Neuroimaging Measures of Error-Processing: Extracting Reliable Signals from Event-Related Potentials and Functional Magnetic Resonance Imaging

Vaughn R. Steele[a,b,c,*], **Nathaniel E. Anderson**[b], **Eric D. Claus**[b], **Edward M. Bernat**[d], **Vikram Rao**[b], **Michal Assaf**[e,f], **Godfrey D. Pearlson**[e,f], **Vince D. Calhoun**[b,c,e], and **Kent A. Kiehl**[b,c]

[a]Neuroimaging Research Branch, National Institute of Drug Abuse, Intramural Research Program, National Institutes of Health, Baltimore, MD, USA

[b]The nonprofit Mind Research Network (MRN) & Lovelace Biomedical and Environmental Research Institute (LBERI)

[c]University of New Mexico

[d]University of Maryland, College Park

[e]Yale University School of Medicine

[f]Olin Neuropsychiatry Research Center, Institute of Living

Abstract

Error-related brain activity has become an increasingly important focus of cognitive neuroscience research utilizing both event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI). Given the significant time and resources required to collect these data, it is important for researchers to plan their experiments such that stable estimates of error-related processes can be achieved efficiently. Reliability of error-related brain measures will vary as a function of the number of error trials and the number of participants included in the averages. Unfortunately, systematic investigations of the number of events and participants required to achieve stability in error-related processing are sparse, and none have addressed variability in sample size. Our goal here is to provide data compiled from a large sample of healthy participants (n = 180) performing a Go/NoGo task, resampled iteratively to demonstrate the relative stability of measures of error-related brain activity given a range of sample sizes and event numbers included in the averages. We examine ERP measures of error-related negativity (ERN/Ne) and error positivity (Pe), as well as event-related fMRI measures locked to false alarms. We find that achieving stable estimates of ERP measures required four to six error trials and approximately 30 participants; fMRI measures required six to eight trials and approximately 40 participants. Fewer trials and participants were required for measures where additional data reduction techniques (i.e., principal component

[*]**Correspondence:** Dr. Vaughn R. Steele, The Mind Research Network, 1101 Yale Boulevard NE, Albuquerque, NM 87106, vaughn.r.steele@gmail.com.

analysis and independent component analysis) were implemented. Ranges of reliability statistics for various sample sizes and numbers of trials are provided. We intend this to be a useful resource for those planning or evaluating ERP or fMRI investigations with tasks designed to measure error-processing.

Keywords

ERP; fMRI; error-processing; reliability; time-frequency analysis; principal component analysis; independent component analysis

1.0 Introduction

A critical aspect of designing human psychophysiological experiments is optimizing the quantity of data collected for adequately testing hypotheses. This includes both the number of participants and number of trials necessary to extract a reliable signal of interest. This is particularly important when procedures involve measurement of brain activity as these techniques require significant time and resources to collect. Prior work has often considered the numbers of trials needed to reliably measure stimulus-locked brain activity with event related potentials (ERPs) and functional magnetic resonance imaging (fMRI). Fewer investigations have addressed the stability of response-locked neural measures and the investigations that do, have focused on error-related brain activity in ERPs. The proliferation of ERP and fMRI studies examining error-related brain activity has underscored the need for definitive stability estimates for these measures. Extant studies vary widely with respect to the number of participants and the number of error trials averaged within participant. As such, ongoing research will benefit from better estimates of required numbers of trials and participants needed for stabile brain measures.

Several sources have estimated an appropriate number of trials required for stable stimulus-locked brain responses at around 20 to 50 trials. For instance, Cohen and Polich (1997) show that averages of 20 events are sufficient for attaining stability of the P300, a large ERP component related to target detection in continuous performance tasks. Stability is highly dependent upon signal to noise ratio. Luck (2005) has suggested 30 to 60 events are sufficient for robust signals such as the P300, but also suggests that smaller components such as the N2 and P1 may require hundreds of averaged trials for reliable measures. Guidelines for fMRI studies typically recommend a minimum of 20 to 30 trials (Desmond and Glover, 2002; Huettel and McCarthy, 2001); however, these estimates are based on limited data and should be expected to vary based on the specific cognitive tasks and specifics in data collection.

In contrast to stimulus-locked brain measures, response-locked events, such as error-related brain activity, are relatively stable with far fewer trial averages (Maurer et al., *in press*; Meyer et al., 2013; Olvet and Hajcak, 2009; Pontifex et al., 2010; Rietdijk et al., 2014). Error processing is often assessed with ERPs and fMRI during a response inhibition task (e.g., Go/NoGo, Stroop, Stop-Signal, Flanker, Wisconsin Card Sorting Task, and Task-Switching), or any variety of speeded, continuous performance tasks likely to produce erroneous responses (for review, see (Niendam et al., 2012). The most relevant error-driven

ERP components include the error negativity (Ne; (Falkenstein et al., 1991) or error-related negativity (ERN; (Gehring et al., 1993) and the error positivity (Pe; (Falkenstein et al., 1991). These are response-locked ERP components elicited by an erroneous response to NoGo stimuli (i.e., False Alarms (FA)). The ERN/Ne is a negative deflection likely generated in the rostral cingulate zone, potentially within the caudal anterior cingulate cortex (cACC; (Carter et al., 1998; Kiehl et al., 2000; Miltner et al., 2003) and peaks between 50 and 100 ms after an incorrect response (Falkenstein et al., 1990, 1991; Gehring et al., 1993; Holroyd and Coles, 2002). The ERN/Ne is believed to be associated with cognitive detection of the response error (Edwards et al., 2012; Falkenstein et al., 1991), incorrect response tendencies (Carbonnell and Falkenstein, 2006) or to reflect initial response conflict processing aimed at increasing cognitive control (Yeung et al., 2004; Yeung and Summerfield, 2012). The Pe is a positive deflection generated from at least one source within the rostral ACC (rACC; (Edwards et al., 2012; van Veen and Carter, 2002) and follows the ERN/Ne, peaking between 200 and 400 ms after an incorrect response. The Pe is believed to index further error processing, conscious evaluation of the error, response strategy adjustment and/or affective assessment of the error (Endrass et al., 2007; Falkenstein et al., 1990, 1991; Leuthold and Sommer, 1999; Nieuwenhuis et al., 2001; Overbeek et al., 2005; Ullsperger et al., 2010; Yeung and Summerfield, 2012) and has been found to be negatively related with rACC activation (Edwards et al., 2012). Conscious awareness of an error is necessary for both the ERN/Ne and Pe though the ERN/Ne can be modulated by uncertainty and task parameters (Shalgi and Deouell, 2012, 2013). Successful error monitoring, as indexed by increased ERN/Ne amplitude, should lead to modulation of response strategies designed to reduce errors in the future, as indexed by reduced Pe amplitude.

The first systematic investigation of the number of trials required for a stable ERN/Ne and Pe found largely stable and reliable measures after six to eight trials among 53 young adults performing a Flanker task (Olvet and Hajcak, 2009). Rietdijk et al. (2014) found stable and internally consistent ERN/Ne and Pe with eight trials among 70 participants also performing a Flanker task. Pontifex et al. (2010) examined potential differences in reliability of error – processing across the lifespan, again with a Flanker task. The authors corroborated that six trials were sufficient for achieving stability and internal consistency of ERN/Ne and Pe in preadolescents and young adults. However, older adults may require up to eight trials to achieve the same stability. It should also be noted that while preadolescent and young adult groups had over 50 participants, the older adult group had half the participants (n=26), which may have impacted the number of trials required for stability of ERPs averaged across participants. In order to extend these findings to other common error-inducing tasks, Meyer et al. (2013), examined reliability differences in ERN/Ne from three different paradigms. The authors report stable ERN, averaged across 43 participants, with six to eight trials for Flanker and Go/NoGo tasks. More errors (> 20) were required to achieve acceptable reliability with the Stroop task. Each of these tasks examined averages across a set number of participants; however, reliability measures may also vary as a function of sample size (e.g., the smaller sample of older adults required more trails than the larger sample of younger adults to achieve a reliable signal). There are currently no reports that

systematically examine stability of error-related activity across sample sizes and the number of events simultaneously.

The growing interest in ERP measures of error-related activity has been accompanied by an increase in MRI-based functional neuroimaging studies of these processes. While temporally less precise than scalp-recorded electrical potentials, fMRI provides more specific information about the anatomical loci supporting neural signals otherwise measured at the scalp. So, although fMRI may not temporally distinguish between the ERN/Ne and the Pe, it can measure activity in specific brain regions critical for these cognitive events with greater spatial segregation. Source localization studies from scalp-recorded ERPs suggest differentiable neural origins for the ERN/Ne and Pe. As mentioned above, the ERN/Ne arises from dorsal/caudal portions of the ACC and the Pe arises from activity in more anterior/rostral portions of the ACC (van Veen and Carter, 2002). Investigations using joint ERP and fMRI measures to evaluate neural activity error-related processes have also supported these findings, showing differentiable networks related to error processing in the rostral and caudal ACC (Edwards et al., 2012). Very little work is available that has systematically evaluated the stability of fMRI measures across sample sizes and trial numbers (Desmond and Glover, 2002; Huettel and McCarthy, 2001). None have specifically addressed stability of blood oxygenation level dependent (BOLD) activation in error-processing networks using event-related fMRI. Signal stability becomes particularly important in fMRI as inadequate power may result in a failure to identify important regions of activity and/or the mischaracterization of noise as signal of interest (see (Huettel and McCarthy, 2001). Furthermore, these parameters will vary with each cognitive task and each anatomical region of interest. Stability estimates will also depend on the experimental design (block or event-related; random or fixed effects).

Desmond and Glover (2002) provided guidelines for relative power in BOLD signal using fMRI data from passive, resting state scans (to estimate within-participant variability) and a working memory task (to estimate between-participant variability). They used these data to simulate generalizable power curves based on the number of time-points per condition and percent signal change in any given region of the brain (block design, random-effects). Results indicated that, given a percent signal change of 0.5%, a minimum of 12 participants are needed to insure 80% power at a liberal alpha of 0.05. Twice as many participants were required to achieve this power at more conservative thresholds typical of controlling for multiple tests.

Murphy and Garavan (2004) used a Go/NoGo task to examine the number of participants needed for an event-related fMRI design. They noted improvements in power and signal to noise ratio with increasing numbers of participants (from n = 4 to n = 58). The authors note that statistical power remained relatively low around n = 20 compared to the full sample, but low power was driven largely by type II errors rather than false-positives. For this study, the percentage of the signal evident at n = 58 was used to evaluate power differentials as proportions of active voxels relative to the full sample. Thus, the power achieved at n = 58 is assumed as an absolute ceiling. This technique does not account for effects that may remain unstable at n = 58, nor does it account for varying numbers of trials across participants.

The number of event trials per participant and the number of participants included in group averages are both critical variables to consider. It should be clear that there is a subtle trade-off between the number of trials averaged per participant and the number of participants involved in the study. With a larger number of participants, fewer trials may be necessary to get a reliable overall signal average. With fewer participants, a similarly reliable average might be obtained by increasing the number of trials per participant. The number of trials plays a particularly important role in studies examining response-locked error-related processing. Each participant's performance freely varies, and so it is expected that a range of error-trials will be available across participants. Although the number of errors committed by an individual is difficult to control, task parameters can be manipulated to produce a desired range of error trials across participants.

Five primary data analysis techniques are implemented to extract neural correlates with error-processing, three for ERPs and two for fMRI. The three analysis techniques for ERPs include classic time-domain (TD) windowed analysis, principal component analysis (PCA), and time-frequency (TF) analysis. For classic TD windowed analysis, a time window is fit around an ERP component of interest (e.g., the ERN/Ne or Pe) amplitude (i.e., mean and/or peak) within that time-window is extracted (Handy, 2005; Luck, 2014). For PCA, separable PCs are extracted then related to ERP windowed components (Dien, 1998; Dien et al., 2007). For TF analysis, the ERP is generally converted to a TF measure using either a wavelet or Cohen's class transformation (Bernat et al., 2005; Demiralp et al., 2001). Additional steps are generally carried out to extract specific TF segments related to a task (e.g., error-processing) and an ERP TD windowed component (e.g., ERN/Ne or Pe; c.f., (Bernat et al., 2005).

The two analysis techniques for fMRI include region of interest (ROI) analysis and independent component analysis (ICA). ROIs are generated anatomically using a-priori regions supported by previous research on the specific cognitive function or task. ICA of fMRI is used to identify maximally independent regions that can be ultimately related to the task (Calhoun et al., 2001). A region is identified with ICA and the time-course of each IC can be correlated with the design matrix (i.e., task relevant stimuli) to relate an IC to a specific trail type (e.g., a FA) and thus a specific cognitive function (e.g., error-processing).

The goal of the current report is to examine the varying stability of response-locked, error-related processing given varying numbers of trials and participants. We use a simple Go/NoGo task and examine error-related activity (elicited by FAs) using ERP and fMRI measures, resampling from large numbers of participants. This is, to the best of our knowledge, the first report to systematically evaluate the effects of varying both the number of trials and participants for ERP and fMRI measures of error-related neural activity. The resulting data tables will be helpful for supporting the relative stability of ERN/Ne and Pe measures, and their respective BOLD signal, for future studies examining these cognitive features, based on their sample characteristics.

2.0 Methods

2.1 Participants

Participants consisted of 180 healthy adults (95 men) ranging in age from 15 to 60 years (M = 30.53, SD = 10.97) drawn from the Olin Neuropsychiatry Research Center at the Institute of Living at Hartford Hospital. They were recruited from the surrounding community of Hartford, CT via advertisements, presentations at local universities, and word-of-mouth. EEG data were available for 100 participants and fMRI data were available for 97 participants. Twenty-six had both EEG and fMRI data. The overall sample contains EEG and previously published fMRI (Steele et al., 2013a; Steele et al., 2014a) data with the constraint that each participant committed at least 14 response errors. In each sample, overall error rates were negatively correlated with age (ERP: $r = -.248$, $p = .013$; fMRI: $r = -.206$, $p = .043$). In the EEG sample, 10% were left handed, 13% self-identified as Hispanic, 80% as Not Hispanic, 72% as White, 8% as Black/African American, 4% as Asian, and 16% as Other. In the fMRI sample, 7% were left handed, 8% self-identified as Hispanic, 79% as not Hispanic, 68% as White, 10% as Black/African American, 8% as Asian, and 13% as Other. Using the Structured Clinical Interview for the DSM-IV, all participants were free of any history of psychiatric illness (Axis I; (First et al., 1997) and reported no history of psychosis in first-degree relatives. All participants reported normal hearing and normal or corrected to normal visual acuity by corrective lenses or MR compatible glasses. Protocols were approved by the Institutional Review Board of Hartford Hospital and participants provided written informed consent.

2.2 Experimental Go/NoGo Task

Participants completed a previously published Go/NoGo paradigm (Kiehl et al., 2000) containing two experimental runs, each comprising 245 visual stimuli, presented to participants using the computer-controlled visual presentation software package, Presentation. As outlined below, only a few experimental protocol differences were necessary between EEG and fMRI data collection. Each stimulus appeared for 250 ms in white text on a black background within a continuously displayed rectangular fixation box and were approximately 3×5 visual degrees. In the scanner, stimuli were displayed on a rear-projection screen mounted at the rear entrance to the magnet. Participants viewed the screen by means of a mirror system attached to the head coil. During EEG data collection, participants were seated in a comfortable chair 60 cm away from a computer monitor on which task stimuli were presented. Participants were instructed to respond as "quickly and accurately as possible" with their right index finger every time the target ("Go") stimulus (a white "X") appeared, and to withhold a response when the distracter "No/Go" stimuli (a white "K") appeared. In the scanner, behavioral responses were recorded using a commercially available MRI-compatible fiber optic response device (Lightwave Medical, Vancouver, BC). Targets appeared with higher frequency (84%, 412 trials with 206 for each run) than distractors (16%, 78 trials with 39 for each run) to establish a strong stimulus-response mapping on "Go" trials. Two K's were never presented sequentially. The relatively high probability of targets was necessary to build a pre-potent response set and elicit a sufficient number of errors to justify their independent examination. The interstimulus interval was pseudo-randomly jittered (1–3 s stimulus-onset asynchrony (SOA; Laaksoa et

al.); averaging 1.5 s). Prior to data collection, each participant performed a block of 10 practice trials to ensure the instructions were understood. The SOA between Go stimuli varied to the constraint that three Go stimuli were presented within each consecutive 6 s period. The NoGo stimuli were interspersed among the Go stimuli in a pseudorandom manner subject to two constraints: the minimum SOA between a Go and NoGo stimulus was 1000 ms; the SOA between successive NoGo stimuli was in the range of eight to 14 s. Both Go and NoGo stimuli had an equal likelihood of occurring at 0, 500 or 1000 ms after the beginning of a 1.5 s fMRI acquisition period, or repeat time (TR). By jittering stimulus presentation relative to the acquisition time, the hemodynamic response to the stimuli of interest was sampled effectively at 500 ms intervals. "Hits" were defined as Go ('X' stimuli) events that were followed by a button press; "Misses" were defined as Go events where the participant did not respond' "Correct Rejections" were determined by the absence of a motor response to a NoGo stimulus; "False Alarms" were defined as a motor response following a NoGo stimulus.

2.3 Electroencephalography Recording and Reduction

Electrophysiological data were collected using two Windows-compatible computers and a 64-channel BioSemi ActiveTwo amplifier. The first computer used Presentation software to deliver the stimuli, accept responses, and send digital triggers to the other computer indicating when a stimulus or response occurred. The second computer acquired electroencephalographic data using BioSemi software and amplifier. All signals collected with this BioSemi system were low-pass filtered using the standard fifth order sinc filter with a half-power cutoff of 204.8 Hz then digitized at 1024 Hz during data collection. EEG activity was recorded using sintered Ag-AgCl active electrodes placed in accordance with the 10–20 International System (Jasper, 1958). The participant's nose was used as the reference. Six electrodes were placed on the participants face to measure electro-oculogram. These electrodes were placed above, below, and lateral to the canthus of each eye. All offsets were kept below 10 kΩ. After placement of the electrodes, participants were seated in a comfortable chair 60 cm away from a computer monitor on which task stimuli were presented, and were instructed to refrain from excessive blinking or moving during data acquisition. Participants then performed the Go/NoGo task described above.

Pre-processing included down sampling to 512 Hz, bad channel detection and replacement, epoching, eye-blink removal, and low-pass filtering at 15 Hz. Bad channels were identified as having activity four standard deviations away from the mean of all other non-ocular channels. These channels were replaced using the mean of surrounding electrodes. This method, used previously (Anderson et al., 2015; Maurer et al., *in press*; Steele et al., 2015; Steele et al., 2014b; Steele et al., *in press*) was implemented to identify very large artifacts and remove them from the data before applying more stringent data cleaning steps in post-processing. ERP epochs were defined in relation to responses, from 1000 ms pre- to 2000 ms post-response. The epoched data were eye-blink corrected using an independent component analysis (ICA) technique. The ICA utility in the EEGLab software (Delorme and Makeig, 2004) was used to derive components then, using an in-house template matching algorithm (Jung et al., 2000), blink components were identified and removed from the data.

We investigated the two response-locked components defined relative to a FA: the ERN, a negative deflection that occurred between −10 ms and 110 ms, and the Pe, a positive deflection that occurred between 150 ms and 450 ms (Figure 1). ERP components were baseline corrected using a −200 ms to −50 ms window, relative to a FA. Within each trial, individual electrodes in which activity exceeded ± 100 µV were omitted from analyses. Applying these criteria, 2.49 % of electrode trials were excluded. A low-pass filter at 20 Hz was applied to the data before analysis. A 4-component TD-PCA was extracted for FA trials (Figure 1) that accounted for 91.79% of the variance. Two TF-PCA were performed, first for theta frequency power (bandpass filtered at 3–7 Hz) and second for delta frequency power (low-pass filtered at 3 Hz). A 4-component TF-PCA solution was extracted for FA trials for theta (Figure 2) and a 5-component solution was extricated for delta power (Figure 3) accounting for 69.73% and 98.09% of the variance, respectively. To reduce ambiguity among PCA measure, theta and delta PCs will include either a "t" or "d" preceding the PC number, respectively. Mean measurements from FCz were extracted and used for time-domain (TD), time-domain PCA (TD-PCA; (Chapman and McCarry, 1995), and time-frequency PCA (TF-PCA; (Bernat et al., 2005) analysis described below.

2.4 Functional Magnetic Resonance Imaging Data Collection and Analysis

Imaging data were collected on a Siemens' Allegra 3T system located at the Olin Neuropsychiatry Research Center, Hartford, CT. Each participant's head was firmly secured and head motion was restricted using a custom cushion inside the head coil. Localizer images were acquired to determine functional image volumes. The echo planar image (EPI) gradient-echo pulse sequence (TR/TE = 1500/28 ms; flip angle = 65°; FOV = 24 × 24 cm; 64 × 64 matrix; 3.4 × 3.4 mm in plane resolution; 5 mm effective slice thickness; 30 total slices) effectively covered the entire brain (150 mm) in 1.5 s. Each of the two runs lasted just over 7 minutes, or 281 scans. A 9 s rest period was included prior to the start of each run to allow for T_1 effects to stabilize. The six initial images from the stabilization period were discarded before post-processing.

Functional images were reconstructed offline at 16-bit resolution and manually reoriented to the anterior commissure/posterior commissure (AC/PC) plane. Functional image runs were motion corrected using an algorithm based on the principle of M-estimation which reduces the influence of large local intensity changes (INRIAlign; (Freire and Mangin, 2001; Freire et al., 2002) as implemented in the SPM software (Wellcome Trust Centre for Neuroimaging).

A mean functional image volume was constructed for each run from the realigned image volumes. The mean EPI image from each run was normalized to the EPI template. The spatial transformation into standard MNI space was determined using a tailored algorithm with both linear and nonlinear components (Friston et al., 1994). The normalization parameters determined for the mean functional volume were then applied to the corresponding functional image volumes for each participant. The normalized functional images were smoothed with a 9 mm full width at half-maximum (FWHM) Gaussian filter. Event-related responses were modeled using a synthetic hemodynamic response function composed of two gamma functions. The first gamma function modeled the hemodynamic

response using a peak latency of 6 s. A term proportional to the derivative of this gamma function was included to allow for small variations in peak latency. The second gamma function and associated derivative was used to model the small "overshoot" of the hemodynamic response on recovery. A latency variation amplitude-correction method was used to provide a more accurate estimate of hemodynamic response for each condition that controlled for differences between slices in timing and variation across regions in the latency of the hemodynamic response (Calhoun et al., 2004). High-pass (cutoff period 116 s) and low-pass (cutoff period 0.23 s) filters were applied to remove any low- and high-frequency confounds, respectively. Condition-specific derivative terms (Calhoun et al., 2004) were extracted for analysis.

Two standard analyses techniques were performed on these fMRI data by extracting of BOLD signal from 1) a-priori regions of interest (ROIs) 2) the IC most related to FA. In each case, first-level general linear models (GLMs) were implemented which included regressors to model motion (six parameters), Hits, Misses, Correct Rejections, and False Alarms and their temporal derivatives. BOLD signal from the contrast FA > mean was extracted from four a-prior ROIs (Kiehl et al., 2000) by fitting 10 mm sphere centered on the cACC ($x = 4$, $y = 22$, $z = 40$; in MNI space), medial frontal gyrus ($x = -4$, $y = 4$, $z = 60$), rACC ($x = 12$, $y = 36$, $z = 12$), and left middle frontal gyrus ($x = -34$, $y = 41$, $z = 20$).

Group ICA was performed on the preprocessed fMRI time-series data (Calhoun et al., 2001). ICA is a data-driven multivariate analysis method that identifies distinct groups of brain regions with the same temporal pattern of hemodynamic signal change. These methods are implemented in the MATLAB toolbox for ICA of fMRI (GIFT) and have been detailed previously (Calhoun and Adali, 2012; Calhoun et al., 2001). A two-stage PCA data reduction step was implemented, at the single participant level 150 PCs were extracted, followed by a group PCA step using 75 components, lower than the first step, as recommended in Erhardt et al. (2011). The data reduction was followed by a group spatial ICA, performed on the participants' aggregate data, resulting in the final estimation of our ICs. The algorithm used in this process was the infomax algorithm, which attempts to minimize the mutual information of network outputs (Bellemann et al., 1995). This was followed by a back reconstruction of single-participant time courses and spatial maps from the raw data using the group solution to accurately depict the participant-to-participant variability that existed in the data (Erhardt et al., 2011). The resulting single-participant time-course amplitudes were then intensity prenormalized (Calhoun et al., 2001). These spatial maps and time courses were then regressed against the same design matrix used in the SPM analysis. These partial correlations were used to identify non-artifactual networks that were related to FAs. A single component (Figure 4) was identified to be highly correlated with FAs across participants (i.e., correlation coefficients for this IC within each participant were significantly different from zero; $t(96) = 17.40$, $p < .001$; Figure 4).

2.5 Reliability Testing

Similar procedures to test internal reliability of neural signals were carried out for both ERP and fMRI. Each participant committed at least 14 False Alarms (ERP mean = 23.56 SD = 7.29, range: 14–49; fMRI mean = 31.09, SD = 10.54, range: 14–61). Sets of 2 trials between

2 and 14 trials were randomly selected, without replacement, for each participant. These sets of random trials were then compared using Cronbach's Alpha to the average ERP or fMRI signal measured to all FAs within each participant. Means and standard deviations were extracted for each ERP (ERN/Ne, Pe, PCA, Theta, and Delta) and fMRI measure (4 ROIs and 1 IC). These were bootstrapped 10,000 times then compared to the overall average signal related to all FAs within each measure. Once these measures were calculated, sets of participants from n = 10 to n = max number of participants in each modality (100 for ERP and 97 for fMRI) were randomly selected, without replacement. Much like randomly selecting sets of trials to identify the minimum number of errors needed to measure a reliable signal, randomly selecting sets of participants were used to identify the minimum number or participants needed to measure a reliable signal. These two randomly selected sets were then combined to identify the number of trials by number of participants needed to measure a reliable error-related signal with ERP and fMRI measures. Correlation coefficients and Cronbach's Alphas were calculated within each trial and participant set. Similar to previous ERP reports testing reliability the ERN/Ne and Pe (Maurer et al., *in press*; Meyer et al., 2013; Olvet and Hajcak, 2009; Pontifex et al., 2010; Rietdijk et al., 2014), Cronbach's Alphas above .70 were considered reliable.

3.0 Results

3.1 Event-Related Potentials

Separate linear regressions were computed predicting ERN/Ne and Pe amplitude with the TD-PCA and TF-PCA components. The ERN/Ne was predicted by TD-PC1, TD-PC4, tPC3, dPC1, dPC2, dPC3, and dPC4 ($p's < .01$; Table S1). The Pe was predicted by TD-PC1, TD-PC2, TD-PC3, TD-PC4, tPC2, dPC1, and dPC2 ($p's < .05$; Table S2). Obviously, these measures overlap in describing the ERP component windows but using these regressions and comparing temporal and special similarities, PCs that primarily represent each component were identified. ERN/Ne was primarily described by TD-PC4, tPC3, and dPC3. Pe was primarily described with TD-PC1, TD-PC3, tPC2, dPC1, and dPC2.

Replicating previous findings, four to six errors were necessary for a reliable TD signal in both the ERN/Ne and Pe window (Figure 5; Table S3). As the number of trials increased, mean amplitude and standard deviations stabilized and correlations with overall signal increased (Table S4). Similar effects were apparent in TD-PCA (Figure 6; Tables S3 & S4) analysis and TF-PCA analyses of theta (Figure 7; Table S1 & S2) and delta (Figure 8; Table S3 & S4). Summarized in Table 1, the minimum number of trials needed for each ERP measure did not deviate drastically beyond 30 participants. This suggests 30 participants with four to six errors are required to extract reliable error-related ERP signals.

3.2 Functional Magnetic Resonance Imaging

Extending previous findings, between four and eight trials were needed for a reliable signal for the five fMRI measures (Figure 9; Tables S3 & S4). The cACC measure was reliable with as few as four trials. Summarized in Table 1, the minimum number of errors and participants necessary for each fMRI measure did not deviate drastically beyond 40

participants. This suggests 40 participants with six to eight errors are required to extract a reliable error-related BOLD signal.

4.0 Discussion

The number of participants and trials required for a reliable error-related signal were identified for common ERP and fMRI data analysis measures. Reliability testing replicated and extended previous findings suggesting ERP measures require between four and six trials and about 30 participants for a reliable signal. Similarly, each of the fMRI measures required six to eight trials and about 40 participants for a reliable signal. By identifying both the number of trials and number of participants needed to extract a reliable error-related signal in both ERP and fMRI, stability of measured signals can be evaluated. Also, appropriate numbers of trials and participants can be established prior to beginning data collection. Researchers and reviewers alike can be confident in their interpretations of findings based on the number of participants and number of error trials measured.

The goal of the current report was to examine the varying stability of response-locked, error-related processing given varying numbers of participants and trial numbers. We used a simple Go/NoGo task to examine error-related activity (elicited by FAs) using ERP and fMRI measures, sampling from large numbers of participants. This is the first report to systematically evaluate the effects of varying both the number of trials and the number of participants for ERP and fMRI measures of error-related neural activity. The resulting data tables will be helpful for supporting the relative stability of ERN/Ne and Pe measures, and their respective BOLD signal, for future studies examining these cognitive functions. Researchers conducting or evaluating these future studies now have an index to compare, based on sample characteristics, to know if the study has sufficient trials and participants to measure a reliable error-signal.

Based on the results described above, we present a general rule of minimum number of trials and participants needed for measuring a reliable error-related signal. However, by measuring number of errors and number of participants needed (see Table 1) across several subsets, it is possible to understand the necessary trials needed for the specific measure of interest and number of participants collected. A reliable signal may be possible with as few as 20 participants in fMRI if each participant commits at least 14 errors. It is apparent that the random sampling and bootstrapping techniques implemented here did not produce results that were completely consistent. For example, the required number of trials for the ERN window with 30 participants is greater than any other set of participants (see Table 1). Replication of these effects is necessary to fully appreciate small idiosyncrasies in these data. We expect the general rule we outline to facilitate implementation of quality methods and interpretation of analyses with specific number of trials and participants in mind.

Finally, some recommendations for which measures to extract are outlined. TD component measures are common though PCA measures have been argued to be more appropriate for representing sequential and overlapping neural signals (Dien, 1998; Dien et al., 2007). We have previously reported the advantages of TD-PCA (Anderson et al., 2015; Maurer et al., *in press*; Steele et al., 2015; Steele et al., 2014b; Steele et al., *in press*) as well as TF-PCA

(Bernat et al., 2015; Bernat et al., 2008; Bernat et al., 2011; Gilmore et al., 2010; Hall et al., 2007; Harper et al., 2014; Nelson et al., 2011; Steele et al., 2013b) measures for condition and group comparisons. As apparent in Table 1, the TD-PCA and TF-PCA measures were reliable with similar or fewer trials and participants than the TD measures. As argued previously, the PCA measures are potentially a better neural correlate of the underlying cognitive processes than the TD measures alone. In the fMRI analyses, the advantage of ICA is more striking (Table 1). As has been argued before (Calhoun et al., 2001; Calhoun et al., 2005; Xu et al., 2013a), ICA represents neural correlates between task conditions and groups more accurately than GLM-based ROI analyses (Malinen et al., 2007; Tie et al., 2008; Xu et al., 2013b). The ICA measure of ACC activation was reliable with four trials and 10 participants when the ROI measures required six trials and 40 participants for similar reliability. It is likely that six trials would be better to ensure reliability for this ICA ACC measure, considering six trials were required for all other participant sets. Nonetheless, the ICA measure of the error signal required fewer trials and participants to achieve reliability than did the ROI measure. In both PCA and ICA data reduction steps, the better representation of the neural signal is likely due to the data reduction properties of these techniques that are not inherent in the TD or GLM-based ROI measures of comparison. Therefore, it is recommended researchers perform an additional step (PCA with ERP data or ICA with fMRI data) to better represent the neural correlates of error-processing.

4.1 Limitations and Future Directions

In this study, as with any scientific experiment, there are a few limitations. First, we extensively analyzed a single Go/NoGo task rather than several tasks that elicit similar error-related signals. The number of trials necessary for an ERP error signal has been found to be similar across tasks such as Flanker and Go/NoGo (Meyer et al., 2013). This suggests error-generating signals exhibit some robustness across tasks. Nevertheless, replications of these effects are necessary and extending to additional tasks is likely needed. It is also possible that participants who make many errors exhibit different error-related neural signals than individuals who make only a few errors. This is an area open for future testing that could extend the findings presented here.

Second, the analyses presented here were based on a relatively young and healthy sample so generalizing to additional samples (e.g., clinical samples) may require additional analyses. We point out that the first attempt to identify the number of trials needed for a reliable ERP error signal was also performed in healthy participants (Olvet and Hajcak, 2009). This is just to say we acknowledge that a useful next step would be to extend these analyses into clinical populations as well as older samples. Clinical samples have been explored previously with ERP measures (Foti et al., 2013) but not all of the ERP and fMRI measures presented here were analyzed. As one may expect, the clinical population required many more trials for a reliable ERP error-related signal measured by TD components (Foti et al., 2013). However, as few as four error trials were necessary for a reliable signal in an incarcerated sample with elevated psychopathic traits (Maurer et al., *in press*). Group comparisons may also require more trials for reliable error-related signals. Age-related reliability measures have also been explored (Pontifex et al., 2010) suggesting older adults (i.e., older than 60 years old) require more trials for a reliable signal. In the current analysis, no participants were older than 60

years old and only a few, four in the ERP sample and six in the fMRI sample, were older than 50 years old. Considering our overall error-rates were negatively correlated with age in both samples, additional samples of older adults will be necessary to replicate the previous finding. Researchers comparing young and old adults should be aware of potential differences in error-rates and trials needed for reliable signals related to age. Additional testing would be required to identify those benchmarks.

Third, here we used a subset of ERP and fMRI measures although additional alternative ERP and fMRI analyses may yield different results. We chose the most common measurements for each modality in an attempt to generalize to the majority of researchers while also striving for brevity. Future explorations should perform similar analysis as presented here on phase coherence, inter-trial coherence, functional connectivity, and fusion of ERP and fMRI measures. Implementing similar reliability measurements with these analysis techniques would be a useful extension of the present study.

4.2 Conclusion

We have replicated and extended previous ERP findings suggesting the minimum number of trials needed to extract a reliable error-related signal. We extend these findings beyond ERP TD measures to ERP PCA and TF measures as well as fMRI ROI and fMRI ICA measures. In all cases, four to eight trials are needed for a reliable signal. Also, we tested the minimum number of participants needed to extract a reliable signal. Thirty participants with four to six error trials are needed to extract a reliable error-related signal in ERPs and 40 participants with six to eight trials are needed for fMRI error-related signals. As number of error trials measured and participants collected fluctuates, so do the minimum requirement for measuring a reliable signal. We present a thorough review of these measures to be used in development of research protocols and evaluation of one's own, and others', work related to error-processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Anderson NE, Steele VR, Maurer JM, Bernat EM, Kiehl KA. Psychopathy, attention, and oddball target detection: New insights from PCL-R facet scores. Psychophysiology. 2015

Bellemann ME, Spitzer M, Brix G, Kammer T, Loose R, Schwartz A, Fuckel F. Neurofunctional MRI imaging of higher cogntive performance of the human brain. Radiologe. 1995; 35:272–282. [PubMed: 7597161]

Bernat EM, Nelson LD, Baskin-Sommers AR. Time-frequency theta and delta measures index separable components of feedback processing in a gambling task. Psychophysiology. 2015; 52:626–637. [PubMed: 25581491]

Bernat EM, Nelson LD, Holroyd CB, Gehring WJ, Patrick CJ. Separating cognitive processes with principal components analysis of EEG time-frequency distributions. Advanced Signal Processing Algorithms, Architectures, and Implementations. 2008; 7074:707401–7074010.

Bernat EM, Nelson LD, Steele VR, Gehring WJ, Patrick CJ. Externalizing psychopathology and gain-loss feedback in a simulated gambling task: dissociable components of brain response revealed by time-frequency analysis. Journal of abnormal psychology. 2011; 120:352–364. [PubMed: 21319875]

Bernat EM, Williams WJ, Gehring WJ. Decomposing ERP time-frequency energy using PCA. Clinical Neurophysiology. 2005; 116:1314–1334. [PubMed: 15978494]

Calhoun VD, Adali T. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. Biomedical Engineering, IEEE Reviews. 2012; 5:60.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Human brain mapping. 2001; 14:140–151. [PubMed: 11559959]

Calhoun VD, Adali T, Stevens MC, Kiehl KA, Pekar JJ. Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. NeuroImage. 2005; 25:527–538. [PubMed: 15784432]

Calhoun VD, Stevens MC, Pearlson GD, Kiehl KA. fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage. 2004; 22:252–257. [PubMed: 15110015]

Carbonnell L, Falkenstein M. Does the error negativity reflect the degree of response conflict? Brain research. 2006; 1095:124–130. [PubMed: 16712810]

Carter CS, Braver TS, Barch DM, Botvinick M, Noll D, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science. 1998; 280:747–749. [PubMed: 9563953]

Chapman RM, McCarry JW. EP component identification and measurment by principal component analysis. Brain and cognition. 1995; 27:288–310. [PubMed: 7626278]

Cohen J, Polich J. On the number of trials needed for P300. International Journal of Psychophysiology. 1997; 25:249–255. [PubMed: 9105949]

Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004; 134:9–21. [PubMed: 15102499]

Demiralp T, Ademoglu A, Istefanopulos Y, Basar-Eroglu C, Basar E. Wavelet analysis of oddball P300. International Journal of Psychophysiology. 2001; 39:221–227. [PubMed: 11163899]

Desmond JE, Glover GH. Estimating sample size in functional MRI (fMRI) neuroimaging studies: Statistical power analyses. Journal of Neuroscience Methods. 2002; 118:115–128. [PubMed: 12204303]

Dien J. Addressing misallocation of variance in principal components analysis of event-related potentials. Brain topography. 1998; 11:43–55. [PubMed: 9758391]

Dien J, Khoe W, Mangun GR. Evaluation of PCA and ICA of simulated ERPs: Promax vs. Infomax rotations. Human Brain Mapping. 2007; 28:742–763. [PubMed: 17133395]

Edwards BG, Calhoun VD, Kiehl KA. Joint ICA of ERP and fMRI during error-monitoring. NeuroImage. 2012; 59:1896–1903. [PubMed: 21930218]

Endrass T, Reuter B, Kathmann N. ERP correlates of conscious error recognition: Aware and unaware errors in an antisaccade task. European Journal of Neuroscience. 2007; 26:1714–1720. [PubMed: 17880402]

Erhardt EB, Rachakonda S, Bedrick EJ, Allen EA, Adali T, Calhoun VD. Comparison of multi-subject ICA methods for analysis of fMRI data. Human brain mapping. 2011; 32:2075–2095. [PubMed: 21162045]

Falkenstein, M.; Hohnsbein, J.; Hoormann, J.; Blanke, L. Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In: Brunia, CHM.; A.W.K.G; Kok, A., editors. Psychophysiological brain research. Tilburg: Tilburg University Press; 1990. p. 192-195.

Falkenstein M, Hohnsbein J, Hoormann J, Blanke L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalography and Clinical Neurophysiology. 1991; 78:447–455. [PubMed: 1712280]

First, MB.; Spitzer, RL.; Gibbon, MWJB.; Williams, JBW. Structured clinical interview for DSM-IV axis I (SCID-I), clinician version. Washington DC: American Psychiatric Press; 1997.

Foti D, Kotov R, Hajcak G. Psychometric considerations in using error-related brain activity as a biomarker in psychotic disorders. Journal of abnormal psychology. 2013; 122:520–531. [PubMed: 23713506]

Freire L, Mangin JF. Motion correction algorithms may create spurious brain activations in the absence of subject motion. NeuroImage. 2001; 14:709–722. [PubMed: 11506543]

Freire L, Roche A, Mangin JF. What is the best similarity measure for motion correction in fMRI time series? IEEE Transactions on Medical Imaging. 2002; 21:470–484. [PubMed: 12071618]

Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Human brain mapping. 1994; 2:189–210.

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E. A neural system for error detection and compensation. Psychological science. 1993; 4:385–390.

Gilmore CS, Malone SM, Bernat EM, Iacono WG. Relationship between the P3 event-related potential, its associated time-frequency components, and externalizing psychopathology. Psychophysiology. 2010; 47:123–132. [PubMed: 19674392]

Hall JR, Bernat EM, Patrick CJ. Externalizing psychopathology and the error-related negativity. Psychological science. 2007; 18:326–333. [PubMed: 17470258]

Handy, TC. Event-related potentials: A methods handbook. Cambridge, MA: MIT Press; 2005.

Harper J, Malone SM, Bernat EM. Theta and delta band activity explain N2 and P3 ERP component activity in a go/no-go task. Clinical Neurophysiology. 2014; 125:124–132. [PubMed: 23891195]

Holroyd CB, Coles MGH. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychological Review. 2002; 109:679–709. [PubMed: 12374324]

Huettel SA, McCarthy G. The effects of single-trial averageing upon the spatial extend of fMRI activation. Neuroreport. 2001; 12:2411–2416. [PubMed: 11496120]

Jasper HH. The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology. 1958; 10:371–375.

Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clinical Neurophysiology. 2000; 111:1745–1758. [PubMed: 11018488]

Kiehl KA, Liddle PF, Hopfinger JB. Error processing and the rostral anterior cingulate: An event-related fMRI study. Psychophysiology. 2000; 37:216–223. [PubMed: 10731771]

Laaksoa MP, Vaurioc O, Koivstod E, Savolainenc L, Eronenc M, Aronenb HJ, Hakolac P, Repoc E, Soininena H, Tihonenc J. Psychopathy and the posterior hippocampus. Behavioural brain research. 2001; 118:187–193. [PubMed: 11164516]

Leuthold H, Sommer S. ERP correlates of error processing in spatial S-R compatibility tasks. Clinical Neurophysiology. 1999; 110:342–357. [PubMed: 10210624]

Luck SJ. Ten simple rules for designing ERP experiments. Event-related potentials: A methods handbook 262083337. 2005

Luck, SJ. An introduction to the event-related potential technique. Cambridge, MA: MIT Press; 2014.

Malinen S, Hlushchuk Y, Hari R. Towards natural stimulation in fMRI--issues of data analysis. NeuroImage. 2007; 35:131–139. [PubMed: 17208459]

Maurer JM, Steele VR, Edwards BG, Bernat EM, Calhoun VD, Kiehl KA. Dysfunctional error-related processing in female psychopathy. Social cognitive and affective neuroscience. in press.

Meyer A, Riesel A, Proudfit GH. Reliability of the ERN across multiple tasks as a function of increasing errors. Psychophysiology. 2013; 50:1220–1225. [PubMed: 24730035]

Miltner WHR, Lemke U, Weiss T, Holroyd C, Scheffers MK, Coles MGH. Implementation of error-processing in the human anterior cingulate cortex: A source analysis of the magnetic equivalent of the error-related negativity. Biological psychology. 2003; 64:157–166. [PubMed: 14602360]

Murphy K, Garavan H. An empirical investigation into the number of subjects required for an event-related fMRI study. NeuroImage. 2004; 22:879–885. [PubMed: 15193618]

Nelson LD, Patrick CJ, Collins P, Lang AR, Bernat EM. Alcohol impairs brain reactivity to explicit loss feedback. Psychopharmacology. 2011; 218:419–428. [PubMed: 21559803]

Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. Cognitive, affective & behavioral neuroscience. 2012; 12:241–268.

Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GPH, Kok A. Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. Psychophysiology. 2001; 38:752–760. [PubMed: 11577898]

Olvet DM, Hajcak G. The stability of error-related brain activity with increasing trials. Psychophysiology. 2009; 46:957–961. [PubMed: 19558398]

Overbeek TJM, Nieuwenhuis S, Ridderinkhof KR. Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. Journal of psychophysiology. 2005; 19:319–329.

Pontifex MB, Scudder MR, Brown ML, O'Leary KC, Wu CT, Themanson JR, Hillman CH. On the number of trials necessary for stabilization of error-related brain activity across the life span. Psychophysiology. 2010; 47:767–773. [PubMed: 20230502]

Rietdijk WJ, Franken IH, Thurik AR. Internal consistency of event-related potentials associated with cognitive control: N2/P3 and ERN/Pe. PLoS One. 2014; 9:e102672. [PubMed: 25033272]

Shalgi S, Deouell LY. Is any awareness necessary for an Ne? Frontiers in human neuroscience. 2012; 6:124. [PubMed: 22593739]

Shalgi S, Deouell LY. Is there any electrophysiological evidence for subliminal error processing? Front Neurosci. 2013; 7:150. [PubMed: 24009548]

Steele VR, Aharoni E, Munro GE, Calhoun VD, Nyalakanti P, Stevens MC, Pearlson G, Kiehl KA. A large scale (N = 102) functional neuroimaging study of response inhibition in a Go/NoGo task. Behavioural brain research. 2013a; 256:529–536. [PubMed: 23756137]

Steele VR, Bernat EM, van den Broek P, Collins PF, Patrick CJ, Marsolek CJ. Separable processes before, during, and after the N400 elicited by previously inferred and new information: Evidence from time-frequency decompositions. Brain Researh. 2013b; 1492:92–107.

Steele VR, Claus ED, Aharoni E, Munro GE, Calhoun VD, Nyalakanti P, Stevens MC, Pearlson G, Kiehl KA. A large scale (N = 102) functional neuroimaging study of error-processing in an Go/NoGo task. Behavioural brain research. 2014a; 268:127–138. [PubMed: 24726752]

Steele VR, Claus ED, Aharoni E, Vincent GM, Calhoun VD, Kiehl KA. Multimodal imaging measures predict rearrest. Frontiers in human neuroscience. 2015; 9:425. [PubMed: 26283947]

Steele VR, Fink BC, Maurer JM, Arbabshirani MR, Wilber CH, Jaffe AJ, Sidz A, Pearlson GD, Calhoun VD, Clark VP, Kiehl KA. Brain potentials measured during a Go/NoGo task predict completion of substance abuse treatment. Biological psychiatry. 2014b; 76:75–83. [PubMed: 24238783]

Steele VR, Maurer JM, Bernat EM, Calhoun VD, Kiehl KA. Error-related processing in adult males with elevated psychopathic traits. Personality Disorders: Theory, Research, and Treatment. in press.

Tie Y, Whalen S, Suarez RO, Golby AJ. Group independent component analysis of language fMRI from word generation tasks. NeuroImage. 2008; 42:1214–1225. [PubMed: 18621548]

Ullsperger M, Harsay HA, Wessel JR, Ridderinkhof KR. Conscious perception of errors and its relation to the anterior insula. Brain Structure & Function. 2010; 214:629–643. [PubMed: 20512371]

van Veen V, Carter CS. The timing of action-monitoring processes in the anterior cingulate cortex. Journal of Cognitive Neuroscience. 2002; 14:593–602. [PubMed: 12126500]

Wellcome Trust Centre for Neuroimaging. U.C.L.

Xu J, Potenza MN, Calhoun VD. Spatial ICA reveals functional activity hidden from traditional fMRI GLM-based analyses. Front Neurosci. 2013a; 7:154. [PubMed: 23986654]

Xu J, Zhang S, Calhoun VD, Monterosso J, Li CS, Worhunsky PD, Stevens M, Pearlson GD, Potenza MN. Task-related concurrent but opposite modulations of overlapping functional networks as revealed by spatial ICA. NeuroImage. 2013b; 79:62–71. [PubMed: 23611864]

Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: Conflict monitoring and the error-related negativity. Psychological Review. 2004; 111:931–959. [PubMed: 15482068]

Yeung N, Summerfield C. Metacognition in human decision-making: Confidence and error monitoring. Philisophical Transactions of the Royal Society B. 2012; 367:1310–1321.

Highlights

- ERP and fMRI measures of error-related signals with reliability testing

- Required number of trials and participants identified for ERP and fMRI measures

- Thirty participants with four to six error trials are needed for ERPs

- Forty participants with six to eight trials are needed for fMRIs

- Findings are useful in developing and evaluating work related to error-processing
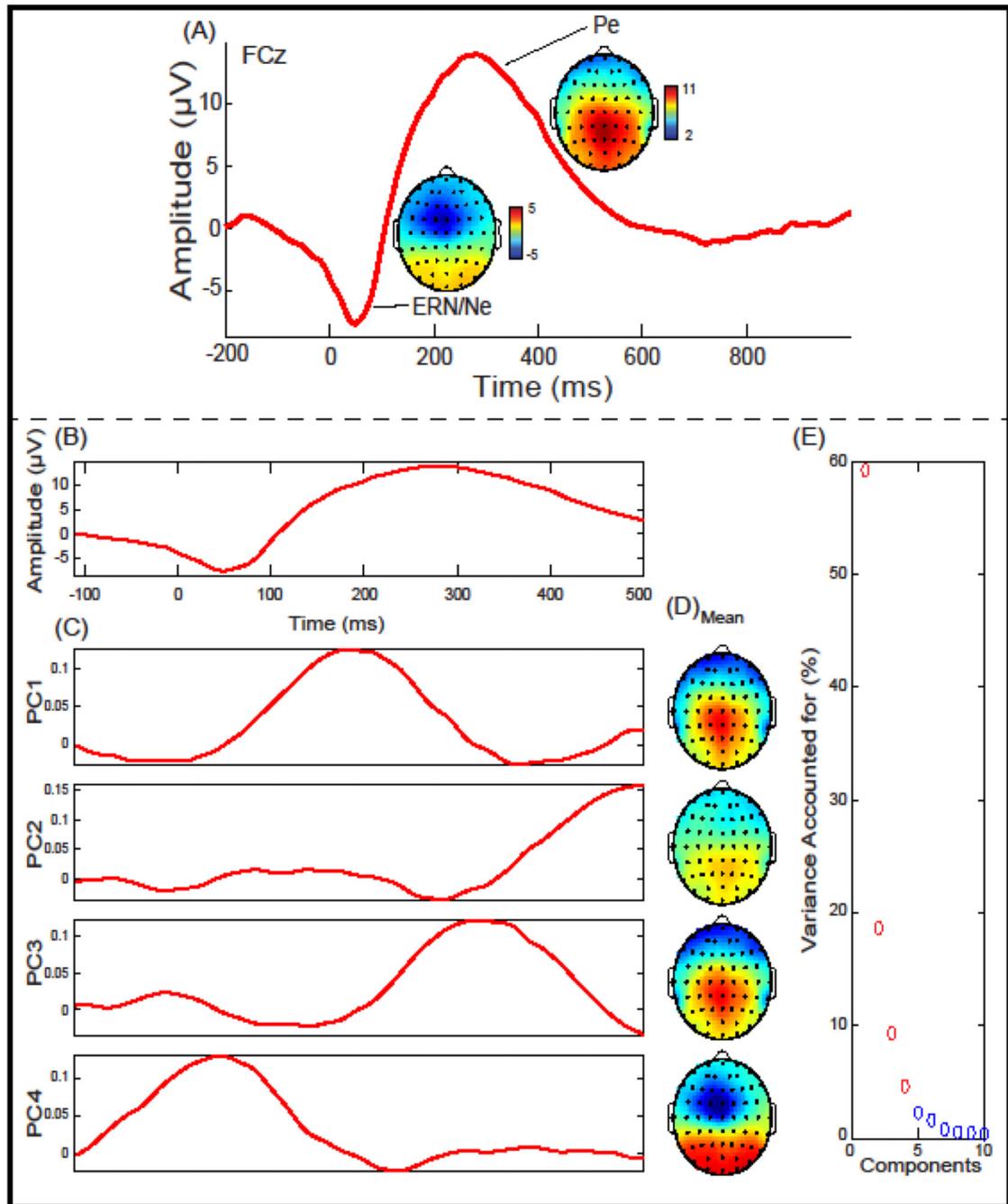
Figure 1.
Response-locked event-related potential (ERP) and Principal Component Analysis (PCA) for False Alarms plotted at FCz: (A) Grand average ERP waveform plotted at FCz. ERP components of interest (ERN/Ne & Pe) and topographical depictions are identified. (B) The Grand average waveform and window used in the principal component analysis. (C) Principal components extracted accounting for 91.79% of the variance. (D) Topographical depiction of the mean spatial distribution for each principal component. (E) Scree plot of singular values which was used to determine a four-component solution.
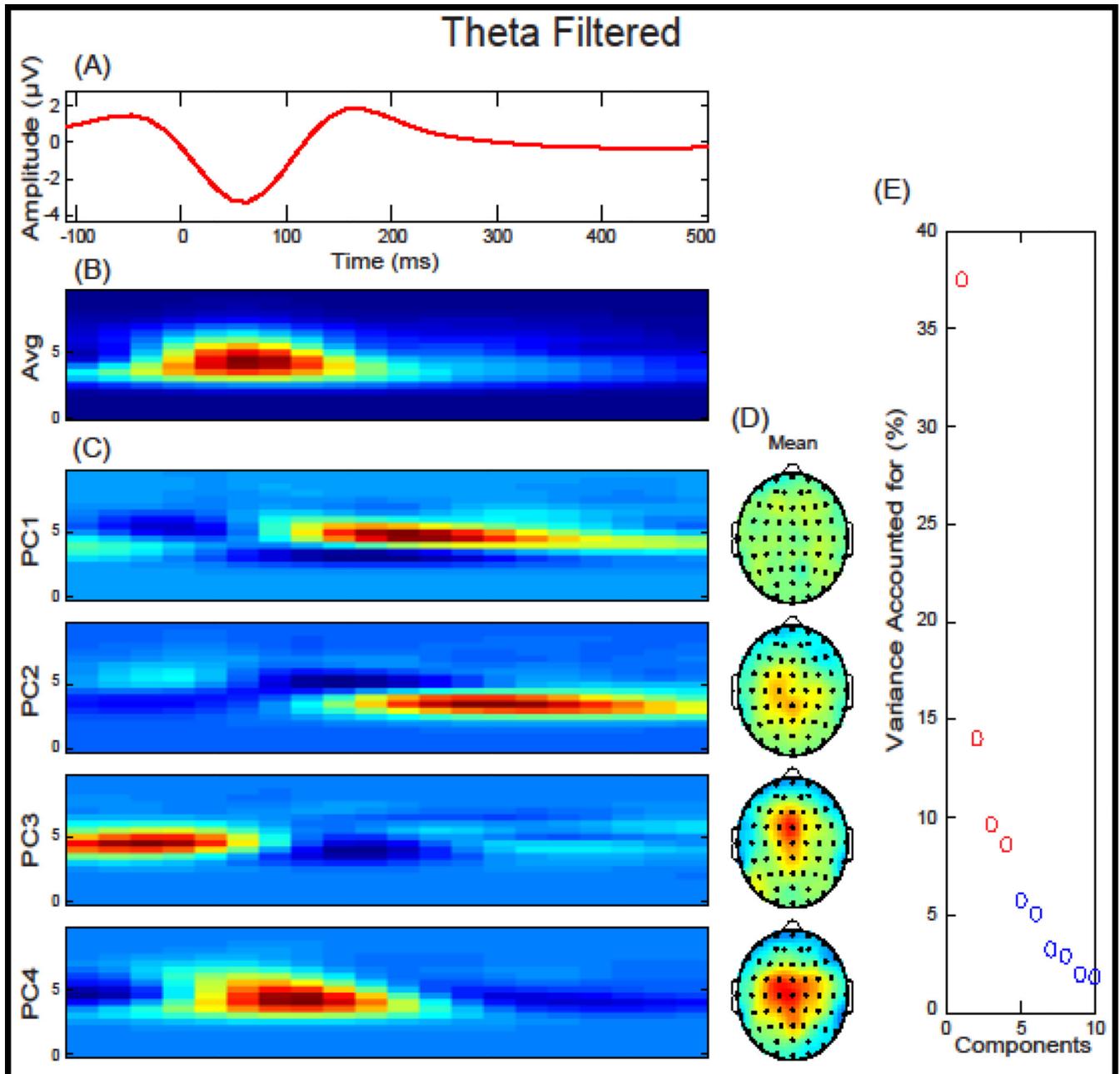
Figure 2.
Theta-filtered response-locked event-related potential (ERP) and time-frequency Principal Component Analysis (TF-PCA) for False Alarms plotted at FCz: (A) Theta-filtered grand average ERP waveform plotted at FCz. (B) The average time-frequency surface used in the principal component analysis. (C) Principal components extracted accounting for 69.73% of the variance. (D) Topographical depiction of the mean spatial distribution for each principal component. (E) Scree plot of singular values which was used to determine a four-component solution.
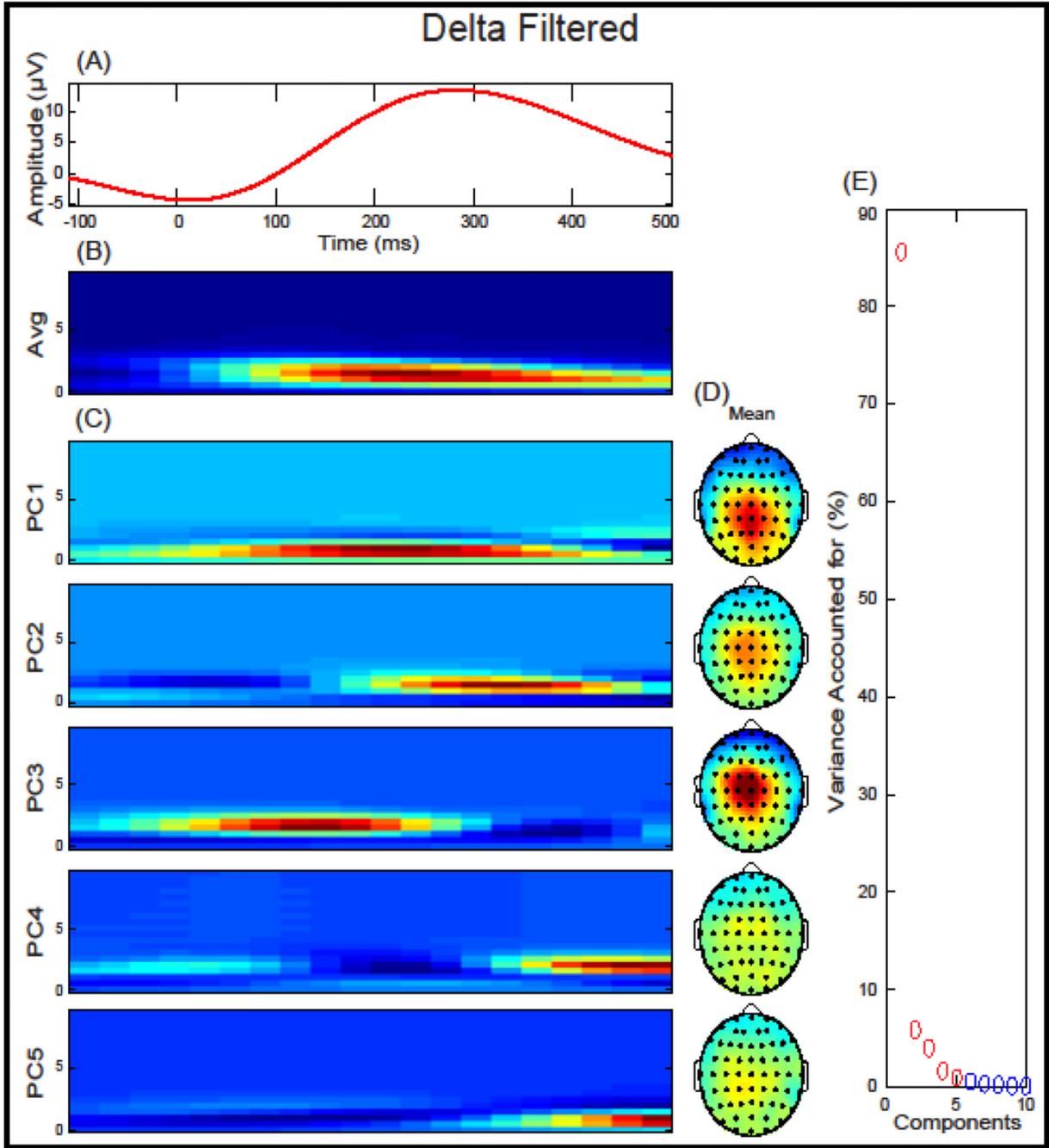
Figure 3.
Delta-filtered response-locked event-related potential (ERP) and time-frequency Principal Component Analysis (TF-PCA) for False Alarms plotted at FCz: (A) Delta-filtered grand average ERP waveform plotted at FCz. (B) The average time-frequency surface used in the principal component analysis. (C) Principal components extracted accounting for 98.09% of the variance. (D) Topographical depiction of the mean spatial distribution for each principal component. (E) Scree plot of singular values which was used to determine a five-component solution.
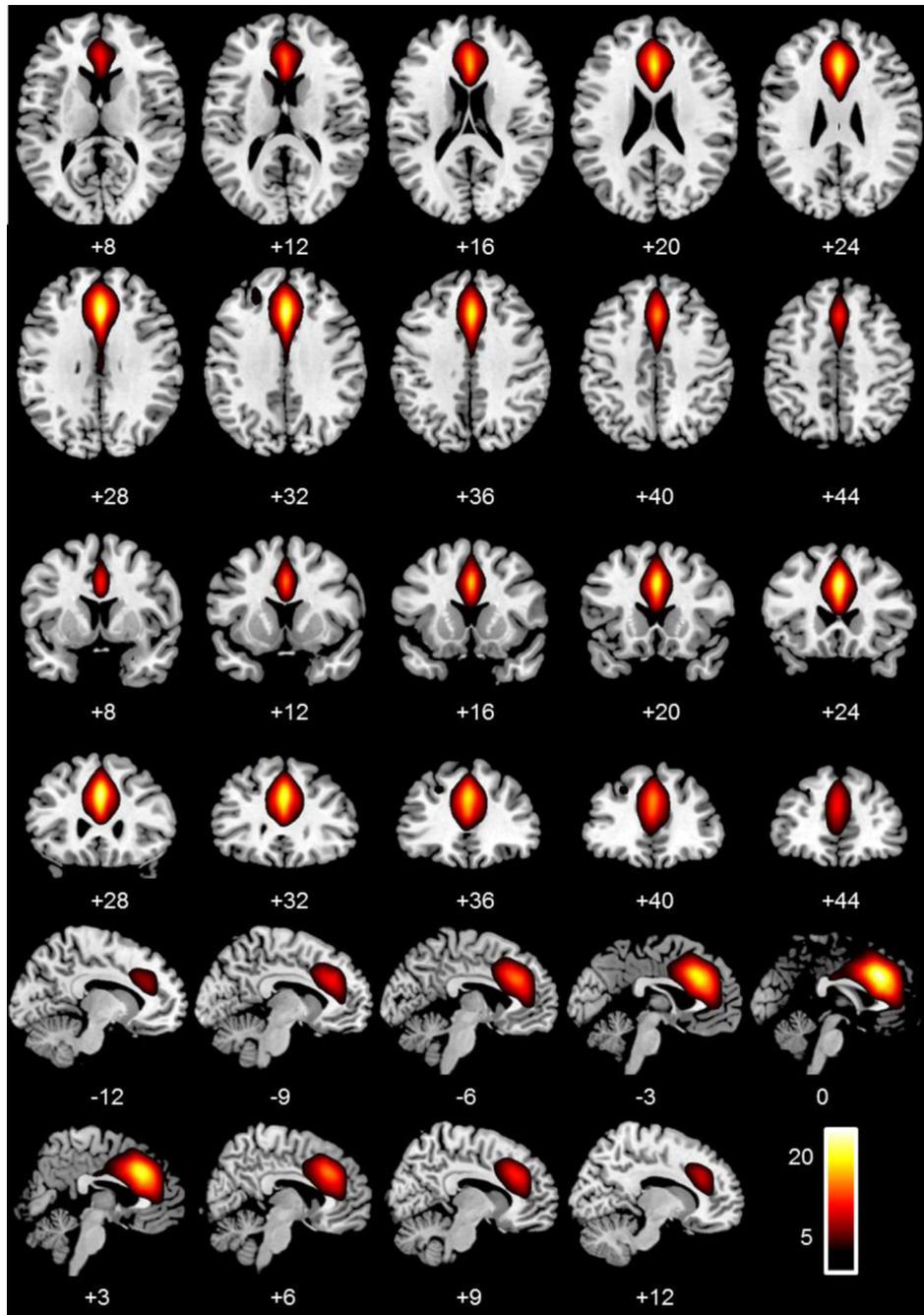
Figure 4.
Axial, coronal, and sagittal slices depicting the independent component, centered on the anterior cingulate cortex, of interest, most related to False Alarms, and used in the analyses.
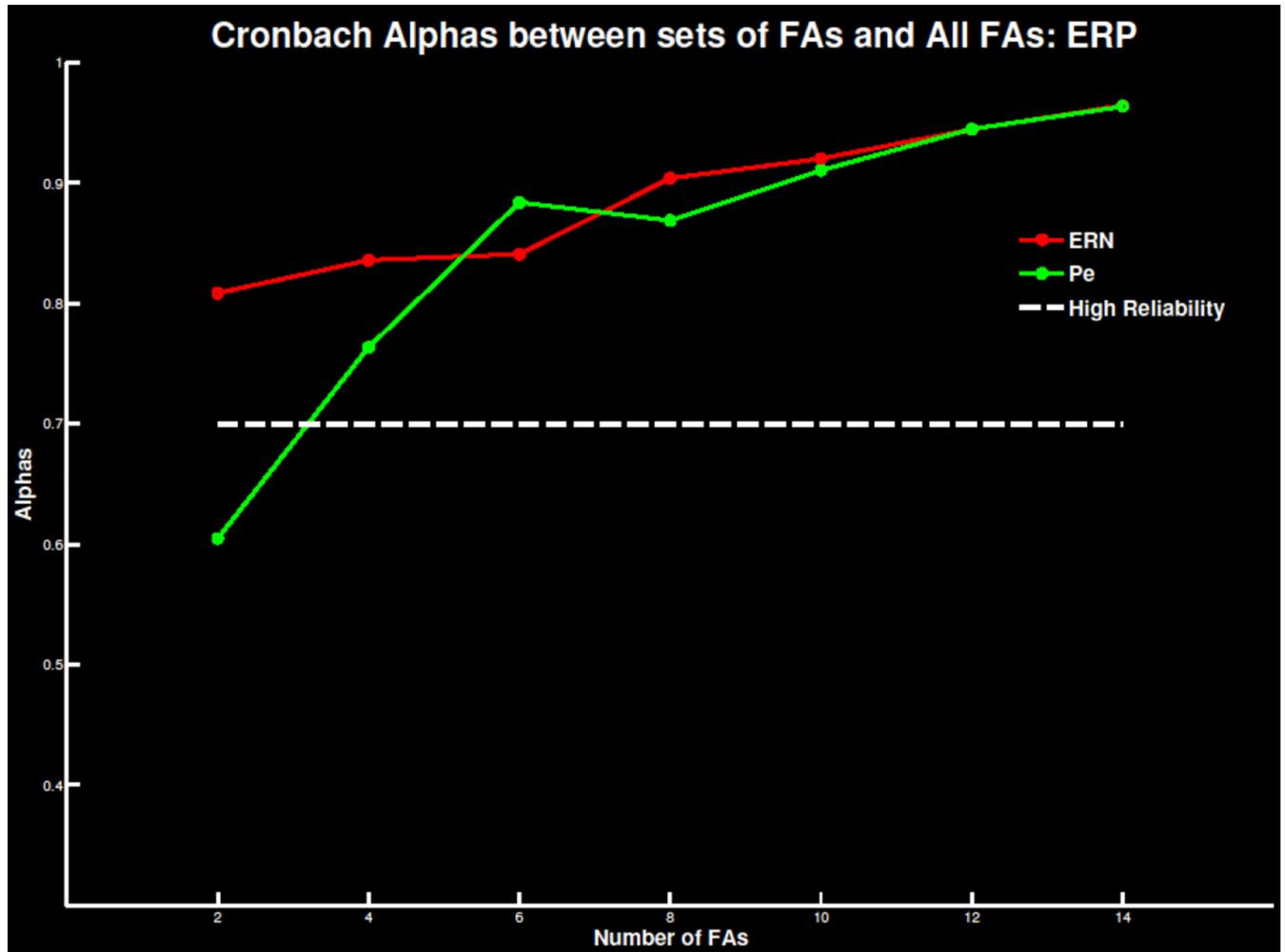
Figure 5.
Cronbach's alpha for the ERN/Ne and Pe at FCz as increasing number of trials are extracted and compared to the overall error signal. The threshold of .7 of highly reliable is also plotted for reference.
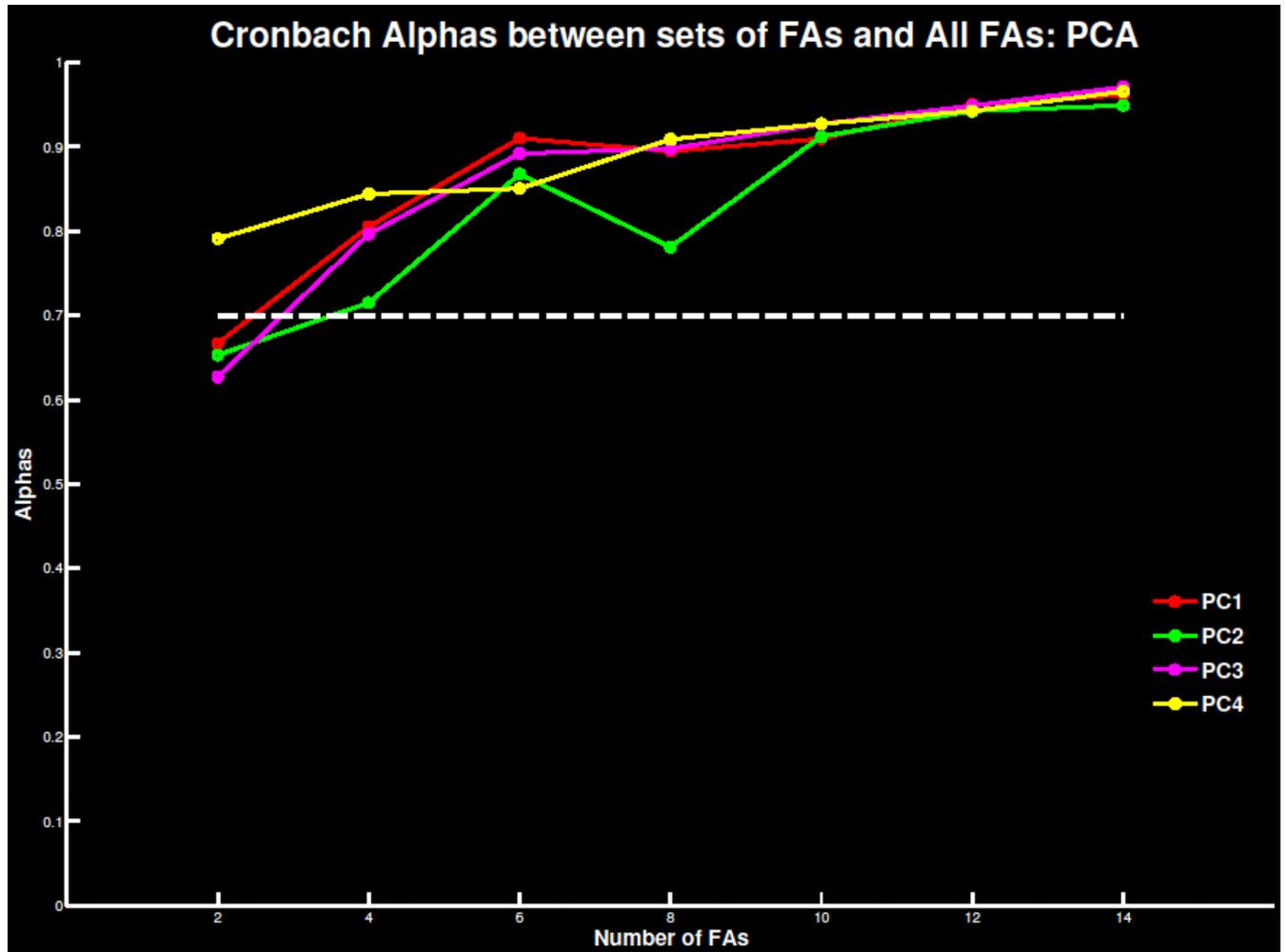
Figure 6.
Cronbach's alpha for the principal component solution with 4 components extracted at FCz as increasing number of trials are extracted and compared to the overall error signal. The threshold of .7 of highly reliable is also plotted for reference.
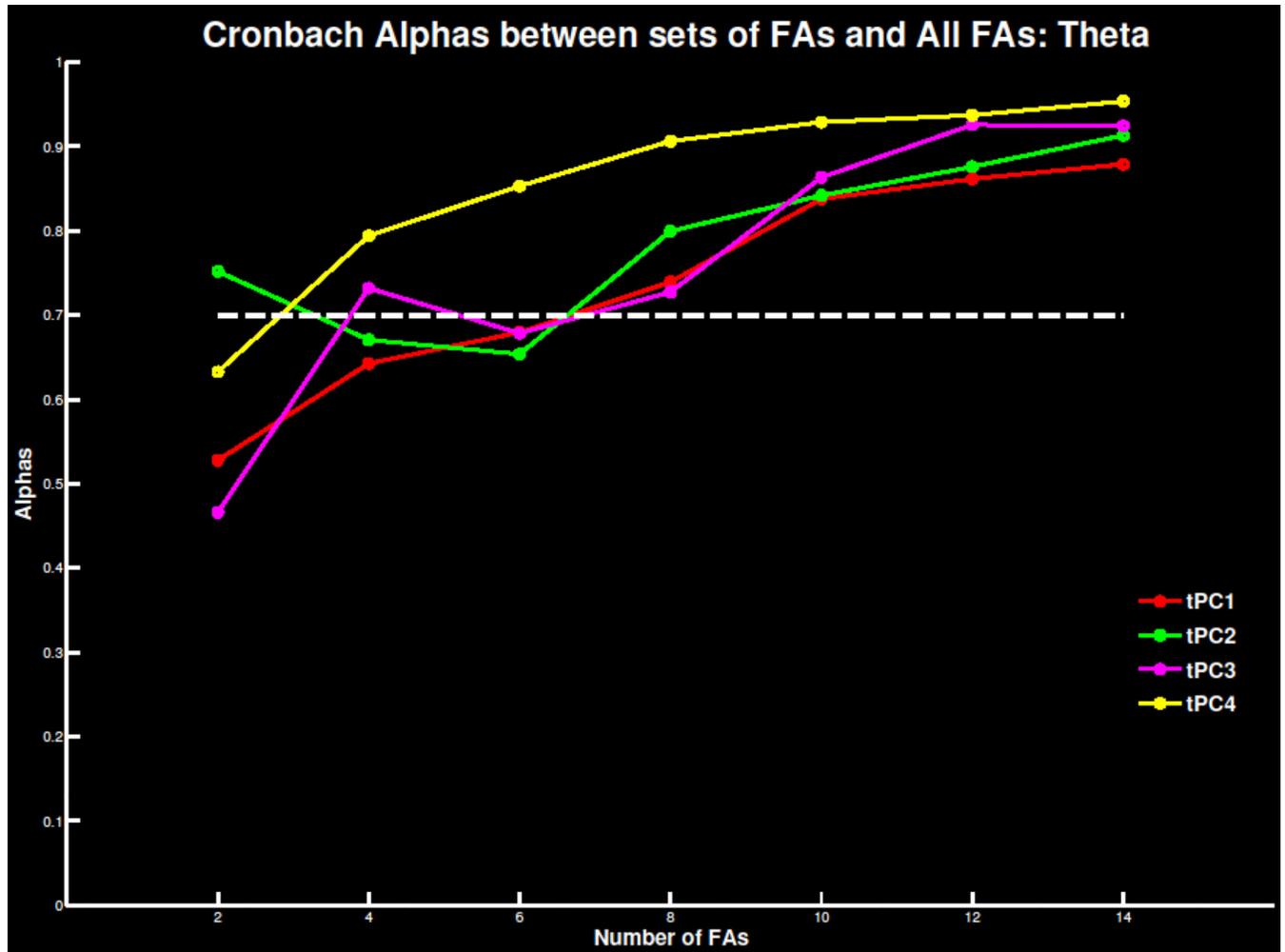
Figure 7.
Cronbach's alpha for the theta filtered time-frequency principal component solution with 4 components extracted at FCz as increasing number of trials are extracted and compared to the overall error signal. The threshold of .7 of highly reliable is also plotted for reference.
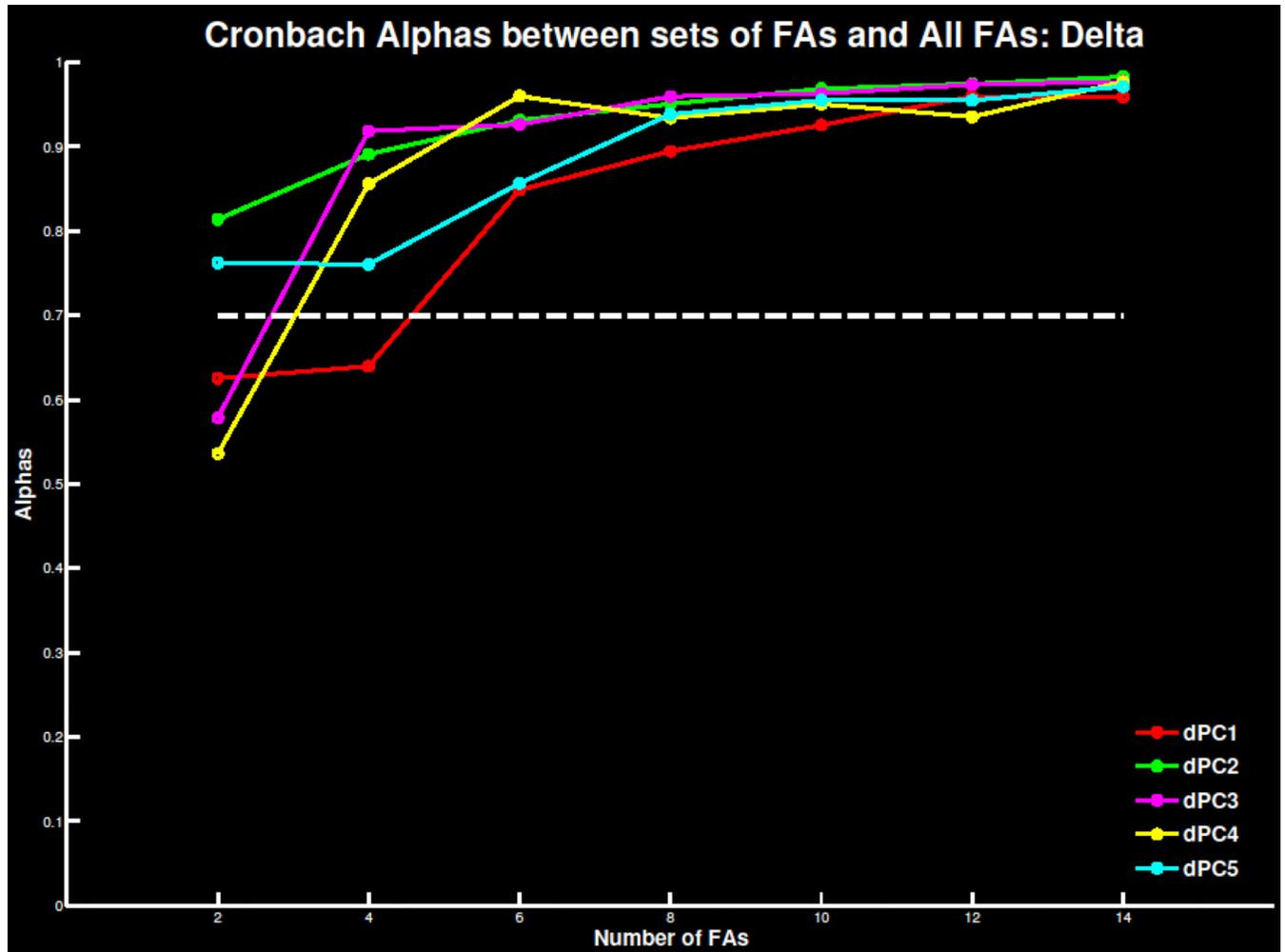
Figure 8.
Cronbach's alpha for the delta filtered time-frequency principal component solution with 5 components extracted at FCz as increasing number of trials are extracted and compared to the overall error signal. The threshold of .7 of highly reliable is also plotted for reference.
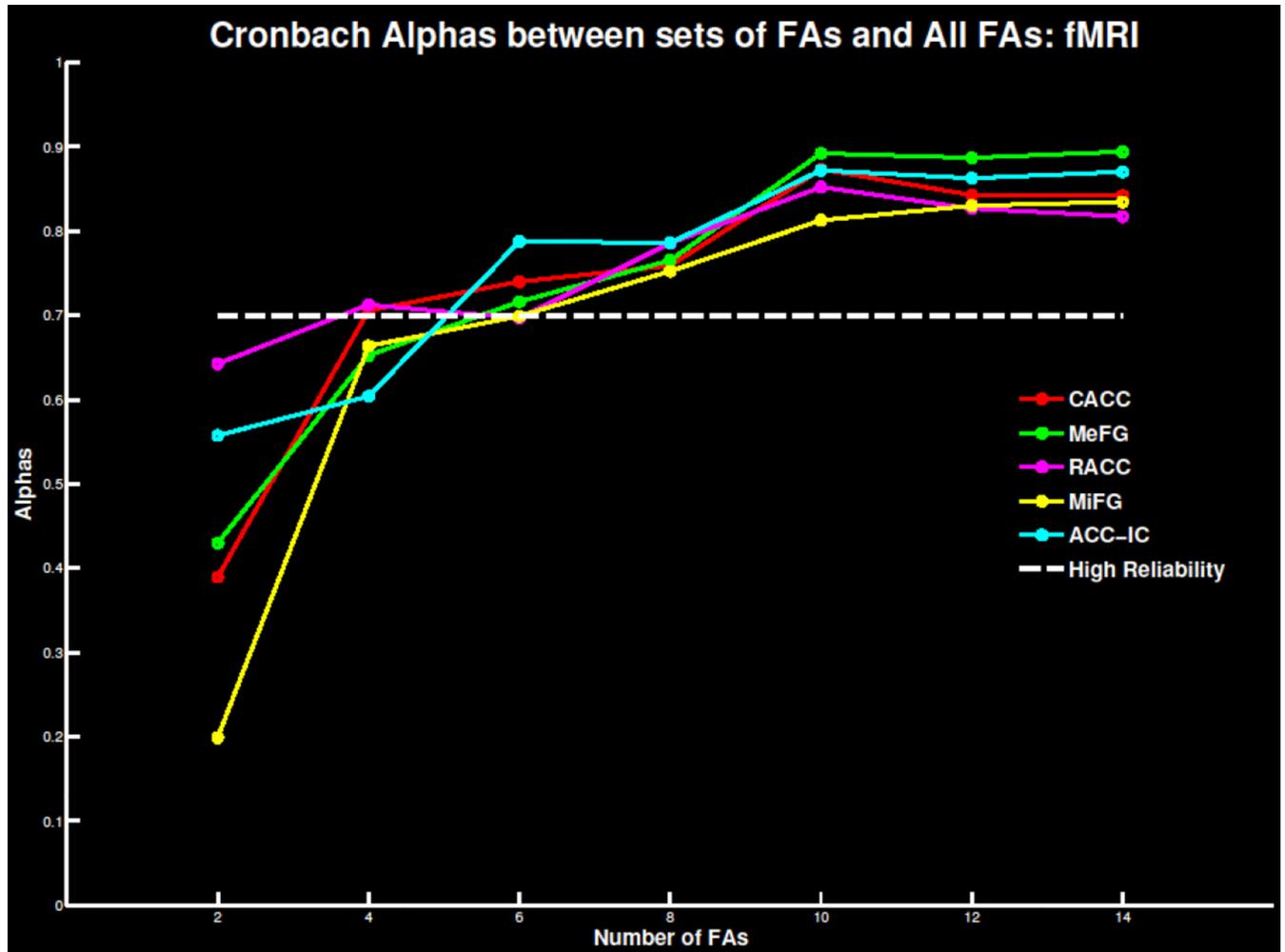
Figure 9.
Cronbach's alpha for the 4 fMRI regions of interest and the independent component centered on the anterior cingulate cortex as increasing number of trials are extracted and compared to the overall error signal. The threshold of .7 of highly reliable is also plotted for reference.

Table 1

Number of Trials by Number of Participants needed for a Reliable Signal by Neuroimaging Measure

Measure	N = 10	N = 20	N = 30	N = 40	N = 50	N = 60	N = 70	N = 80	N = 90	All
ERN/Ne	2	2	6	2	2	2	2	2	2	2
Pe	10	4	4	4	4	6	4	4	4	4
PC1	2	4	4	2	4	2	2	4	4	4
PC2	4	10	6	6	2	6	6	4	2	4
PC3	6	4	2	6	4	4	4	4	2	4
PC4	2	2	4	2	2	2	2	2	2	2
tPC1	14	4	8	4	10	4	6	8	8	8
tPC2	12	12	2	2	8	6	8	8	8	8
tPC3	N/A	6	10	6	4	8	8	4	8	8
tPC4	4	2	4	10	2	2	4	4	4	4
dPC1	2	4	2	2	4	2	2	2	2	2
dPC2	2	4	2	2	4	2	2	2	2	2
dPC3	2	4	2	2	4	4	4	2	4	4
dPC4	14	6	4	4	4	10	4	4	4	4
dPC5	6	6	6	6	4	6	2	2	2	2
CACC	N/A	8	14	6	4	4	6	4	4	4
MeFG	8	10	8	6	4	6	6	6	6	6
RACC	N/A	N/A	14	6	4	8	6	8	8	8
MiFG	N/A	8	14	6	4	10	6	6	8	8
ACC-IC	4	6	6	6	6	6	6	6	6	6

Note: A summary of reliability tests are presented by reporting the minimum number of trials needed to reach .70 reliability for each randomly selected number of participants. For example, the ERN/Ne requires only 2 trials if all participants are included in the analysis but as many as 6 trials if 30 participants are included in the analysis.