International Journal of Psychophysiology 109 (2016) 9–20

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of residue iteration decomposition

Guang Ouyang [a,b,c], Werner Sommer [d], Changsong Zhou [a,b,e,f,*]

[a] Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
[b] Centre for Nonlinear Studies and The Beijing–Hong Kong–Singapore Joint Centre for Nonlinear and Complex Systems (Hong Kong), Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong
[c] Department of Psychology, Ernst Moritz Arndt Universität Greifswald, Germany
[d] Department of Psychology, Humboldt-Universität zu Berlin, D-10099 Berlin, Germany
[e] Research Centre, HKBU Institute of Research and Continuing Education, Virtual University Park Building, South Area Hi-tech Industrial Park, Shenzhen, China
[f] Beijing Computational Science Research Center, Beijing, China

ARTICLE INFO

Article history:
Received 26 April 2016
Received in revised form 12 September 2016
Accepted 25 September 2016
Available online 28 September 2016

Keywords:
ERP
Latency variability
Latency correction
Method
Residue iteration decomposition

ABSTRACT

Stimulus-locked averaged event-related potentials (ERPs) are among the most frequently used signals in Cognitive Neuroscience. However, the late, cognitive or endogenous ERP components are often variable in latency from trial to trial in a component-specific way, compromising the stability assumption underlying the averaging scheme. Here we show that trial-to-trial latency variability of ERP components not only blurs the average ERP waveforms, but may also attenuate existing or artificially induce condition effects in amplitude. Hitherto this problem has not been well investigated. To tackle this problem, a method to measure and compensate component-specific trial-to-trial latency variability is required. Here we first systematically analyze the problem of single trial latency variability for condition effects based on simulation. Then, we introduce a solution by applying residue iteration decomposition (RIDE) to experimental data. RIDE separates different clusters of ERP components according to their time-locking to stimulus onsets, response times, or neither, based on an algorithm of iterative subtraction. We suggest to reconstruct ERPs by re-aligning the component clusters to their most probable single trial latencies. We demonstrate that RIDE-reconstructed ERPs may recover amplitude effects that are diminished or exaggerated in conventional averages by trial-to-trial latency jitter. Hence, RIDE-corrected ERPs may be a valuable tool in conditions where ERP effects may be compromised by latency variability.

1. Introduction

Event-related brain potential (ERP) components are important and frequently employed tools in Cognitive Neuroscience, in both basic and applied settings. The components of averaged ERP waveforms can be related to specific mental sub-processes; questions usually concern amplitude or latency differences between experimental conditions (e.g., Rugg and Coles, 1995), populations (e.g., Polich and Herbst, 2000), or individuals (e.g. Kaltwasser et al., 2014). Recently, latency variability of certain ERP components has received growing attention in single trial studies (e.g. Saville et al., 2014). However, despite the prevalence of the average ERP protocol in cognitive brain research, the existence and consequences of trial-to-trial latency variability (latency jitter) is a long-standing but still under-explored problem (cf., Jung et al., 2001; Luck, 2005; Möcks et al., 1988; Picton et al., 1984; Woody, 1967). The present paper concerns the interpretation of measured amplitude effects as a mixture of amplitude variation and trial-to-trial latency jitter (denoted as 'latency jitter' in the following). Latency jitter smears or blurs averaged ERPs and – depending on the affected condition(s) – may obscure or mimic amplitude differences. For example, schizophrenic patients show consistently smaller P3 amplitudes than healthy controls (e.g., Jeon and Polich, 2003); however the effect may be partly accounted for by different extents of latency jitter as patients also show larger reaction time variability (Ford et al., 1994; Röschke et al., 1996; Roth et al., 2007). Such ambiguities may pervade any study where differences between conditions or population samples are confounded with different degrees of latency jitter, for example aging or brain damage (Fjell et al., 2011; Patterson et al., 1988; Walhovd et al., 2008). In principle, all amplitude variations in average ERP waveform across conditions are mixtures of true amplitude variations and different extents of latency jitter. As we will demonstrate, if strong enough even identical amounts of latency jitter across conditions may suppress

* Corresponding author at: Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.
E-mail address: cszhou@hkbu.edu.hk (C. Zhou).

http://dx.doi.org/10.1016/j.ijpsycho.2016.09.015

true condition effects in amplitude. Therefore, it is imperative to investigate the effects of latency jitter and to find ways of correcting for it. The mixing problem and its solution are still under-explored probably because tackling this issue requires handling highly noisy single trial ERP signals.

Notably, in addition to trial-to-trial latency jitter, there are other causes of ERP waveform blurring, for examples, volume conduction, trial-to-trial variability in amplitude and morphology, inter-individual variability, etc. The present report exclusively addresses the blurring problem due to trial-to-trial latency jitter, first on a theoretical level and then by suggesting a solution based on a novel application of residue iteration decomposition (RIDE; e.g., Ouyang et al., 2015b).

1.1. The consequences of latency jitter in ERPs

ERPs are obtained by averaging many epochs – typically 30 to 60 – of EEG from single trials, assuming that the stimulus- or response-related signal embedded in the EEG is identical from trial to trial but that background noise varies independently of the signal around a mean of zero across trials. But in fact, ERP components may strongly vary in latency across trials, as recognized by ERP researchers for a long time (e.g., Jung et al., 2001; Kutas et al., 1977; Leuthold and Sommer, 1998; Pfefferbaum et al., 1980; Verleger, 1997; Woody, 1967).

The trial-to-trial latency variability of ERP components may have two major consequences on condition effects, as illustrated in Fig. 1. Firstly, (Fig. 1, Case 1) trial-to-trial latency variability blurs the waveforms, attenuating both ERP component amplitudes and amplitude differences between conditions. The reduced amplitude differences may not be large enough to outweigh the noise, diminishing the size of experimental effects and statistical test parameters.

As a second consequence different extents of latency variability across conditions may mimic amplitude effects in ERPs (Fig. 1, Case 2). If conditions with identical amplitudes differ in variability of single trial latencies, the average ERP will show amplitude differences across conditions that might become statistically significant. In this case, amplitude differences may be erroneously attributed to different strengths of activities generated by the underlying neural systems rather than to different degrees of temporal variability of the neural activities across single trials.

In reality, between-condition ERP differences might be affected by a combination of both cases. This problem could lead to the erroneous conclusions that there is an amplitude difference when there is only a difference in latency variability or that there is no amplitude difference, when latency variability obscures a true amplitude effect. Therefore it is highly desirable to solve the ambiguities caused by latency variability.

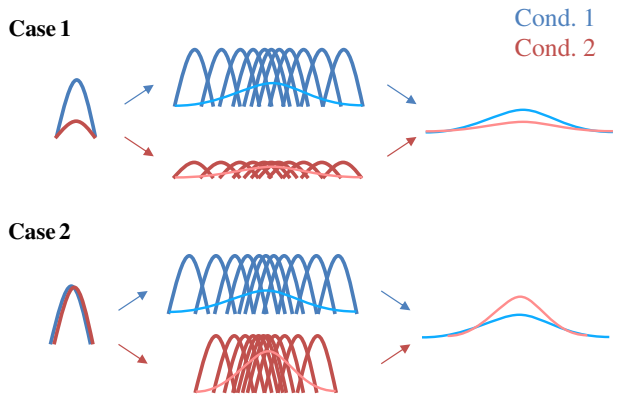

Fig. 1. Illustration of the smearing effect by trial-to-trial latency variability. The blue and red sinus half-waves represent ERP components from two conditions for different cases. Case 1: Two components show the same latency variability but differ in amplitude. The amplitude difference in the average ERPs for two conditions (red and blue) is diminished by trial-to-trial latency variability. Case 2: Two components are the same in amplitude but differ in trial-to-trial latency variability, mimicking an amplitude difference between condition averages.

1.2. Previous attempts

Although the problem of latency jitter in ERP data has been long recognized (Kutas et al., 1977; Woody, 1967), as explained next, a satisfactory solution to the problem has not yet been established. In the following we will briefly review previous suggestions to solve the latency jitter problem and their limitations.

A traditional approach to tackle trial-to-trial latency variability is response-locked averaging, with the idea that late ERP components, blurred in stimulus-locked averaging, will come out more clearly when synchronized to the response. While this assumption is true for response-related components, response-locking has several limitations: 1) It compromises the stimulus-locked components (Fig. 2). 2) There may be components that are neither locked to stimulus onsets nor to RTs, which would be smeared by both stimulus and response locking. 3) A great number of experiments do not require immediate responses or any responses, which precludes response-locked averaging. In this case, an alternative method, latency-locked averaging, is to average single trial ERP to the estimated latencies of a dominant component, for example, the P300 (Ahmadi and Quiroga, 2013; Woody, 1967; Tuan et al., 1987). This approach, however, still suffers from the above limitation since ERPs are not solely composed of a unitary dominant component or component cluster but are rather composed of multiple component clusters with varying inter-component delays (Hansen, 1983; Jung et al., 2001; Verleger, 1997).

Concisely speaking, stimulus-locked, response-locked, or abovementioned latency-locked averaging schemes share the same property of increasing the resolution of a certain component by sacrificing the resolution of other components (Fig. 2), which Poli et al. (2010) metaphorically termed magnifying-glass effect. These authors proposed a reaction-time binning method in order to partly avoid the magnifying-glass effect and suggested to average single-trial ERP from bins with similar RTs. They separated single trials into three bins, each with 30% of the trials after discarding 10% of the trials on the long tail of the RT distribution. Although this procedure is likely to improve the ERP components, the smearing effects are still present due to the spread of RTs, especially in Bin 1 and 3. The results did show that the temporal resolution of ERP in each bin was increased, especially in Bin 2 where RT jitter around the most probable RT value was relatively small. However, by discarding many trials, only part of the data was analyzed, that is, a great amount of information was lost. Since trials with long reaction times were discarded, effects that might be exclusively localized in slow responses might have been diminished. For example, the so-called worst performance rule (Larson and Alderton, 1990) shows that general intelligence is better reflected in the extreme responses (for a review, see Coyle, 2003). In addition, the RT-binning method requires large trial numbers and the analysis is limited to datasets with recorded RTs.

Jung et al. (2001) applied independent component analysis (ICA) to single trial ERP data and identified some independent components (ICs) that seem to have variable latency and correlate with RT. This approach requires to evaluate and select the ICs with variable latency. Since ICA can separate a great number of components (the same as the number of electrodes), there is a large cluster of ICs that are not easy to classify as being locked to the stimuli or RTs, or neither. Clear and applicable criteria are missing for the selection of the ICs.

Another class of methods dealing with trial-to-trial variability is the time marker-based separation of ERP components based on the assumption that ERPs exclusively consist of marker-locked components (Hansen, 1983; Knuth et al., 2006; Yin et al., 2009; Zhang, 1998; Takeda et al., 2008). The markers refer to external events such as stimulus onsets, response times or other cues. Although implemented by different algorithms, there is a common theory underlying these methods – the General Linear Model in which the markers serve as

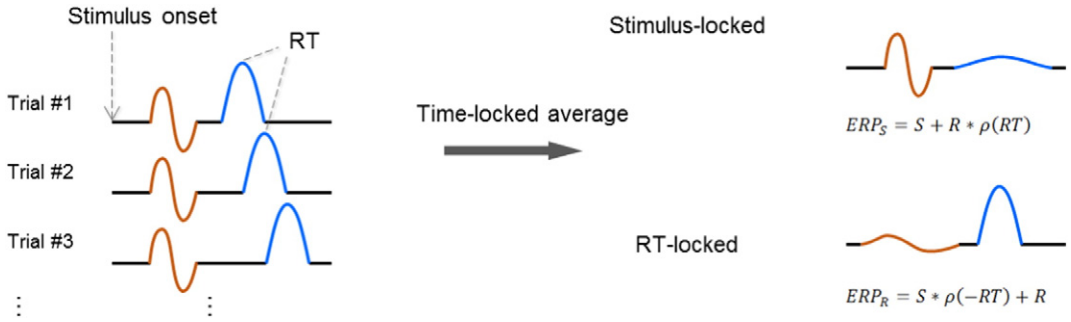

Fig. 2. Illustration of stimulus-locked and RT-locked averaged ERPs – each method smears the other component, which becomes the convolution of the component with the distribution of reaction time (RT).

regressors and the components are coefficients to be derived. Another novel method that was not based on the approach of general linear model was proposed by Kayser et al. (2007) in which they applied principal component analysis (PCA) to concatenated stimulus- and response-locked ERP waveforms, where some principal components are exclusively stimulus-related, some are exclusively response-related, and some show relations to both. The major limitation of the marker-based approaches consists in the assumption that ERPs exclusively consist of marker-locked components. A side issue is that the General Linear Model approach inherently suffers from distortion of ERP components by low-frequency noise as reported in Ouyang et al. (2015b). In addition, when ERPs are separated into two or more components, the amplitude effects will probably be segregated into several parts, which brings about a much higher level of complexity in the analysis as compared to traditional ERP averaging. Therefore, a reconstruction of ERPs from the separated, latency-corrected components is desirable, as will be introduced in the following.

To overcome the limitations of previous approaches to solving the latency jitter problem we essentially suggest to identify the various component clusters and their single trial latencies – not merely relying on external time markers – and to re-synchronize each component cluster to its respective latency (see Fig. 3). After compensating for trial-to-trial latency variability, the de-blurred waveforms and amplitudes of ERP components between different conditions can be compared, disentangling the confounded effects of differences in mean amplitude, mean latency and latency variability across experimental conditions. Recently, we have developed a new method – residue iteration decomposition (RIDE) – for separating ERPs into different component clusters with or without external time markers (Ouyang et al., 2011, 2015a, 2015b), which can also avoid the low-frequency distortion inherent in several other methods (Ouyang et al., 2015b). In previous applications of RIDE we have focused on component decomposition. Here, we systematically demonstrate for the first time that RIDE can also be used to reconstruct jitter-compensated ERPs and, hence, disambiguate the confounded effects of amplitude variation and latency jitter on ERP amplitudes. In the following we will explain the principles of RIDE and how it can be used to reconstruct jitter-compensated ERPs.

1.3. Re-constructed ERPs by RIDE

RIDE was developed to decompose ERPs into different component clusters with specific latency variabilities relative to stimulus onset (Ouyang et al., 2011, 2015a, 2015b). We showed that RIDE can un-mix overlapping component clusters that may be associated to different cognitive sub-processes (Ouyang et al., 2013; Stürmer et al., 2013; Verleger et al., 2014). Though developed as a tool for decomposing ERP component clusters and extracting their single trial variability information, RIDE has not been systematically explored on how it can improve the analysis of cross-condition effects by compensating for trial-to-trial latency variability. In this paper we demonstrate that by decomposing and reconstructing ERPs with RIDE, we can restore condition effects in amplitudes that are smeared by trial-to-trial latency variability, without blurring either stimulus-, response-locked or intermediate components as in previous averaging methods, using just one time marker at a time (stimulus onset or reaction time or estimated latency).

Based on a general framework, RIDE decomposes the ERP from each single trial into a stimulus-locked component cluster S, a central component cluster C, and a response-locked component cluster R. The complete algorithm can be found in Ouyang et al. (2015a) and a toolbox can be obtained from *http://cns.hkbu.edu.hk/RIDE.htm*. A brief overview of the method is given in the Method section below.

Fig. 3 illustrates the idea of latency-correcting ERPs with RIDE. For ease of understanding, noise is not shown. Schematically, we assume three component clusters in each single trial, shown in blue, red, and

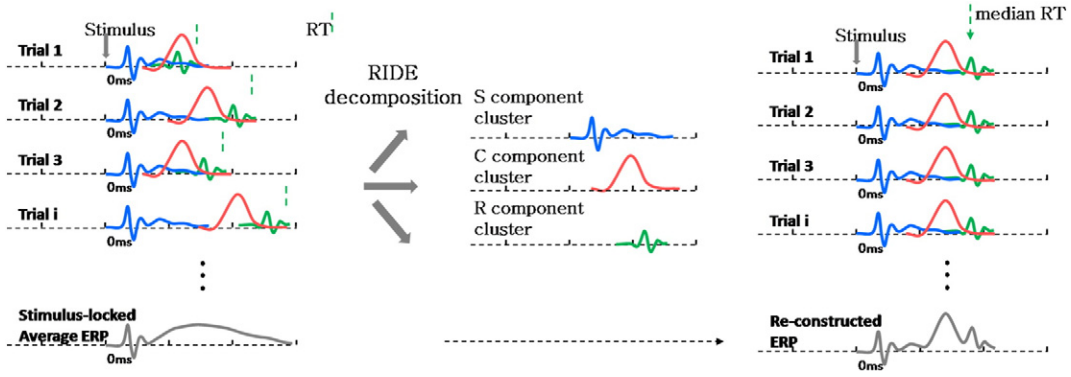

Fig. 3. Schematic illustration of conventional stimulus-locked average ERP (left) from single trials with latency jitter, decomposition by RIDE (middle) and reconstructed ERP after correcting for latency variability (right). Please note, in reality, ERPs may differ with respect to the waveform patterns, polarities, inter-component temporal relationships and the extent of latency variability, etc.

green. The blue component is locked to the stimulus, whereas the red and green components vary in their latencies relative to the stimulus from trial to trial. The green component is assumed to be locked to reaction time (RT). Due to the trial-to-trial latency variabilities of the red and green components, the stimulus-locked averaged ERP (bottom) blurs the representations of single trial ERPs and diminishes the true average amplitudes. The idea of ERP re-construction is shown in the right column; it is achieved by first synchronizing each component cluster (separated by RIDE) to its most probable latency across single trials, followed by averaging. Therefore, the amplitudes of the various components will not be attenuated, and the reconstructed ERP will show the most probable neural response in single trials (Fig. 3, right, bottom).

In the present paper, we will apply RIDE to datasets from several experiments to obtain reconstructed ERPs, corrected for trial-to-trial latency variability. Then we will compare the amplitude effects from standard ERPs and RIDE-reconstructed ERPs to see how latency variability affects amplitude differences. We will also conduct a validation analysis to show that the improvement of amplitude effects in reconstructed ERPs is not an artifact of the method.

2. Method

2.1. Simulation

The simulation aimed to demonstrate the attenuation of ERP amplitude effects between two conditions due to trial-to-trial latency variability. Only a single EEG channel was simulated; therefore, the implications of the electrode montage in real data are not considered. Based on the illustration in Fig. 1, we simulated single-trial EEG data by a simple half sine wave (representing an ERP component) with 1/f noise added, where f is frequency and spectrum power density is proportional to 1/f ("pink noise"), which closely resembles the spectrum pattern of EEG data. Thus, each single trial consists of a single half sine wave (Fig. 5) with a given latency (varying across trials) and added noise. Then, we assigned different parameters (e.g., latency variability and noise strength) to different conditions to examine the smearing effect on the statistical results. Simulation parameters were as follows. Each single trial epoch represented 500 ms, involving 500 sampling points; ERP components were represented by half sine waves with a span of 300 ms. The peak latency of the component was Gaussian-distributed across single trials, centered at t = 250 ms with a standard deviation denoted as σ_1. We added background noise with 1/f spectrum and a standard deviation denoted as σ_2. Conditions 1 and 2 were set to grand average amplitudes of 10 and 11 for the ERP components, respectively. Across-participant standard deviation of ERP amplitudes was $\sigma_3 = 1$. For each participant, the amplitude was constant across single trials, which is an idealized setting that is unrealistic for empirical data but sufficient to model the problem of trial-to-trial latency jitter. There were 20 fictitious participants, each with 100 trials per condition. The blurring effect on the simulation data was investigated across the parameter range of $[20-300]$ for σ_1 and $[1-12]$ for σ_2, which covers the transition of the condition effect from significant to insignificant without correction of latency jitter (Fig. 5).

Single trial ERP data was simulated by the summation of components with assigned random latencies and noise. Since the amplitude of the components differed between conditions an amplitude effect was expected to be revealed in statistical testing, for example, by *t*-test or analysis of variance, given noise under certain level. Here, *t*-tests were performed on the mean amplitudes between the two conditions for the time window 100–400 ms covering the simulated ERP component.

Since the simulated single trial ERPs vary in latencies, the cross-condition amplitude effect will be attenuated if the statistical analysis is conducted on the stimulus-locked ERPs. The main idea of the present work is that the effect will be closer to the true amplitude effect when the statistical analysis is conducted on latency-locked versions of ERPs, which are obtained by re-synchronizing single trials to the retrieved true latencies. Resynchronization of single trials was based on the latencies detected by peak-picking after low-pass filtering at 3 Hz. Specifically, all single trials were synchronized to the detected single trial latencies by temporally shifting each trial with the relative lag between its latency and the median of all latencies across trials. In this way the long-latency trials were shifted backward and the short-latency trials were shifted forward in time. The *t*-tests were performed again after resynchronization of the data.

2.2. Experimental data

In the empirical part we used five published datasets from different experiments concerning various psychological questions. Here, the datasets are dedicated to examine the smearing problem existing in generic ERP data that would affect the analysis of between-condition differences. The five exemplary datasets covering various cognitive tasks serve to demonstrate the validity of the method but are not exhaustive in reflecting all potential entanglements of amplitude and latency effects in different cognitive psychological studies. Since only the methodological issue is addressed in the present work, the specific psychological issue in each dataset is not of particular interest; for each dataset we compared only selected conditions since we intended to demonstrate some typical scenarios about the smearing problem. The datasets were provided by different labs with different setups, for example reference selection, and different data preprocessing. For all datasets that were fed into RIDE their original conditions after preprocessing as described in the respective published paper were maintained.

2.2.1. Dataset 1: face recognition

This data was taken from Herzmann and Sommer (2010) where complete experimental details can be found. The experiment concerned priming effects in face recognition. Briefly, 21 participants made familiarity judgments by key-pressing about famous, unfamiliar, and experimentally learned faces. All stimuli were preceded either by a different face or by the same face; the latter case is a repetition priming condition. Of interest here was whether and to what extent the ERP amplitude variation (particularly the early component sensitive to repetition) due to face priming is affected (Schweinberger et al., 1995) by trial-to-trial latency jitter. Therefore, only data for primed (Condition 1) and unprimed (Condition 2) familiar faces were used here. Trials with ocular artifacts (blinks or saccades) and other artifacts (voltage steps exceeding 50 μV/ms or a difference of more than 100 μV within an interval of 200 ms) and incorrect behavioral responses were discarded. ERPs were aligned to a 100-ms baseline before target onset, digitally low-pass filtered at 30 Hz with zero phase shift, and recalculated to average reference. EEG was recorded from 65 sites covering the whole scalp.

2.2.2. Dataset 2: visual oddball

Complete experimental details can be found in Valsecchi et al. (2009) who kindly provided the data. The experiment was a typical oddball task. Here we assessed whether and to what extent the ERP amplitude variation on the P3 component due to oddball frequency (Squires et al., 1975; Duncan-Johnson and Donchin, 1977) is affected by trial-to-trial latency jitter. Participants were 12 healthy adults. They were instructed to minimize eye blinks and to maintain fixation on a small (0.48° visual angle) white point that was continuously displayed on an otherwise empty black screen. Once per second, a red or green disc with a diameter of 2.04° appeared for 100 ms around the fixation point. The participant's task was to silently count the stimuli with the prespecified target color. The assignment of target color (red or green) was counterbalanced over participants. In three experimental blocks of 500 trials each, the frequency of target stimuli was 20, 50, or 80%. Here, target conditions 20% (Condition 1) and 80% (Condition 2) were used; 86% of all trials were free of eye blinks and submitted to analyses. Data were recorded at a sampling rate of 250 Hz and a bandpass from

0.1 to 70 Hz. All 40 channels, covering the whole scalp, were referenced against left mastoid and converted offline to average reference.

2.2.3. Dataset 3: Simon task

Data are from Böckler et al. (2011) where complete experimental details can be found. The experiment concerned the compatibility effect elicited by a Simon task. Here we investigated whether and to what extent the ERP amplitude variation on P3 component due to compatibility (e.g., Sommer et al., 1993) is affected by trial-to-trial latency jitter. Sixteen students participated in the experiment. Compatible and incompatible Simon task conditions were orthogonally combined with three arousal conditions. A tone of 65 dB was presented either 500 or 200 ms before the visual stimulus or no tone was presented. The EEG from 63 channels covering the whole scalp was referenced to the left mastoid and converted to average reference offline. Bandpass was 0.01 to 70 Hz; sampling rate was 250 Hz. Trials containing blinks were corrected off-line with Brain Electrical Source Analysis. Remaining artifacts were eliminated according to visual inspection. Here, data from the compatible and incompatible conditions without tones were contrasted (Conditions 1 and 2).

2.2.4. Dataset 4: reading

This data were provided by Wang et al. (2015) where complete experimental details can be found. The experiment concerned semantic and syntactic violation effects in reading tasks. Here we assessed whether and to what extent the ERP amplitude variation on P600 component due to the violation syntactic rules (Kuperberg, 2007) is affected by trial-to-trial latency jitter. Eighteen Chinese native speakers read Chinese sentences shown one word at a time. The sentences were either regular (control condition, CON), contained semantic violations (SEM), or both semantic and syntactic violations (SEM + SYN). In each condition there were 40 trials per participant. The EEG signals from 40 channels covering the whole scalp were digitized online with a sampling rate of 500 Hz and filtered offline with a bandpass of 0.02 to 30 Hz. Epochs with amplitudes exceeding ± 75 μV were excluded from the averages. Online EEG was referenced to the left mastoid and re-referenced offline to the average of both mastoids. The CON and SEM + SYN were used here.

2.2.5. Dataset 5: word recognition

This data was taken from Bayer et al. (2012). The experiment concerned the emotion effect in word processing. Here we tested whether and to what extent the ERP amplitude variation on the LPC (late positive component, Schacht and Sommer, 2009) due to the emotion manipulation is affected by trial-to-trial latency jitter. Briefly, 23 native German speakers completed two tasks, reading and lexical decisions (LDT). The stimuli consisted of 180 German words, which were positive, neutral, or negative in valence and of either high or low arousal. For the lexical decision task, 180 orthographically legal and pronounceable pseudowords were constructed from other nouns by replacing one letter at a random position within the word. During the LDT, participants had to indicate by button presses whether or not a given letter string represented a German word. All 62 channels covering the whole scalp were recorded with left mastoid as initial reference, filtered with a bandpass of 0.03–70 Hz, and sampled at 500 Hz. Offline, the EEG was converted to average reference. Epochs containing artifacts and incorrect responses were discarded. ERP waveforms were referred to a 100 ms pre-stimulus baseline. Here, the data from positive (Condition 1) and negative (Condition 2) words in the LDT were used.

2.3. RIDE decomposition of ERPs

In the present paper we employed the RIDE algorithm from Ouyang et al. (2015b). As illustrated in Fig. 3, RIDE aims to separate a stimulus-locked component cluster S, RT-locked component cluster R and an intermediate component cluster C that is neither locked to stimulus onset nor to RT. If no response is present, the R cluster is not derived. In generic cases, C captures the major late component with variable trial-to-trial latencies, for example, P3b. RIDE also obtains single trial latency information of C. All information about the component clusters S, C, and R, and about C latencies in single trials is obtained by the RIDE algorithm consisting of a decomposition module as an inner iteration loop and a latency estimation module as outer iteration loop described in the following. A flow-chart is provided in Fig. 4.

2.3.1. Decomposition module

Given that the single trial latencies of S, C, and R (notated as L_S, L_C, and L_R) are known (with L_C being conveyed from the latency estimation module), this inner iteration module separates the ERP into the three component clusters. Initially $S(t) = C(t) = R(t) = 0$ was set. To estimate S, RIDE subtracts C and R from each single trial and aligns the residuals of all trials to the latency L_S in order to obtain S as the *median waveform* over all time points. The same procedure is applied to obtain C and R. The whole procedure is iterated till convergence.

2.3.2. Latency estimation module

While the latencies of S and R are known – stimulus onsets and reaction times, respectively – the latency of the assumed central component cluster C is unknown. RIDE uses a self-optimized iteration scheme for latency estimation, starting with an approximate initial estimation of L_C from the raw data. Woody's method is used to obtain an initial estimate of L_C: cross-correlation time courses between the ERP and single trials are calculated for each single electrode and averaged across all electrodes and low-pass filtered at 3 Hz. The lag of the maximum in the scalp-averaged cross-correlation time course for each single trial is taken as the single trial latency (L_C). Starting with this L_C estimate, the analysis is subjected to the following iteration: (1) Use L_S, L_C, and L_R to decompose S, C, and R using the *Decomposition Module* till convergence. The convergence is effectively defined as the difference between the values of two successive iterations being much smaller ($<10^{-3}$) than that between the two initial iterations (Ouyang et al., 2015b). (2) Remove S and R (by subtraction) from each single trial and calculate the cross-correlations between the residue and the C component cluster to re-estimate the latency of L_C. (3) Return to (1) and (2) and iterate until convergence of both the latency L_C and the component clusters S, C, and R.

The C component cluster is latency-variable and has no overt latency information corresponding to external time markers (e.g, stimulus onsets or response times). Due to volume conduction, components are spread across the scalp, with dominant amplitudes at some electrodes and weaker or close-to-zero amplitude at other electrodes. The average reference scheme will further induce an inverse waveform of the dominant component at the weak-amplitude electrodes. For example, P3b is usually dominant at Pz/CPz and vanishes or inverses at frontal sites, depending on the reference scheme. Therefore in estimating the single trial latency of C by the latency estimation module, the data of the whole scalp is used. That is, the cross-correlation time series are calculated for each electrode and the peak latency is identified from the average time series. This approach automatically emphasizes the contribution of the dominant electrodes (for P3b it would be Pz and its nearby electrodes) where the cross-correlation time series would be bell-shaped; at the same time, the procedure weakens the contributions of the irrelevant sites where the cross-correlations fluctuate around zero.

Signals at all electrodes are processed by RIDE with a common set of stimulus-onsets, C latencies and RTs in order to obtain the averaged waveforms of S, C and R. The topographies of each RIDE component can then be plotted for each time point. Using a common C latency for all electrodes is largely justified based on the effect of volume conduction: A neural generator of a given ERP component will affect many recording sites simultaneously. Of course, RIDE can be applied to each electrode separately with an electrode-specific C latency estimated

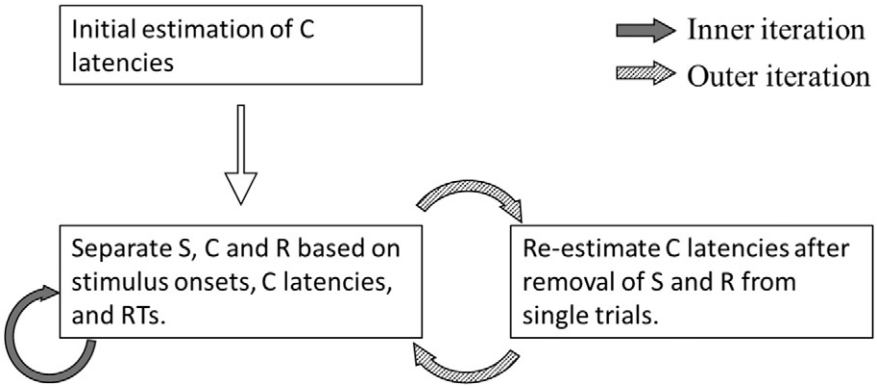

Fig. 4. Flow-chart of RIDE decomposition. For explanation see text.

only from that particular electrode. However, for electrodes with poor signal-to-noise ratio or very weak C component, the latency estimation would be less reliable and from a practical point of view, the topography of C cannot be obtained without using a common latency for all electrodes.

The RIDE method is extendable to other schemes, for example, no R cluster when there is no response trigger, or more than one C cluster if appropriate (e.g., Wang et al., 2015). When more than one C cluster is to be separated, the initial latency estimations for different C clusters are done in different time windows. For the experimental Datasets 1, 2, 3, and 5 with available response times, we followed the general scheme of separating ERP into the stimulus-locked S cluster, the central C cluster (mainly capturing the P3 complex) and the response-locked R cluster. For Dataset 4 concerning linguistic processing, we separated two C clusters dedicated to capture the N400 and P600 components, but no R cluster since there was no immediate overt response to the stimuli.

2.4. RIDE reconstruction of ERP

As explained in the introduction, non-stimulus-locked ERP components are blurred by conventional stimulus-locked averaging, and ERP reconstruction attempts to average each ERP component cluster in a more appropriate way, that is, by synchronizing to their own latencies rather than to stimulus onset. This can be done after RIDE decomposition. To obtain the ERP that shows the most probable waveforms in single trials, we need to relocate each RIDE component cluster to its most probable latency. We use the robust median value as an approximation of the most probable latency. When each RIDE component cluster is synchronized to its median latency, the ERP can be reconstructed in a de-blurred version (Fig. 3, right). This can be done on each electrode separately but with a common latency of C for all electrodes. The reconstructed ERP may hence show larger amplitudes and richer waveform patterns (Fig. 3, right, bottom) as a function of the amount of trial-to-trial latency jitter across the recorded trials.

3. Results

3.1. Simulation of the smearing effect

In the first simulation we set the ERP amplitudes for the two conditions to 10 and 11, respectively. In principle, if noise is relatively small, the significance test will reveal a condition effect in conventionally averaged ERPs, whereas, the effect will be overridden as noise increases. Apart from the effect of noise, as stressed in the introduction, the latency variability of the component may diminish experimental effects on ERP amplitudes. We gradually increased noise strength σ_2 and latency variability σ_1 and checked the changes of effect size (Cohen's *d*) and *p*-value in *t*-tests (Fig. 5D, E). The *p* value was converted to binary levels with the customary threshold of 0.05. Indeed, latency variability and noise strength did reduce effect size. These two variables exert effects on the between-condition and within-condition variance, respectively.

Whereas noise (within-condition variance) is an inevitable contribution to EEG signals, the between-condition difference can be properly restored if it is depressed by latency variability. Once we compensated for the latency variability of the simulated ERP components across single trials by synchronizing the single trials to the median of the component latencies detected by peak-picking, the result (effect size matrix and binary *p* value in Fig. 5F, G) improved in line with expectations. The difference between the effect size from original data (Fig. 5D) and that from latency-corrected data (Fig. 5F) is shown in Fig. 5H, indicating increase of effect size after compensating for the trial-to-trial latency jitter. The lattices in σ_1-σ_2 panel where the *p* values are >0.05 in original data (Fig. 5E) but <0.05 in latency-corrected data (Fig. 5G) are shown in Fig. 5I, indicating the region where significant effects were hidden by trial-to-trial latency jitter. This simulation illustrates how trial-to-trial latency variability can diminish between-condition amplitude differences and obscure true condition effects.

3.2. Real data

3.2.1. Single trial variability, RIDE separation, and reconstructed ERPs

Fig. 6 shows the scenario of RIDE separation and reconstruction of data from a typical participant of Dataset 1. The layout of the figure is similar to the illustration in Fig. 3. The data for panel A is from the single electrode Pz where the P3 is most pronounced. Left panels show the single trial ERPs sorted by response times, and the average ERP, respectively. The plot shows a late component cluster with variable latency that is somewhat correlated with RT. The RIDE-separated component clusters (Fig. 6, middle panels) show distinct wave shapes. After synchronizing all component clusters at their own most probable latencies (Fig. 6, right panel), the latency variability is greatly diminished (Fig. 6A, right) and the reconstructed ERP (Fig. 6A, right) is enhanced in amplitude and more detailed in structure than the conventional ERP (Fig. 6A, left). To better show the stimulus-locked component cluster we also show the data from channel PO8 where the N170 component is most prominent. However, the latency-variable cluster is much weaker. The difference between panel A and B indicates that different RIDE component clusters have different scalp topographies.

To better show the features of both stimulus-locked and latency-variable component clusters, we applied a spatial filter to the data (Fig. 6C). We applied PCA to each RIDE component cluster and obtained the first principal component (PC) for each RIDE component. Each PC represents the topography on which each RIDE component is dominantly distributed. The PC weights for all three RIDE-derived component clusters were added to form a single spatial filter that was applied on the spatio-temporal patterns in each single trial. It is like a virtual channel integrating the prominent features of all RIDE component clusters. Based on this approach of spatial filtering we can now show both stimulus-locked and

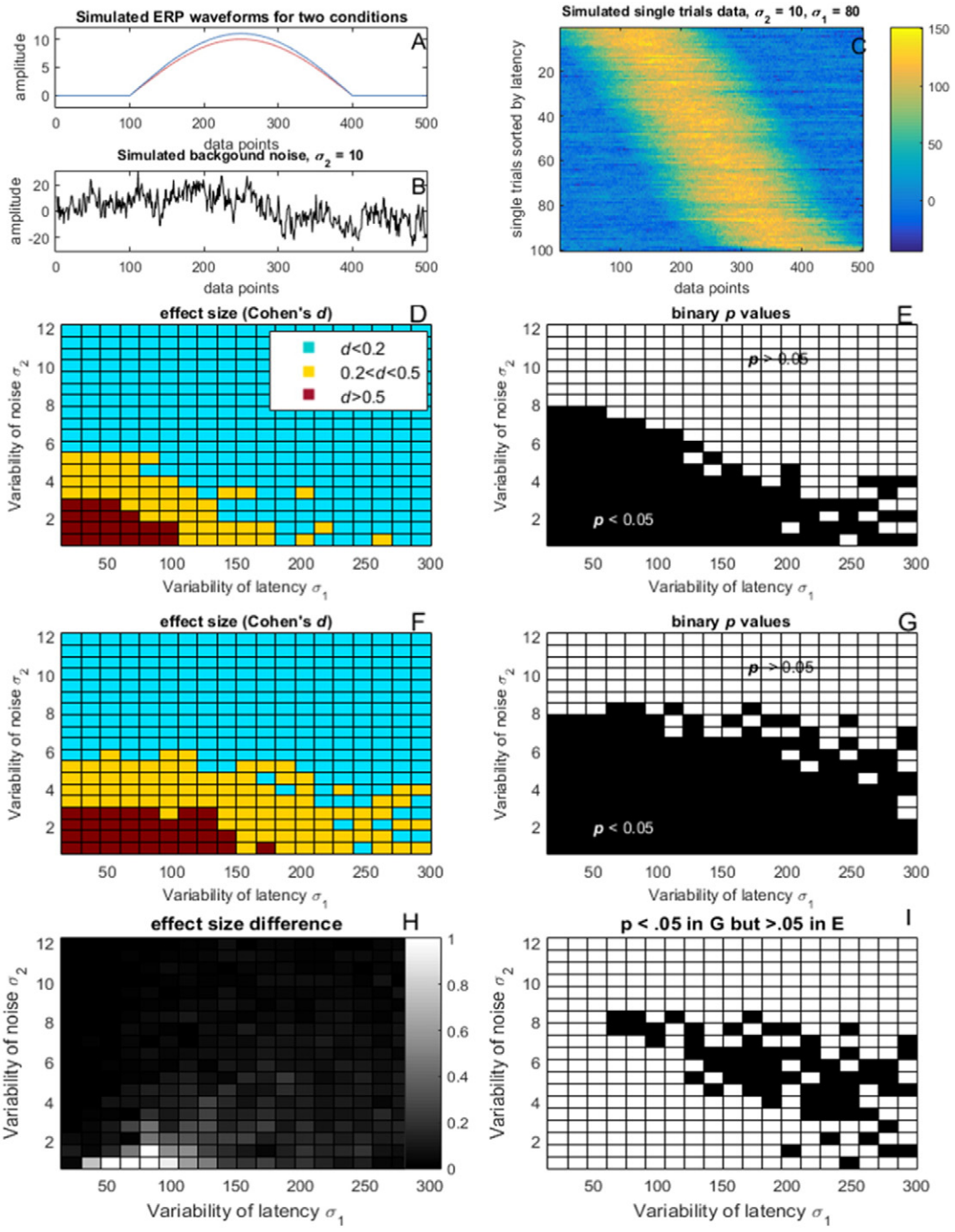

Fig. 5. Illustration of the simulation data (A–C) and the amplitude effects across different extents of latency variability (σ_1) and noise strength (σ_2) (D–G). **A:** Simulated ERP waveform. **B:** Simulated background noise when $\sigma_2 = 10$. **C:** Simulated single trials data when $\sigma_1 = 80$, $\sigma_2 = 10$. **D:** The effect size of the amplitude difference between conditions. **E:** Binary *p* values being larger or smaller than 0.05 across different parameters. **F:** The updated effect size for latency-synchronized ERP. **G:** Binary *p* values for latency-synchronized ERP. **H:** The difference of effect size between **D** and **F**. **I:** Region of restored conditional effects shown by the lattices (black) where $p < 0.05$ in **G** but >0.05 in **E**.

latency-variable component clusters (Fig. 6C). One can compare Figs. 3 and 6 to understand how RIDE separates and reconstructs ERP data. Here we will focus on the smearing effect on amplitude differences.

3.2.2. Statistical testing on conventional and reconstructed ERPs

In this section we present the statistical results for the five datasets from different experimental paradigms to see the change of between-condition effects in selected time windows from standard ERPs to reconstructed ERPs (Fig. 7). As expected, the amplitudes of the reconstructed ERPs on the right panel are always larger than those of standard ERPs. This is due to the correction of trial-to-trial latency variability. On average, the increase of amplitudes of the largest peaks in the reconstructed ERP is around 2 μV. But note that the increase of the amplitude only indicates the correction for trial-to-trial latency jitter. When it comes to the condition effect, the associated variance across participants also plays an important role (Fig. 1). This will be revealed in the following statistical analysis.

For each dataset, we superimposed waveforms at Pz for the conditions considered here and for both ERPs and reconstructed ERPs. The amplitude differences from this electrode between the two conditions were *t*-tested across participants. The quantitative results for *t* and *p* are shown directly in Fig. 7. The amplitudes were averaged within the

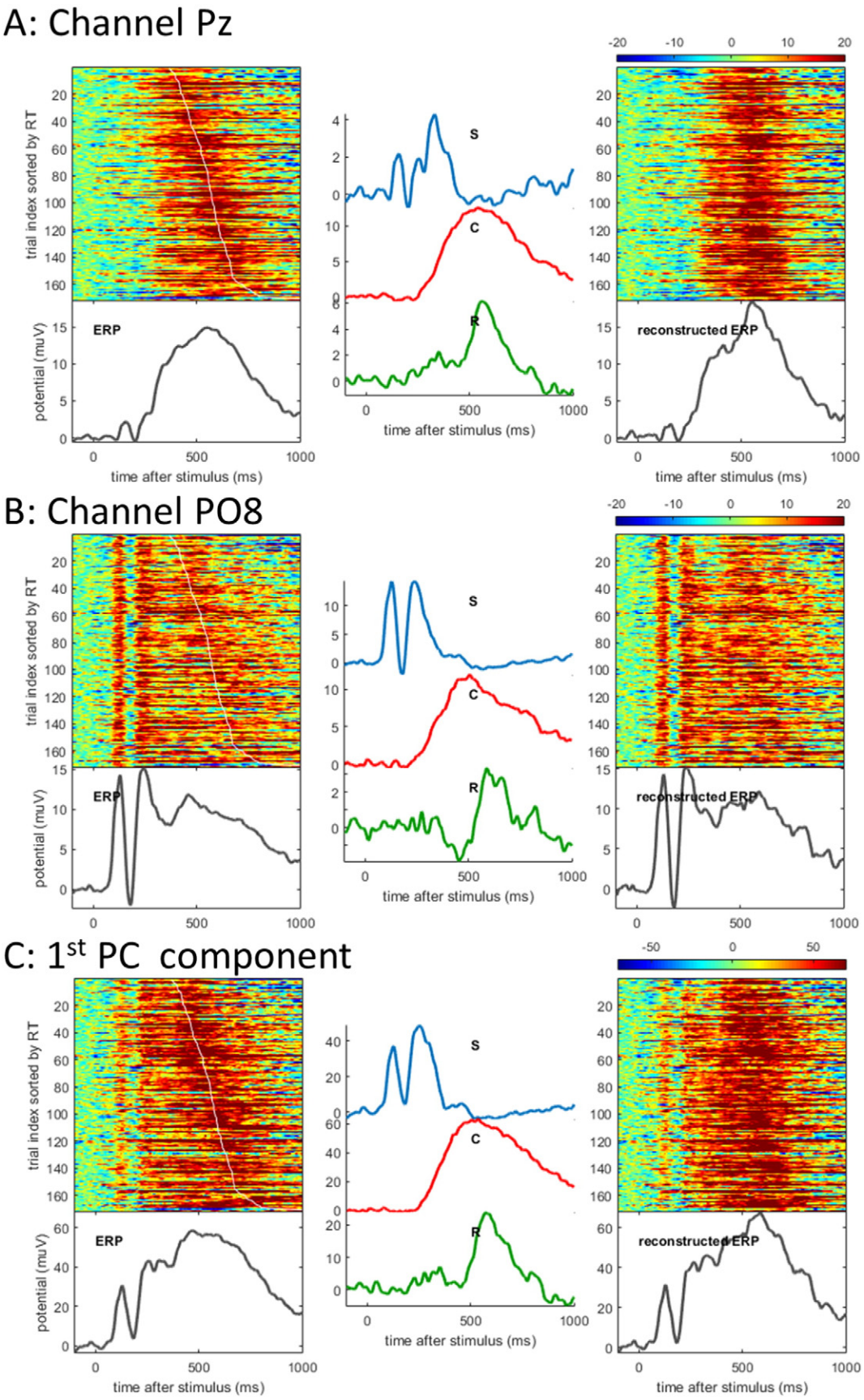

Fig. 6. Corresponding to the illustration in Fig. 3, but for real data from one participant of Dataset 1 at electrode Pz, PO8 and the first PC component. Left: single trial ERPs and the conventional stimulus-locked average ERP. Middle: The separated S, C and R component clusters, synchronized to their own latency (most probable latency across single trials). Note the different scales of the y-axis for the components. Right: Reconstructed single trial ERPs and their average. The waveform data was referenced to the average of all recording sites.

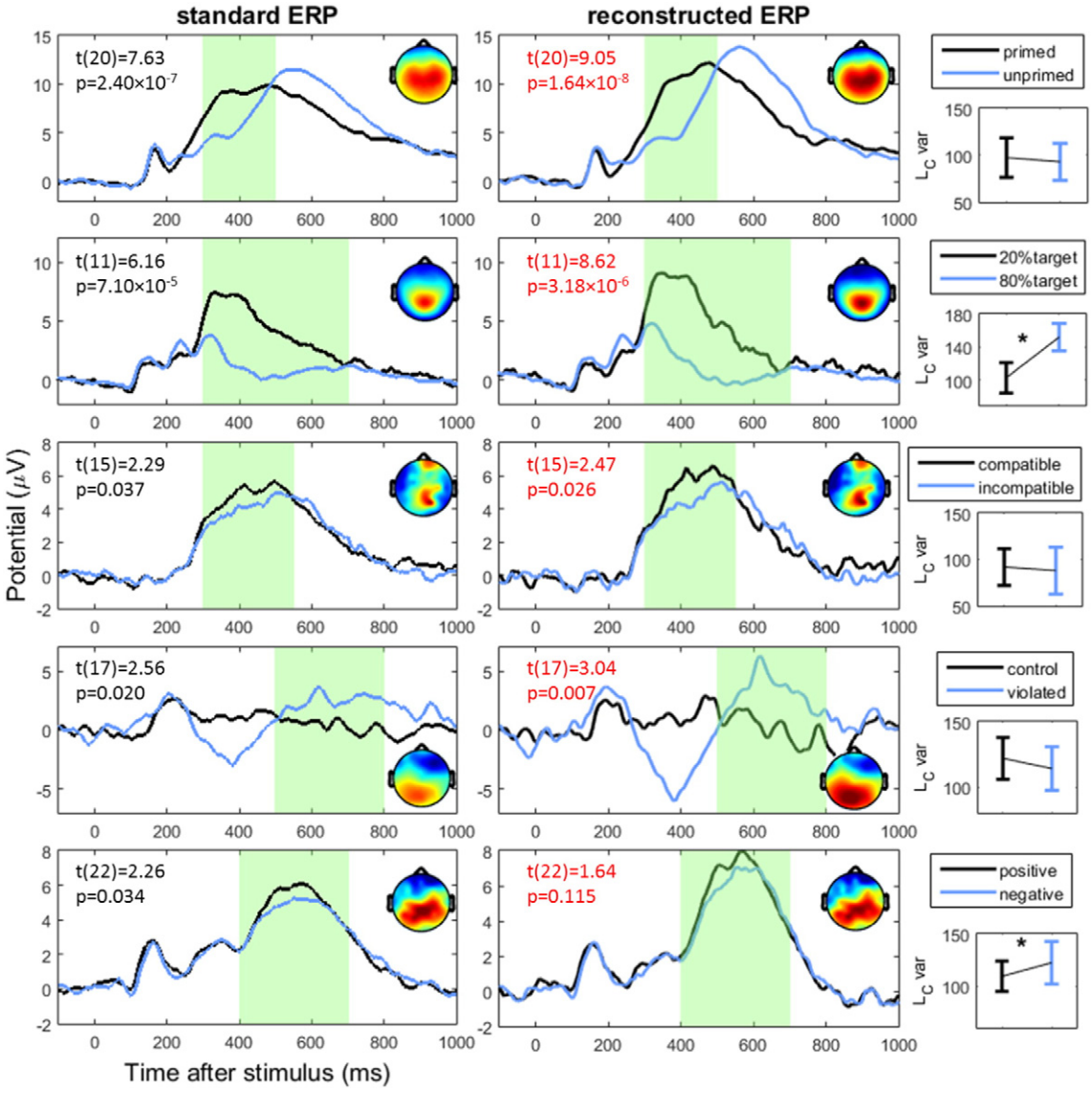

Fig. 7. Comparison of condition effects in both standard and reconstructed ERPs at the Pz electrode. The *t*-test was performed on the time window indicated by the light green area, covering the condition difference in question. The topographies show the average difference between two conditions in the tested window, with the same color scale for ERPs and reconstructed ERPs. (A, B) Dataset 1: face recognition task. (C, D) Dataset 2: oddball task. (E, F) Dataset 3: Simon task. (G, H) Dataset 4: reading task. (I, J) Dataset 5: Lexical decision task. The conditions are indicated in the legends. Right column: The mean latency variabilities of the C-component and error bars (± SD) across participants for the two conditions.

time windows of interest which basically cover the prominent waveform differences. The time windows of interest are highlighted in green in Fig. 7. The topographies of the amplitude differences across all electrodes are shown for the time window of interest.

As already mentioned in the Method section, the five ERP components to be examined for different datasets are 1) Priming effect (N250r/ERE, early repetition effect) in face recognition, 2) classical oddball effect in the P3., 3) Compatibility effect on P3 in Simon task, 4) semantic/syntactic violation effect on P600 in reading, 5) emotion effect on LPC (late positive component) in processing emotional words. There are different patterns of results obtained from reconstructing ERPs. (1) In the first two datasets, the condition effects were already highly significant in the conventional ERP as seen from the large waveform differences and the statistical results. Yet the effects are still enlarged after compensating for trial-to-trial latency variability in the reconstructed ERPs. (2) In Dataset 3, reconstruction did not essentially change statistical results. Here the time window was chosen to test the compatibility effect in the amplitude of the P3 range (e.g. Sommer et al., 1993). (3) Very interesting is Dataset 4, where the condition effect was barely significant in the conventional ERP but became highly significant in the reconstructed ERPs. In this dataset, the P600 was the component where the effect was tested. The P600 reflects linguistic processing and is elicited by syntactic and sometimes also by semantic violations (e.g. Kuperberg, 2007).

In Dataset 5 it is the other way around: The significant effect vanished in the reconstructed ERP. Here we tested the emotional effect in the LPC component (e.g., Schacht and Sommer, 2009). The result in Dataset 5 therefore should correspond to Case 2 proposed in the introduction (Fig. 1), that is, a significant amplitude difference is largely due to the difference in latency variability between conditions.

To corroborate these assumptions, we calculated the latency variabilities (as standard deviation) of the C component clusters for both conditions of each dataset. The error bars for both conditions (indicating ±1 SD across participants) are shown on the right panel of Fig. 7 to visually indicate the fluctuation of the latency variability across participants. In Dataset 5, for the positive- and negative-word conditions, the average variability of C latency across participants was 109.3 ± 14.5 ms and 122.3 ± 21.0 ms, respectively. Indeed, the results

confirmed our assumption: The paired *t*-test for the C latency variability between the two conditions was significant, $t(22) = 3.09$, $p < 0.01$. However, it is worthwhile to note that the error bars shown in Fig. 7 are not indicative of the significance of difference of latency variability between conditions, because the samples are paired, in which case only the SD and the mean of the difference are relevant. Notably, Dataset 2 showed a significant difference in both amplitude and latency variability.

3.3. Validations

A critical question regarding the magnified or uncovered effects is whether they are real or just artifacts of the method. To address this question, we conducted two validations, using a surrogate from real data, where the trials from the conditions are uniformly mixed, and using simulated data without built-in condition effects.

3.3.1. Surrogate from real data

We conducted a permutation test: In each permutation, we mixed the single-trial data of each participant of both conditions and randomly split them into two halves labeled 'Condition 1' and 'Condition 2'. In these permutated data, no condition effect is supposed to be systematically found in the reconstructed ERP. We permutated Dataset 1 and tested the same time window as in Fig. 7 (A, B). We used *t*-test to test the between-condition effect in the reconstructed ERP from 100,000 permutations. The 100,000 realizations were divided into 100 bins. In each bin we counted how many positive effects (α was set to 0.05) were found. Fig. 8 (left panel) shows the probability of finding positive effects in each bin. The results showed that the reconstructed ERPs did not yield a significant effect at higher probability than chance. The results were the same for the other four datasets.

3.3.2. Simulated data without built-in condition effect

A more straightforward test of whether RIDE may generate artificial effects is applying it to noise data without a structured component. In this test, we generated 1/f noise, resembling the spectrum pattern of EEG data. The data for two conditions (the numbers of participants and trials are the same as in Dataset 1) are identical except for the random number seed. We constructed 100,000 datasets with two conditions; in each the random number seed was different, and RIDE was applied with the same parameters as in Dataset 1 to generate the reconstructed ERPs. Finally we *t*-tested the reconstructed data and found that the obtained number of positive effects did not outnumber the random probability across the 100 sets of 1000 permutations (Fig. 8, right).

4. Discussion

In the present paper we explored the problem of confounding trial-to-trial latency variability with amplitude effects across conditions. The latency variability problem obscures the analysis of ERP amplitudes when researchers investigate the effects of variables of interest. We reviewed the limitations of previous attempts to tackle this problem and proposed the RIDE method as a possible solution. Essentially, RIDE decomposes ERPs into different component clusters that have different latency variabilities or are time-locked to different events (across single trials). After decomposing ERPs into different component clusters, RIDE synchronizes each component cluster to its own latency across single trials, reconstructing a new ERP that is corrected for trial-to-trial latency variability. The reconstructed ERPs usually show enhanced amplitudes and appear to have a richer structure, owing to de-blurring by correcting for trial-to-trial latency jitter. Moreover, reconstructed ERPs show different scenarios concerning the condition effects. Reconstructed ERPs have diminished latency variability and thus may be supposed to reveal purer amplitude effects. The applications to various datasets shows that, in some situations, the existing amplitude effects are obscured by smearing (e.g., Dataset 4), whereas in some other situations, amplitude differences in conventional ERP may be due to differences in latency variability (Dataset 5) and each effect may be significant by itself or in combination (Dataset 2).

We performed validations to clarify that the changes in amplitude after latency correction are not spurious or method artifacts. We applied RIDE to surrogates derived from real data, and to simulated data without true amplitude effects. In both cases, no condition effects were expected, and we found indeed that the RIDE-reconstructed ERPs did not show significant amplitude effects at a higher rate than chance in a vast number of realizations. Therefore, if there are no significant differences between conditions in terms of amplitude or latency of ERP components other than what would be expected from chance alone, there is no reason to assume that RIDE will create an effect that is not present in the data.

4.1. Novelty of RIDE

The entanglement of latency variability and amplitude effects in ERPs has been pointed out and discussed in previous literature (Luck, 2005; McDowell et al., 2003; Pfefferbaum et al., 1980; Poli et al., 2010; Roth et al., 2007; Verleger et al., 2005; Walhovd et al., 2008; Woody, 1967). In these previous works, researchers identified the trial-to-trial latency variability as a confounding effect for between-condition amplitude differences. Simple adjustment, for example, estimating the P3 complex as a whole (Tuan et al., 1987; Walhovd et al., 2008; Woody, 1967), has been proposed and applied to retrieve the single trial latency variability and to correct ERP waveforms. However, treating the whole

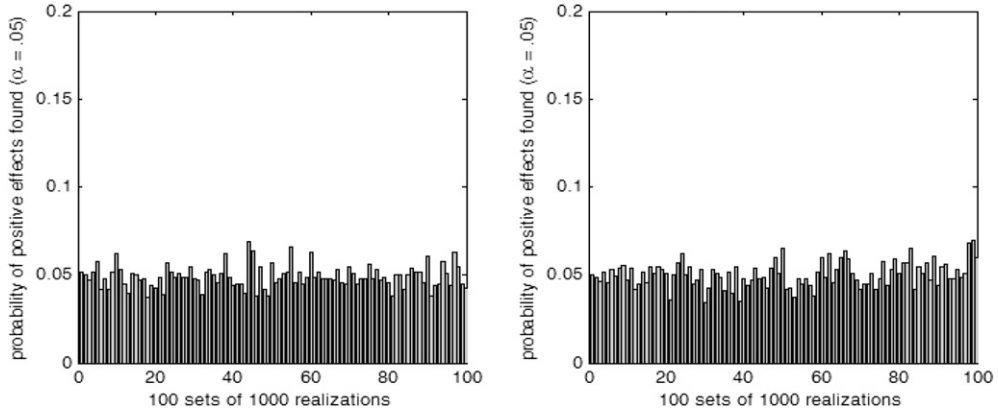

Fig. 8. Validations of reconstructed ERPs. **Left:** the probability of finding positive effect (α was set to be 0.05) from 100,000 permutations (divided into 100 sets) of conditionally mixed and split data from experimental Dataset 1. **Right:** the probability of finding a positive effect across the 100 sets of 1000 simulation datasets with two conditions.

ERP as a unitary latency-variable component cluster and synchronizing it, will lead to blurring of early and late components. Other approaches (Hansen, 1983; Knuth et al., 2006; Yin et al., 2009; Zhang, 1998; Takeda et al., 2008) based on General Linear model require explicit time markers and suffer from low-frequency distortions (Ouyang et al., 2015b). In contrast, RIDE first separates ERP into component clusters with different types of latency variability, that is, being either locked to stimulus onsets, responses, or neither and aligns the waveforms of each component cluster by synchronizing them to their own most probable single trial latencies in order to reconstruct un-blurred ERPs.

4.2. Limitations and challenges

RIDE aims to synchronize all ERP component clusters to their own single trial latency in order to generate an un-blurred waveform. Effectively, the reconstructed ERPs show enhanced structure and amplitudes. Nevertheless, this method involves the estimating the latency of C by cross-correlation. This step is inevitably affected by background noise and involves some estimation error. While the noise effect is unavoidable, it can be alleviated by increasing the trial number (to better shape the C component) or by enhancing the signal-to-ratio through careful data pre-processing. In this sense, depending on the component, there may be a trade-off between stimulus-locked averaging and performing latency estimation with RIDE. For example, it is not advisable to treat the early ERP components (e.g., the visual P1-N1 complex) as latency-variable component cluster and attempt to estimate and compensate its variability. Because the early component clusters are relatively stable in latency and their waveforms resemble background alpha waves, the latency estimation would possibly introduce more errors. Therefore, stimulus-locked averaging is preferable for these early components. Because the late P3 family shows stronger latency variability and is more distinguishable from background noise due to its slow waveform, latency estimation is more feasible for these components.

The number of three component clusters is based on a general consideration about typical ERP experiment settings (with RTs); accordingly, there are at least three types of component clusters: stimulus-locked, response-locked, and not locked to either stimulus or response. This assumption is inevitably a simplification of the ERP structure. One of the main limitations of this assumption is that there might be more than one cluster of C components that vary in latency somewhat independently from each other. In this case considering only one C cluster remains an imprecise model of the data. However, separating more than one C cluster requires that different component clusters are distinct from each other in morphology, such that estimating the single trial latencies of each one will not affect the other. This renders great challenges in separating some endogenous ERP components, for example, the P3 complex, into sub-C component clusters. But in certain situations such separations are possible; for example, in psycholinguistic studies N400 and P600 components can be treated as two C clusters (Wang et al., 2015). The early ERP components like P2, N2, and even P1, N1, might also be considered as latency-variable depending on the time scale one is referring to. But in the RIDE context it is difficult to treat them as additional C component clusters because their single trial latency is greatly affected by the background alpha noise, leading to close-to-arbitrary estimation of single trial latency. Therefore the RIDE method works best in separating large, late endogenous components like P3, N400, P600, etc.

RIDE requires prior knowledge about the time windows in which the component clusters are assumed (Ouyang et al., 2015b). This is similar to previous simpler approaches like the Woody filter (Woody, 1967; Kutas et al., 1977) where the time window needs to be specified before applying the method. The time window specification is usually based on visual inspection of (grand) average ERP waveforms (Ouyang et al., 2015a). This introduces a certain amount of subjectivity but facilitates hypothesis-driven applications. In principle, the sensitivity of the results with respect to the model parameters needs to be examined, especially when the statistical results supporting the relevant conclusions are marginally significant.

Inter-individual variability is another factor that blurs the between-condition amplitude effect. Similar to within-person trial-to-trial latency variability, there must be also person-to-person latency variability of ERP sub-components. This partly explains why the differences between the grand average ERP and RIDE-reconstructed ERP is usually not as dramatic as Fig. 3 indicates. For R cluster it is easy to compensate the person-to-person latency variability – simply synchronizing R to the grand mean (or median) of individual mean (or median) RTs, just treating each subject like a single trial. But the major challenge stems from the treatment of the C component cluster for which no external time marker is available. In principle we might use a common template to estimate the relative single-subject latencies of C and synchronize them to the grand average. But the problem is that there is great morphology variability across participants where some may show very large components whereas some show almost none, rendering the identification of the relative latency of single subject C component very difficult or, if done, very un-reliable. A simpler approach of peak-picking may suffer from similar problem of individual differences as some subjects' C components have a single and clear peak but some do not. For readers' reference we tried to compensate the person-to-person latency variability of RIDE components (for C it was based on peak-picking) in our previous work (Ouyang et al., 2015b) for the purpose of revealing how seriously the ERP waveform could be blurred due to the combination of intra- and inter-subject latency variability. But due to the relatively larger variability of morphology across individuals which might lead to questionable synchronization of the C component across participants, in the present work we do not attempt to compensate for inter-individual variability in the context of blurring of conditional effects. This challenge should be taken up in future work. An interesting line of research may consist in trying to use them as neurophysiological correlates of individual differences in performance.

Likewise, we should point out that latency variability does not only affect amplitudes and amplitude differences. There is also an effect on certain chronometric measures of the averages, in particular onsets and offsets of ERP components (e.g., Sommer et al., 1996). Thus, given invariant mean peak latency, increasing variability of peak latency in single trials will advance average onset latencies and delay average offset latencies. RIDE should in principle be able to deal with this problem and future work should systematically explore this option.

The effects of ERP preprocessing on RIDE results are also of importance. It has to be noted that reference settings for the five different datasets used here were not always consistent (average reference for datasets #1, #2, #3, and #5 and linked mastoids for #4). The inconsistency arose because the datasets were from different labs and we had to make sure the datasets were in the original states as described in the respective published paper before feeding them into RIDE for further analysis. In principle, any steps in ERP preprocessing could affect RIDE results to some extent. In essence, changing the reference causes re-distribution of the zero level across electrodes, therefore, should affect the relative contribution of a given channel to the cross-correlation curve averaged over all electrodes to determine the latency. However, since RIDE uses the spatiotemporal pattern from the whole-scalp data to determine the single trial C latencies, results will probably remain stable across different references. After all, RIDE is a secondary data processing algorithm that is applied on 'clean' ERP data that were already baselined, filtered, referenced, artifact rejected. RIDE aims to address a more specific problem how the trial-to-trial latency jitter blurs the ERP waveforms and affects condition effects. That being said, any problems in ERP analysis that stem from preprocessing steps would not really benefit from RIDE processing, since RIDE was not designed to solve them. For example, different reference settings were shown to affect statistical results (Brunet et al., 2011). The degrading of ERP results in standard ERP analysis due to sub-optimal choice of reference would not be resolved by RIDE. But it is possible that ERP data from different

kinds of reference settings would all benefit from RIDE reconstruction because the existence of trial-to-trial latency-jitter is independent of reference selection. In general, a better choice of reference, as well as other preprocessing steps, are recommended before applying RIDE, as it will affect RIDE results like it does with standard ERP analysis.

Author contributions

G.O., W.S., and C.Z. discussed the work and plan; G.O. conducted the data analysis, prepared the figures and drafted the manuscript; G.O., W.S., and C.Z edited and revised the manuscript.

Acknowledgements

This work was partially supported by Hong Kong Baptist University (HKBU) Strategic Development Fund, the HKBU Faculty Research Grant (FRG2/14-15/025), the Hong Kong Research Grant Council (RGC) (HKBU12302914), Germany-Hong Kong Joint Research Scheme (G-HK012/12), the National Natural Science Foundation of China (Grant No. 11275027) to G.O. and C.Z., and the Germany-Hong Kong Joint Research Scheme (PPP 56062391) to W.S., and by the German Research Foundation (DFG, Project SO177/26-1 to W.S.). This research was conducted using the resources of the High Performance Cluster Computing Centre, Hong Kong Baptist University, which receives funding from RGC, University Grant Committee of the HKSAR and HKBU.

References

Ahmadi, M., Quiroga, R.Q., 2013. Automatic denoising of single-trial evoked potentials. NeuroImage 66, 672–680.
Bayer, M., Sommer, W., Schacht, A., 2012. P1 and beyond: functional separation of multiple emotion effects in word recognition. Psychophysiology 49, 959–969.
Böckler, A., Alpay, G., Stürmer, B., 2011. Accessory stimuli affect the emergence of conflict, not conflict control: a Simon-task ERP study. Exp. Psychol. 58 (2), 102.
Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. 2.
Coyle, T.R., 2003. A review of the worst performance rule: evidence, theory, and alternative hypotheses. Intelligence 31 (6), 567–587.
Duncan-Johnson, C.C., Donchin, E., 1977. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology 14 (5), 456–467.
Fjell, A.M., Westlye, L.T., Amlien, I.K., Walhovd, K.B., 2011. Reduced white matter integrity is related to cognitive instability. J. Neurosci. 31 (49), 18060–18072.
Ford, J.M., White, P., Lim, K.O., Pfefferbaum, A., 1994. Schizophrenics have fewer and smaller P300s - a single-trial analysis. Biol. Psychiatry 35 (2), 96–103.
Hansen, J.C., 1983. Separation of overlapping waveforms having known temporal distributions. J. Neurosci. Methods 9 (2), 127–139.
Herzmann, G., Sommer, W., 2010. Effects of previous experience and associated knowledge on retrieval processes of faces: an ERP investigation of newly learned faces. Brain Res. 1356, 54–72.
Jeon, Y.W., Polich, J., 2003. Meta-analysis of P300 and schizophrenia: patients, paradigms, and practical implications. Psychophysiology 40 (5), 684–701.
Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2001. Analysis and visualization of single-trial event-related potentials. Hum. Brain Mapp. 14 (3), 166–185.
Kaltwasser, L., Hildebrandt, A., Recio, G., Wilhelm, O., Sommer, W., 2014. Neurocognitive mechanisms of individual differences in face cognition: a replication and extension. Cogn. Affect. Behav. Neurosci. 14, 861–878.
Kayser, J., Tenke, C.E., Gates, N.A., Bruder, G.E., 2007. Reference-independent ERP old/new effects of auditory and visual word recognition memory: joint extraction of stimulus- and response-locked neuronal generator patterns. Psychophysiology 44 (6), 949–967.
Knuth, K.H., Shah, A.S., Truccolo, W.A., Ding, M., Bressler, S.L., Schroeder, C.E., 2006. Differentially variable component analysis: identifying multiple evoked components using trial-to-trial variability. J. Neurophysiol. 95 (5), 3257–3276.
Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Res. 1146, 23–49.
Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197 (4305), 792–795.
Larson, G.E., Alderton, D.L., 1990. Reaction time variability and intelligence: a "worst performance" analysis of individual differences. Intelligence 14 (3), 309–325.
Leuthold, H., Sommer, W., 1998. Postperceptual effects and P300 latency. Psychophysiology 35 (1), 34–46.
Luck, S.J., 2005. An Introduction to the Event-related Potential Technique. MIT Press, Cambridge, MA.
McDowell, K., Kerick, S.E., Santa Maria, D.L., Hatfield, B.D., 2003. Aging, physical activity, and cognitive processing: an examination of P300. Neurobiol. Aging 24 (4), 597–606.
Möcks, J., Köhler, W., Gasser, T., Pham, D.T., 1988. Novel approaches to the problem of latency jitter. Psychophysiology 25 (2), 217–226.
Ouyang, G., Herzmann, G., Zhou, C., Sommer, W., 2011. Residue iteration decomposition (RIDE): a new method to separate ERP components on the basis of latency variability in single trials. Psychophysiology 48, 1631–1647.
Ouyang, G., Schacht, A., Zhou, C., Sommer, W., 2013. Overcoming limitations of the ERP method with Residue Iteration Decomposition (RIDE): a demonstration in go/no-go experiments. Psychophysiology 50 (3), 253–265.
Ouyang, G., Sommer, W., Zhou, C., 2015a. A toolbox for residue iteration decomposition (RIDE)—a method for the decomposition, reconstruction, and single trial analysis of event related potentials. J. Neurosci. Methods 250, 7–21.
Ouyang, G., Sommer, W., Zhou, C., 2015b. Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE). Psychophysiology 52 (6), 839–856.
Patterson, J.V., Michalewski, H.J., Starr, A., 1988. Latency variability of the components of auditory event-related potentials to infrequent stimuli in aging, Alzheimer-type dementia, and depression. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 71 (6), 450–460.
Pfefferbaum, A., Ford, J.M., Roth, W.T., Kopell, B.S., 1980. Age differences in P3-reaction time associations. Electroencephalogr. Clin. Neurophysiol. 49 (3), 257–265.
Picton, T.W., Hink, R.F., Perez-Abalo, M., Linden, R.D., Wiens, A.S., 1984. Evoked potentials: how now? Am. J. Electroneurodiagnostic Technol. 10, 177–221.
Poli, R., Cinel, C., Citi, L., Sepulveda, F., 2010. Reaction-time binning: a simple method for increasing the resolving power of ERP averages. Psychophysiology 47 (3), 467–485.
Polich, J., Herbst, K.L., 2000. P300 as a clinical assay: rationale, evaluation, and findings. Int. J. Psychophysiol. 38 (1), 3–19.
Röschke, J., Wagner, P., Mann, K., Fell, J., Grözinger, M., Frank, C., 1996. Single trial analysis of event related potentials: a comparison between schizophrenics and depressives. Biol. Psychiatry 40 (9), 844–852.
Roth, A., Roesch-Ely, D., Bender, S., Weisbrod, M., Kaiser, S., 2007. Increased event-related potential latency and amplitude variability in schizophrenia detected through wavelet-based single trial analysis. Int. J. Psychophysiol. 66 (3), 244–254.
Rugg, M.D., Coles, M.G.H., 1995. The ERP and cognitive psychology: conceptual issues. In: Rugg, M.D., Coles, M.G.H. (Eds.), Electrophysiology of Mind. Event-related Potentials and Cognition. Oxford University Press, Oxford, pp. 27–39.
Saville, C.W.N., Lancaster, T.M., Stefanou, M.E., Salunkhe, G., Lourmpa, I., Nadkarni, A., ... Feige, B., 2014. COMT Val 158 Met genotype is associated with fluctuations in working memory performance: converging evidence from behavioural and single-trial P3b measures. NeuroImage 100, 489–497.
Schacht, A., Sommer, W., 2009. Time course and task dependence of emotion effects in word processing. Cogn. Affect. Behav. Neurosci. 9, 28–43.
Schweinberger, S.R., Pfütze, E.M., Sommer, W., 1995. Repetition priming and associative priming of face recognition: Evidence from event-related potentials. J. Exp. Psychol. Learn. Mem. Cogn. 21 (3), 722.
Sommer, W., Leuthold, H., Hermanutz, M., 1993. Covert effects of alcohol revealed by event-related potentials. Percept. Psychophys. 54, 127–135.
Sommer, W., Ulrich, R., Leuthold, H., 1996. The lateralized readiness potential as psychophysiological approach to the investigation of cognitive processes. Psychol. Rundsch. 47, 1–14.
Squires, N.K., Squires, K.C., Hillyard, S.A., 1975. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalogr. Clin. Neurophysiol. 38 (4), 387–401.
Stürmer, B., Ouyang, G., Zhou, C., Boldt, A., Sommer, W., 2013. Separating stimulus-driven and response-related LRP components with Residue Iteration Decomposition (RIDE). Psychophysiology 50 (1), 70–73.
Takeda, Y., Yamanaka, K., Yamamoto, Y., 2008. Temporal decomposition of EEG during a simple reaction time task into stimulus-and response-locked components. NeuroImage 39 (2), 742–754.
Tuan, P.D., Möcks, J., Köhler, W., Gasser, T., 1987. Variable latencies of noisy signals: estimation and testing in brain potential data. Biometrika 74 (3), 525–533.
Valsecchi, M., Dimigen, O., Kliegl, R., Sommer, W., Turatto, M., 2009. Microsaccadic inhibition and P300 enhancement in a visual oddball task. Psychophysiology 46 (3), 635–644.
Verleger, R., 1997. On the utility of P3 latency as an index of mental chronometry. Psychophysiology 34, 131–156.
Verleger, R., Jaśkowski, P., Wascher, E., 2005. Evidence for an integrative role of P3b in linking reaction to perception. J. Psychophysiol. 19 (3), 165–181.
Verleger, R., Metzner, M.F., Ouyang, G., Smigasiewicz, K., Zhou, C., 2014. Testing the stimulus-to-response bridging function of the oddball-P3 by delayed response signals and residue iteration decomposition (RIDE). NeuroImage 100C, 271–280.
Walhovd, K.B., Rosquist, H., Fjell, A.M., 2008. P300 amplitude age reductions are not caused by latency jitter. Psychophysiology 45 (4), 545–553.
Wang, F., Ouyang, G., Zhou, C., Wang, S., 2015. Re-examination of chinese semantic processing and syntactic processing: Evidence from conventional ERPs and reconstructed ERPs by residue iteration decomposition (RIDE). PLoS One 10 (1).
Woody, C.D., 1967. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. Biol. Eng. 5 (6), 539–554.
Yin, G., Zhang, J., Tian, Y., Yao, D., 2009. A multi-component decomposition algorithm for event-related potentials. J. Neurosci. Methods 178 (1), 219–227.
Zhang, J., 1998. Decomposing stimulus and response component waveforms in ERP. J. Neurosci. Methods 80 (1), 49–63.