developed in Zhang (1998) with full mathematical formulation. In Yin et al. (2009), this method was extended to deal with three or more markers in a single trial, with recovered individual ERP components time-locked to those markers. In Yin and Zhang (in press), noise reduction techniques were applied so that the method can robustly recover component waveforms despite variances in the sampling of reaction-time distributions. This Chapter provides a tutorial to this method, with demonstrations using both simulated and real data from two experiments—a Go/NoGo experiment about image classification and recognition and an attention pre-cuing experiment involving three behavioral markers.

CA email: junz@umich.edu

1 Introduction

ERP are evoked activities of the brain that are averaged over many trials to remove trial-by-trial noise and to improve signal-to-noise ratio. When the individual trials are aligned with respect to the stimulus onset, the ensemble average is called stimulus-aligned ERP, with waveforms demonstrating such components like P300 (e. g. , Squires et al. , 1975) and N400 (Kutas and Hillyard, 1980). When the ensemble averaging is performed by aligning individual trials to behavioral response onset, we call it response-aligned ERP, with waveforms demonstrating such components like error-related negativity (ERN) (Gehring et al. , 1993; Falkenstein et al. , 1995) and lateralized readiness potential (LRP) after inter-hemispheric subtraction. In experimental paradigms where a behavioral response is required in response to the stimulus presentation, stimulus-aligned ERP average and response-aligned ERP average may yield different waveforms, since the subject does not respond with uniform reaction time across the ensemble of trials. This creates a non-trivial problem of interpreting these ensemble averages, since either stimulus-aligned or response-aligned ERP average may contain both components related to the encoding and analysis of stimulus (and hence better time-locked to the stimulus onset event) and components related to the preparation and execution of response (and hence more tightly coupled with the moment of behavioral response onset). The overlapping event-related components in the ERP average waveforms can cause difficulty and confusions in the interpretation of underlying neural mechanisms for ERP. For example, in the study of response inhibition by Go/Nogo paradigm, a question arises whether the N2/P3 differences between Go and Nogo trials in the overall ERP mainly comes from the accompanying motor response in the Go trials versus the inhibitory processes associated with the Nogo trials.

Traditional methods for decomposing multiple components in ERP/EEG and fMRI dataset, such as Independent Component Analysis (ICA) and Principal Component Analysis (PCA) , have strong constraints in their application. For example, ICA assumes the sources to be statistically independent, while PCA is built upon the premise that various sub-components of a waveform are orthogonal to one another. Independence and orthogonality are very strong mathematical assumptions—there is no reason to expect that neurally distinct modes in the data should be orthogonal to or in-

dependent of one another (Mitra et al., 1999). To complicate the matter, the non-uniformness of response/reaction time (RT), defined as the time elapse between the stimulus onset and the behavioral response onset for each trial, implies a kind of non-stationarity across the ensemble of trials in an ERP experiment. It is all too common to calculate ERP averages with trials aligned with respect to the onset of stimulus (response) and to conclude that the resulting ERP waveforms reflect stimulus (respectively, response) process, as if non-stationarity can be “averaged” out. To repeat, one cannot conclude that the stimulus-aligned (or response-aligned) ERP average contains purely those elements that are time-locked to stimulus (response, respectively) onset, since both these component waveforms contribute towards the ERP average waveforms, and are hence cross-contaminated therein.

Furthermore, quite a number of psychological experiments involve more than two behavioral events in a single trial. For example, in the study of visual spatial attention where a peripheral or central cue is utilized, a single trial will involve three event-related components related to cue presentation, stimulus presentation and behavioral response, respectively. A similar situation occurs in the study of ERN when EMG measurement is applied (Gehring & Fenesik, 1999)—in that case, the onset of EMG, along with the stimulus onset and response onset, constitute three independent behavioral events with variable lapses between their onset times. This provides an opportunity to understand underlying mechanisms of ERN, but also presents a challenge to unconfounding the ERP component waveforms individually time-locked to those events. We need a general methodology to decompose ERP components given those various behavioral events.

This Chapter reviews a methodology and algorithm to recover the stimulus-and response-locked waveforms first advanced in Zhang (1998). There, given stimulus-aligned and response-aligned ERP averages, as well as the empirically obtained reaction time distribution, a unique recovery into stimulus-locked and response-locked component waveforms is provided. The algorithm can be implemented through either the Fourier transform method or an iterative procedure that is guaranteed to converge. Following the same logic, the method of Zhang (1998) has now been extended in Yin et al. (2009) to separate N components (N event related waveforms) in ERP recordings given N behavioral markers with associated event time distributions (especially, the $N = 3$ case useful for many paradigms). One of the key challenges in applying these methods to real ERP data sets is the issue of input noise, especially when

the amount of trials in an experiment is limited. Recently, we investigated this issue thoroughly, and advanced Wiener deconvolution approach for noise control (Yin & Zhang, 2011). In this Chapter, we will demonstrate our decomposition method and denoising techniques using both simulated and real ERP data.

2 Method

We present here a method to decompose ERP component waveforms that takes, as input, ERP ensemble averages that are aligned to behavioral event points (the most common ones are stimulus onset and response onset). We will illustrate this method with $N = 2$ event points (behavioral markers) first, then proceed to $N = 3$ event points, and finally to a general N . We then describe a method of noise control that will make our method robust against both noise in ERP recordings and insufficient sampling (i.e., total trial number) of reaction-time distribution.

2.1 Decomposition for Stimulus and Response Components with $N = 2$ markers

Let us assume that the ERP waveform on an individual trial consists of an underlying stimulus-locked component waveform, denoted by $f_s(t)$, with $t = 0$ referring to stimulus onset, and an underlying response-locked component waveform, denoted by $f_r(t)$, with $t = 0$ referring to response onset, respectively, plus some noise waveform ξ_i that varies across different trials labeled by i . Denote the reaction time (RT) on trial i as t_i . Then the stimulus-aligned ERP average waveform, denoted by $F_s(t)$, and the response-aligned waveform, denoted by $F_r(t)$, are constructed, respectively

$$F_s(t) = \frac{1}{n} \sum_i (f_s(t) + f_r(t - t_i) + \xi_i(t)) \approx f_s(t) + \frac{1}{n} \sum_i f_r(t - t_i) \quad (1)$$

$$F_r(t) = \frac{1}{n} \sum_i (f_s(t + t_i) + f_r(t) + \eta_i(t)) \approx f_r(t) + \frac{1}{n} \sum_i f_s(t + t_i) \quad (2)$$

Here, in the stimulus-locked calculation, where time zero is taken to be the stimulus onset, the response component is offset by an amount equal to t_i (the reaction time, which is the amount of time elapse since the onset of the stimulus); in the response-locked calculation, where time zero is taken to be the response onset, the stimulus component is offset by an amount $-t_i$ (which is when the stimulus onset was, from the response reference-zero perspective). The variables ξ_i and η_i represent white

noise, measured with respect to stimulus onset and response onset, on trial i (total trial number is n), with $\xi_i(t) = \eta_i(t - t_i) \iff \xi_i(t + t_i) = \eta_i(t)$; they are averaged “out” during ensemble averaging assuming that the noise is “white”. But the summation terms in the expression of $F_s(t)$ and $F_r(t)$ do not vanish in general; they represent residual contaminations: $F_s(t) \neq f_s(t)$, $F_r(t) \neq f_r(t)$.

From Equations (1) and (2), and casting them in the limiting case of continuous time with infinite trials (n goes to infinity), we can derive the following two mathematical equations

$$F_s(t) = f_s(t) + \frac{1}{n} \sum_{i=1}^n f_r(t) * \delta(t - t_i) = f_s(t) + f_r(t) * g(t) \quad (3)$$

$$F_r(t) = f_r(t) + \frac{1}{n} \sum_{i=1}^n f_s(t) * \delta(t + t_i) = f_r(t) + f_s(t) * g(-t) \quad (4)$$

where we defined $g(t) = \frac{1}{n} \sum_{i=1}^n \delta(t - t_i)$, the reaction-time distribution. Its

Fourier transform is (we use the tilde sign \sim to denote the frequency domain representation)

$$\tilde{g}(w) = \int g(t) e^{-i2\pi wt} dt.$$

Applying Fourier transform to the above two equations, we get

$$\tilde{F}_s(w) = \tilde{f}_s(w) + \tilde{f}_r(w) \tilde{g}(w)$$

$$\tilde{F}_r(w) = \tilde{f}_r(w) + \tilde{f}_s(w) \tilde{g}(-w)$$

Solving for the frequency components $\tilde{F}_s(w)$, $\tilde{F}_r(w)$ and then applying inverse-Fourier transform, Zhang (1998) obtained a closed-form solution:

$$f_s(t) = \frac{1}{2\pi} \int \frac{\tilde{F}_s(w) - \tilde{F}_r(w) \tilde{g}(w)}{1 - |\tilde{g}(w)|^2} e^{iwt} dw \quad (5)$$

$$f_r(t) = \frac{1}{2\pi} \int \frac{\tilde{F}_r(w) - \tilde{F}_s(w) \tilde{g}(-w)}{1 - |\tilde{g}(w)|^2} e^{iwt} dw \quad (6)$$

where \sim represents the Fourier frequency spectrum. The solution can also be obtained by iterative methods (Zhang, 1998).

To summarize, assume we are given the following experimental data: (a) the stimulus-aligned ERP average waveform denoted $F_s(t)$, (b) the response-aligned ERP average waveform $F_r(t)$, and (c) the distribution of reaction times $g(t)$. We

can recover the stimulus component waveform time-locked to stimulus onset (“S-component”) $f_s(t)$ and the response component waveform time-locked to response onset (“R-component”) $f_r(t)$.

2.2 Decomposition of ERP Components with $N = 3$ markers

We now consider the three-component case ($N = 3$). Imagine an experimental paradigm where subjects perform a stimulus-response association task with stimulus pre-cuing. That is, some time after the start of a trial, a cue appears after which the stimulus appears under a variable delay, and finally the subject responds. The event sequence is: Cue-Stimulus-Response, and the onset times of each form the three behavioral markers. Suppose the evoked waveform of each trial involves three different components $f_c(t)$, $f_s(t)$ and $f_r(t)$, each of which is time-locked to the three different events, namely cue, stimulus and response respectively. Denote the onset time as, t_c , t_s and t_r , with $t_c < t_s < t_r$. Note that fixing any one of the t 's, the others are “random” variables in the sense that they vary across the trials (they may be under experimenter control). Typically, one chooses t_c as the reference zero—in this case the probability distribution of $t_s - t_c$ and that of $t_r - t_c$, which are treated as random variables, is denoted $g_1(t)$ and $g_2(t)$, respectively. When choosing t_s as the reference zero, the distribution of $t_r - t_s$ can be written as $g_1(-t)$ and $g_3(t)$, respectively. When choosing t_r as the reference zero, the distribution of t_c and t_s (relative to t_r) is $g_2(-t)$ and $g_3(-t)$, respectively. It is important to note that only *two* of the three distributions, $g_1(t)$, $g_2(t)$, $g_3(t)$ are independent, and the other can be derived knowing the two distributions. For instance, $g_2(t) = g_1(t) * g_3(t)$, where $*$ denotes convolution (note that this relates to, but is different from the three-marker case one of which is a presumptive decision-related marker, as discussed in Zhang (1998)). Denote $F_c(t)$, $F_s(t)$ and $F_r(t)$ as the ensemble averages of ERP aligned to t_c , t_s and t_r , respectively. Then the relationship between the pure event-related components $f_c(t)$, $f_s(t)$, $f_r(t)$ and the ensemble average signals $F_c(t)$, $F_s(t)$, $F_r(t)$ are

$$F_c(t) = f_c(t) + f_s(t) * g_1(t) + f_r(t) * g_2(t) \quad (7)$$

$$F_s(t) = f_s(t) + f_c(t) * g_1(-t) + f_r(t) * g_3(t) \quad (8)$$

$$F_r(t) = f_r(t) + f_c(t) * g_2(-t) + f_s(t) * g_3(-t) \quad (9)$$

These equations are derived using the same logic as for the two-component case

(with stimulus and response markers only). Equations (7)–(9) can be written in matrix notation:

$$\begin{bmatrix} F_e \\ F_s \\ F_r \end{bmatrix} = \begin{bmatrix} I & A & B \\ A^T & I & C \\ B^T & C^T & I \end{bmatrix} \begin{bmatrix} f_e \\ f_s \\ f_r \end{bmatrix} \quad (10)$$

where

$$A = \begin{bmatrix} g_1(1) & g_1(2) & \cdots & g_1(n) \\ g_1(n) & g_1(1) & \cdots & g_1(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(2) & g_1(3) & \cdots & g_1(1) \end{bmatrix}, \quad B = \begin{bmatrix} g_2(1) & g_2(2) & \cdots & g_2(n) \\ g_2(n) & g_2(1) & \cdots & g_2(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ g_2(2) & g_2(3) & \cdots & g_2(1) \end{bmatrix},$$

$$C = \begin{bmatrix} g_3(1) & g_3(2) & \cdots & g_3(n) \\ g_3(n) & g_3(1) & \cdots & g_3(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ g_3(2) & g_3(3) & \cdots & g_3(1) \end{bmatrix}$$

Here, I denotes the identity matrix (with 1's in diagonal entries and 0's elsewhere), and superscript T denotes matrix transpose. Note in the above matrix-form equations in the time domain, we have imposed the periodic condition in constructing the ERP averages. As a result, $g_1(t)$, $g_2(t)$ and $g_3(t)$ are all periodic when $t > n$ or $t < 0$, so only the values of $g_i(1)$, \dots , $g_i(n)$, $i = 1, 2, 3$ are involved in the computation. In frequency domain, the equations are

$$\begin{bmatrix} \tilde{F}_e(w) \\ \tilde{F}_s(w) \\ \tilde{F}_r(w) \end{bmatrix} = \begin{bmatrix} I & \tilde{g}_1(w) & \tilde{g}_2(w) \\ \tilde{g}_1(-w) & I & \tilde{g}_3(w) \\ \tilde{g}_2(-w) & \tilde{g}_3(-w) & I \end{bmatrix} \begin{bmatrix} \tilde{f}_e(w) \\ \tilde{f}_s(w) \\ \tilde{f}_r(w) \end{bmatrix} \quad (11)$$

2.3 Decomposition of ERP Components with N Event Points with $N > 3$

This formulation for the three-component case can be easily extended to $N > 3$ situations. A single trial ERP waveform is the superposition of N component waveforms that are time-locked to N corresponding behavioral events, with occurrence of these events marked by N variables t_1, \dots, t_N , ($t_1 < \dots < t_N$), among which $N - 1$ are random variables (in the sense discussed earlier). Let $f_i(t)$ ($i = 1 \dots N$) denote each of the N component waveforms, and $F_i(t)$ ($i = 1 \dots N$) denote each of the N ensem-

ble averages of ERP aligned with respect to those behavioral events. Denote $g_{i,j}(t)$ ($i = 1 \dots N$, $j = 1 \dots N$, $i \neq j$) as onset time distribution of event j aligned to a fixed event i . They are not all independent—only $N - 1$ of them are. We can calculate the time distribution of an arbitrary event onset time point relative to the other event onset time points. An analogous equation can be obtained

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix} = \begin{bmatrix} I & G_{1,2} & \cdots & G_{1,N} \\ G_{2,1} & I & \cdots & G_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N,1} & G_{N,2} & \cdots & I \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} \quad (12)$$

where

$$G_{i,j} = \begin{bmatrix} g_{i,j}(1) & g_{i,j}(2) & \cdots & g_{i,j}(n) \\ g_{i,j}(n) & g_{i,j}(1) & \cdots & g_{i,j}(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ g_{i,j}(2) & g_{i,j}(3) & \cdots & g_{i,j}(1) \end{bmatrix}$$

and $G_{i,j} = G_{j,i}^T$. In frequency domain, the matrix-form equation is

$$\begin{bmatrix} \tilde{F}_1(w) \\ \tilde{F}_2(w) \\ \vdots \\ \tilde{F}_N(w) \end{bmatrix} = \begin{bmatrix} I & \tilde{G}_{1,2}(w) & \cdots & \tilde{G}_{1,N}(w) \\ \tilde{G}_{2,1}(-w) & I & \cdots & \tilde{G}_{2,N}(w) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{G}_{1,N}(-w) & \tilde{G}_{2,N}(-w) & \cdots & I \end{bmatrix} \begin{bmatrix} \tilde{f}_1(w) \\ \tilde{f}_2(w) \\ \vdots \\ \tilde{f}_N(w) \end{bmatrix} \quad (13)$$

2.4 Methodological Limitation

While the above methods are rigorous and exact, it relies on a stationary reaction time distribution (or distributions, for $N > 2$ cases), which is an empirically observed quantity during the experiments. Let us focus on the $N = 2$ case for ease of illustration. RT distribution $g(t)$ is estimated using the empirical distribution (frequency count of proportion of trials associated with a particular RT value t as measured from a fixed stimulus onset moment, denoted as $t = 0$, to the moment of behavioral response on individual trials). It is convenient to work in the frequency (as opposed to time) domain. Define the matrix $H(w)$ as

$$H(w) = \begin{bmatrix} 1 & \tilde{g}(w) \\ \tilde{g}(-w) & 1 \end{bmatrix} \quad (14)$$

Equations (3) and (4) can now be written in matrix form

$$y = Hx; \quad y = \begin{bmatrix} \tilde{F}_s(w) \\ \tilde{F}_r(w) \end{bmatrix}, \quad X = \begin{bmatrix} \tilde{f}_s(w) \\ \tilde{f}_r(w) \end{bmatrix} \quad (15)$$

Note that the complex conjugate (denoted by c) of $g(w)$ is

$$\tilde{g}^c(w) = \int g(t) e^{iwt} dt = \int g(t) e^{-i(-w)t} dt = \tilde{g}(-w)$$

This means $|\tilde{g}(w)| = |\tilde{g}(-w)|$, which implies that the H matrix in (15) is Hermitian, meaning that its conjugated transpose (denoted by $\#$) equals itself $H^\#(w) = H(w)$. From well-known facts in matrix analysis, the H matrix has real eigenvalues, with associated eigenvectors orthogonal to each other. To see that, we examine the eigen-equation, $Ht = \lambda t$. The eigenvalues can be solved from:

$$\text{Det } |H - \lambda I| = (1 - \lambda)^2 - |\tilde{g}(w)|^2 = 0$$

The solutions (eigenvalues of H) are

$$\lambda_1 = 1 + |\tilde{g}(w)|, \quad \lambda_2 = 1 - |\tilde{g}(w)|$$

Substituting into the eigen-equation yields the corresponding orthonormal eigenvectors

$$t_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ e^{i\phi} \end{bmatrix} \quad t_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -e^{i\phi} \end{bmatrix} \quad (16)$$

Denote the condition number of H as the ratio of its two eigenvalues λ_1, λ_2

$$\text{Cond}(H) = \frac{\lambda_1}{\lambda_2} = \frac{1 + |\tilde{g}(w)|}{1 - |\tilde{g}(w)|}$$

The condition number will determine how input (at the corresponding frequency component) will be amplified—the larger the value of $\text{Cond}(H)$, the greater the amplification of small fluctuating noise. Since $0 \leq |\tilde{g}(w)| \leq 1$, $\text{Cond}(H)$ is monotonically related to $|\tilde{g}(w)|$ and bounded below by 1 and unbounded from above. Figure 1 shows the relationship between $|\tilde{g}(w)|$ and $\text{Cond}(H)$, it can be seen that as frequency goes to zero, $|\tilde{g}(w)|$ approaches 1, resulting in the explosion of the condition number of H , and hence disproportionate amplification of input noise (Figure 1).

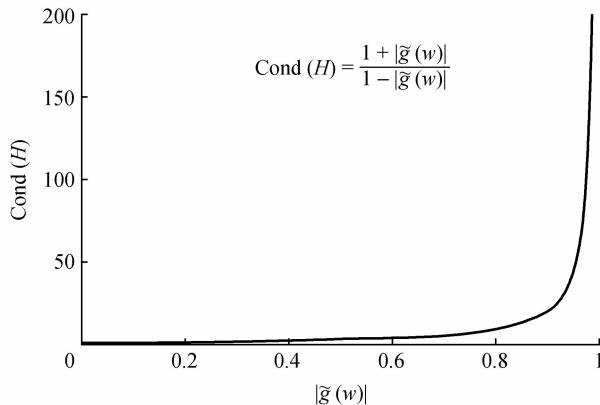


Figure 1 The relationship between $|\tilde{g}(w)|$ and $\text{Cond}(H)$

The explosion of condition number at low frequencies turns out to be quite detrimental (if not fatal) in recovering component waveforms time-locked to stimulus and to response, because even small noise (at low frequency end) can be amplified, resulting in change of overall shape of the ERP waveform. Figure 2 demonstrates how artifacts may be introduced, using the vanilla formulae from Zhang (1998) when different amounts of noise are injected. We first create the ground-truth waveforms $f_s(t)$, $f_r(t)$ with a known RT distribution. Then we generate an ensemble of trials, and from which, the s-aligned and r-aligned ERP average waveforms. Plugging these s-and r-aligned waveforms as well as the RT distribution as input to Equations 5 and 6 allows us to calculate $f_s(t)$, $f_r(t)$. With different low frequency noise added (SNR = 20dB, 30dB, 40dB), Equations (5) and (6) gives different outputs of $f_s(t)$, $f_r(t)$; the recovered waveforms have more distortions with more low-frequency noise introduced. We see that the recovered S-waveform and R-waveform appear to contain corruptions that are nevertheless themselves “correlated”! These distortions, when treated as fictitious S-component and R-component waveforms, will cancel themselves in the stimulus-aligned and response-aligned ensemble averages—that is, they result in zero amplitude (or noise). These fictitious components live in the “null space” of the H operator (which is determined by the RT distribution). In short, naïve application of the S-R decomposition algorithm of Zhang (1998) suffers from disproportionate noise amplification at low frequency—low-frequency distortions in the recovered waveforms may occur if noise is not handled properly.

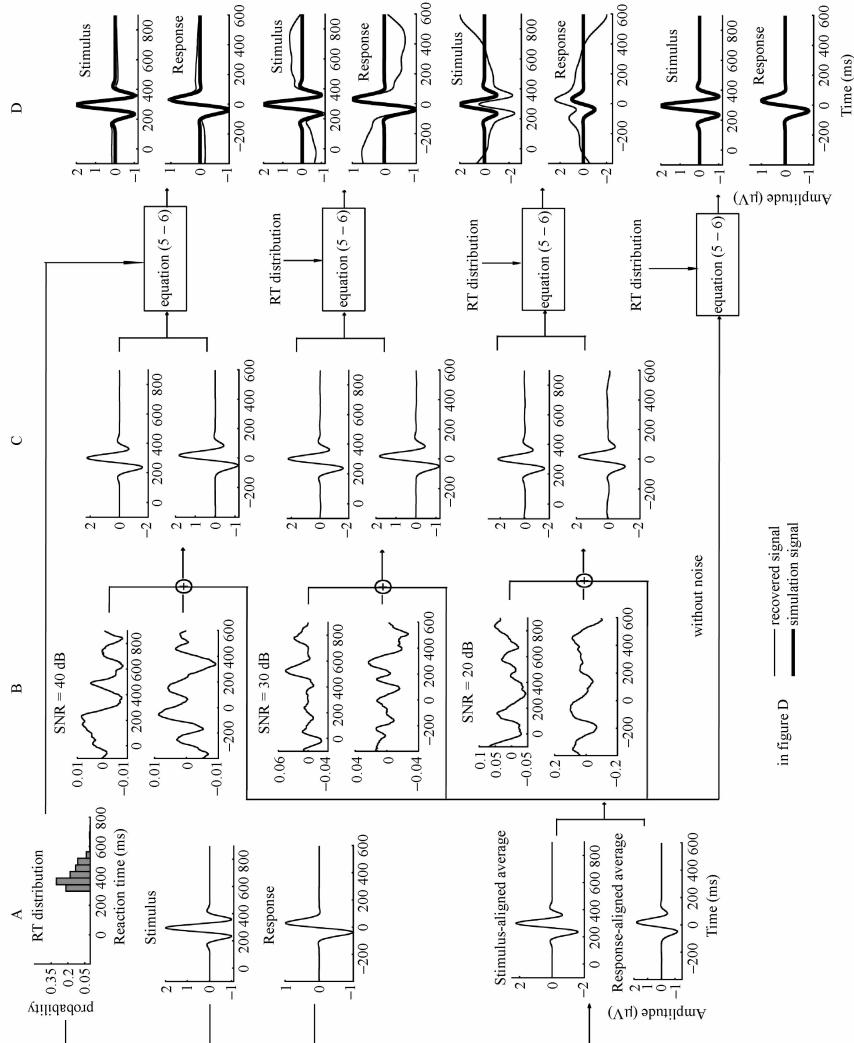


Figure 2 Illustration of how artifacts in stimulus-response decomposition may arise. **A:** Simulated S-and R-component waveforms created, along with an RT distribution and resulting stimulus-and response-aligned waveforms. **B:** Samples of low frequency noise waveform at different SNR level (SNR = 20dB, 30dB, 40dB). Note the difference in ordinate scales. **C:** With different amount of low frequency noise added into the ensemble of trials, resulting stimulus-and response-aligned waveforms change only slightly. **D:** Recovered S-and R-component waveforms (dotted line) from Equations. 5 and 6 under different SNR condition compared with original S-and R-component waveforms (thick lines). Clearly, Equations. (5) and (6) lead to faithful recovery only when there is no noise added, which made them impractical in dealing with real data.

2.5 Wiener Deconvolution for Noise Control

Wiener deconvolution refers to the application of Wiener filter to deal with noise inherent in a linear input-output system. The rationale is as follows. Given a linear input-output system

$$y(t) = h(t) * x(t) + n(t)$$

where $x(t)$ is an unknown input signal, $h(t)$ is the known response, $n(t)$ is an unknown additive noise independent of $x(t)$, and $y(t)$ is an observed signal, we want to find some $d(t)$ so that $x(t)$ can be estimated as:

$$\hat{x}(t) = d(t) * y(t) \quad (17)$$

Here $\hat{x}(t)$ (with a hat) denotes the estimate of $x(t)$, in the sense that it minimizes the mean square error (MSE)

$$\text{MSE} = E[\|\hat{x}(t) - x(t)\|^2] \quad (18)$$

The solution for $d(t)$ is provided by the so-called Wiener deconvolution filter which, in frequency domain, has the form:

$$\tilde{d}(w) = \frac{\bar{h}^c(w)}{|\bar{h}(w)|^2 + \frac{n(w)}{s(w)}} \quad (19)$$

Here $s(w)$, $n(w)$ are, respectively, the mean power spectrum of the signal $x(t)$ and of the noise $n(t)$, and $\bar{h}^c(w)$ is the complex conjugate of the Fourier spectrum $\bar{h}^c(w)$ of the transfer function $h(t)$. In real life applications, one can i) find an initial estimate of signal $x^{(0)}(t)$ (by assuming zero noise and recovering from Equations (5) and (6), for instance) and calculate its signal power spectrum $s^{(0)}(w)$; ii) subtract the signal $x^{(0)}(t)$ from trial-by-trial data to obtain an estimate of noise power spectrum $n^{(0)}(w)$; iii) substitute the signal and noise power spectra into the Wiener filter Equation (19) to obtain an estimate of the new signal $x^{(1)}(t)$; iv) repeat steps i) to iii) until convergence. The final solution $x(t)$ has the property that $E[\|n(w)\|^2]$ is minimized.

There is a slight complication in applying Wiener filters to our problem, because our “input signal” consists of both a stimulus-aligned waveform and a response-aligned waveform—they are not totally independent, due to given RT distribution. So we need to adapt the spirit of Wiener deconvolution filter to the current setting. Note that our signal model assumes that for every single trial, noise is independent of signal. With noise considered, Equations (3) and (4) become, in matrix form:

$$\begin{bmatrix} \tilde{F}_s(w) \\ \tilde{F}_r(w) \end{bmatrix} = \begin{bmatrix} 1 & \bar{g}(w) \\ \bar{g}(-w) & 1 \end{bmatrix} \begin{bmatrix} \tilde{f}_s(w) \\ \tilde{f}_r(w) \end{bmatrix} + \begin{bmatrix} \tilde{\xi}_s(w) \\ \tilde{\xi}_r(w) \end{bmatrix} \quad (20)$$

where $\tilde{\xi}_s(w)$ and $\tilde{\xi}_r(w)$ are the Fourier transforms of noise waveforms time-locked to stimulus onset and to response onset, respectively. In vector notation, it is

$$y = Hx + \xi \quad (21)$$

Note that the orthonormal eigenvectors of H satisfy

$$t_1^\# \cdot t_1 = t_2^\# \cdot t_2 = 1, \quad t_1^\# \cdot t_2 = t_2^\# \cdot t_1 = 0$$

In the above notations, the signal model Eqn (21) breaks into two equations

$$\begin{aligned} y_1 &= Hx_1 + \xi_1 & (t_1^\# \cdot y)t_1 &= \lambda_1 x_1 + (t_1^\# \cdot \xi)t_1 \\ y_2 &= Hx_2 + \xi_2 & (t_2^\# \cdot y)t_2 &= \lambda_2 x_2 + (t_2^\# \cdot \xi)t_2 \end{aligned} \quad (22)$$

The left-hand sides are what we can calculate from the data, and x_1 and x_2 are unknown two-dimensional vectors. Now, we can apply the Wiener deconvolution method to each of its component separately to obtain

$$x_i = \frac{\lambda_i}{\lambda_i^2 + (\text{SNR}_i)^{-1}} (t_i^\# \cdot y)t_i \quad i = 1, 2 \quad (23)$$

where $\text{SNR}_i = E[\|x_i\|^2]/E[\|\xi_i\|^2]$, estimated separately for $i = 1, 2$ components. This is what we call Wiener deconvolution with “uncoupled filters”, because we treat the projections of the noise onto t_1 and t_2 directions distinctly and apply Wiener filters separately, based on the noise decomposition given by Equation (14). Note that since λ_1 is always greater than or equal to 1, the denominator $(\text{SNR}_1)^{-1}$ can (in most cases of noise range) be dropped in the estimate of x_1 . The more crucial application of Wiener filter is to obtain the x_2 component robustly against input noise (see Yin & Zhang, 2011). Figure 3 is the detail flow chart of the algorithm.

Wiener deconvolution, though a popular method for noise control in solving the inverse problem for linear input-output systems, is one of the several tools one can use for noise control. Alternative methods include error regularization (see Yin & Zhang, 2011, for details). The main advantage is that there is no need to estimate the regularization parameter; instead, one only needs to estimate power spectra for both signal and noise. In the above paper, we have compared various approaches to noise control, and come down with the Wiener deconvolution method as the most robust and objective method to use for decomposing ERP waveforms. The implementa-

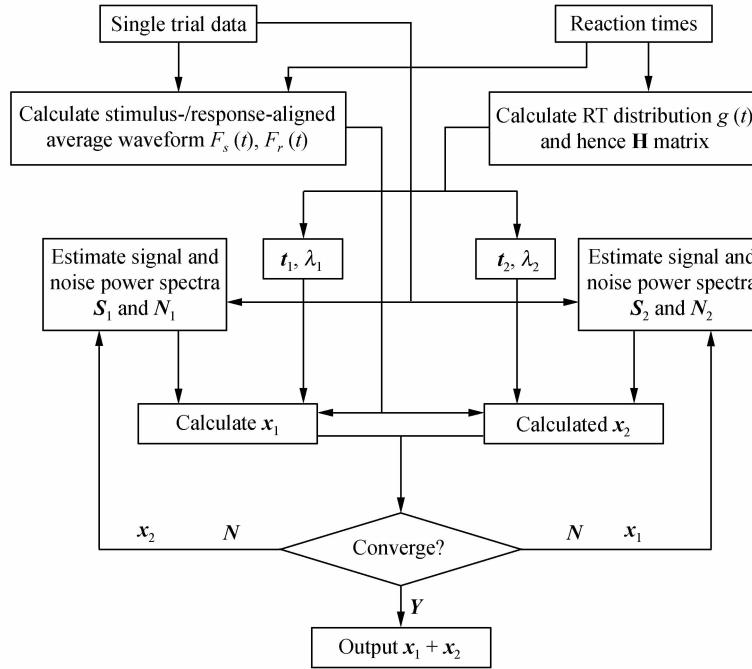


Figure 3 Flow diagram of S-R waveform decomposition with noise control. During the first pass, x_1 , x_2 is calculated without estimating signal and noise spectra. During the second and subsequent passes, Wiener deconvolution filter is used for calculating x_1 , x_2 (as indicated by the open arrow, distinct from all other arrows).

tion steps outlined above have overcome two main challenges of i) adapting the Wiener filter to only the low-frequency components; 2) estimating signal and noise power spectra in two-dimensional correlated signals (stimulus-aligned and response-aligned waveforms).

3 Validation and Application

In this section, we provide some examples both to validate our decomposition algorithm and to illustrate its utility in addressing empirical questions raised by ERP data. We will use both simulation ERP data with known ground-truth (with $N = 2$ and $N = 3$ markers) and real ERP recordings from two experiments: a Go/Nogo experiment ($N = 2$ markers) and an attentional precuing experiment ($N = 3$ markers). For more details, see Yin et al. (2009) and Yin and Zhang (2011).

3.1 Simulated ERP Waveforms

In our simulation, we model ERP waveforms on each trial as consisting of two ($\beta = s, r$) or three ($\beta = c, s, r$) distinct component waveforms, one time-locked to the cue onset (in the case of $N = 3$), one time-locked to the stimulus onset, and one time-locked to the response onset. The exact shapes of these component waveforms are according to the following equations

$$f_{\beta}(t_i) = A_{\beta} \exp\left(-\left(2\pi\lambda_{\beta}\frac{t_i - \tau_{\beta}}{\gamma_{\beta}}\right)^2\right) \cos(2\pi\lambda_{\beta}(t_i - \tau_{\beta}) + \alpha_{\beta})$$

$$i = 1, \dots, n; \quad (\beta = s, r \text{ or } \beta = c, s, r) \quad (24)$$

Here, A_{β} is the amplitude of the stimulus-, response-, and the cue-locked (for $N = 3$ only) component waveform and τ_{β} is the random variable representing trial-by-trial onset time point for stimulus event or response event. The value of α_{β} is fixed, and λ_{β} and γ_{β} are constants that are related to the shape of the component waveforms. ERP waveform on any individual trial is taken to be the sum of the above two (or three) components, plus background noise, which were simulated by two autoregressive (AR) processes with white noise injected into each trial. See Krieger et al. (1995).

3.1.1 Simulation Result for $N = 2$ Time Markers

Figure 4 shows an example of decomposition S-and R-components ($N = 2$) using the Wiener deconvolution method, under the following simulation parameters: $\gamma_s = 1.2$ and $\gamma_r = 0.8$, $\lambda_s = 5.9$ and $\lambda_r = 4.7$, $\alpha_s = 0.36$ and $\alpha_r = -0.42$, SNR = -10dB , and the standard deviation of the RT distribution is 20 ms, and the number of trials 100. With stimulus time point chosen as the reference zero, the response time (τ_r) relative to the stimulus time (τ_s) is taken from a Gamma distribution with a mean of 300 ms and four different values of standard deviations. Four different SNR values were used (see below). For each condition of RT distribution and SNR value, the simulation is repeated 50 times to yield statistical averages. In every repeat, α_{β} is randomly selected within $0-2\pi$, A_{β} is also a random value between $1.0-2.0$, λ_s and λ_r are random values between $5.0-7.0$ and $4.0-6.0$ respectively. ERP waveform on any individual trial is taken to be the sum of the stimulus and the response component waveform plus an AR process-like background noise.

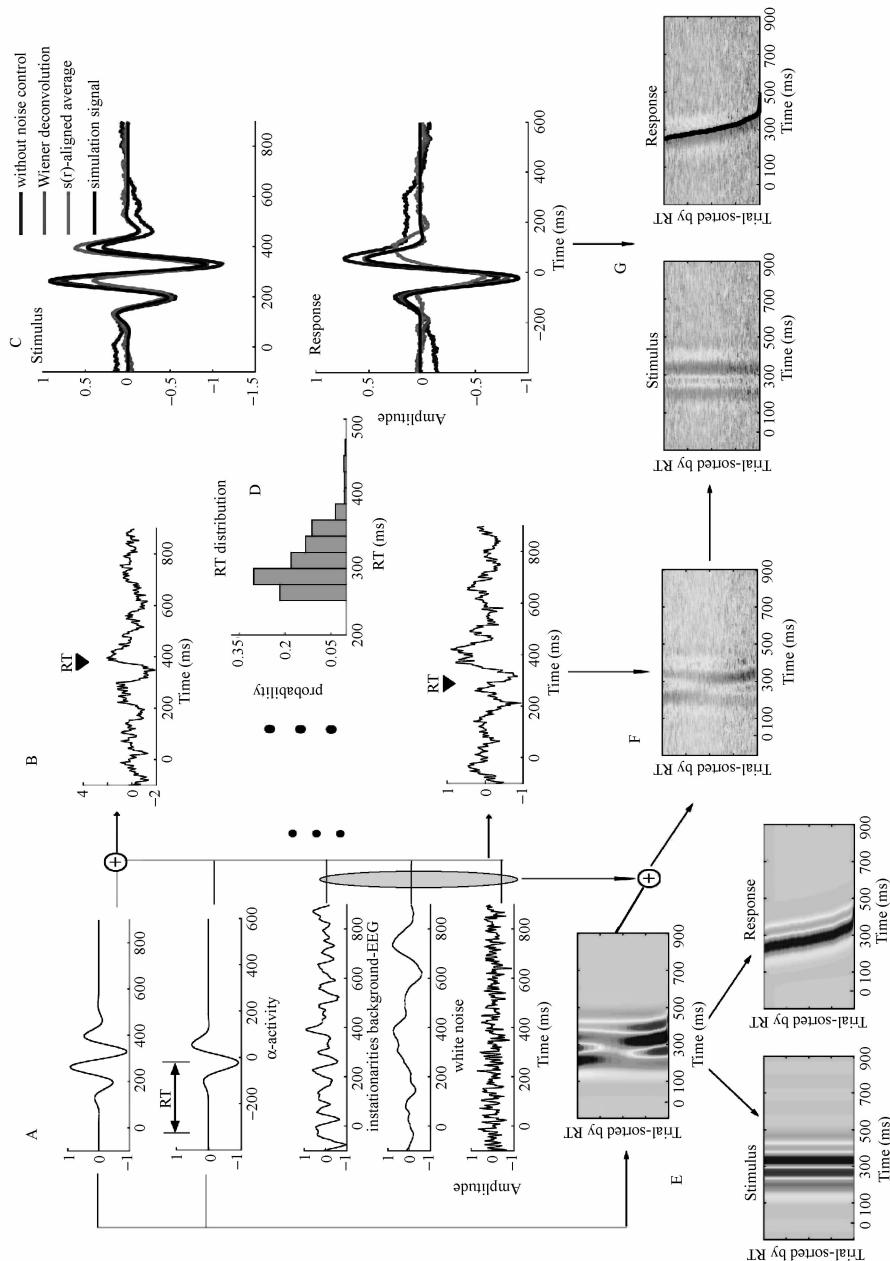


Figure 4 An illustration of the ERP decomposition algorithm with or without noise control by the method of Wiener deconvolution, for the $N = 2$ case. A: S-and R-component waveforms and noise waveforms (α -activity, non-stationary background-EEG, and white noise) used for simulation. B: Simulated single trial waveforms created through summing together S-and R-component waveforms, according to different RT, plus noise. C: Comparison of original S-and R-component waveforms, s-and r-aligned ensemble average waveforms, and recovered S-and R-components either without noise control (calculated by Equation 5 and 6) or Wiener deconvolution method. D: RT distribution. E: ERP-image of the ensemble of trials for original stimulus-aligned data (top) and for either S-component (lower left) or R-component (lower right) after trials sorted according to RT, with no noise injected. F: ERP-image of the ensemble of trials with noise added. This is the input to the S-R decomposition algorithm. G: ERP-image of “stimulus” part and “response” part, in which on each single trial, recovered R-component waveform (or S-component waveform) is subtracted, respectively. Here different colors in an ERP image encode different amplitude values. The black line in “response” part is RT curve.

3.1.2 Simulation Result for $N = 3$ Time Markers

Figure 5 shows the simulation results for the $N = 3$ markers situation, as in the attentional precuing paradigm, where the three time markers are the onset times of

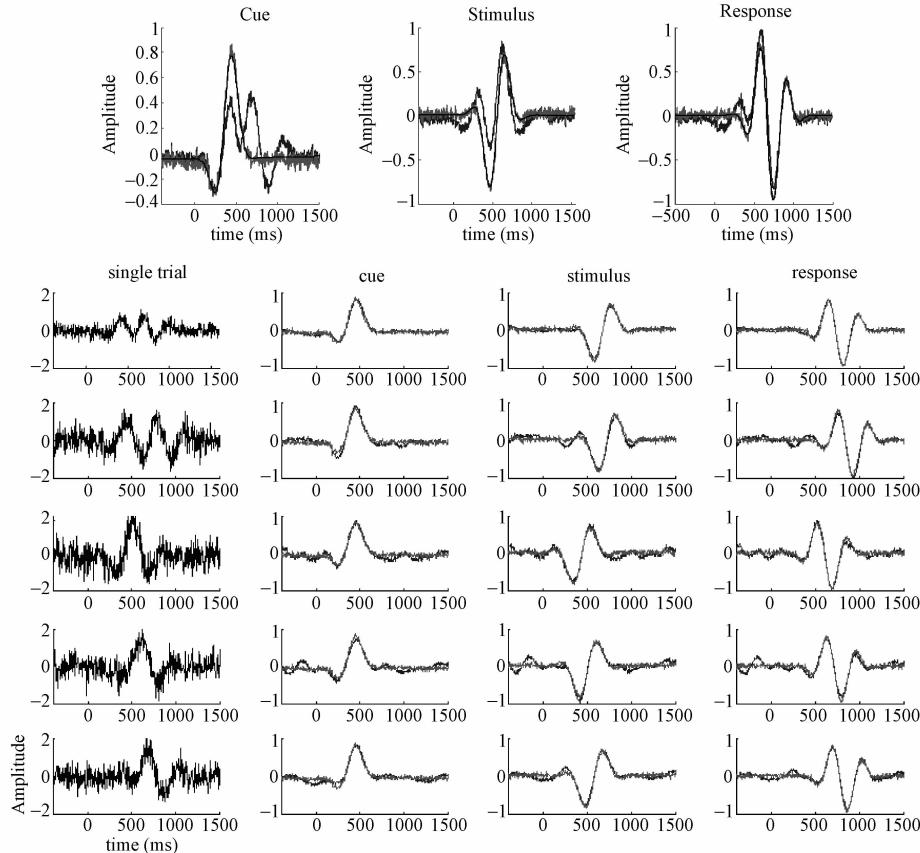


Figure 5 Illustration using simulated waveforms, for the $N=3$ marker situation. ERP is assumed to consist of three component waveforms, time-locked to the onset of cue, stimulus, and response, respectively. A: Recovered cue/stimulus/response-locked component waveform (red curve), original cue/stimulus/response-aligned average ERPs (blue curve), and ground-truth cue/stimulus/response component waveform (black curve), plotted in the same graph. The reference zero point represents cue onset (left), stimulus onset (middle), and response onset (right), respectively. B: Recovered cue/stimulus/response component waveform (from Wiener filter) plus single-trial noise (blue curve), and original ground-truth of cue/stimulus/response-locked component waveform (red curve). Different rows correspond to different single trials (randomly selected a total of 100 trials), with single-trial signal-plus-noise ERP waveform displayed on the left (black). In all graphs, 0 represents the time of cue onset, and the onset times of stimulus and of response are of variable delay (that differ across trials).

cue (c), stimulus (s), and response (r). The parameter values of equation (24) are $\gamma_c = 3.4$, $\gamma_s = 2.9$, $\gamma_r = 5.4$ and $\lambda_c = 1.5$, $\lambda_s = 2.0$, $\lambda_r = 2.7$. For each trial, $\tau_\beta(\beta = c, s, r)$ is the random variable representing onset time point for event β . Without loss of generality, we take τ_c to be fixed at 200, and $\tau_s = 250 \pm \text{Rand}$, $\tau_r = 300 \pm \text{Rand}$, where Rand represent a uniformly distributed random number with mean 0 and variance 100. The value of α is taken to be fixed between 0 and 2π . ERP waveform on any individual trial is taken to be the sum of the above three components plus AR process-like background noise.

3.2 Data From Real Experiments

Here we illustrate the application of our technique to two real experiments, a Go/NoGo task ($N = 2$ markers) and an attention pre-cuing task ($N = 3$ markers).

3.2.1 A Go/NoGo Experiment

Our S-R decomposition algorithm has been applied to a Go/NoGo task, with experimental data obtained from an internet open source (http://www.sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html).

EEG was recorded with a 32-channel system at a sampling rate of 1000 Hz, and Cz was taken as the reference. Epochs contaminated with excessive eye movements, blinks, muscle artifact, or amplifier blocking were manually removed by the data provider, and the ERPs were re-referenced to the grand average during off-line data processing. The supplied data-set deal with two cognitive tasks, a recognition task and a classification task (see the above online link for details).

Data from recognition task. ERP data was taken (from a single subject fsa) at electrode Pz with a total 248 trials under analysis, and results are shown as Figure 6. Our Wiener deconvolution algorithm (using uncoupled filters) is seen to offer a visually remarkable improvement over the direct calculation method since, without noise control, the latter method disproportionately magnified low frequency noise and introduced significant distortions, causing an artifact even in the baseline regions of the waveforms. Visualizing the results using the ERP-image method (Jung et al., 2001) shows that our algorithm can recover the underlying S-and R-component waveforms with fidelity.

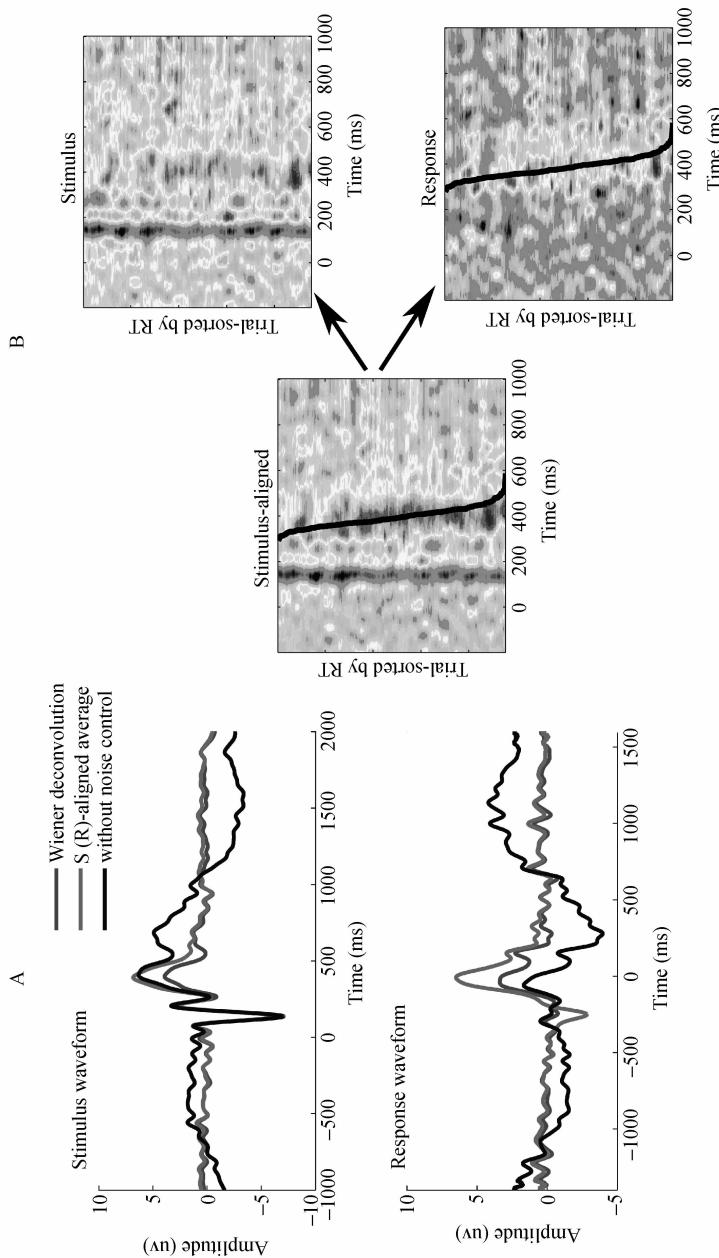


Figure 6 Result of S-R decomposition in a Go/NoGo task (recognition subtask). Data from Pz channel of one subject is shown. **A:** The recovered S-and R-component waveform, obtained with Wiener deconvolution for noise control (red curve) or without noise control (blue curve), is compared with s-aligned ERP average, where 0 marks stimulus (or response) onset. **B:** ERP-images of the empirical data both before and after S-R decomposition. Left: ERP-image of all single trials sorted according to ascending RT; black curve represents the RTs across the ensemble of trials. Right: ERP-image of “stimulus” part and of “response” part of the ensemble of trials component recovered by Wiener deconvolution noise control, where the R-component or S-component, respectively, has been subtracted from every single trial. For better visual display, all ERP-images are smoothed across trials with a ten-trial moving-average.

Data from classification task. Figure 7 (A left) depicts the pure S-component waveform extracted from Go trials, and compares it to stimulus-aligned ERP averages (across all subjects) on both Go and NoGo trials on typical channels Fz, Pz and Oz. It can be seen that, for Fz channel, the N2 component is smaller than that in the uncorrected s-aligned average, while the P3 component in Go trials resembles that in NoGo trials. Our results thus revealed a significant amount of contamination, by response-locked ERP component, in stimulus-aligned ERP average on Go trials.

In order to demonstrate how our decomposition algorithm, when applied to all electrodes simultaneously, improves the read-out of ERP topography, we choose three typical ERP time windows (P1 patch, about 100 ms after stimulus onset; P3 patch, about 350 ms after stimulus onset, response patch, about 25 ms after response onset) between original averages and recovered signals, and compare the ERP topography before and after S-R decomposition. Figure 7B shows the results: i) for the P1 patch, original waveforms and the recovered waveforms have very similar spatial distribution; ii) for the P3 and response patches, significant differences are revealed in the prefrontal area and posterior parietal areas, with much larger amplitudes, due to S-R cross-contamination (statistical test results will be shown in a separate report). In Figure 8, we also plot the difference waveforms and ERP typography at N2 and P3 time points, before and after applying our decomposition algorithm. Note the conspicuous decrease (almost absence) in the P3 peak of the difference waveform in B (after SR decomposition) compared with that in A (before SR decomposition).

The importance of our methodology for proper interpretation of experimental data is readily demonstrated by this example of Go/NoGo task. In Go/NoGo paradigm, considerable interests have been devoted to a positive wave peaking within a latency range of 300—500 ms (P3 component), which is reported to have larger amplitude for frontal channels in the NoGo trials than in the Go trials (Fallgatter & Strik, 1999; Jodo & Kayama, 1999). Our analysis showed that there is a big difference in P3 between the original stimulus-and response-aligned EEG averages and recovered stimulus-and response-locked component waveforms for channels Fz and Pz—P3 amplitude in the original stimulus-/response-aligned ERP averages are larger than the true underlying S-component/R-component waveforms. On the other hand, there appears to be no difference in the P3 time patch between recovered S-component waveform on Go trials and the stimulus-aligned ERP average on NoGo trials (Figure 8B). This im-

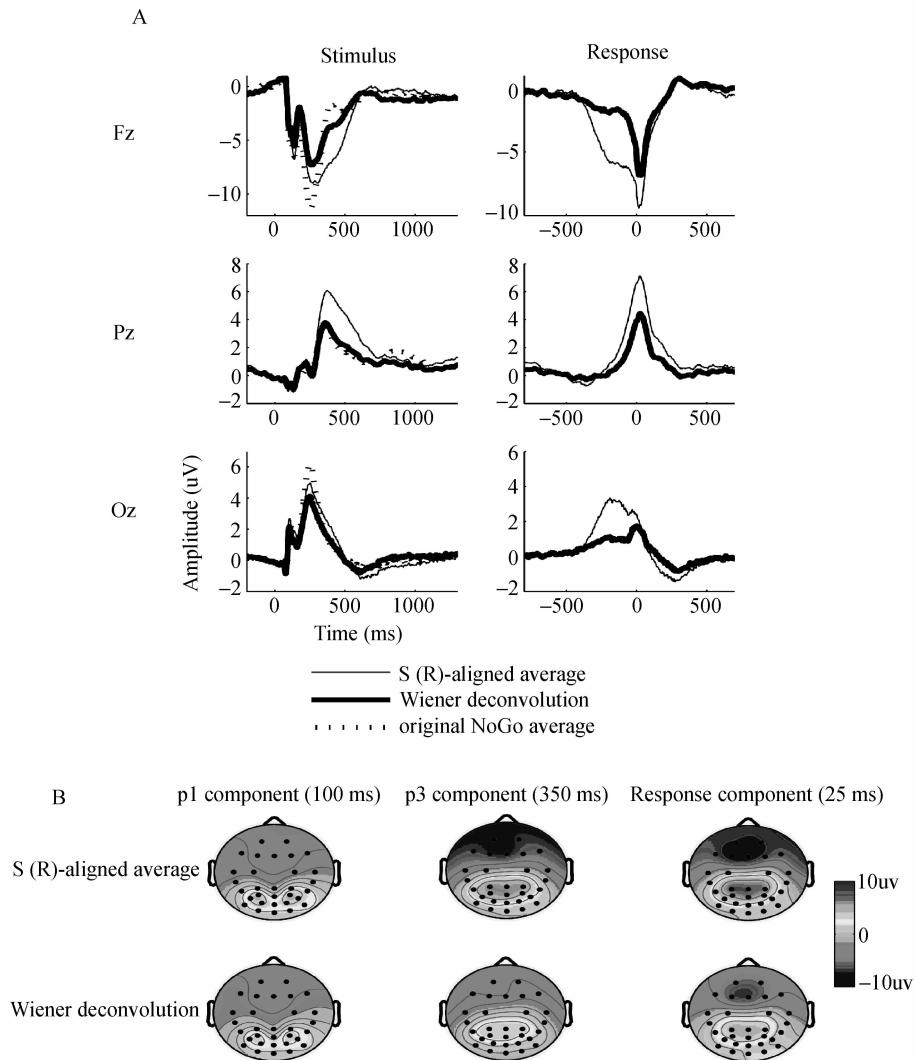


Figure 7 Result of S-R decomposition in a Go/NoGo task (classification subtask). A: Extracted pure S-and R-component waveforms for Go trials, and the stimulus-aligned ERP averages for both Go and NoGo trials on channels Fz, Pz and Oz across all subjects. B: Topography of the P1 time patch (100 ms), the P3 time patch (350 ms), and the response time patch (25 ms) of original ERP average and of recovered waveforms, respectively, for Go trials.

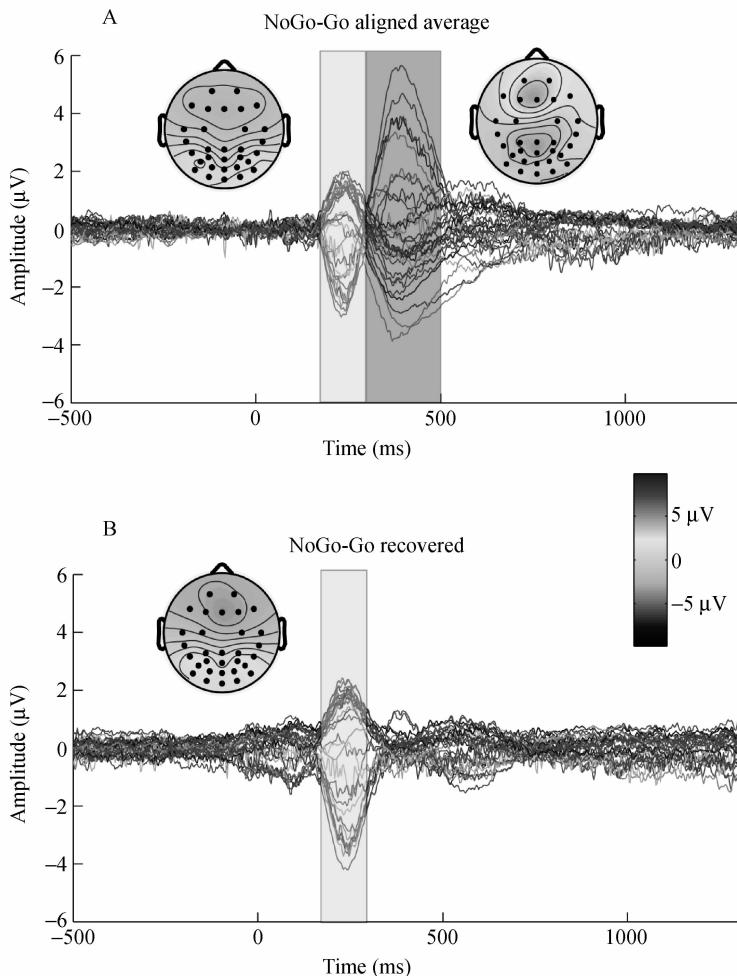


Figure 8 The topographic distribution and ERP difference waveform between Go and NoGo task in N2 and P3 time segment (for the classification subtask of the Go/NoGo experiment). **A:** Without applying decomposition algorithm. Topographic distribution and ERP difference waveforms (of stimulus-aligned averages between NoGo and Go trials) for the two time patches: N2 time patch (head on the left and pink-colored stripe below) and P3 time patch (head on the right and grey-colored stripe below). **B:** After applying decomposition algorithm. Topographic distribution for the N2 patch and ERP difference waveforms (between the stimulus-aligned waveform on NoGo trials and the stimulus-locked component recovered from Go trials). For both A and B, different waveforms (color-coded) correspond to different recording channels as shown by the head model.

plies that the difference in P3 between Go and NoGo trials may result from the presence of a response-locked component in Go trials, rather than any difference in stimulus processing between Go and NoGo trials. That the P3 component (peaking around 350ms post stimulus onset) and the response component (peaking around 50 ms post response onset) suffer from stronger cross-contamination is understandable since the mean reaction time is only 429 ms. In contrast, the earlier P1 component suffers little cross-contamination, as evidenced by the close resemblance of the spatial distributions of the original stimulus-aligned ERP average and the recovered stimulus-locked waveform (Figure 7B). A more detailed report, including tests of statistical significance for these effects, is forthcoming (Yin & Zhang, in preparation).

3.2.2 An Attention Pre-Cuing Experiment

We also applied our S-R decomposition algorithm to experiment data on an attention pre-cuing task (Yin et al., 2009). The stimuli used for the pre-cuing experiment is as follows (Figure 9). Each of the two rectangle-outlined boxes was briefly presented, and one of them was highlighted (brightened) as “cuing”. After a variable delay of cue onset, the target, either a short vertical line or a long vertical line, was presented within either of the two boxes with equal probability.

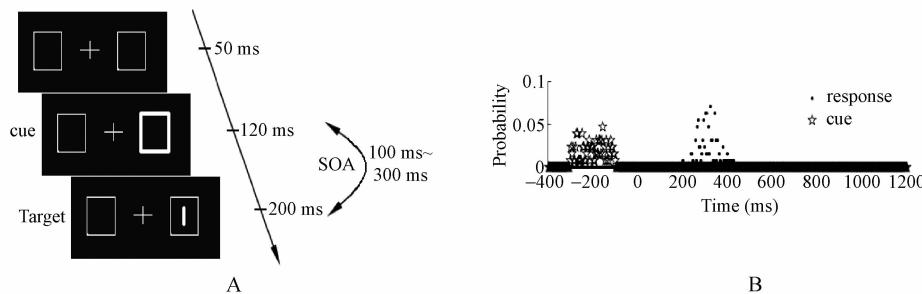


Figure 9 Experimental paradigm of the attentional pre-cuing task, with $N = 3$ time markers. A: The cue is the brightening of an unfilled box centered on one of the two peripheral target locations; the target is the appearance of a short or long vertical line in one of the boxes. Interval between cue and stimulus (SOA) is a randomly variable between 100 ms to 300 ms. B: The two event time distributions, one for cue and one for response, are given with respect to stimulus onset (with stimulus onset time at 0 ms).

Our method is applied to decompose the cue-, stimulus-, and response-locked component waveforms in the ensemble averaged ERP (averaged across all seven subjects) as well as to the ERP averages of individual subjects. The data of all the cor-

rect trials from one subject—the cue is located in the left, the stimulus is a short vertical line (“target”) located in left, and the subject responded correctly by a left-hand key-press. ERP data was taken at electrode Pz with a total of 127 trials under analysis. The cue-aligned, stimulus-aligned and response-aligned ERP averages (denoted F_c , F_s and F_r) were constructed by aligning the waveforms on individual trials with respect to cue-, stimulus-and response-onset respectively and summing over those waveforms bin-by-bin (each bin is 1 ms). The event time (cue, stimulus, response) distributions were simply constructed from these 127 trials whose durations are available around cue-onset, stimulus-onset and response-onset respectively. The evoked component waveform related to (i. e., time-locked) to cue, stimulus, and response, respectively, are denoted f_c , f_s and f_r . They are recovered by our decomposition algorithm along with Wiener deconvolution noise control.

Figure 10 show the results of decomposition of the cue-, stimulus-, and response-locked component waveforms from an analysis of ERP ensemble (grand) averages across seven subjects, under different alignment of individual trials for, respectively, the cue-valid (A) and cue-invalid conditions (B). The top panels reveal that the reconstructed cue-locked component waveform f_c is consistent with the cue-aligned ERP average F_c during the first 300 ms after cue onset, but tends to flatten out 500–1100 ms after cue-onset, in contrast to the amplitude of the F_c which remains large during this late period (this latter activity reflects a cross-contamination from the stimulus-related and response-related neural processes that have not been “averaged out” in cue-aligned averaging). The middle panels reflects a difference between stimulus-aligned ERP average F_s and the recovered stimulus-locked component waveform f_s during 300 ~ 800 ms after stimulus onset, due to possible cross-contamination from cue-locked and response-locked components. The bottom panels reveal larger amplitude for the response-aligned ERP average F_r compared with the response-locked component f_r in the period of 0—200 ms after behavioral response onset, again presumably due to cross-contamination of stimulus-locked component. The above observations are true for both cue-valid and cue-invalid experimental conditions.

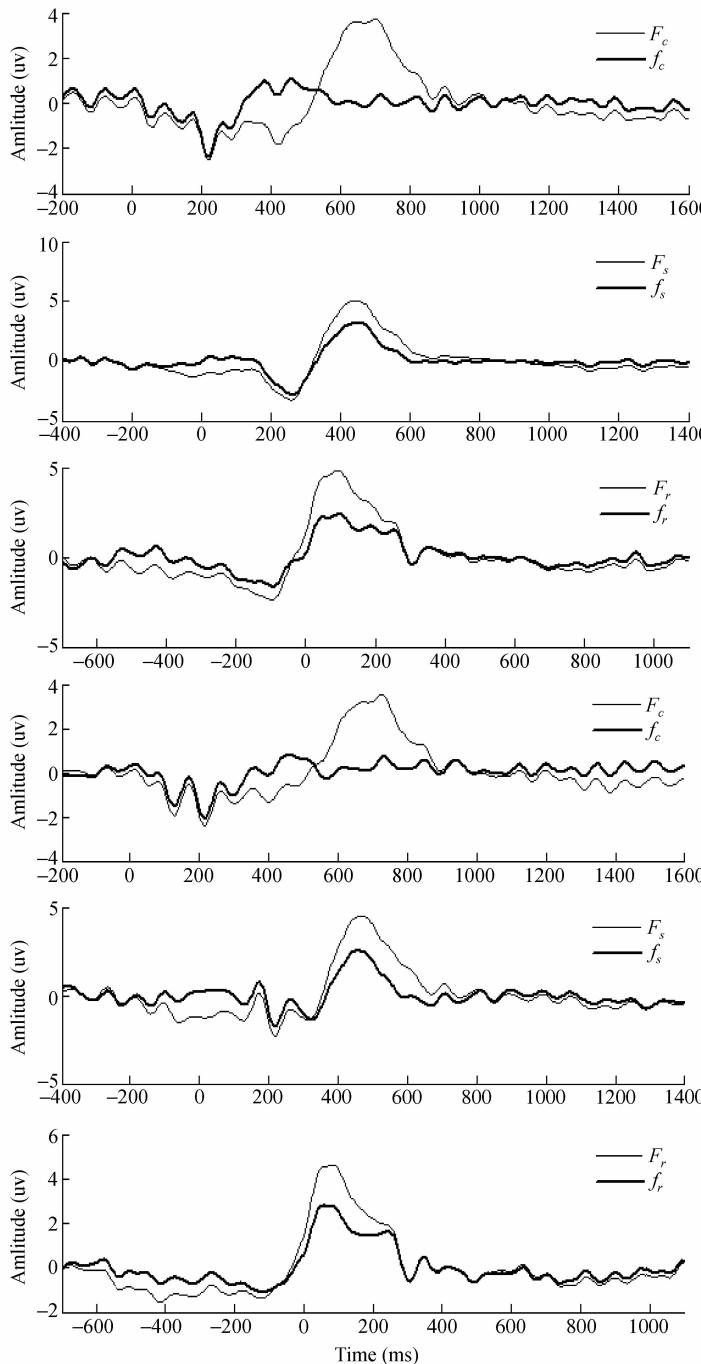


Figure 10 Recovered cue-locked (top panel), stimulus-locked (middle panel), and response-locked (bottom panel) component waveforms (f_c , f_s , and f_r) from ERP ensemble averages (aligned to cue, stimulus, and response (F_c , F_s and F_r), respectively for top, middle, and bottom panels) across all seven subjects. Thick lines: recovered cue-, stimulus-, response-component waveforms. Thin lines: cue-, stimulus-, response-aligned ERP ensemble averaged across all seven subjects. **A:** Experimental condition when cue is valid—cue is located in the left, stimulus is a short line (“target”) located in the left, and response is left key-press. **B:** Experimental condition when cue is invalid—cue is located on the right, stimulus is a short line (“target”) located in the left, and response is left key-press.

In Figure 11, corresponding waveforms of the cue-valid and the cue-invalid conditions are plotted directly on top of each other. Visual inspection shows that the cue-locked waveforms, as well as the response-locked waveforms, are virtually identical for the two cue-conditions (valid versus invalid). The only stage that a valid cue and an invalid cue differ is with respect to the processing of the stimulus, where the two waveforms differ significantly (middle panel), since the validity of a cue on a trial (i.e., priming effect) is operationally only meaningful with respect to the stimulus on that trial.

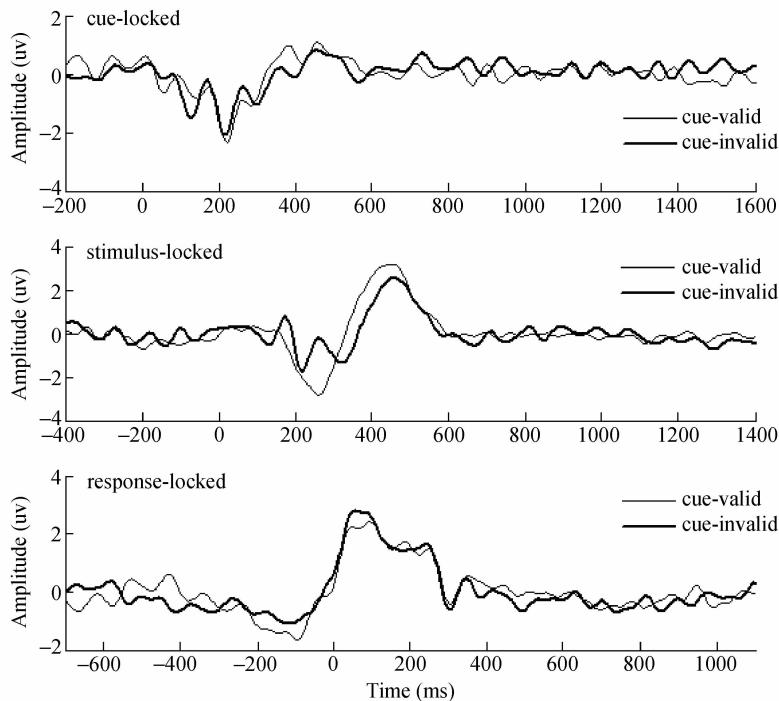


Figure 11 Comparison the recovered cue-locked, stimulus-locked, and response-locked component waveforms across the cue-valid and cue-invalid conditions (c.f. Figure 10).

To demonstrate sensitivity of our decomposition algorithm, we show the results for all seven individual subjects (Figure 12). The three columns represent, respectively, cue/stimulus/ response-aligned ERP averages (thin lines) and recovered cue/stimulus/response-locked component waveforms (thick lines), for “cue-valid” experimental condition (A) and for “cue-invalid” experimental condition (B). The

difference between cue-aligned waveform and cue-locked component is quite pronounced across all subjects.

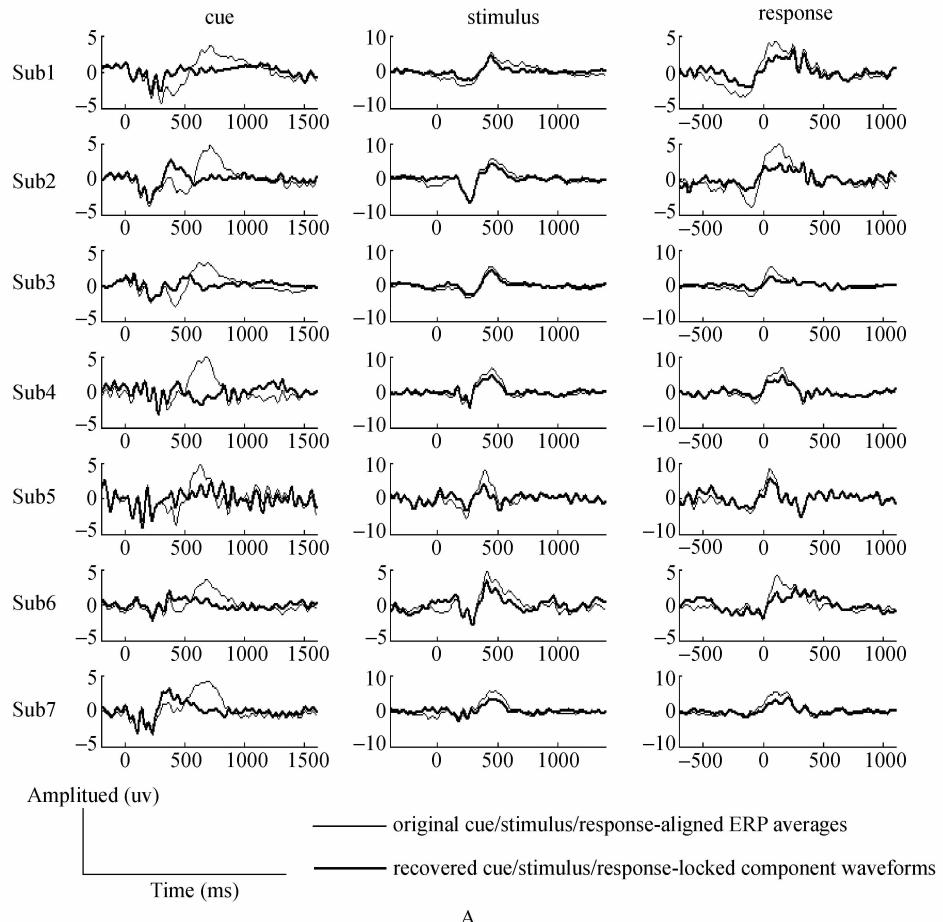
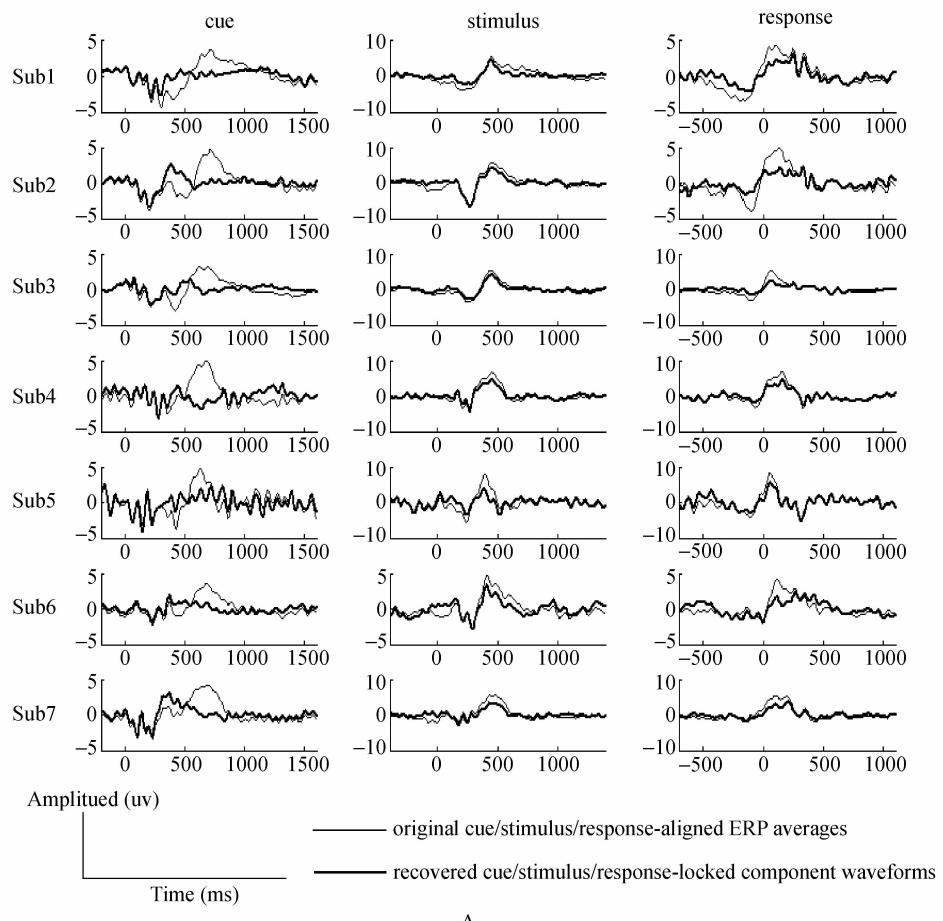


Figure 12 Decomposition results for each individual subject at electrode Pz. Thin curves are cue/stimulus/response-aligned ERP averages, and thick curves are recovered cue/stimulus/response-locked component waveforms. **A:** Experimental condition when the cue is valid—cue is located in the left, stimulus is short-line (“target”) located in left, and response is left key-press. **B:** Experimental condition when cue is invalid—cue is located on the right, stimulus is a short line (“target”) located in the left, and response is left key-press.

**Figure 12 (continue)**

To test the stability of our decomposition algorithm, we split all trials into two piles, one pile with relatively longer reaction-times and another with relatively shorter reactions times, and then perform decomposition separately. Results are shown in Figure 13, for cue/stimulus/response-aligned ERP averages and recovered cue/stimulus/response-locked component waveforms, when all trials are lumped together (A) or separated according to reaction times (B). Despite of obvious difference in ERP averages for short-and long-RT trials (green versus black lines in B), recovered waveforms (blue and red lines in B) are remarkably similar, demonstrating the robustness of our method.

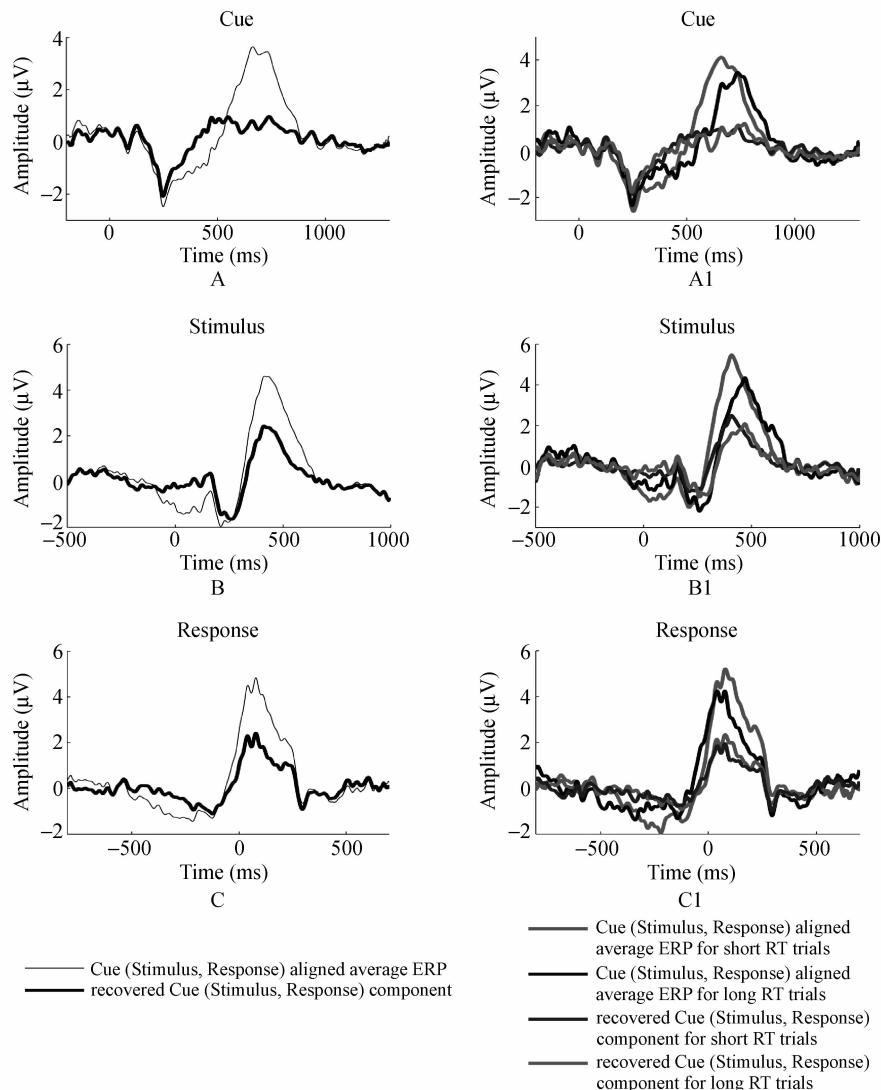


Figure 13 Comparison of original cue/stimulus/response-aligned average waveform and recovered cue/stimulus/response component waveform, on Oz channel for the cue-valid condition. Left: All trials are used. Right: Trials are divided into two halves, those with below-median reaction times (“short RT” trials) and those with above-median reaction times (“long RT” trials). A and A1: Cue-related waveforms; B and B1: Stimulus-related waveforms; C and C1: Response-related waveforms.

Finally, the effectiveness of our decomposition method can be appreciated when we perform decomposition across all channels and then construct topographic distribution. Figure 14 compares the topographic distribution of the original cue-aligned ERP average with that of the recovered cue-locked waveform, at the time point of 250ms, 700ms after cue onset. It can be readily concluded that posterior activities at 700ms after cue onset, as seen in the cue-aligned ERP typography, is an artifact (due to possibly contamination from stimulus-locked processing).

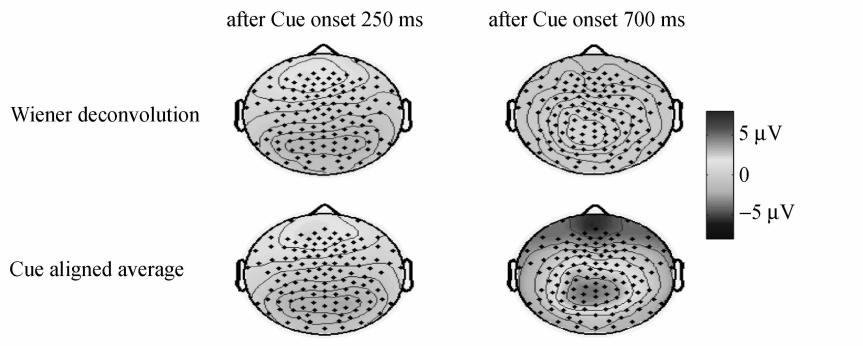


Figure 14 Comparison of topographic distribution between recovered cue-locked component (top) and original cue-aligned average (bottom), at the time point of 250 ms and 700 ms after cue onset.

4 Discussion

Common methods for decomposition of multiple components in ERP/EEG and fMRI data include Independent Component Analysis ICA (Amari, Cichocki & Yang, 1996; Bell & Sejnowski, 1995; Comon, 1994; Jutten & Herault, 1991; Makeig et al., 1996) and Principal Component Analysis PCA (Pearson, 1901; also known as Karhunen-Loeve transform), among others. PCA is a classical technique in statistical data analysis, feature extraction and data reduction, aiming at decomposing observed signals into a linear combination of orthogonal components based on the variance-covariance matrix, whereas ICA is a technique of array processing and data analysis, aiming at recovering underlying “sources” from their observed mixtures based on an assumption of mutual independence between the subcomponents. The exact number of recovered components is often determined in an ad hoc fashion—in PCA, it is determined by comparing the amount of variance captured versus the residual variance,

whereas in ICA, the maximal number of recovered components is further constrained by the total number of observed channels. Although ICA/PCA have been widely applied in ERP/EEG and fMRI dataset analysis for artifact remove, these two methods made strong assumptions which often may constrain their applications. For example, ICA assumes that the underlying sources are statistically independent, whereas for PCA, the underlying components are assumed to be orthogonal to one another. Independence and orthogonality may be too strong assumptions to be neurobiologically realistic. From a methodological point of view, a more serious problem is that both methods assume each trial to be a stationary point process despite of change in key behavioral events, such as reaction time. The trial-by-trial variation of behavioral markers even causes problem to properly align EEG/fMRI recordings across trials, which is a precursor for ICA/PCA type analysis.

Working under this last constraint is crucial for successful decomposing EEG/fMRI components in many psychological experiments which, typically, involve two or more behavioral events in a single trial. This problem is now completed solved (Zhang, 1998, for the $N = 2$ case and Yin et al, 2009, for arbitrary N). There, ERP averages aligned to behavioral events are being used, along with the inter-event time distributions, to uniquely recover event-locked component waveforms. This method, as summarized in the present Chapter, can be used to supplement the more commonly used multi-component decomposition method, such as PCA/ICA, by overcoming the hurdle of non-stationarity of time series (i. e., ERP/fMRI recordings across trials) whose trial-by-trial variation is of behavioral significance.

Our simulation results demonstrate that the decomposition method we advanced is effective and practicable in disentangling the contributions of different components. However, we also note that the effectiveness of the method relies heavily on the “proper shape” of event time distribution across the ensemble of trials. Here “proper shape” means that the RT distribution (event time distribution in general) is unimodal with variance significantly different from zero. Actually, the less variance the RT (event-time) distribution is, the less effective our method becomes. Mathematically, when the RT variance is small (e.g., all trials have nearly uniform reaction time), its Fourier transform approaches unity and the recovery error generated by applying Equation (5) or (6) becomes large. At the other extreme, when the RT distribution has too large a variance that exceeds the *average* reaction time, event-aligned ERP waveforms will have included late or early contributions outside the region of interest.

Technically, to get the event-aligned ERPs across individual trials, we need segment the raw EEG recording into single trials all with same length and then impose periodic condition in order to perform Fourier transform—the variance of event distribution need to be small enough so that the segmented single trial recording still contains the region of interest.

Notwithstanding such technical constraints, our method actually addresses and solves a problem that was not addressed by conventional methods, and is thus a completely complementary technique in ERP multi-component analysis. Though our simulations are based on ERP context, the basic mathematical technique behind our method can also easily be adapted to deal with event-related signals in other neuron-imaging techniques, especially fMRI.

There are certain limitations of our method. One of those is that our mathematical model makes the fundamental assumption that the true, underlying S-and R-components (or other components, in $N > 2$ case) are the same in each and every single trial, in terms of timing and overall amplitude of the waveform. The latency variation of an event-locked ERP component across single trials can be easily addressed: in this case, the recovered waveforms are simply the jittered version of the underlying S- and R-component waveforms. Future research will address the more serious constraint imposed by the assumption of equal amplitude of the underlying waveform across individual trials (which amounts to a multiplication of the RT distribution by the RT-dependent amplitude modulation function and hence will change the RT distribution). A possible approach is to combine the template matching technique (Truccolo et al., 2003) with our S-R decomposition technique to develop an adaptive technique to extract underlying waveforms that give rise to trial-by-trial evoked potential.

The other limitation of our method is that it assumed all event-time distributions are known. For instance, our signal model (Eqn 21) assumed that single-trial waveforms result entirely from a stimulus-locked component waveform and a response-locked component waveform, plus noise. Based on this assumption, unique recovery of these component waveforms is mathematically guaranteed. However, by scrutinizing the results (ERP images in Figure 7), one finds that the recovered S-component waveform and R-component waveform share considerable common activities (on a window from about 400 ms after stimulus onset). Such common activity seems to reflect a component waveform that may neither be time-locked to stimulus nor to response (perhaps to some stimulus-response “decision”, with an unknown transition

time distribution); yet this component has been, unwarrantedly, decomposed into the S-and R-component waveforms. In Zhang (1998), the possibility of a “decision-related” component, which is time-locked to the putative transition of a stimulus processing and a response processing stage, has been investigated, whereby the unknown event time distribution (in this case, the time marker related to “decision”) is estimated from data. Future research will implement the solution strategy and develop noise control techniques tailored to such estimation.

5 Conclusion

In this Chapter, we provided a robust method for decomposition of a stimulus-locked and a response-locked component ($N = 2$ markers), or multiple components time-locked to various behavioral events ($N > 2$ markers), in ERP recordings. With our technique, trial-by-trial variation in behavioral reaction time is no longer an obstacle but rather an advantage for isolating the underlying neurocognitive processes mediating a task. The method is robust because we applied denoising techniques based on Wiener deconvolution. When applied to those (low) frequency components that fall under the null space of an operator (controlled by the reaction time distribution), this noise control technique gives robust results of recovery, despite of considerable variation in trial-by-trial reaction times and the limited total number of trial. Our algorithm, when combined with other traditional methods of source separation and localization, will provide reliable tools for pinning down the underlying neural substrate for cognitive processes interrogated by neuroimaging methodology.

References

- Amari, S., Cichocki, A., & Yang, H. H. (1996). A new learning algorithm for blind source separation. *Advances in neural information processing systems 8*, 757—763
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6), 1129—1159.
- Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3), 287—314.
- Falkenstein, M., Koshlykova, N., Kiroj, V., Hoormann, J., & Hohnsbein, J. (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(1), 36—43.

- Fallgatter, A. J., & Strik, W. K. (1999). The NoGo-anteriorization as a neurophysiological standard-index for cognitive response control. *International Journal of Psychophysiology*, 32, 233–238.
- Gehring, W. J., & Fencsik, D. (1999). Slamming on the brakes: An electrophysiological study of error response inhibition. Paper presented at the Poster presented at the annual meeting of the Cognitive Neuroscience Society (pp. 11–13). Washington, DC.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, 4(6), 385–390.
- Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In: E. Plum (Ed.), *Handboook of Physiology: The nervous system* (pp. 519–584). Bethesda, Maryland: American Physiological Society.
- Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and clinical neurophysiology*, 82(6), 477–482.
- Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human brain mapping*, 14(3), 166–185.
- Jutten, C., & Herault, J. (1991). Blind separation of sources, Part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Process*, 24(1), 1–10.
- Krieger, S., Timmer, J., Lis, S., & Olbrich, H. (1995). Some considerations on estimating event-related brain signals. *Journal of neural transmission*, 99(1), 103–129.
- Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11(2), 99–116.
- Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems* 8, 145–151.
- Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical journal*, 76(2), 691–708.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11), 559–572.
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4), 387–401.
- Truccolo, W., Knuth, K. H., Shah, A., Bressler, S. L., Schroeder, C. E., & Ding, M. (2003). Estimation of single-trial multicomponent ERPs: Differentially variable component

analysis (dVCA). *Biological cybernetics*, 89(6), 426—438.

Yin, G., Zhang, J., Tian, Y., & Yao, D. Z. (2009). A multi-component decomposition algorithm for event-related potentials. *Journal of neuroscience methods*, 178 (1), 219—227.

Yin, G., & Zhang, J. (2011). On decomposing stimulus and response waveforms in event-related potentials recordings. *IEEE Trans Biomed Eng*, 58(6), 1534—1545.

Zhang, J. (1998). Decomposing stimulus and response component waveforms in ERP. *Journal of neuroscience methods*, 80(1), 49—64.