ELSEVIER

Journal of Neuroscience Methods 80 (1998) 49–63

JOURNAL OF NEUROSCIENCE METHODS

Decomposing stimulus and response component waveforms in ERP

Jun Zhang *

Department of Psychology, The University of Michigan, 525 East University Avenue, Ann Arbor, MI 48109-1109, USA

Received 11 February 1997; received in revised form 14 November 1997; accepted 16 November 1997

Abstract

Event-related potentials (ERPs) are evoked brain potentials that are averaged across many trial repetitions with individual trials aligned (i.e. time-locked) to a specific behavioral event, typically the onset of the stimulus (s-lock) or the onset of the behavioral response (r-lock). These evoked potential averages may reflect brain activities during the stimulus encoding/analyzing stage (stimulus component waveform, or 'S-component'), during the response preparation/production stage (response component waveform, or 'R-component'), or a combination thereof. In the stimulus-locked average of the ensemble of the recorded waveforms (i.e. in the s-locked ERP), the contribution of an Rcomponent will be convoluted, due to the trial-by-trial variance in reaction time (RT); so will an S-component in the r-locked ERP. It is shown here that the knowledge of (1) the s-locked and r-locked ERP waveforms constructed from the same ensemble of trials and (2) the RT distribution of this ensemble allows us to determine whether the recorded potential results from a single S-component, a single R-component, or a single intermediate ('decisional' or D-) component related to the transition of the two stochastically independent stages. If it can be assumed that the evoked potential is the result of a linear summation of an S-component and an R-component, then there is a unique recovery into these two components, such that the reconstructed waveform on an individual trial is a superposition of the two components with their relative offset determined by the RT of that trial and the ensemble average is the experimentally obtained s-locked and r-locked ERP waveforms. Two independent methods can be used to recover those components, one based on Fourier transform techniques which was first proposed by Hansen (1983) in the context of ERP component isolation and the other based on a recursive iteration approach through which the contamination of the R or S-component is successively removed from the s-locked or r-locked ERP waveforms, respectively. The iterative procedure is analytically proven to converge to the Fourier-based solution, demonstrating the equivalence of the two approaches. Finally, if the condition of a single intermediate D-component is satisfied, then one can recover this component waveform along with the probability distributions of the relative durations of the two underlying linear stages; however, there is always an equivalent pair of S- and R-component which also satisfy the same data set (s-locked and r-locked ERP waveforms and the overall RT distribution). In this case, the S/R-component assumption and the D-component assumption cannot be distinguished solely on the ground of the available data set. The technique developed here outlines the assumptions and the boundary conditions upon which ensemble ERP waveforms are to be analyzed and interpreted in terms of processing mechanisms related to stimulus, to response, or to the transition between the two.

Keywords: Event-related potentials; Evoked potentials; Stimulus-lock; Response-lock; Stimulus component; Response component; Fourier transform

1. Introduction

Event-related potentials (ERPs) are small, phasic brain potentials elicited in conjunction with sensory, motor, or cognitive events and can be extracted from the ongoing electroencephalogram (EEG) recordings from scalp electrodes. An essential step to reduce the variability and nonstationarity of evoked potential is the signal averaging technique (Dawson, 1951) through off-line computation of ERP averages across many

* Tel.: +1313 763 6161; fax: +1313 763 7480; e-mail: junz@umich.edu

0165-0270/98/$19.00
PII S0165-0270(97)00194-5

repeated trials to reduce noise and to detect event-related components. The sequence of ERP peaks and troughs are assumed to reflect an ordered sequence of neural events associated with information processing stages of the brain. The basic premise for the signal averaging technique is the assumption that the waveform to be averaged is the sum of a signal waveform which is stationary across trial repetitions with negligible variance and a noise waveform which is a stationary random process and independent of the signal waveform. Although various adaptive filters techniques have been developed to deal with latency jitter in single-trial waveform thereby providing better estimates of various peak components (Woody, 1967; McGillem and Aunon, 1977; Yu and McGillem, 1983; McGillem et al., 1985; Yu et al., 1994a,b; Gupta et al., 1996), the primary form of data analysis and presentation by psychophysiologists is still the ensemble averaged ERP waveform. Often, the ERP averages are calculated with individual trials (in an ensemble of many repetitions) aligned to the stimulus-onset; in this way, ERP components related to the processing of stimulus, such as P300 (Squires et al., 1975; Hillyard and Picton, 1987) and N400 (Kutas and Hillyard, 1980), can be demonstrated. However, ERP averages have also been calculated with individual trials aligned to the onset of behavioral response to reveal mental processes involved in movement preparation and initiation, such as the LRP or lateralized readiness potential (Coles, 1989) and ERN or error-related negativity (Gehring et al., 1993; Falkenstein et al., 1995). However, stimulus-locked ERP average may contain, in addition to those events or components that are more tightly coupled with the moment of stimulus onset (collectively called 'stimulus component waveform' or 'S-component'), residual contaminations from those events or components that are more tightly coupled with the moment of response onset (collectively called 'response component waveform' or 'R-component') and vice versa for the response-locked ERP. These component waveforms have yet to be properly separated in order for any peak or trough to be accurately attributed to stimulus-related or response-related events. Previously, Hansen (1983) outlined a general method for isolating EPR component waveforms from recordings of linear sums of these time-invariant but overlapping components based on known distributions of trial-by-trial time-shift of each component. Based on the Fourier transform technique, this inverse-filtering method will apply to any case where the number of recording trials is greater than the number of underlying components to be recovered (assuming all of their temporal distributions are known and assuming no noise). Woldorff (1993) deviced a technique that specifically deals with distortion of ERP averages due to overlapping evoked potential responses to rapidly and successively presented stimuli. An iterative procedure is suggested for the estimation and removal of the residual distortion in ERP averages due to stimulus sequence with short interstimulus intervals (ISI). In a slightly different context which also deals with the problem of identifying the stimulus/response contribution to neural activities, Seal and Commenges (1985) proposed (for their analysis of single neuron data) to quantitatively compare the temporal relationships between the time of peak neuronal response (i.e. peak latency) and the time of stimulus onset or movement initiation in order to determine whether the neuronal response is attributable to stimulus processing or response processing. For each neuron, the distributions of peak latency with reference to either the stimulus onset or the movement onset can be obtained and a comparision of the relative variances for the stimulus-locked distribution and the response-locked distribution across the ensemble of trials should suggest whether the peak neuronal activities are more tightly coupled to the stimulus or to the behavioral response. Of course, this analysis cannot be directly applied to ERP data simply because there are multiple peaks and troughs that are not apparent in single-trial data due to relatively poor signal-to-noise ratio.

In this report, the basis and assumptions of attributing and decomposing ERP waveforms into underlying component waveforms (that are related to stimulus, response, or possibly an intermediate 'decisional' stage) is formally investigated. First, constraints will be derived on the stimulus-locked ERP average, the response-locked ERP average and the reaction time distribution for these averages to be generated by (i.e. attributable to) a pure stimulus-coupled waveform (S-component), a pure response-coupled waveform (R-component), or by a waveform time-locked to the transition between the stimulus and response stages (decisional or D-component). Secondly, under the assumption of linear combination of an S-component and an R-component (dual component assumption), there is a unique recovery of those component waveforms; both the Fourier transform technique and the recursive iteration procedure will lead to the same solution. An example will be given to illustrate the two decomposition approaches which are analytically proven to be equivalent. Finally, it is shown that in the case where the single D-component assumption is satisfied, the D-component waveform can be successfully recovered; however, there is always an equivalent pair of S- and R-component waveforms (under the dual component assumption) that could give rise to the exact same set of data (i.e. ERP averages and RT distribution). Therefore, the solution

to the decomposition problem (given ERP averages and RT distribution) may not be unique without specifying the premises (assumptions) underlying ERP decomposition.

The intent of this report is to define the assumptions and boundary conditions for relating ERP averages to the underlying brain processes (i.e. related to stimulus, to response, or to the 'decision' that connects the two) when knowledge of behavioral events are very limited—only the times of stimulus onset and of the onset of behavioral response are known on each trial. In most ERP experiments, stimulus-locked and response-locked average waveforms are obtained along with the reaction-time distribution of the ensemble of trials. Given these available waveforms and distribution, the question to be addressed is what sort of conclusions regarding stimulus and response processing (which are of interest to psychologists) one can draw purely from a signal processing perspective. The results of this paper, therefore, provide constraints for properly interpreting ERP average waveforms arising from the signal averaging technique so widely used in ERP research.

2. Mathematical analysis

2.1. Preliminary

Mathematically, the question that is addressed in this report is as follows: Given the experimental data of (a) the s-locked ERP average[1] waveform $\mathscr{F}_s(t)$; (b) response-locked ERP average waveform $\mathscr{F}_r(t)$ and (c) the distribution rt(t) (in terms of percentage of trials) of reaction times t across the ensemble of trials, how could one infer whether the evoked potential waveform results from a single stimulus-locked component waveform (S-component), from a single response-locked component waveform (R-component), from both, or otherwise?

It is helpful to first clarify the meaning of each of these available data. The ERP average waveforms are averaged across many trial repetitions that are behaviorally indistinguished except for their reaction times, which vary from trial to trial. The single-trial evoked potential waveform is a time series (a sequence of numbers) representing voltage fluctuations against a reference ground-level. Each trial is associated with two marks on the time axis, an S-mark indicating the stimulus onset and an R-mark indicating the response onset. These are the only two behavioral marks or reference points for the single-trial waveform (time-series). The reaction time (RT) is simply a measure of the time lapse between these two marks. Of course, the R-mark of every trial in the ensemble can be arbitrarily shifted by a fixed amount (e.g. toward the right or 'postponed' by a certain amount, say 150 ms), thereby changing the meaning of this mark to something like '150 ms after the onset of behavioral response'. The S-mark may be shifted arbitrarily as well. As a result, the time lapse in between, which operationally defines the 'reaction time', also changes. But those changes are nominal and only have trivial consequence in the following formulation.

Single-trial waveforms from the same ensemble of trials have been averaged in two distinct ways: once where the S-marks of all individual trials are aligned to produce the stimulus-locked ERP average $\mathscr{F}_s(t)$ and once where the R-marks of all individual trials are aligned to produce the response-locked ERP average $\mathscr{F}_r(t)$. Here, we adopt the convention that the reference zero on the time axis refers to either (a) the moment of stimulus onset for the s-locked ERP waveform, that is, $\mathscr{F}_s(0)$ is the amplitude of evoked potential at stimulus onset time, orb) to the moment of response onset for the r-locked ERP waveform, that is, $\mathscr{F}_r(0)$ is the amplitude of evoked potential at response onset time. Everything that occurs in-between the S-mark and the R-mark is coded with $t>0$ for $\mathscr{F}_s(t)$ and $t<0$ for $\mathscr{F}_r(t)$.

Often, it is convenient to discuss ERP waveforms in terms of their temporal frequency components. Let

$$\tilde{\mathscr{F}}_s(k)=\int \mathscr{F}_s(t)\,\mathrm{e}^{-\mathrm{ikt}}\,\mathrm{d}t, \tag{1}$$

$$\tilde{\mathscr{F}}_r(k)=\int \mathscr{F}_r(t)\,\mathrm{e}^{-\mathrm{ikt}}\,\mathrm{d}t \tag{2}$$

be the complex Fourier transforms of $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$, respectively, with k denoting the temporal frequency in the Fourier domain and the integral taken from $-\infty<t<\infty$. In this Fourier (or frequency) domain representation, $\tilde{\mathscr{F}}_s(k)$ and $\tilde{\mathscr{F}}_r(k)$ are both complex numbers that compactly represent the cosine and the sine transforms of $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$, respectively. These complex-valued components can be equivalently represented by the amplitude and phase spectra conjointly, i.e.

$$\tilde{\mathscr{F}}_s(k)=\left|\tilde{\mathscr{F}}_s(k)\right|\exp(i\Phi\{\tilde{\mathscr{F}}_s(k)\}), \tag{3}$$

$$\tilde{\mathscr{F}}_r(k)=\left|\tilde{\mathscr{F}}_r(k)\right|\exp(i\Phi\{\tilde{\mathscr{F}}_r(k)\}). \tag{4}$$

A few words about the reaction time (RT) distribution. Commonly, RT distribution rt(t) describes the probability density (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. This is the stimulus-referenced representation (with S-marks aligned). Its Fourier domain representation is introduced:

[1] In this report, 'average' can refer to either within subject average over repeated trials, or, in addition to average over trial repetitions, across subject as well (the so-called grand-average). This difference in interpretation does not affect the formal exposition.

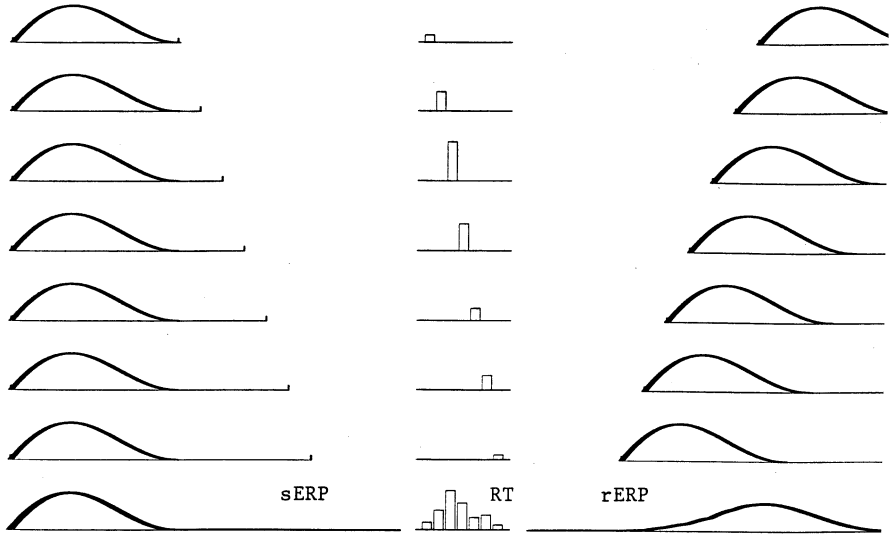

Fig. 1. Schematic representation of the contribution of a stimulus component waveform (S-component) to s-locked ERP average (left panel) and r-locked ERP average (right panel). Each trace represents symbolically the S-component waveform of a single trial, with stimulus onset and response onset represented by small vertical marks at its boundaries. In the left panel, the moment of stimulus onset in individual trials has been aligned, whereas in the right panel, the moment of response onset for the same trials has been aligned. The duration of a trial, as indicated by the length in-between the two vertical marks, is determined by the reaction time, which is given in the central panel. The height of a bar represents the proportion of trials with that particular value of RT, while the overall RT distribution across the ensemble of trials given as the bar histogram in the bottom row. In the responselocked representation (i.e. right panel) of this S- component, there is a leftward shift of the waveform in an amount specified by the RT on the corresponding trial(s). The s-locked and r-locked ERP averages in the bottom row (denoted sERP and rERP) are calculated by summing all individual waveforms above as weighted by the corresponding height of the RT-bar. As can be seen straightforward, the sERP is identical to the S-component, while rERP is related to the S-component in a convoluted manner.

$$\tilde{\mathrm{rt}}(k) = \int \mathrm{rt}(t)\, \mathrm{e}^{-\mathrm{ikt}}\, \mathrm{d}t. \tag{5}$$

On the other hand, the RT of a trial can be equivalently described in terms of how long ago the stimulus appeared before a fixed response onset moment. This response-referenced representation of the RT (with R-marks aligned) is mirror-symmetric to the stimulus-referenced representation.

2.2. Pure S-component or pure R-component

The case of a pure S- or R-component waveform is considered first. Suppose the recorded ERP is generated (apart from any background white noise that is stochastically independent from this signal generation process) by a single underlying component that is time-locked to stimulus (S-component), with its waveform given by $f_s(t)$. This is to say that, on every single trial, the evoked potential waveform is identical after the moment of stimulus onset has been aligned[2]. Thus, the s-locked ERP waveform averaged across the ensemble of trials is simply

$$\mathscr{F}_s(t) = f_s(t). \tag{6}$$

However, the relationship between $f_s(t)$ and the r-locked ERP average waveform is not that simple. When the moment of response onset on each trial has been aligned, its waveform will be shifted along the time axis by an amount that depends on the RT of that trial. Furthermore, the number of trials whose waveforms are shifted by that exact amount is the number of trials (as given by the RT-distribution data) that contain the same RT value. The response-locked sum and, after divided by the total number of trials, the averaged waveforms (across the ensemble of trials) is related to both $f_s(t)$ and $\mathrm{rt}(t)$. Formally, since the percentage of trials with stimulus onset at $t = -\tau$ (prior to the fixed response onset moment) is $\mathrm{rt}(\tau)$, the waveform on those trials, after a leftward shift in the amount of τ along the time axis to become $f_s(t+\tau)$, contributes toward $\mathscr{F}_r(t)$ by the amount

$$f_s(t+\tau)\mathrm{rt}(\tau),$$

Summing over all possible reaction times,

$$\mathscr{F}_r(t) = \int f_s(t+\tau)\mathrm{rt}(\tau)\, \mathrm{d}\tau. \tag{7}$$

Combining Eq. (6) with Eq. (7) yields

$$\mathscr{F}_r(t) = \int \mathscr{F}_s(t+\tau)\mathrm{rt}(\tau)\, \mathrm{d}\tau. \tag{8}$$

[2] The variance in reaction time is assumed to affect neither the time-course nor the peak amplitude of this waveform. If this cannot be assumed, or there is random, trial-by-trial jitter between stimulus onset and the evoked potential waveform, then the S-component under discussion refers to the average waveform (averaged over its dependency on RT latency or stimulus-onset jittering). The same argument applies to the Rcomponent to be discussed.

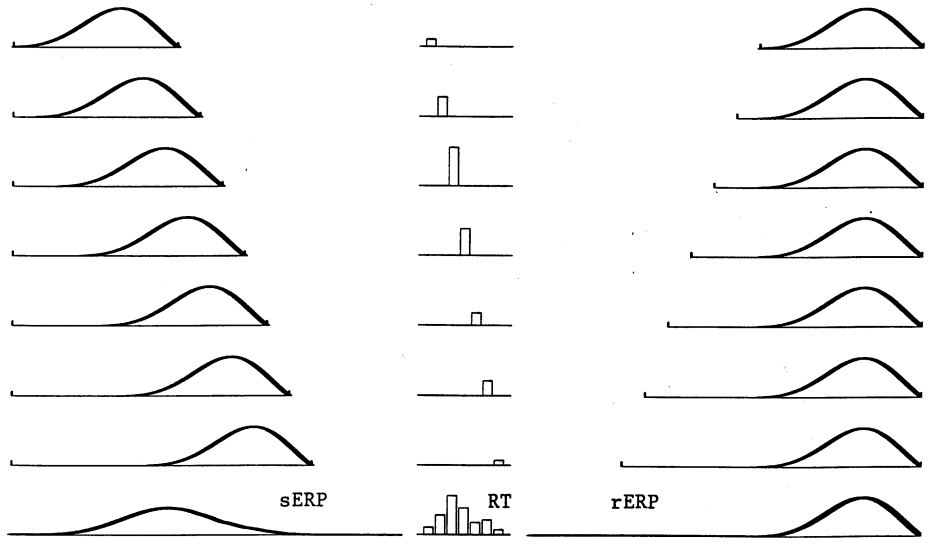

Fig. 2. Schematic representation of the contribution of a response component waveform (R-component) to s-locked ERP average (left panel) and r-locked ERP average (right panel). Each trace represents symbolically the R-component waveform of a single trial, with stimulus onset and response onset represented by small vertical marks at its boundaries. In the left panel, the moment of stimulus onset in individual trials has been aligned, whereas in the right panel, the moment of response onset for the same trials has been aligned. The duration of a trial, as indicated by the length in-between the two vertical marks, is determined by the reaction time, which is given in the central panel. The height of a bar represents the proportion of trials with that particular value of RT, with the overall RT distribution across the ensemble of trials given as the bar histogram in the bottom row. In the stimulus-locked representation (i.e. left panel) of this R- component, there is a rightward shift of the waveform in an amount specified by the RT on the corresponding trial(s). The s-locked and r-locked ERP averages in the bottom row (denoted sERP and rERP) are calculated by summing all individual waveforms above as weighted by the corresponding height of the RT-bar. As can be seen straightforward, rERP is identical to the R-component straightforwardly, while sERP is related to the R-component in a convoluted manner.

This equation determines a relationship among $\mathscr{F}_s(t)$, $\mathscr{F}_r(t)$ and $\mathrm{rt}(t)$ that must be obeyed if the recorded ERP is generated by a single, S-component waveform (Fig. 1).

An analogous relationship can be derived if the recorded ERP is generated by a single, R-component waveform. In this case, the r-locked ERP average $\mathscr{F}_r(t)$ is simply the R-component waveform denoted by $f_r(t)$:

$$\mathscr{F}_r(t) = f_r(t). \tag{9}$$

The contribution of $f_r(t)$ to the s-locked ERP waveform $\mathscr{F}_s(t)$ has to take into account the distribution of RTs across the ensemble of trials. Since the percentage of trials with response onset at $t=\tau$ (following the fixed stimulus onset moment) is $\mathrm{rt}(\tau)$, the waveform on those trials, when the time axis has been aligned with respect to the moment of stimulus onset, is shifted rightward by an amount τ to become $f_r(t-\tau)$. These trials contribute to $\mathscr{F}_s(t)$ by the amount

$$f_r(t-\tau)\mathrm{rt}(\tau).$$

When all trials are considered,

$$\mathscr{F}_s(t) = \int f_r(t-\tau)\mathrm{rt}(\tau)\,\mathrm{d}\tau. \tag{10}$$

Combining Eqs. (9) and (10) gives the constraint on $\mathscr{F}_s(t)$, $\mathscr{F}_r(t)$ and $\mathrm{rt}(t)$ that must be satisfied if the recorded ERP contains a single, R-component waveform (Fig. 2):

$$\mathscr{F}_s(t) = \int \mathscr{F}_r(t-\tau)\mathrm{rt}(\tau)\,\mathrm{d}\tau. \tag{11}$$

Though Eqs. (8) and (11) are cast in integral form, they nevertheless involve simple convolution operation[3]:

$$\mathscr{F}_r(t) = \mathscr{F}_s(t) * \mathrm{rt}(-t),$$

$$\mathscr{F}_s(t) = F_r(t) * \mathrm{rt}(t).$$

Convolutions are easily handled if the mathematical representation is switched to the frequency domain. Since the Fourier transform of a convolution of two functions equals the product of Fourier transforms of these two functions,

$$\tilde{\mathscr{F}}_r(k) = \tilde{\mathscr{F}}_s(k)\cdot \tilde{\mathrm{rt}}(-k) \text{ for a pure S}-\text{component}, \tag{12}$$

[3] The convolution operator * between two functions $p(t)$ and $q(t)$ is defined by

$$p(t)*q(t) = \int p(t-\tau)q(\tau)\,\mathrm{d}\tau = \int p(\tau)q(t-\tau)\,\mathrm{d}\tau = q(t)*p(t)$$

so that

$$p(t)*q(-t) = \int p(t-\tau)q(-\tau)\,\mathrm{d}\tau = \int p(t+\tau)q(\tau)\,\mathrm{d}\tau$$

$$p(-t)*q(t) = \int q(t-\tau)p(-\tau)\,\mathrm{d}\tau = \int p(\tau)q(t+\tau)\,\mathrm{d}\tau$$

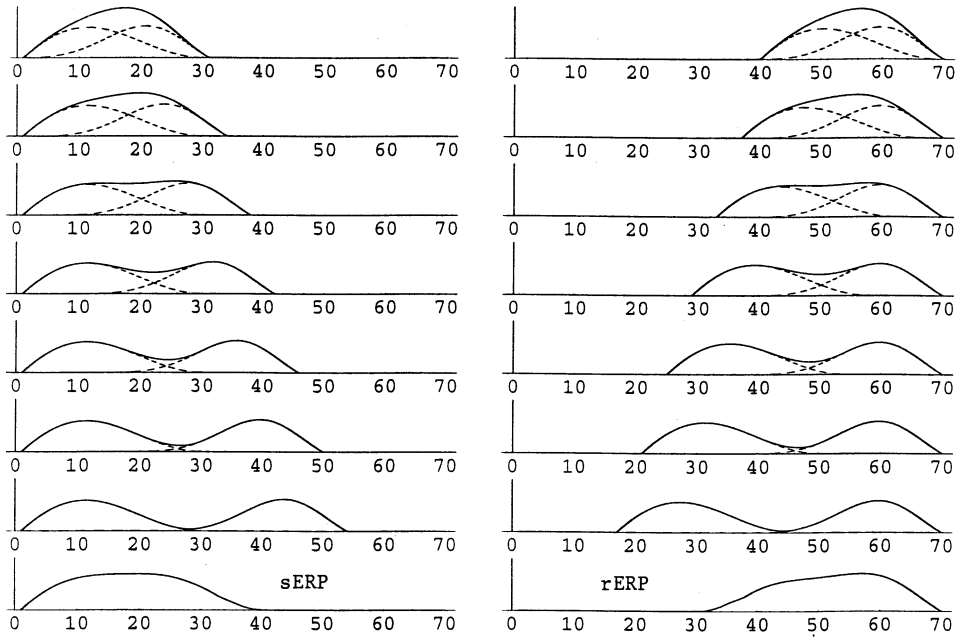

Fig. 3. Schematic representation of the contribution of an S-component waveform and an R-component waveform to s-locked and r-locked ERP averages. The dotted trace represents symbolically the S- and R- component waveforms, and the solid trace the overall waveform on a single trial. Individual trials are aligned with respect to stimulus onset (left panel) or to response onset (right panel). The numbers are arbitrary chosen and symbolically represent a time series, with stimulus onset as bin 0 in the stimulus-locked representation, and response onset as bin 70 in the response-locked representation. The relative offset in time between the S-component and the R-component is determined by the RT on corresponding trials (not shown). The s-locked and r-locked average waveforms in the bottom row are weighted sums of all individual waveforms. The two individual components (dotted waveforms) can be recovered from the s-locked ERP average (sERP), r-locked ERP average (rERP), and RT distribution using either a Fourier-based procedure or an iterative procedure (explained in the text).

$$\tilde{\mathscr{F}}_s(k) = \tilde{\mathscr{F}}_r(k)\cdot \tilde{rt}(k) \text{ for a pure R} - \text{component.} \quad (13)$$

2.3. An S-component with an R-component

Next to be considered is the case where the evoked potential waveform is contributed to by two underlying components simultaneously, an S-component $f_s(t)$ that is always timelocked to the onset of stimulus and an R-component $f_r(t)$ that is always time-locked to the onset of behavioral response. They are assumed to linearly sum (without interaction) to give rise to the overall evoked potential waveform. The complication here is, of course, that their relative offset in time varies across individual trials in an amount given by the RT distribution $\mathrm{rt}(t)$ (Fig. 3).

Because of the linearity assumption, the contribution of $f_s(t)$ and $f_r(t)$ can be considered separately. When the s-locked average waveform $\mathscr{F}_s(t)$ is concerned, the contribution from $f_s(t)$ is straightforward and identical to Eq. (6), whereas the contribution from $f_r(t)$ is convoluted and has the form of Eq. (10). The overall equation is

$$\mathscr{F}_s(t) = f_s(t) + \int f_r(t-\tau)\mathrm{rt}(\tau)\,\mathrm{d}\tau. \quad (14)$$

For the r-locked average waveform $\mathscr{F}_r(t)$, the contribution from $f_s(t)$ is convoluted and takes the form of Eq. (7), whereas the contribution from $f_r(t)$ is straightforward and identical to Eq. (9):

$$\mathscr{F}_r(t) = f_r(t) + \int f_s(t+\tau)\mathrm{rt}(\tau)\,\mathrm{d}\tau. \quad (15)$$

Both can be cast in the convolution notation

$$\mathscr{F}_s(t) = f_s(t) + f_r(t) * \mathrm{rt}(t), \quad (16)$$

$$\mathscr{F}_r(t) = f_r(t) + f_s(t) * \mathrm{rt}(-t). \quad (17)$$

Through working in the Fourier representation as detailed in the Appendix, a unique, closed-form solution can be found[4]

$$f_s(t) = \frac{1}{2\pi}\int \frac{\tilde{\mathscr{F}}_s(k) - \tilde{\mathscr{F}}_r(k)\cdot \tilde{rt}(k)}{1 - |\tilde{rt}(k)|^2}\, e^{ikt}\,\mathrm{d}k, \quad (18)$$

$$f_r(t) = \frac{1}{2\pi}\int \frac{\tilde{\mathscr{F}}_r(k) - \tilde{\mathscr{F}}_s(k)\cdot \tilde{rt}(-k)}{1 - |\tilde{rt}(k)|^2}\, e^{ikt}\,\mathrm{d}k. \quad (19)$$

When Eq. (12) or Eq. (13) is satisfied, $f_r(t)$ or $f_s(t)$ respectively becomes zero. This reduction to the single component case demonstrates the internal consistency of the above derivations.

[4] This solution, which is independently derived here, was first developed in Hansen (1983), but there is an error in his final form. Unfortunately, Hansen's result has almost been ignored, as evidenced by the few citations generated since its publication.

An alternative approach to solve the coupled intergral Eqs. (16) and (17) is through the use of iteration method. Rearranging the terms

$$f_s(t) = \mathscr{F}_s(t) - f_r(t) * \mathrm{rt}(t),$$

$$f_r(t) = \mathscr{F}_r(t) - f_s(t) * \mathrm{rt}(-t)$$

and mutually substituting,

$$f_s(t) = \mathscr{F}_s(t) - \mathscr{F}_r(t) * \mathrm{rt}(t) + (f_s(t) * \mathrm{rt}(-t)) * \mathrm{rt}(t),$$

$$f_r(t) = \mathscr{F}_r(t) - \mathscr{F}_s(t) * \mathrm{rt}(-t) + (f_r(t) * \mathrm{rt}(t)) * \mathrm{rt}(-t).$$

Due to the associative nature of convolution operation, the above equations can be simplified as

$$f_s(t) = C_s(t) + f_s(t) * A(t), \tag{20}$$

$$f_r(t) = C_r(t) + f_r(t) * A(t), \tag{21}$$

where

$$A(t) = \mathrm{rt}(t) * \mathrm{rt}(-t) = \int \mathrm{rt}(t+\tau)\mathrm{rt}(\tau)\,\mathrm{d}\tau$$

$$= \int \mathrm{rt}(\tau)\mathrm{rt}(t-\tau)\,\mathrm{d}\tau = \mathrm{rt}(-t) * \mathrm{rt}(t) \tag{22}$$

is the auto-correlation of the RT distribution and the auxiliary functions $C_s(t)$ and $C_r(t)$ stand for

$$C_s(t) = \mathscr{F}_s(t) - \mathscr{F}_r(t) * \mathrm{rt}(t), \tag{23}$$

$$C_r(t) = \mathscr{F}_r(t) - \mathscr{F}_s(t) * \mathrm{rt}(-t). \tag{24}$$

The linear, integral Eqs. (20) and (21) are now uncoupled for the two unknown functions $f_s(t)$ and $f_r(t)$. According to a well-known theorem in integral equation (Fredholm Theorem), each has a unique solution. One can employ an iterative procedure (the so-called Neumann series) to find such solution. Choose

$$f_s^{(0)}(t) = C_s(t), \tag{25}$$

$$f_r^{(0)}(t) = C_r(t), \tag{26}$$

and proceed to iteratively obtain ($n = 1, 2, \ldots$)

$$f_s^{(n+1)}(t) = C_s(t) + f_s^{(n)}(t) * A(t), \tag{27}$$

$$f_r^{(n+1)}(t) = C_r(t) + f_r^{(n)}(t) * A(t), \tag{28}$$

until the difference between $f^{(n+1)}(t)$ and $f^{(n)}(t)$ (for both subscripts) is close to zero (under some prespecified criterion). Taking the limit of $n \to \infty$ on both sides of Eqs. (27) and (28) leads to Eqs. (20) and (21). Therefore the series $\{f^{(n)}(t),\ n = 0, 1, 2, \ldots\}$ (with subscript s or r) will eventually converge to the solution of Eqs. (18) and (19). Motivated by this guarantee of convergence, a slight variation of the interation procedure can be adopted, which is directly based on equations Eqs. (16) and (17):

$$f_s^{(n+1)}(t) = \mathscr{F}_s(t) - f_r^{(n)}(t) * \mathrm{rt}(t),$$

$$f_r^{(n+1)}(t) = \mathscr{F}_r(t) - f_s^{(n)}(t) * \mathrm{rt}(-t),$$

with

$$f_s^{(0)}(t) = \mathscr{F}_s(t),$$

$$f_r^{(0)}(t) = \mathscr{F}_r(t).$$

This recursive, cross-iteration gradually removes the contamination of the R- (S-) component from the s-locked and r-locked ERP waveforms and eventually gives rise to the desired solution.

To summarize, it has been shown that if the single-trial evoked potential waveform consists of two independent and linearly additive components, one time-locked to the stimulus onset (S-component) and one time-locked to the response onset (R-component), then, there is a unique recovery of these two components given the s-locked and r-locked ERP averages $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$ and the reaction time distribution $\mathrm{rt}(t)$. One may proceed with the recovery either through working in the Fourier domain via Eqs. (18) and (19), or through a recursive procedure via Eqs. (25)–(28).

2.4. A single component locked to S–R transition

In the above, the component waveforms underlying ERP are assumed either to be timelocked to the stimulus onset, or to the response onset. There is yet the possibility that the ERP waveform is contributed to by a third kind of neural process that is neither time-locked to stimulus onset nor to response onset. Rather, the component waveform is time-locked, on a trial-to-trial basis, to the moment of transition between two successive (e.g. stimulus encoding and response preparation) stages. The actual moment of stage transition on individual trials, which depends on the actual durations of each stage, is unknown but assumed to be random and stochastically independent. The component timelocked to the stage transition (tentatively called 'decisional' or D-component) gives rise to distinct s-locked and r-locked ERP waveforms due to trial-by-trial variations in the moment of transition.

For purpose of derivation, the 'reaction time' distributions for these two linearly additive, stochastically independent stages are denoted as $\mathrm{rt}_1(t)$ and $\mathrm{rt}_2(t)$ (i.e. they are probability distributions of the durations subsumed for each stage). According to linear stage model (Sternberg, 1969), the distribution of overall RT $\mathrm{rt}(t)$ is given by

$$\mathrm{rt}(t) = \mathrm{rt}_1(t) * \mathrm{rt}_2(t) = \int \mathrm{rt}_1(\tau)\mathrm{rt}_2(t-\tau)\,\mathrm{d}\tau, \tag{29}$$

(this can be easily understood if we consider RT distributions as probability density functions and that Prob(RT = t) for a trial with two stages equals the product of Prob(RT = τ) for stage I and Prob(RT = $t - \tau$) for stage II, summed over all possible values of τ). In Fourier domain, this is

$$\tilde{\mathrm{rt}}(k) = \tilde{\mathrm{rt}}_1(k) \cdot \tilde{\mathrm{rt}}_2(k). \tag{30}$$

Of course, the exact forms of $\mathrm{rt}_1(t)$, $\mathrm{rt}_2(t)$, or their Fourier counterparts $\tilde{\mathrm{rt}}_1(k)$, $\tilde{\mathrm{rt}}_2(k)$, remain unknown.

Now imagine that there is a mark on the time axis for each trial indicating the moment of transition between stage I and stage II and that all trials are aligned with respect to these imaginary marks of stage transition (called 'transitional marks' for short). The decision component $f_d(t)$ is assumed to be time-locked to this transitional mark at every single trial. For stage I, the distribution of time lapses between stimulus onset and the transitional mark across the ensemble of trials is simply $\mathrm{rt}_1(t)$. Being time-locked to the end of stage I (the moment of transition from stage I to stage II), the D-component appears as if it is an R-component with respect to stage I where the RT distribution is $\mathrm{rt}_1(t)$. Therefore, according to Eq. (10), it is related to $\mathscr{F}_s(t)$, the average ERP time-locked to the beginning of stage I, via

$$\mathscr{F}_s(t) = \int f_d(t-\tau)\mathrm{rt}_1(\tau)\,\mathrm{d}\tau. \tag{31}$$

By similar arguments, the D-component, being time-locked to the beginning of stage II (the moment of transition to stage II from stage b, appears as if it is an S-component with respect to stage II where the RT distribution is $\mathrm{rt}_2(t)$. According to Eq. (7), it is related to $\mathscr{F}_r(t)$, the average ERP time-locked to the end of stage II, via

$$\mathscr{F}_r(t) = \int f_d(t+\tau)\mathrm{rt}_2(\tau)\,\mathrm{d}\tau. \tag{32}$$

The above two equations may be cast in convolution terms

$$\mathscr{F}_s(t) = f_d(t) * \mathrm{rt}_1(t),$$

$$\mathscr{F}_r(t) = \mathrm{f}_d(t) * \mathrm{rt}_2(-t),$$

or in Fourier domain

$$\tilde{\mathscr{F}}_s(k) = \tilde{f}_d(k) \cdot \tilde{\mathrm{rt}}_1(k), \tag{33}$$

$$\tilde{\mathscr{F}}_r(k) = \tilde{f}_d(k) \cdot \tilde{\mathrm{rt}}_2(-k). \tag{34}$$

Therefore, combining Eqs. (30), (33) and (34),

$$|\tilde{f}_d(k)|^2 = \tilde{f}_d(k) \cdot \tilde{f}_d(-k) = \frac{\tilde{\mathscr{F}}_s(k) \cdot \tilde{\mathscr{F}}_r(-k)}{\tilde{\mathrm{rt}}_1(k) \cdot \tilde{\mathrm{rt}}_2(k)} = \frac{\tilde{\mathscr{F}}_s((k)\,\tilde{\mathscr{F}}_r(-k)}{\tilde{\mathrm{rt}}(k)}. \tag{35}$$

This is the equation that must be satisfied if the ERP averages are generated by a single Dcomponent time-locked to the transition of two sequential stages with random, stochastically independent durations. Since the left-hand side of Eq. (35) is the Fourier power spectrum of the D-component ($|\tilde{f}_d(k)|^2$) and therefore is always real, it implies that the imaginary part (to be denoted Im) on the right-hand side must vanish. The single D-component hypothesis therefore gives rise to a strong restriction between the waveforms of $\mathscr{F}_s(t)$, $\mathscr{F}_r(t)$ and $\mathrm{rt}(t)$:

$$\mathrm{Im}\left\{\frac{\tilde{\mathscr{F}}_s(k)\tilde{\mathscr{F}}_r(-k)}{\tilde{\mathrm{rt}}(k)}\right\} = 0. \tag{36}$$

Denote the Fourier phase as $\Phi\{\cdot\}$, the above constraint can be explicitly written

$$\Phi\{\tilde{\mathscr{F}}_s(k)\} - \Phi\{\tilde{\mathscr{F}}_r(k)\} - \Phi\{\tilde{\mathrm{rt}}(k)\} = 0 \ (\mathrm{mod}\ 2\pi). \ \text{(Phase Condition)}. \tag{37}$$

When this 'phase condition' is satisfied, the D-component can be recovered from Eq. (35) up to an arbitrary phase spectrum. As for $\tilde{\mathrm{rt}}_1(k)$ and $\tilde{\mathrm{rt}}_2(k)$, the following may be derived

$$|\tilde{\mathrm{rt}}_1(k)|^2 = \tilde{\mathrm{rt}}_1(k) \cdot \tilde{\mathrm{rt}}_1(-k) = \frac{\tilde{\mathscr{F}}_s(k)}{\tilde{\mathscr{F}}_r(k)}\,\tilde{\mathrm{rt}}(-k), \tag{38}$$

$$|\tilde{\mathrm{rt}}_2(k)|^2 = \tilde{\mathrm{rt}}_2(k) \cdot \tilde{\mathrm{rt}}_2(-k) = \frac{\tilde{\mathscr{F}}_r(k)}{\tilde{\mathscr{F}}_s(k)}\,\tilde{\mathrm{rt}}(k). \tag{39}$$

The phase condition Eq. (37) guarantees that the right-hand sides of Eq. (38) and Eq. (39) are real

$$\mathrm{Im}\left\{\frac{\tilde{\mathscr{F}}_s(k)\tilde{\mathrm{rt}}(-k)}{\tilde{\mathscr{F}}_r(k)}\right\} = \mathrm{Im}\left\{\frac{\tilde{\mathscr{F}}_r(k)\tilde{\mathrm{rt}}(k)}{\tilde{\mathscr{F}}_s(k)}\right\} = 0, \tag{40}$$

so $\tilde{\mathrm{rt}}_1(k)$ and $\tilde{\mathrm{rt}}_2(k)$ can be recovered as well from Eq. (38), Eq. (39), again up to a freedom in the phase spectrum. When the pure S-component assumption Eq. (12) is satisfied, $|\tilde{\mathrm{rt}}_1(k) \equiv 1$ from Eq. (38), hence $\mathrm{rt}_1(t) = \delta(t-\tau_1)$ with constant τ_1. Thus, the stage I in this two-stage model has a fixed duration τ_1 and the recovered D-component is time-locked to the beginning of stage II with variable duration-the D-component is indeed equivalent to a pure S-component. Similarly, when the pure R-component assumption Eq. (13)is satisfied, $|\tilde{\mathrm{rt}}_2(k) \equiv 1$ from Eq. (39), hence $\mathrm{rt}_2(t) = \delta(t-\tau_2)$ with constant τ_2. The recovered D-component is time-locked to the ending of the variable-duration stage I, thus indeed equivalent to a pure R-component. The reduction to the cases of pure S- or R-component again demonstrates the internal consistency of the current framework.

2.5. Relationship between hypotheses

To summarize, four different hypotheses have been proposed and dealt with regarding the underlying source for generating single-trial evoked potential waveform: (a) the pure S-component hypothesis (b) the pure R-component hypothesis; (c) the pure D-component hypothesis and (d) the S-plus-R-component (dual component) hypothesis. Each of the single component hypotheses (a–c above) generates a prediction of the relationship between the s-locked and r-locked ERP

averages and the RT distribution function. The corresponding equations Eqs. (12), (13) and (37), specified the strong constraint on the set of experimental data that would be consistent for a certain hypothesis. It should be noted that the constraint for the D-component hypothesis is (relatively speaking) not as strong as the constraint for either the S-component or the R-component hypothesis: Eq. (37) associated with the D-component hypothesis merely specified a relationship between Fourier phase spectra between $\tilde{\mathscr{F}}_s(k)$, $\tilde{\mathscr{F}}_r(k)$, $\wedge$ $r\tilde{t}(k)$, whereas Eqs. (12) and (13) specified a relationship between their Fourier power spectrum in addition to the same phase relationship. As a result, one might always start by testing whether Eq. (37) holds and, if so, proceed to recover $|r\tilde{t}_1(k)|$ and $|r\tilde{t}_2(k)|$ according to Eqs. (38) and (39). If, furthermore, the power spectrum of one of them turns out to be flat (i.e. $\equiv 1$), then the S-component hypothesis or the R-component hypothesis is upheld.

As far as the S-plus-R-component (dual component) hypothesis is concerned, the method described above allows a unique recovery of the two underlying components given the data of s-locked ERP average, r-locked ERP average and RT distribution. However, the method itself does not provide a way of directly verifying or falsifying this hypothesis itself. If it turns out that one of the recovered component is zero, then this hypothesis reduces to the single S- component or single R-component hypothesis. Therefore, as before, one might always start by assuming two (an S- and an R-) components and proceed to recover them both. The single S- or R-component hypothesis could be evaluated under the context of recovered S/R-component Fig. 4.

In cases where the phase condition Eq. (37) is satisfied, this method allows either unique recovery of the S-component plus the R-component under the dual-component hypothesis, or recovery of the D-component along with the two RT distributions under the two-stage D-component hypothesis. A natural question is whether there is a way to compare and evaluate (from within the framework itself) the relative validity of these two hypotheses. The answer is negative, for it can be shown (Appendix A) that for each set of $f_d(t)$, $rt_1(t)$ and $rt_2(t)$ recovered under the D-component hypothesis, there is an equivalent pair of $f_s(t)$ and $f_r(t)$ under the dual-component hypothesis that can generate exactly the same data set (s-locked and r-locked ERP averages and RT distribution):

$$f_s^{\text{equiv}}(t) = \frac{1}{2\pi}\int \frac{1-|r\tilde{t}_2(k)|^2}{1-|r\tilde{t}(k)|^2}\, r\tilde{t}_1(k)\tilde{f}_d(k)\, e^{ikt}\, dk \qquad (41)$$

$$f_r^{\text{equiv}}(t) = \frac{1}{2\pi}\int \frac{1-|r\tilde{t}_1(k)|^2}{1-|r\tilde{t}(k)|^2}\, r\tilde{t}_2(-k)\tilde{f}_d(k)\, e^{ikt}\, dk \qquad (42)$$

It is thus not possible to differentiate, without supplying more criteria, the dual-component hypothesis and the D-component hypothesis from within the framework itself.

3. An illustrative example

For demonstration purpose, the above decomposition methods have been applied to sample data from a study of error-related negativity (ERN) in choice-RT paradigm (Gehring et al., 1993; Falkenstein et al., 1995). The ERPs under analysis were from error trials in a speed-accuracy trade-off paradigm which is briefly summarized as follows (data previously published in Gehring et al., 1993): A warning signal was presented 1 s before the onset of the stimulus array, which consists of one of the following five-letter arrays: HHHHH, SSSSS, HHSHH, SSHSS, each with 0.25 prior probability. Subjects were required to respond by squeezing left or right hand according to the identity (H or S) of the letter at the center of this five-letter array. Overt response (and RT) was defined as a squeeze that exceeded 25% of maximum squeze force (Gehring et al., 1993 for more details of the experiment).

The analysis and the result from the decomposition analysis is shown in Fig. 5. ERP data was taken from one subject, with a total of 71 error trials under analysis. The sampling frequency of the voltage waveform is 100 Hz (each dot represents a 10 ms bin). The stimulus-

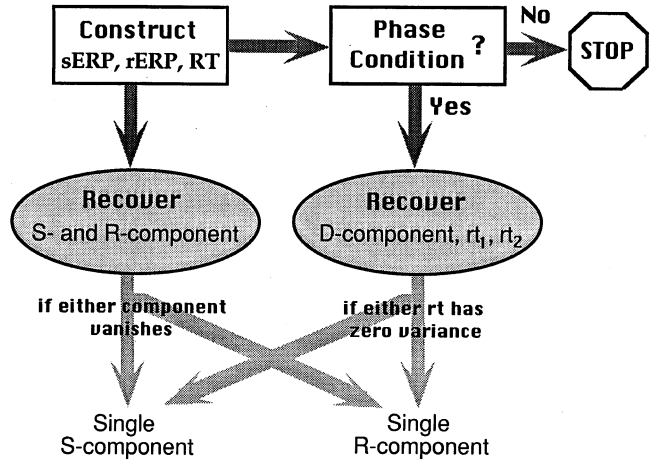

Fig. 4. Flow diagram describing the procedures for recovering S-, R-, or D-component waveforms and the relationship between different hypotheses. From the experimental data set (s-locked ERP average, r-locked ERP average, RT distribution), one can always recover the S-component waveform and the R-component waveform under the dual component hypothesis (left path). If the data satisfy the phase condition (Eq. (37)), then one can also recover the D-component waveform (along with the two RT distributions) under the two-stage, D-component hypothesis (right path). Either hypothesis will reduce to the single component (S- or R-) hypothesis under appropriate conditions.

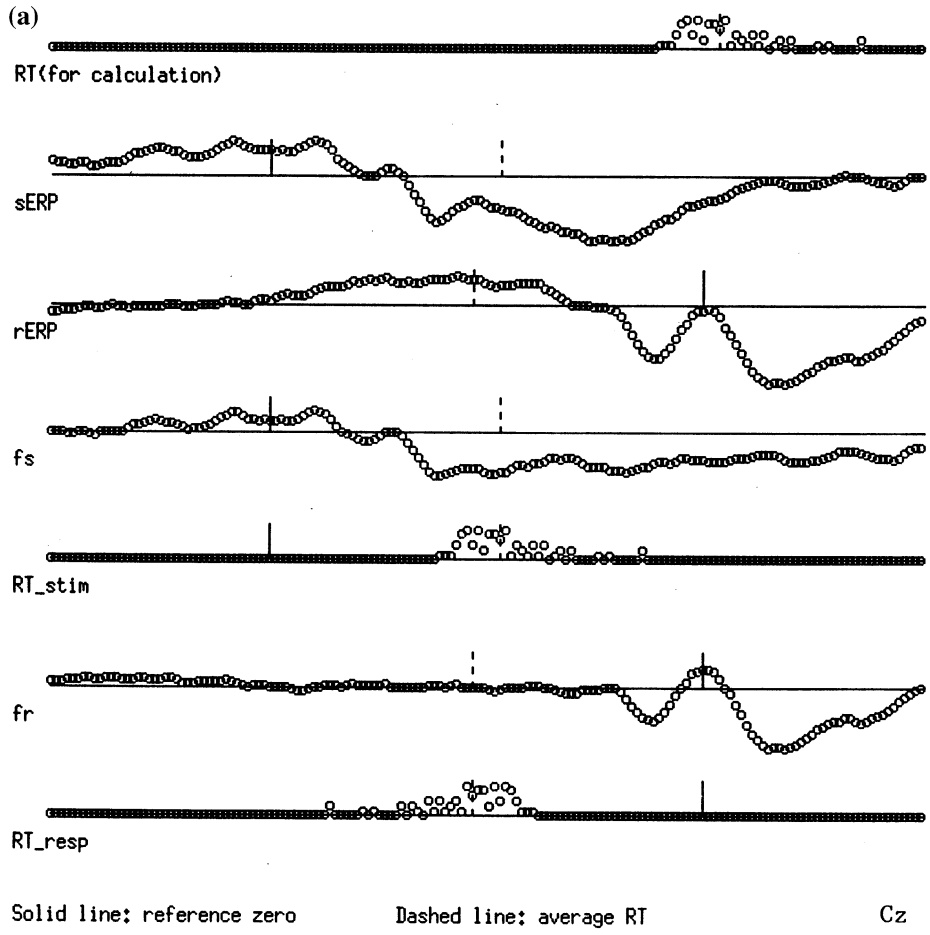

Fig. 5. Recovery of S- and R-component waveforms, denoted as f_s and f_r respectively, from the s-locked and r-locked ERP averages, denoted as sERP and rERP respectively, using the Fourier transform method. RT distributions are given both with reference to stimulus onset and with reference to response onset. The solid line indicates the fixed reference mark (stimlus onset or response onset), whereas the dashed line indicates the average RT with respect to it. The original data was from a published study of error-related negativity (ERN) in a choice-AT paradigm (Gehring et al., 1993). Each dot represents 10 ms time-bin. Negative is upward. Data courtesy of Bill Gehring. (a) Electrode C_z; (b) Electrode P_z.

locked and the response-locked ERP averages were calculated by aligning trials with respect to stimulus- or response-onset and summing over waveforms on individual trials bin-by-bin. Periodic boundary conditions were invoked during the summation to ensure the applicability of the Fourier method. The reaction-time distribution rt(t) was simply constructed from these 71 trials whose durations between stimulus-onset and response- onset are available. The S-component waveform, $f_s(t)$ and the R-component waveform, $f_r(t)$, are recovered using the Fourier transform method Eqs. (18) and (19), using the s-locked ERP average, r-locked ERP average and the RT distribution[5]. As can be seen, the reconstructed S- and T-component waveforms differ from the ERP averages because of the cross-contamination of these pure waveforms in the computation of ERP averages.

A remarkable observation is that the reconstructed R-component waveform turns out to be almost flat until some 150 ms before the response onset. This, of course, is not what the method would have required of the data; there is no restrictions from the decomposition technique that would lead to the flatness of this portion of the waveform, which is sensible since the waveform associated with response processing necessarily lags behind stimulus processing. This may, in turn, provide evidence to support the dual component hypothesis underlying the data of Gehring et al. (1993), i.e. the single-trial waveform is indeed contributed to by an S-component and an R-component in linear additive fashion.

[5] Since their recovery depends heavily on the form of RT distribution, an accurate estimate of rt(t) is crucial. In practice, so long as the reaction times were distributed unimodally and with a central tendency, the method wiH work pretty well. Also, to sensibly approximate the integral (convolution operation) by the summation over a discrete time series, one needs to consider boundary effects, i.e. distortions of the waveform due to truncation of the convolution kernels. We dealt with this effectively by adopting a temporal window (i.e. the upper and lower limits of the summation) large enough to cover the entire region of our interest — the window extends to a time-mark considerably before the stimulus-onset and after the response-onset.

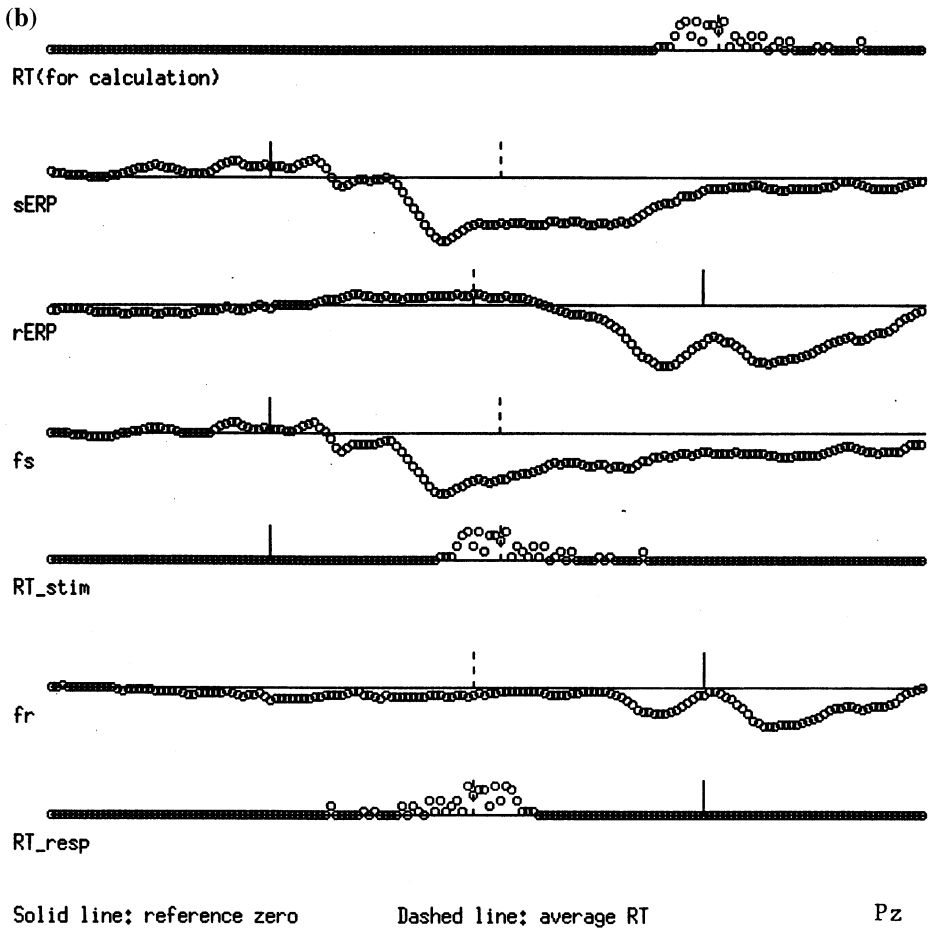

Fig. 5. (*Continued*)

To compare the Fourier method with the iterative procedure of reconstruction, the same ERP averages are recovered using Eqs. (25)–(28). Fig. 6 shows the steps of $n = 1$, $n = 2$, $n = 5$, $n = 10$, $n = 50$ during iteration, illustrating the gradual removal of the cross-contamination of the S-component waveform to r-locked ERP and R-component waveform to s-locked ERP. The final, reconstructed S- and R-component waveforms are virtually indistinguishable from those reconstructed using the Fourier method (given in Fig. 5).

The advantage of the iterative method is that the logic behind it is very clear: Estimates of the S-component and R-component waveforms are first generated and their expected cross-contaminations are removed from the ERP averages. This gives new estimates of component waveforms and new expectation of their cross-contaminations and so on. The estimates of component waveforms gradually approach their actual shapes.

4. Discussion

This report is concerned with the question of how to properly interprete ERP data when the ERP average waveforms are produced by summing experimental data of individual trials time-aligned with respect to stimulus-onset or response-onset. The results are summarized as follows: Given the experimental data of (1) the stimulus-locked (stimulus-aligned) average of evoked potential waveform $\mathscr{F}_s(t)$; (2) response-locked (response-aligned) average of the evoked potential waveform $\mathscr{F}_r(t)$ and (3) the probability distribution $\mathrm{rt}(t)$ of reaction times t across the ensemble of trials, there is a unique recovery, based on either the Fourier approach or the iterative approach, of the underlying stimulus-coupled and response-coupled component waveforms that will give rise to the exact same ensemble-averaged waveforms of $\mathscr{F}_s(t)$, $\mathscr{F}_r(t)$ while satisfying the RT distribution $\mathrm{rt}(t)$. The recovery procedure is directly related and reducible to the conditions of a single S- or R-component waveform. Conditions have also been provided to determine whether the ERP is generated by a single 'decisional' component time-locked to the moment of transition between two successive, stochastically independent stages. Furthermore, it is shown that if this condition is satisfied, any decisional component is associated with a pair of equivalent S- and R-components that can generate the same ERP waveforms. Therefore, with only stimulus-locked and

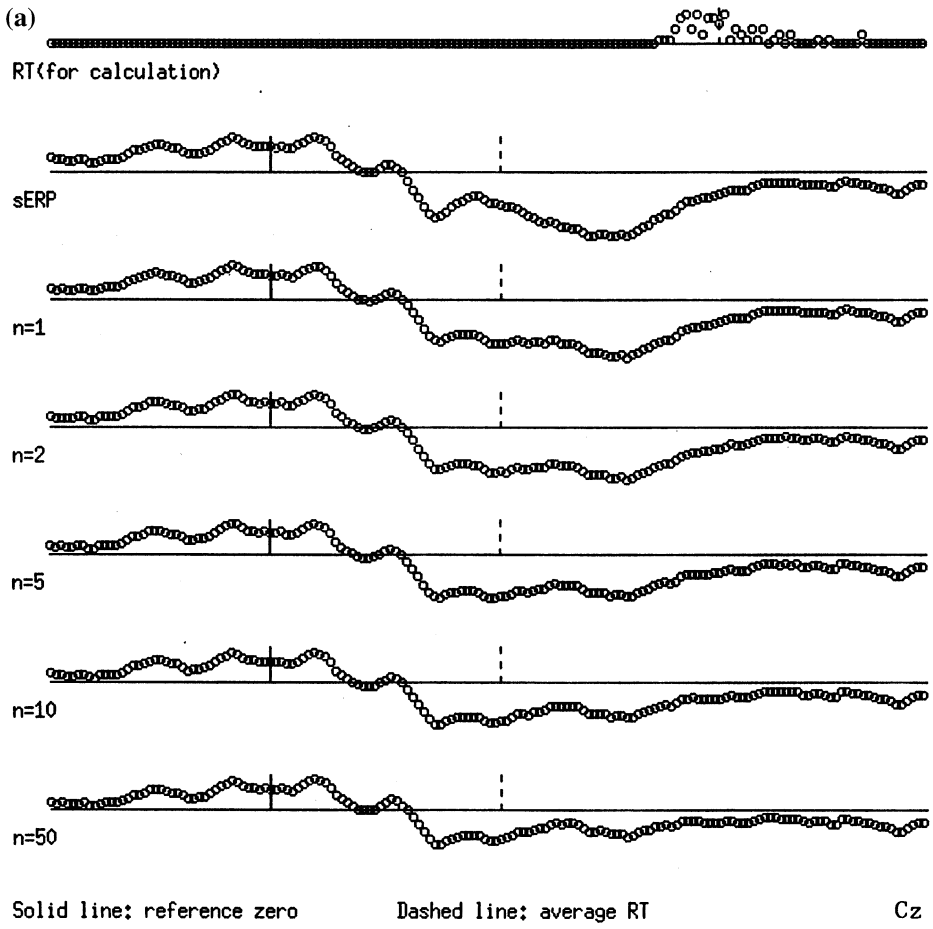

Fig. 6. Recovery of S- and R-component waveforms using the iterative procedure. The original s-locked or r-locked ERP average are shown as the top trace ($n = 0$). The waveforms during successive iterations are shown for $n = 1$, $n = 2$, $n = 5$, $n = 10$, $n = 50$ steps. Data refer to electrode C_z. (a) S-component waveform; (b) R-component waveform.

response-locked ERP averages, it is fruitless to extract any 'covert' component(s) that mediate the transition from stimulus to response. An illustrative example was given to demonstrate the recovery procedure under the dual (stimulus and response) component hypothesis.

It should be pointed out that the stimulus and response component waveforms are recoverable only in the sense that they will generate exactly the same ensemble-averaged waveforms of $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$ under the same RT distribution $\mathrm{rt}(t)$. Specifically, if one were to recontruct, on a trial-by-trial basis, the ensemble of waveforms by linearly summing, in a pointwise fashion (i.e. bin by bin along time axis), the reconstructed component waveforms $f_s(t)$ and $f_r(t)$ such that their relative offset in time is determined appropriately by the RT of that single trial, then this ensemble of single-trial waveforms will give rise to identical $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$. On individual trials, the reconstructed and the original waveform need not match. This is due to large background EEG variation (e.g. alpha activity) in the singletrial waveform. In fact, this method may provide a template for the signal component—$f_s(t)$ plus $f_r(t)$—of the single-trial waveform and thereby allow proper study of the source of these intrinsic (background) variations.

In practical applications, the effectiveness of the method relies on the proper behavior of RT distribution across the ensemble of trials. The less variance the RT distribution is, the less effective the method becomes. Mathematically, when the RT variance is small (i.e. all trials have nearly uniform reaction time), its Fourier transform $\tilde{\mathrm{rt}}(k)$ approaches unity and the recovery error generated by applying Eqs. (18) and (19) becomes large (the denominator approaches zero). This limitation is intrinsic to this method, in the sense that the extent to which $\mathscr{F}_s(t)$ and $\mathscr{F}_r(t)$ differ depends on the relative spread of the RT distribution. If all trials have uniform reaction times, then there is no difference between the e-lock and the r-locked ERP averages. Consequently, those underlying components whose dynamics are distinctly coupled with the stimulus onset or with the response onset cannot be differentiated on the basis of their temporal properties. Surprisingly or not, this method is best applied to a RT distribution which

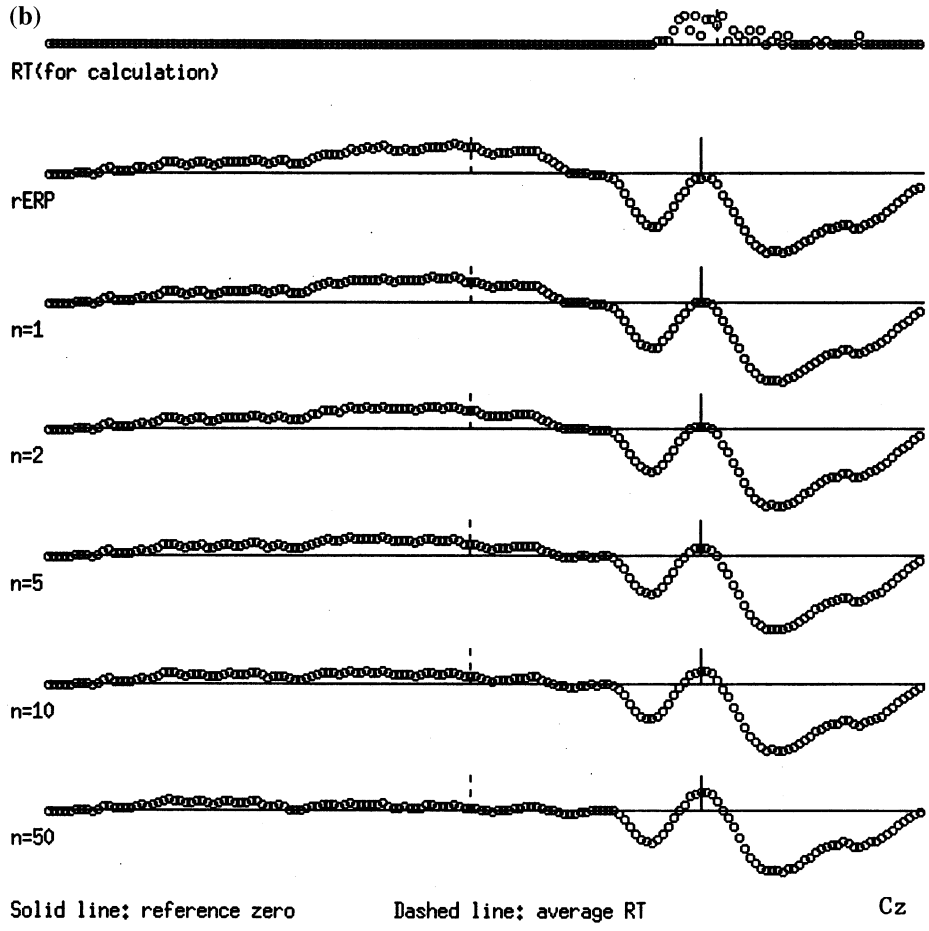

Fig. 6. (*Continued*)

demonstrates a central tendency but with relatively large spread (variance).

Despite of these limitations, the method provides a self-contained framework for unconfounding functional components (stimulus, response, or decisional process) in ERP and defining the boundary conditions under which conclusions about functional components may be drawn. It provides the maximal information one can extract from the ERP average data (and RT distribution) allowed under signal averaging methodology. The dichotomous separation of an S-component and an R-component in the overall ERP is possible under the assumptions that each component is strictly time-locked to stimulus or response onset and its amplitude and time-course are independent of the RT on an individual trial. When these assumptions are violated (e.g. small intrinsic jitter in time-locking and small amplitude modulation across trials), the recovered S- and R-components will be the smoothed or averaged components (with respect to the jitter and modulation), which are still well-defined. Again, this is the best one can do with the kind of ERP average data. When recordings from multiple electrodes are examined in combination, these ERP component waveforms will provide important information to the sensorimotor nature of the underlying generators. The application of this method can be especially useful in conjunction with the conventional analysis of the topography of scalp voltage distributions to localize current sources and sinks as in current source density analysis (Nunez, 1981, 1989; Mackay, 1983), or even the localization of equivalent dipoles (Scherg, 1989; Scherg and Picton, 1991; Dale and Sereno, 1993) related to stimulus and response processing. With this method, the S-component and the R-component waveforms at each electrode would be reconstructed first so that each waveform is separately subject to conventional source localization techniques. Because sources for stimulus components and response components are mostly likely to have different anatomical loci and temporal dynamics that should not be simply averaged, our method would allow more accurate estimation about brain structures and mechanisms generating event-related potentials.

Nomenclature

t	time variable in a time-series	
Eqs. (1) and (2)		
$\mathscr{F}_s(t)$	stimulus-locked ERP average waveform	Eq. (1)
$\mathscr{F}_r(t)$	response-locked ERP average waveform	Eq. (2)
$rt(t)$	reaction-time distribution (with reference to stimulus onset	Eq. (5)
$f_s(t)$	stimulus component waveform in ERP	Eq. (6)
$f_r(t)$	response component waveform in ERP	Eq. (9)
$f_d(t)$	'decisional' component waveform in ERP	Eqs. (31) and (32)
$rt_1(t)$	reaction-time distribution for stage I in the two-stage model	Eq. (29)
$rt_2(t)$	reaction-time distribution for stage II in the two-stage model	Eq. (29)
$A(t)$	auto-correlation of reaction-time distribution	Eq. (22)
$C_s(t)$	auxilary waveform used in the iterative substraction method	Eq. (23)
$C_r(t)$	auxilary waveform used in the iterative substraction method	Eq. (24)
k	frequency component of a time-series (Fourier representation)	Eqs. (1) and (2)
$\tilde{\mathscr{F}}_s(k)$	complex-valued Fourier transform of $\mathscr{F}_s(t)$	Eq. (1)
$\tilde{\mathscr{F}}_r(k)$	complex-valued Fourier transform of $\mathscr{F}_r(t)$	Eq. (2)
$\tilde{rt}(k)$	complex-valued Fourier transform of $rt(t)$	Eq. (5)
$\tilde{f}_s(k)$	complex-valued Fourier transform of $f_s(t)$	Eq. (43)
$\tilde{f}_r(k)$	complex-valued Fourier transform of $f_r(t)$	Eq. (44)
$\tilde{f}_d(k)$	complex-valued Fourier transform of $f_d(t)$	Eqs. (33) and (34)
$\tilde{rt}_1(k)$	complex-valued Fourier transform of $rt_1(t)$	Eq. (30)
$\tilde{rt}_2(k)$	complex-valued Fourier transform of $rt_2(t)$	Eq. (30)
$\|\|$	Fourier amplitude	Eqs. (3) and (4)
Φ	Fourier phase	Eqs. (3) and (4)
*	convolution operator	Fn. 3
Im	imaginary part	Eqs. (36) and (40)

Acknowledgements

An abstract of this report has been presented at the Third Annual Meeting on Cognitive Neuroscience (1996). The author acknowledges technical assistance from Min Chang for implementation of the decomposition algorithms. The ERP data to test these algorithms were kindly provided by Bill Gehring.

Appendix A

To solve Eqs. (16) and (17) by Fourier transform techniques, simply define

$$\tilde{f}_s(k) = \int f_s(t)\, e^{-ikt}\, dt, \tag{43}$$

$$\tilde{f}_r(k) = \int f_r(t)\, e^{-ikt}\, dt. \tag{44}$$

Eqs. (43) and (44) become the following algebraic equations

$$\tilde{\mathscr{F}}_s(k) = \tilde{f}_s(k) + \tilde{f}_r(k)\cdot \tilde{rt}(k),$$

$$\tilde{\mathscr{F}}_r(k) = \tilde{f}_r(k) + \tilde{f}_s(k)\cdot \tilde{rt}(-k),$$

whose solutions can be immediately obtained

$$\tilde{f}_s(k) = \frac{\tilde{\mathscr{F}}_s(k) - \tilde{\mathscr{F}}_r(k)\cdot \tilde{rt}(k)}{1 - \tilde{rt}(k)\cdot \tilde{rt}(-k)} \tag{45}$$

$$\tilde{f}_r(k) = \frac{\tilde{\mathscr{F}}_r(k) - \tilde{\mathscr{F}}_s(k)\cdot \tilde{rt}(-k)}{1 - \tilde{rt}(k)\cdot \tilde{rt}(-k)} \tag{46}$$

(the condition $\tilde{rt}(k)\cdot \tilde{rt}(-k) = |\tilde{rt}(k)|^2 < 1$ is guaranteed so long as the RT distribution is not a δ function, i.e. with non-zero variance). Inverse Fourier transform leads to the unique solution in Eq. (18) and Eq. (19). The solution Eq. (45), Eq. (46) was first derived in Hansen (1983) in the context of general conditions for ERP component isolation (with a minor, typographic error).

If, in fact, the $\tilde{\mathscr{F}}_s(k)$ and $\tilde{\mathscr{F}}_r(k)$ are contributed to by a single component $\tilde{f}_d(k)$ which is time-locked to transition of two successive stages, from Eqs. (45) and (46), the equivalent set of $\tilde{f}_s(k)$, $\tilde{f}_r(k)$ can be calculated as follows

$$\tilde{f}_{s}^{equiv}(k)=\frac{\tilde{\mathscr{F}}_{s}(k)-\tilde{r}(k)\tilde{\mathscr{F}}_{r}(k)}{1-\tilde{r}(k)\tilde{r}(-k)}$$

$$=\frac{\tilde{r}_{1}(k)\tilde{f}_{d}(k)-\tilde{r}_{1}(k)\tilde{r}_{2}(k)\tilde{r}_{2}(-k)\tilde{f}_{d}(k)}{1-\tilde{r}(k)\tilde{r}(-k)}$$

$$=\frac{1-|\tilde{r}_{2}(k)|^{2}}{1-|\tilde{r}(k)|^{2}}\tilde{r}_{1}(k)\tilde{f}_{d}(k)=\frac{1-|\tilde{r}_{2}(k)|^{2}}{1-|\tilde{r}(k)|^{2}}\tilde{\mathscr{F}}_{s}(k)$$

$$\tilde{f}_{r}^{equiv}(k)=\frac{\tilde{\mathscr{F}}_{r}(k)-\tilde{r}(-k)\tilde{\mathscr{F}}_{s}(k)}{1-\tilde{r}(k)\tilde{r}(-k)}$$

$$=\frac{\tilde{r}_{2}(-k)\tilde{f}_{d}(k)-\tilde{r}_{1}(-k)\tilde{r}_{2}(-k)\tilde{r}_{1}(k)\tilde{f}_{d}(k)}{1-\tilde{r}(k)\tilde{r}(-k)}$$

$$=\frac{1-|\tilde{r}_{1}(k)|^{2}}{1-|\tilde{r}(k)|^{2}}\tilde{r}_{2}(-k)\tilde{f}_{d}(k)=\frac{1-|\tilde{r}_{1}(k)|^{2}}{1-|\tilde{r}(k)|^{2}}\tilde{\mathscr{F}}_{r}(k)$$

Inverse Fourier transform leads to Eqs. (41) and (42), the dual (S-plus-R) component equivalent to the solution under single (D-component) hypothesis.

References

Coles MGH. Modern mind-brain reading: Psychophysiology, physiology and cognition. Psychophysiol 1989;26:251–69.

Dale AM, Sereno MI. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn. Neurosci 1993;5:163–76.

Dawson GD. A summation technique for detecting small signals in a large irregular background. J Physiol 1951;115:2–3.

Falkenstein M, Hohnsbein J, Hoormann I. Event-related potential correlates of errors in reaction tasks. Perspectives of Event-Related Potentials Research (EEG Suppl. 44), 1995:280–86.

Gehring WJ, Goss B, Coles MGH, Meyer DE, Dochin E. A neural system for error detection and compensation. Psycholog Sci 1993;4:385–90.

Gupta L, Molfese DL, Tammana R, Simos PG. Nonlinear alignment and averaging for estimating the evoked potential. IEEE Trans Biomed Eng BME 1996;43:348–56.

Hansen JC. Separation of overlapping waveforms having known temporal distributions. J Neurosci Methods 1983;9:127–39.

Hillyard SA, Picton TW. Electrophysiology of congition. In Plum E, editor. Handboook of Physiology: The nervous system, Vol. V, Bethesda, Maryland: American Physiological Society, 1987:519–84.

Kutas M, Hillyard SA. Event-related brain potentials to semantically inappropriate and surprisingly large words. Biol Psychol 1980;11:99–116.

Mackay DM. On-line source density computation with a minimum of electrodes. Electroencephalogr Clin Neurophysiol 1983;56:696–8.

McGillem CD, Aunon II. Measurements of signal components in single visually evoked brain potentials. IEEE Trans Biomed Eng BME 1997;24:232–41.

McGillem CD, Aunon JI, Pomalaza CA. Improved waveform estimation procedures for event-related potentials. IEEE Trans Biomed Eng BME 1985;32:371–9.

Nunez PL. Electric Fields of the Brain: The Neurophysics of EEG. New York: Oxford University Press, 1981.

Nunez PL. Estimation of large scale neocortical source activity with EEG surface Laplacians. Brain Topography 1989;2:141–54.

Scherg M. Fundamentals of dipole source potential analysis. In: Grandori F, Romani GL, Hoke M, editors. Auditory evoked potentials and fields: Advances in audiology 6. Basel: Karger, 1989.

Scherg M, Picton TW. Separation and identification of event-related potential components by brain electric source analysis. In Brunia CHM, Mulder G, Verbaten MN, editors. Event-Related Brain Research, Electroencephalography and clinical neurophysiology Suppl. 42. Amsterdam: Elsevier, 1991:24–37.

Seal J, Commenges D. A quantitative analysis of stimulus- and movementrelated responses in the posterior parietal cortex of the monkey. Exp Brain Res 1985;58:144–53.

Squires NK, Squires KC, Hillyard SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroenceph Clin Neurophysiol. 1975;38:387–401.

Sternberg S. The discovery of processing stages: Extensions of Donders' method. In: Koster WG, editor. Attention and Performance, II. Amsterdam: North-Holland, 1969: 276–315.

Woldorff MG. Distortion of ERP averages due to overlap from temporally adjacent ERPs. Analysis and correction. Psychophysiol 1993;30:98–119.

Woody CD. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med Biol Eng 1967;5:539–53.

Yu KB, McGillem CD. Optimal filters for estimating evoked waveforms. IEEE Trans Biomed Eng BME 1983;30:730–7.

Yu XH, He ZY, Zhang YS. Time-varying adaptive filters for evoked potential estimation. IEEE Trans Biomed Eng BME 1994a;41:1062–71.

Yu XH, Zhang YS, He ZY. Peak component latency-corrected average method for evoked potential waveform estimation. IEEE Trans Biomed Eng BME 1994b;41:1072–82.