UC Davis
UC Davis Previously Published Works

Title

ERP CORE: An open resource for human event-related potential research

Permalink

https://escholarship.org/uc/item/9cf2f1xw

Authors

Kappenman, Emily S

Farrens, Jaclyn L

Zhang, Wendy

et al.

Publication Date

2021

DOI

10.1016/j.neuroimage.2020.117465

ERP CORE: An Open Resource for Human Event-Related Potential Research

Emily S. Kappenman[1,2], **Jaclyn L. Farrens**[1], **Wendy Zhang**[1,2], **Andrew X. Stewart**[3], **Steven J. Luck**[3]

[1]San Diego State University, Department of Psychology, San Diego, CA, 92120

[2]SDSU/UC San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, 92120

[3]University of California, Davis, Center for Mind & Brain and Department of Psychology, Davis, CA, 95616

Abstract

Event-related potentials (ERPs) are noninvasive measures of human brain activity that index a range of sensory, cognitive, affective, and motor processes. Despite their broad application across basic and clinical research, there is little standardization of ERP paradigms and analysis protocols across studies. To address this, we created ERP CORE (Compendium of Open Resources and Experiments), a set of optimized paradigms, experiment control scripts, data processing pipelines, and sample data (N = 40 neurotypical young adults) for seven widely used ERP components: N170, mismatch negativity (MMN), N2pc, N400, P3, lateralized readiness potential (LRP), and error-related negativity (ERN). This resource makes it possible for researchers to 1) employ standardized ERP paradigms in their research, 2) apply carefully designed analysis pipelines and use a priori selected parameters for data processing, 3) rigorously assess the quality of their data, and 4) test new analytic techniques with standardized data from a wide range of paradigms.

Keywords

event-related potentials; EEG; data quality; open science; reproducibility

1. Introduction

The event-related potential (ERP) technique is a widely used tool in human neuroscience. ERPs are primarily generated in cortical pyramidal cells, where extracellular voltages produced by thousands of neurons sum together and are conducted instantaneously to the

scalp (Buzsáki et al., 2012; Jackson and Bolger, 2014). ERPs therefore provide a direct measure of neural activity with the millisecond-level temporal resolution necessary to isolate the neurocognitive operations that rapidly unfold following a stimulus, response, or other event. Indeed, many ERP *components* have been identified and validated as measures of sensory, cognitive, affective, and motor processes (for an overview, see Luck and Kappenman, 2012). In recent years, the ERP technique has become accessible to a broad range of researchers due to the development of relatively inexpensive EEG recording systems and both commercial and open source software packages for processing ERP data.

Although some aspects of EEG recording and processing have become relatively standardized (Keil et al., 2014; Pernet et al., 2018), many others vary widely across laboratories and even across studies within a laboratory. For example, the P3 component has been measured in thousands of studies using the *oddball* paradigm (Ritter and Vaughan, 1969), but the task parameters, recording settings, and data processing methods vary widely across studies. In many cases, the protocols are based on decades-old traditions that include confounds in the experimental design (such as a lack of counterbalancing) and analysis procedures that are now known to be flawed or suboptimal (Luck, 2014). When improved protocols are developed, there are no widely accepted methods for demonstrating their superiority or for disseminating them so they become broadly adopted. In addition, many important methodological details are often absent from published journal articles. As a result, a researcher who wishes to start using a given ERP paradigm has no standardized protocol to use and no standardized method for assessing whether the quality of the EEG data falls within normative values and whether the ERP components have been properly quantified.

We addressed these issues by creating the ERP CORE (Compendium of Open Resources and Experiments), a freely available online resource consisting of optimized paradigms, experiment control scripts, data from 40 neurotypical young adults, data processing pipelines and analysis scripts, and a broad set of results (https://doi.org/10.18115/D5JW4R). Following extensive piloting and consultations with experts in the field, we developed six 10-minute optimized paradigms that together isolate seven ERP components spanning a range of neurocognitive processes (see Figure 1): 1) a visual discrimination paradigm for isolating the face-specific N170 response (for reviews, see Eimer, 2011; Feuerriegel et al., 2015; Rossion and Jacques, 2011); 2) a passive auditory oddball paradigm for isolating the mismatch negativity (MMN; for reviews, see Garrido et al., 2009; Näätänen and Kreegipuu, 2011); 3) a visual search paradigm for isolating the N2pc component (for a review, see Luck, 2012); 4) a word-pair association paradigm for isolating the N400 component (for reviews, see Kutas and Federmeier, 2011; Lau et al., 2008; Swaab et al., 2012); 5) an active visual oddball paradigm for isolating the P3 component (for reviews, see Dinteren et al., 2014; Polich, 2007; Polich, 2012); and 6) a flankers paradigm for isolating the lateralized readiness potential (LRP; for reviews, see Eimer and Coles, 2003; Smulders and Miller, 2012); and the error-related negativity (ERN; for reviews, see (Gehring et al., 2012; Olvet and Hajcak, 2008). Each of these ERP components can be isolated from overlapping brain activity using the difference wave procedure (Kappenman and Luck, 2012; Luck, 2014).

The ERP CORE paradigms were implemented in a widely used experiment control package (*Presentation*; Neurobehavioral Systems) and have been extensively tested. They can be run with a free trial license, and many of the parameters can be adjusted from the user interface so that researchers can create variations on the basic paradigms without editing the code (thus minimizing the potential for programming errors). We expect that data from hundreds of additional participants will be added to our current database of 40 participants by other laboratories. We also anticipate the creation of experiment control scripts that can be run with open source software (e.g., Peirce, 2019).

We also developed optimized signal processing and data analysis pipelines for each component using the open source EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014) MATLAB toolboxes. Archival copies of the analysis scripts, raw and processed data for all 40 participants (including BIDS-compatible data), and a broad set of results are available online on the Open Science Framework (https://doi.org/10.18115/D5JW4R). The analysis scripts are also hosted on GitHub (https://github.com/lucklab/ERP_CORE), where updated code can be provided and bugs can be reported.

We anticipate that the ERP CORE will be used in at least eight ways. First, researchers who are setting up a new ERP lab can use the standardized CORE paradigms to test their laboratory set-up and data quality. Second, researchers who are new to the ERP technique may use the CORE paradigms and analysis scripts to enhance their understanding of ERP experimental design and analysis methods, which will serve as a starting point for the development of new paradigms. Third, researchers who would like to add a standardized ERP measure to a multi-method study can take our experiment control scripts and data processing pipelines and "plug them in" to their study with relative ease and confidence. Moreover, they could use the a priori analysis parameters that we provide (e.g., time windows for amplitude and latency quantification), reducing *researcher degrees of freedom* (Simmons et al., 2011). Fourth, we have provided a participant-by-participant quantification of the noise levels in the data, which both new and experienced researchers can use as a comparison against the noise levels in their data. ERP papers rarely provide information about noise levels, making it difficult to know which data collection protocols yield the cleanest data, and the ERP CORE provides a first step toward standardized reporting of noise levels.

Fifth, researchers who would like to create a new variant of a standard ERP paradigm could use our experiment control scripts and data processing pipelines as a starting point, saving substantial time and reducing uncertainty and error. Sixth, researchers could test new hypotheses by reanalyzing the existing data in novel ways. For example, because we have data from seven different ERP components in each participant, it would be possible to ask how the timing and amplitude of one component is correlated with the timing and amplitude of other components. Seventh, newly developed data processing procedures could be applied to the ERP CORE data to test the effectiveness of these procedures across a broad range of paradigms. Finally, educators could use these resources to teach students about the design, implementation, and analysis of ERP experiments. For example, the ERP CORE is already being integrated into a formal curriculum for teaching human electrophysiology (Bukach et al., 2019).

2. Material and Methods

This study was approved by the Institutional Review Board at the University of California, Davis, and all participants provided informed consent. All materials are freely available at https://doi.org/10.18115/D5JW4R.

2.1. Participants

We tested 40 participants (25 female, 15 male; Mean years of age = 21.5, SD = 2.87, Range 18-30; 38 right handed) from the University of California, Davis community. Each participant had native English competence and normal color perception, normal or corrected-to-normal vision, and no history of neurological injury or disease (as indicated by self-report). Participants received monetary compensation at a rate of $10/hour.

In our research with typical young adults, we always exclude participants who exhibit artifacts on more than 25% of trials (Luck, 2014). To maximize the amount of data we were able to retain in each experiment, this criterion was applied separately for each task. We also excluded participants from a task if their accuracy was below 75%, or if fewer than 50% of trials remained in any single experimental condition. To ensure an adequate number of error trials for the ERN, participants were excluded from the error-related negativity (ERN) analysis if fewer than 6 error trials remained after artifact rejection (Boudewyn et al., 2018). These criteria resulted in the exclusion of 1–6 participants per component. The final sample size for each component is listed in Table 1. Additional details of participant exclusions are available in the online resource. All analyses were performed using the final sample for each task, except where noted. The individual-participant data files for excluded participants are provided in the online resource.

2.2. Stimuli and Tasks

Figure 1 shows example trials in each of the six tasks. Here we provide a brief overview of each task; details are provided in the Supplementary Materials and Methods. The N170 was elicited in a face perception task modified from Rossion & Caharel (2011) using their stimuli (which are available in the online resource; see Figure 1A). In this task, an image of a face, car, scrambled face, or scrambled car was presented on each trial in the center of the screen, and participants responded whether the stimulus was an "object" (face or car) or a "texture" (scrambled face or scrambled car). The MMN was elicited using a passive auditory oddball task modeled on Naatanen et al. (2004; see Figure 1B). Standard tones (presented at 80 dB, with p = .8) and deviant tones (presented at 70 dB, with p = .2) were presented over speakers while participants watched a silent video and ignored the tones. The N2pc was elicited using a simple visual search task based on Luck et al. (2006; see Figure 1C). Participants were given a target color of pink or blue at the beginning of a trial block, and responded on each trial whether the gap in the target color square was on the top or bottom. The N400 was elicited using a word pair judgment task adapted from Holcomb & Kutas (1990; see Figure 1D). On each trial, a red prime word was followed by a green target word. Participants responded whether the target word was semantically related or unrelated to the prime word. The P3 was elicited in an active visual oddball task adapted from Luck et al. (2009; see Figure 1E). The letters A, B, C, D, and E were presented in random order (p = .2

for each letter). One letter was designated the target for a given block of trials, and the other 4 letters were non-targets. Thus, the probability of the target category was .2, but the same physical stimulus served as a target in some blocks and a nontarget in others. Participants responded whether the letter presented on each trial was the target or a non-target for that block. The lateralized readiness potential (LRP) and the error-related negativity (ERN) were elicited using a variant of the Eriksen flanker task (Eriksen and Eriksen, 1974; see Figure 1F). A central arrowhead pointing to the left or right was flanked on both sides by arrowheads that pointed in the same direction (congruent trials) or the opposite direction (incongruent trials). Participants indicated the direction of the central arrowhead on each trial with a left- or right-hand buttonpress.

2.3. EEG Recording

The continuous EEG was recorded using a Biosemi ActiveTwo recording system with active electrodes (Biosemi B.V., Amsterdam, the Netherlands). We recorded from 30 scalp electrodes, mounted in an elastic cap and placed according to the International 10/20 System (FP1, F3, F7, FC3, C3, C5, P3, P7, P9, PO7, PO3, O1, Oz, Pz, CPz, FP2, Fz, F4, F8, FC4, FCz, Cz, C4, C6, P4, P8, P10, PO8, PO4, O2; see Supplementary Figure S1). The common mode sense (CMS) electrode was located at PO1, and the driven right leg (DRL) electrode was located at PO2. The horizontal electrooculogram (HEOG) was recorded from electrodes placed lateral to the external canthus of each eye. The vertical electrooculogram (VEOG) was recorded from an electrode placed below the right eye. Signals were incidentally also recorded from 37 other sites, but these sites were not monitored during the recording and are not included in the ERP CORE data set. All signals were low-pass filtered using a fifth order sinc filter with a half-power cutoff at 204.8 Hz and then digitized at 1024 Hz with 24 bits of resolution. The signals were recorded in single-ended mode (i.e., measuring the voltage between the active and ground electrodes without the use of a reference), and referencing was performed offline, as described below.

2.4. Signal Processing and Averaging

Signal processing and analysis were performed in MATLAB using EEGLAB toolbox (version 13_4_4b; Delorme and Makeig, 2004) and ERPLAB toolbox (version 8.0; Lopez-Calderon and Luck, 2014). Here we provide a detailed conceptual description of the analysis procedures; further details of each data processing step are provided in comments within the online MATLAB scripts, which can be used to exactly replicate our analyses.

The event codes were shifted to account for the LCD monitor delay, and the EEG and EOG signals were downsampled to 256 Hz to increase data processing speeds (this decreased sampling rate is well within normative values for these paradigms). For analysis of the MMN, N2pc, N400, P3, LRP, and ERN, the EEG signals were referenced offline to the average of P9 and P10 (located adjacent to the mastoids). We find that P9 and P10 provide cleaner signals than the traditional mastoid sites, but the resulting waveforms are otherwise nearly identical to mastoid-referenced data; for analysis of the N170, the EEG signals were referenced to the average of all 33 sites (because the average reference is standard in the N170 literature). A bipolar HEOG signal was computed as left HEOG minus right HEOG. A bipolar VEOG signal was computed as lower VEOG minus FP2.

The DC offsets were removed, and the signals were high-pass filtered (non-causal Butterworth impulse response function, half-amplitude cut-off at 0.1 Hz, 12 dB/oct roll-off). In preparation for artifact correction, portions of EEG containing large muscle artifacts, extreme voltage offsets, or break periods longer than two seconds were identified by a semi-automatic ERPLAB algorithm and removed. Independent component analysis (ICA) was then performed, and components that were clearly associated with eyeblinks or horizontal eye movements—as assessed by visual inspection of the waveforms and the scalp distributions of the components—were removed (Jung, et al., 2000). Corrected bipolar HEOG and VEOG signals were computed from the ICA-corrected data. The original (pre-correction) bipolar HEOG and VEOG signals were also retained to provide a record of the ocular artifacts that were present in the original data.

The data were segmented and baseline-corrected for each trial using the time windows shown in Table 1. Channels with excessive levels of noise as determined by visual inspection of the data were interpolated using EEGLAB's spherical interpolation algorithm. Note that interpolation was not performed on any channels that were subsequently used to quantify the ERP components and only impacted the topographic maps shown in Supplementary Figure S2.

Segments of data containing artifacts that survived the correction procedure were flagged and excluded from analysis using automated ERPLAB procedures with individualized thresholds set on the basis of visual inspection of each participant's data (see justification for individualized thresholds in Luck (2014)). This included excluding trials with large voltage excursions in any channel. Because ICA does not always correct eye movements perfectly, especially in participants who rarely make eye movements, we also discarded any trials with evidence of large eye movements (greater than 4° of visual angle) in the corrected HEOG. Because horizontal eye movements are more likely to occur in tasks that present stimuli away from fixation, as in the visual search task we used to elicit the N2pc component, we also removed trials that contained horizontal eye movements larger than 0.2° of visual angle that could have impacted the N2pc (i.e., that occurred between −50 and 350 ms relative to stimulus onset). To ensure that blinks and horizontal eye movements did not interfere with perception of the visual stimuli, an additional procedure was performed for the tasks examining stimulus-locked responses to visual stimuli (i.e., N170, N2pc, N400, and P3). Specifically, we excluded trials on which an eyeblink or horizontal eye movement was present in the original (uncorrected) HEOG or VEOG signal during the presentation of the stimulus.

Trials with incorrect behavioral responses were excluded from all analyses, except for the ERN. Trials with excessively fast or slow reaction times (RTs) were also excluded from all analyses, with the acceptable RT range determined by visual inspection of RT probability histograms averaged across participants for each task. This resulted in an acceptable response window of 200 to 1000 ms after the onset of the stimulus in the N170, N2pc, P3, LRP, and ERN analyses, and an acceptable response window of 200 to 1500 ms after the onset of the target word in the N400 analysis; the MMN task required no responses and therefore no exclusions were made on the basis of RT. Accuracy was defined as the proportion of correct trials within the acceptable RT range prior to artifact rejection. Note

that future studies with these tasks can use these time ranges as a priori windows (assuming that a similar population is being tested).

2.5. Component Isolation with Difference Waves

The tasks were designed so that each ERP component could be isolated from overlapping brain activity by means of a difference waveform. Although difference waves have limitations in some contexts (Meyer et al., 2017), they are often valuable because they eliminate any brain activity that is in common to the two conditions, allowing precise assessment of the time course and magnitude of the small set of processes that differ between conditions (see Kappenman and Luck, 2012; Luck, 2014). A difference waveform was created for each component using the procedures specified in the Supplementary Materials and Methods. All amplitude and latency measurements were then performed on the difference waves using measurement procedures described later.

2.6. Quantification of Signals and Noise

2.6.1. Electrode Site Determination—For each component, we determined the electrode site at which that component was largest in the difference wave, and we used that site for all analyses. Data from the other electrode sites are provided in the online resource. Table 1 shows the electrode site chosen to quantify each component. The P3 showed similar amplitudes at CPz and Pz, and we chose to quantify the P3 at Pz because this is the site most widely used in the literature. We recommend the electrode sites in Table 1 as a priori measurement sites for future research using these tasks.

Ordinarily, it would be inappropriate to choose the site at which the effect is largest, because this "cherry picking" would bias the data in favor of the presence of an effect, inflating the Type I error rate (Luck and Gaspelin, 2017). However, the effects examined in the present study are already known to exist, and our goal was to characterize these effects rather than to test for their existence. Researchers often avoid cherry picking by averaging across a cluster of sites rather than choosing a single site. This may also improve data quality (Luck, 2014). However, the cluster approach may decrease the size of the effect, and we know of no formal analyses demonstrating which approach leads to the greatest statistical power. The ERP CORE data would provide an excellent test bed for assessing which approach is best across a range of components, but such an analysis is beyond the scope of the present paper. For the sake of simplicity, and to avoid arbitrary decisions about cluster sizes, the present analyses are based on the single electrode site with the largest amplitude; data from the other sites (see Supplementary Figure S1) are available in the online resource.

2.6.2. Time Window Determination.—The present study also provided an opportunity to determine optimal time periods for each component that can be used as a priori measurement windows in future studies. To accomplish this, we used the Mass Univariate Toolbox (Groppe et al., 2011) to find a cluster of statistically significant time points at the electrode site shown in Table 1. Our procedure began by comparing the mean voltage from the relevant difference wave to zero using a separate one-sample *t* test at each time point. The algorithm then found the largest cluster of consecutive time points that were individually significant and computed the mass of this cluster (the sum of the single-point *t*

values for the cluster). A permutation test with 2,500 iterations was then used to verify that each of these cluster masses was larger than the 95th percentile of values that would be expected by chance. The cluster obtained for each component greatly exceeded this 95th percentile (the lowest value was greater than the 99.99th percentile). The time ranges of these clusters are provided in Table 2.

Although the mass univariate approach does not provide strict control of the Type I error rate for individual time points (Groppe et al., 2011; Sassenhagen and Draschkow, 2019)—and using it to define time windows for subsequent analyses of the same data set would ordinarily be considered "double dipping" (Kriegeskorte et al., 2009)—our goal was simply to provide a set of empirically justified time windows for use in subsequent studies. We used the resulting windows to measure the amplitudes and latencies from the present study, which would generally be inappropriate. However, our goal in these analyses was not to determine whether real differences between conditions were present (because these differences have already been widely replicated). Instead, the goal of these analyses was to provide amplitude and latency values that can be used for comparison by future researchers. Note that, because the data and analysis scripts are available online, researchers can easily measure the amplitudes and latencies in the CORE data using other time windows.

In many studies, multiple components may be present in the same difference wave, and a narrower time window may help isolate the component of interest. We therefore performed an extensive literature review to determine the time windows that are most commonly used in analysis of these components. For each component, we then chose a final time window that (a) was within the range of the statistically significant cluster obtained from the mass univariate approach, and (b) was also within the range of commonly used values. The resulting time windows are shown in Table 2. As in the case of our procedure for determining the channel for measurement, this procedure for choosing time windows would not be appropriate in most studies. However, the present effects are known to exist, and our goal was to determine the best windows for future research. We recommend these as a priori time windows for future studies using these specific tasks or other similar tasks (assuming that a similar population is tested).

2.6.3. EEG Amplitude Spectrum Quantification.—We computed amplitude spectra from the EEG. Specifically, fast Fourier transforms were computed on zero-padded 5-s segments of the continuous high-pass filtered EEG with 50% overlap. The data were analyzed separately for each of the seven components using the data from the relevant task at the measurement site for that component. Break periods and segments containing large artifacts were excluded, and the amplitude spectra were averaged across segments and participants using the full sample (N = 40).

2.6.4. Noise Quantification.—Our first measure of noise focused on the baseline period (see Table 1) in the averaged ERP waveforms. To quantify the baseline noise, we calculated the point-by-point SD of the voltage across the baseline period in the averaged ERP waveform for each component, separately for each participant. Low-pass filtering is often but not always applied to ERPs prior to measuring components (see below), so we performed these calculations before and after applying a 20 Hz low-pass filter (non-causal

Butterworth impulse response function, half-amplitude cut-off at 20 Hz, 48 dB/oct roll-off). We quantified the baseline noise in both the parent waveforms and the difference waves. Note that some systematic activity was present during the baseline periods of the parent waveforms (reflecting preparatory activity and/or overlapping activity from the previous trial), and this systematic activity also contributes to this measure of baseline noise. However, given our experimental designs, all systematic variation was eliminated during the baseline period in the difference waves, so our baseline noise measure is a pure measure of noise for the difference waves.

Our second measure of noise was analogous to the baseline noise measure, but applied during the measurement period for each component. The point-by-point SD during the measurement window would ordinarily reflect both the signal and the noise, so we used plus-minus averaging (Schimmel, 1967) to eliminate the signal but retain the noise. This method eliminates the signal by inverting the polarity of half the single-trial EEG epochs prior to averaging. Specifically, plus-minus averages were computed by dividing the artifact-free EEG epochs for each participant into even-numbered trials and odd-numbered trials separately for each condition, and then applying an algorithm that was mathematically equivalent to inverting the polarity of the even-numbered trials and then averaging the odd-numbered trials together with the inverted even-numbered trials. To quantify the noise in the plus-minus averages, we calculated the standard deviation (SD) of the plus-minus waveforms for each participant during the measurement window for each component. Note that this measure of noise includes trial-by-trial variability in the signal of interest, whereas the baseline noise measure does not. Thus, two these measures of noise provide complementary information.

Our final measure of noise used the plus-minus averages to estimate the amount of noise at each individual time point in the waveforms. In the absence of noise, the plus-minus average for a given participant would be zero at all time points, and any deviation from zero reflects variability (noise) in the data. However, the polarity of this deviation is random, so one cannot quantify the noise by simply taking the mean across participants (which would have an expected value of zero). Instead, we computed the SD across participants at each time point. This yielded a waveform showing noise level at each time point for each of the ERP components.

2.6.5. Amplitude and Latency Measures.—Using the electrodes shown in Table 1 and the time windows shown in Table 2, we measured the mean amplitude, peak amplitude, peak latency, and 50% area latency (the time point that divides the area under the curve into subregions of equal area) from each of the seven difference waves. We also quantified the onset latency of each difference wave with the fractional peak latency measure (the time at which the voltage reaches 50% of the peak amplitude; see Kiesel et al., 2008); the measurement windows were shifted 100 ms earlier for these onset measurements. As secondary analyses, we also quantified the mean amplitude of the individual parent waveforms that were used to create the difference waveforms. The other measurements are typically valid only for difference waveforms (see Luck, 2014), so these measures were not obtained from the parent waveforms. For example, it would be difficult to measure the peak latency of the MMN from the parent waveforms given that many other components are also

present in the same time window. Because peak amplitude, peak latency, and fractional peak latency are all sensitive to high-frequency noise, a low-pass filter (non-causal Butterworth impulse response function, half-amplitude cut-off at 20 Hz, 12 dB/oct roll-off) was applied to the waveforms prior to obtaining these measures.

3. Results

Figure 2 shows the grand average "parent" ERP waveforms from each relevant condition and the ERP difference waveforms between conditions (including the standard error of the mean across participants). Behavioral data are summarized in Supplementary Table S1 in the supplementary materials. The experiment control scripts, raw data, analysis scripts, processed data, and a broad set of results are available in the online resource (https://doi.org/10.18115/D5JW4R).

3.1 Basic ERP Effects

The ERP waveforms in Figure 2 show that the expected experimental effects were observed for all seven components. The N170 component was larger and earlier for faces than for cars, with a maximum effect over right inferotemporal cortex; the latency difference may seem surprising, but it is commonly observed (Carmel, 2002; Itier, 2004; Sagiv and Bentin, 2001). Auditory deviants in the MMN paradigm elicited a negativity peaking near 200 ms with a maximum over medial frontocentral cortex. The N2pc paradigm yielded a more negative (less positive) voltage from 200-300 ms over the hemisphere contralateral to the target compared to the ipsilateral hemisphere. In the N400 paradigm, the second word in a pair elicited a larger negativity from approximately 200-600 ms when it was semantically unrelated to the first word than when it was related, with a maximum over medial centroparietal cortex. The visual P3 paradigm yielded a larger positive voltage for the rare stimulus category than for the frequent stimulus category, with a maximum effect at the parietal midline electrode. In the flankers paradigm, both the lateralized readiness potential (LRP) and the error-related negativity (ERN) were observed. The LRP was a more negative voltage over the motor cortex contralateral to the response hand compared to the ipsilateral side, and the ERN was a more negative voltage over midline frontocentral cortex for incorrect responses compared to correct responses. Topographic maps of each effect are provided in Supplementary Figure S2.

3.2 ERP Quantification

We quantified the magnitude and timing of each component in several different ways. All measurements were performed on the difference waveforms, which was necessary for components that are only well defined in difference waves (e.g., the N2pc and LRP components). Mean amplitude measures for the parent waveforms are provided in Supplementary Table S2. The time windows used for measuring each component are listed in Table 2, and the procedures used for determining the time windows are described in the Materials and Methods. The mean values for each component are provided in Table 3, along with standard deviations (SDs) to quantify the variability across participants. Histograms of the single-participant values are provided in Supplementary Figure S3 so that future studies can compare their single-participant values to the range observed in our data. The range, first

and third quartiles, and the interquartile range are provided in Supplementary Table S3. As shown in Table 3 and Supplementary Figure S3, the size and timing of the ERP effects varied widely across participants, even within this relatively homogenous sample of neurotypical young adults.

We quantified the magnitude of the components using mean amplitude and peak amplitude. The effect sizes are provided in Table 3 (i.e., Cohen's d_z for a one-sample comparison against zero) to aid future researchers in performing a priori statistical power calculations. Both the mean amplitude and peak amplitude measures showed very large effect sizes ($d_z >$ 1) for all components, demonstrating the robustness of the CORE paradigms. Although the observed effect sizes appear larger for peak than mean amplitude, it should be noted that using 0 µV as the chance value would be expected to overestimate the true effect size for peak amplitude, because peaks are statistically biased away from zero (Luck, 2014). Future research is needed to establish the ideal method for quantifying effect sizes for peak amplitude measures.

We quantified the midpoint latency of each effect using the peak latency and the 50% area latency techniques. Peak latency is much more widely used to quantify midpoint latency, but 50% area latency has several advantages (although it is mainly useful when a component is measured from a difference wave; see Luck, 2014). For most components, there was close agreement between the means of these two measures, but the standard deviations were substantially lower for the 50% area latency measure. We also calculated the onset latency[1] of each effect, using the fractional peak latency technique (see Kiesel et al., 2008). Future research could use the ERP CORE data to compare the effect sizes yielded by different algorithms for quantifying ERP amplitudes and latencies.

3.3 EEG Spectral Quantification

Researchers who use the ERP CORE paradigms may wish to compare the frequency-domain characteristics of their data with the CORE data (e.g., to compare alpha-band activity as an index of attentional engagement). We therefore computed the amplitude density of the EEG signal at each frequency ranging from 1 to 100 Hz (see Materials and Methods for details). Figure 3 (left panel) shows the amplitude spectra at the electrode site of interest for each component, averaged across all 40 participants. As usual, the spectra exhibited a gradual falloff as the frequency increased (approximately $1/f$) along with a peak in the alpha band at posterior scalp sites and a small spike at 60 Hz reflecting electrical noise.

3.4 Noise Quantification

In addition to measuring the components (as is standard in ERP studies), we also quantified the noise level of the data (which is not typically provided in ERP papers). The point-by-point standard errors shown in Figure 2 do not provide a good measure of the noise level, because they reflect a combination of EEG noise and true individual differences among

---

[1]Onset latencies can be distorted by low-pass filters (Luck, 2014; Rousselet, 2012), which spread the signal in time (symmetrically for noncausal filters and asymmetrically toward longer latencies for causal filters). However, the measure of onset latency used in the present study is only minimally affected by low-pass filtering (see Figure 12.8C in Luck, 2014). Extreme high-pass filters can also impact onset latencies, but the mild 0.1 Hz cutoff frequency used here has very little impact (Rousselet, 2012, Tanner et al., 2015).

participants. We therefore used three other methods to quantify noise (a detailed justification and explanation of each method is provided in the Materials and Methods). Note that noise is typically defined relative to the signal of interest, and here we define *noise* as any source of variability that impacts the averaged ERP waveforms, even if some of that variability reflects neural signals that might be the focus of study in other contexts (e.g., phase-random alpha-band oscillations).

First, we quantified the noise in the baseline of each averaged ERP waveform by calculating the point-to-point variability (SD) of the voltage across the baseline period listed in Table 1 (separately for each participant). This provides an overall measure of the amount of noise that survives averaging, without influence from the actual ERP signals. Because many researchers would apply a low-pass filter prior to measuring component amplitudes or latencies, we performed this analysis both before and after applying a 20 Hz low-pass filter. Table 4 provides the mean of the baseline noise values across participants, and Figure 3 (middle panel) provides probability histograms to show the range of single-participant noise levels in the baseline of the unfiltered averaged ERP waveforms. The range, first and third quartiles, and the interquartile range in the unfiltered ERP waveforms are provided in Supplementary Table S4. Interestingly, the baseline noise levels varied markedly across participants for some of the components. The baseline noise was only slightly reduced in the filtered waveforms, reflecting the fact that most of the EEG energy was below 20 Hz (see Figure 3, left panel). As expected, the noise level was inversely related to the number of trials being averaged together (for the parent waveforms; e.g., higher for the rare than for the frequent stimuli in the P3 paradigm) and was typically larger for the difference waveforms than for the parent waveforms.

Second, we quantified the noise during the measurement time window of each component. Because the SD across time points in the measurement window would include the signal as well as the noise, we performed plus-minus averaging (Schimmel, 1967) before measuring the SD. This procedure inverts the EEG for half of the trials prior to averaging, which eliminates the signal while retaining the noise (including any variability in the signal itself). The plus-minus averages are shown in Supplementary Figure S4. Table 4 provides the mean noise level across participants during the measurement window for each component, and Figure 3 (right panel) provides probability histograms to show the range of single-participant noise levels in the measurement window. The range, first and third quartiles, and the interquartile range in the unfiltered ERP waveforms are provided in Supplementary Table S5.

Finally, we estimated the noise level at each individual time point in the waveform, quantified as the SD across participants at that time point in the plus-minus averages. These point-by-point estimates of noise are shown in Figure 4. For several components (e.g., P3, N400, MMN), the noise tended to increase over the course of the epoch. For the ERN waveform, noise increased dramatically following the execution of the response (see also the bottom row in Supplementary Figure S4).

Together, these three measures of noise will allow researchers to determine how their noise levels compare to the noise levels in our data. Researchers who are developing new signal

processing procedures can also apply these procedures to the ERP CORE data and assess their impact on the noise level. In addition, our data analysis scripts can be used as a starting point for computing these metrics of noise in other paradigms.

4. Discussion

The ERP CORE—a Compendium of Open Resources and Experiments for human ERP research—consists of optimized paradigms, experiment control scripts, EEG data, analysis pipelines, data processing scripts, and a broad set of results. All materials are freely available at https://doi.org/10.18115/D5JW4R. Despite the widespread use of ERPs in the fields of psychology, psychiatry, and neuroscience, this resource is the first of its kind.

Each paradigm in the ERP CORE successfully elicited the component of interest with a large effect size in a 10-minute recording. Moreover, we provided several measures of the magnitude and timing of the ERP components, noise levels in the data, and variability across participants. These results will serve as a guide to researchers in selecting task and analysis parameters in future studies, and they provide a standard against which future data sets can be compared. We are not claiming that our data quality is optimal or that our effect sizes can be used as formal norms for the broader population. However, no other standard of this nature currently exists, and the ERP CORE provides a starting point for the development of new standards based on larger samples and other subject populations.

We have provided our experiment control and data processing scripts so that other researchers can utilize our tasks and analysis pipelines in their research, saving time and increasing reproducibility. Researchers can also perform quality control by comparing their noise levels with those presented here. ERP studies do not ordinarily present a detailed assessment of noise levels, and the noise quantifications provided here are therefore a unique resource. By providing these data quality measures on our own data and sharing our analysis scripts, we hope to encourage other researchers to provide similar metrics in their publications so that the field can see which experimental procedures and data analysis methods yield the best data quality. Broad adoption of these noise metrics in EEG/ERP publications would also over time allow for field standards regarding acceptable levels of noise to be established.

We also provided several measures of variability across participants, and all of our single-participant data and measures are available online. Although we tested a relatively homogeneous sample of neurotypical young adults, substantial variability was observed (see Supplementary Figure S3). Although the existence of individual differences in ERP waveforms is well known (Kappenman and Luck, 2012), publications typically do not provide such a detailed characterization of the variability in their samples. The present results therefore provide a useful comparison point for researchers who are evaluating variability in their own single-participant ERP waveforms.

The ERP CORE contains data from several ERP paradigms in the same participants, which is quite rare given that a full recording session is typically required to yield a single component. However, the CORE paradigms were designed to produce optimal results with

minimal recording time. Consequently, the ERP CORE is the first published data set that includes such a comprehensive set of ERP components in the same participants. This resource can therefore be used to examine a broad range of empirical and methodological questions, such as assessing the effectiveness of a new signal processing method across multiple components and paradigms. The ERP CORE tasks could also serve as an electrophysiological test battery.

It is also worth noting that our data set contains many other components beyond those analyzed here, including auditory and visual sensory responses, the stimulus-locked LRP, the response-locked P3, and the anterior N2 component that is sensitive to response competition in the flankers task (Kopp et al., 1996; Purmann et al., 2011).

In many cases, researchers will be able to apply our validated paradigms and analysis pipelines directly to their own experiments, which will save time, reduce errors, and decrease researcher degrees of freedom in statistical analyses (Simmons et al., 2011). However, the task and analysis parameters provided in the present study were validated with neurotypical young adult participants and may need to be adjusted for other populations. Nonetheless, these parameters provide a useful starting point for researchers interested in any of the neurocognitive processes encompassed by the ERP CORE.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Boudewyn MA, Luck SJ, Farrens JL, Kappenman ES, 2018. How many trials does it take to get a significant ERP effect? It depends. Psychophysiology. 55, e13049. 10.1111/psyp.13049 [PubMed: 29266241]

Bukach CM, Stewart K, Couperus JW, Reed CL, 2019. Using Collaborative Models to Overcome Obstacles to Undergraduate Publication in Cognitive Neuroscience. Front. Psychol. 10, 549. [PubMed: 30949093]

Buzsáki G, Anastassiou CA, Koch C, 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–420. 10.1038/nrn3241 [PubMed: 22595786]

Carmel D, 2002. Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition. 83, 1–29. 10.1016/S0010-0277(01)00162-7 [PubMed: 11814484]

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

van Dinteren R, Arns M, Jongsma MLA, Kessels RPC, 2014. P300 Development across the Lifespan: A Systematic Review and Meta-Analysis. PLoS One. 9, e87347. 10.1371/journal.pone.0087347 [PubMed: 24551055]

Eimer M, 2011. The Face-Sensitive N170 Component of the Event-Related Brain Potential. Oxford University Press. 10.1093/oxfordhb/9780199559053.013.0017

Eimer M, Coles MGH, 2003. The Lateralized Readiness Potential, in: Jahanshahi M, Hallett M (Eds.), The Bereitschaftspotential: Movement-Related Cortical Potentials. Springer US, Boston, MA, pp. 229–248. 10.1007/978-1-4615-0189-3_14

Eriksen BA, Eriksen CW, 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16, 143–149.

Feuerriegel D, Churches O, Hofmann J, Keage HAD, 2015. The N170 and face perception in psychiatric and neurological disorders: A systematic review. Clin. Neurophysiol. 126, 1141–1158. 10.1016/j.clinph.2014.09.015 [PubMed: 25306210]

Garrido MI, Kilner JM, Stephan KE, Friston KJ, 2009. The mismatch negativity: A review of underlying mechanisms. Clin. Neurophysiol. 120, 453–463. [PubMed: 19181570]

Gehring WJ, Liu Y, Orr JM, Carp J, 2012. The Error-Related Negativity (ERN/Ne), in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 231–292. 10.1093/oxfordhb/9780195374148.013.0120

Groppe DM, Urbach TP, Kutas M, 2011. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology. 48, 1711–1725. 10.1111/j.1469-8986.2011.01273.x [PubMed: 21895683]

Holcomb PJ, Neville HJ, 1990. Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-related Brain Potentials. Lang. Cogn. Process 5, 281–312.

Itier RJ, 2004. N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs. Cereb. Cortex. 14, 132–142. [PubMed: 14704210]

Jackson AF, Bolger DJ, 2014. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us: Neurophysiological bases of EEG. Psychophysiology 51, 1061–1071. [PubMed: 25039563]

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

Kappenman ES, Luck SJ, 2012. ERP components: The ups and downs of brainwave recordings, in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 3–30. 10.1093/oxfordhb/9780195374148.013.0014

Keil A, Debener S, Gratton G, Junghöfer M, Kappenman ES, Luck SJ, Luu P, Miller GA, Yee CM, 2014. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography: Guidelines for EEG and MEG. Psychophysiology 51, 1–21. 10.1111/psyp.12147 [PubMed: 24147581]

Kiesel A, Miller J, Jolicœur P, Brisson B, 2008. Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. Psychophysiology 45, 250–274. 10.1111/j.1469-8986.2007.00618.x [PubMed: 17995913]

Kopp B, Rist F, Mattler U, 1996. N200 in the flanker task as a neurobehavioral tool for investigating executive control. Psychophysiology. 33, 282–294. 10.1111/j.1469-8986.1996.tb00425.x [PubMed: 8936397]

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI, 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540. [PubMed: 19396166]

Kutas M, Federmeier KD, 2011. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annu. Rev. Psychol. 62, 621–647. 10.1146/annurev.psych.093008.131123 [PubMed: 20809790]

Lau EF, Phillips C, Poeppel D, 2008. A cortical network for semantics: (de)constructing the N400. Nat Rev Neurosci. 9, 920–933. 10.1038/nrn2532 [PubMed: 19020511]

Lopez-Calderon J, Luck SJ, 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8, 213. [PubMed: 24782741]

Luck SJ, 2014. An Introduction to the Event-Related Potential Technique. The MIT Press, Cambridge.

Luck SJ, 2012. Electrophysiological Correlates of the Focusing of Attention within Complex Visual Scenes: N2pc and Related ERP Components, in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb.

Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 329–360. 10.1093/oxfordhb/9780195374148.013.0161

Luck SJ, Fuller RL, Braun EL, Robinson B, Summerfelt A, Gold JM, 2006. The speed of visual attention in schizophrenia: Electrophysiological and behavioral evidence. Schizophr. Res. 85, 174–195. 10.1016/j.schres.2006.03.040 [PubMed: 16713184]

Luck SJ, Gaspelin N, 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology. 54, 146–157. 10.1111/psyp.12639 [PubMed: 28000253]

Luck SJ, Kappenman ES, 2012. Oxford handbook of event-related potential components. Oxford University Press, Oxford. 10.1093/oxfordhb/9780195374148.001.0001

Luck SJ, Kappenman ES, Fuller RL, Robinson B, Summerfelt A, Gold JM, 2009. Impaired response selection in schizophrenia: Evidence from the P3 wave and the lateralized readiness potential. Psychophysiology. 46, 776–786. 10.1111/j.1469-8986.2009.00817.x [PubMed: 19386044]

Meyer A, Lerner MD, De Los Reyes A, Laird RD, Hajcak G, 2017. Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches: Difference scores and ERPs. Psychophysiology. 54, 114–122. 10.1111/psyp.12664 [PubMed: 28000251]

Näätänen R & Kreegipuu K, 2012. The Mismatch Negativity (MMN). in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 143–157.

Näätänen R, Pakarinen S, Rinne T & Takegata R, 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiol. 115, 140–144. [PubMed: 14706481]

Olvet D, Hajcak G, 2008. The error-related negativity (ERN) and psychopathology: Toward an endophenotype. Clin. Psychol. Rev. 28, 1343–1354. 10.1016/j.cpr.2008.07.003 [PubMed: 18694617]

Peirce J, 2019. PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods 51, 195–203. [PubMed: 30734206]

Pernet CR, Garrido M, Gramfort A, Maurits N, Michel C, Pang E, … Puce A (2018, 8 9). Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG. 10.31219/osf.io/a8dhx

Polich J 2007. Updating P300: An integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. [PubMed: 17573239]

Polich J, 2012. Neuropsychology of P300. in Oxf, in: Luck SJ, Kappenman ES (Eds.), Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 159–188. 10.1093/oxfordhb/9780195374148.013.0089

Purmann S, Badde S, Luna-Rodriguez A, Wendt M, 2011. Adaptation to Frequent Conflict in the Eriksen Flanker Task: An ERP Study. J. Psychophysiol. 25, 50–59. 10.1027/0269-8803/a000041

Ritter W, Vaughan HG, 1969. Averaged Evoked Responses in Vigilance and Discrimination: A Reassessment. Science. 164, 326–328. 10.1126/science.164.3877.326 [PubMed: 5776645]

Rossion B, Caharel S, 2011. ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. Vision Res. 51, 1297–1311. [PubMed: 21549144]

Rossion B, Jacques C, 2012. The N170: Understanding the Time Course of Face Perception in the Human Brain, in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 115–141. 10.1093/oxfordhb/9780195374148.013.0064

Rousselet GA (2012). Does Filtering Preclude Us from Studying ERP Time-Courses? Frontiers in Psychology, 3, 131. 10.3389/fpsyg.2012.00131 [PubMed: 22586415]

Sagiv N, Bentin S, 2001. Structural Encoding of Human and Schematic Faces: Holistic and Part-Based Processes. J. Cogn. Neurosci. 13, 937–951. 10.1162/089892901753165854 [PubMed: 11595097]

Sassenhagen J, Draschkow D, 2019. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology. 56, e13335. [PubMed: 30657176]

Schimmel H, 1967. The (±) Reference: Accuracy of Estimated Mean Components in Average Response Studies. Science. 157, 92–94. [PubMed: 6026675]

Simmons JP, Nelson LD, Simonsohn U, 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychol. Sci. 22, 1359–1366. [PubMed: 22006061]

Smulders FTY, Miller JO, 2012. The Lateralized Readiness Potential, in: Luck SJ, Kappenman ES (Eds.), Oxf. Handb. Event-Relat. Potential Compon. Oxford University Press, Oxford, pp. 209–230. 10.1093/oxfordhb/9780195374148.013.0115

Swaab TY, Ledoux K, Camblin CC, Boudewyn MAL-RERPC in O.H., 2012. Event-Relat. Potential Compon. Oxford University Press, Oxford. 10.1093/oxfordhb/9780195374148.013.0197

Tanner D, Morgan-Short K, & Luck SJ (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology, 52, 997–1009. [PubMed: 25903295]
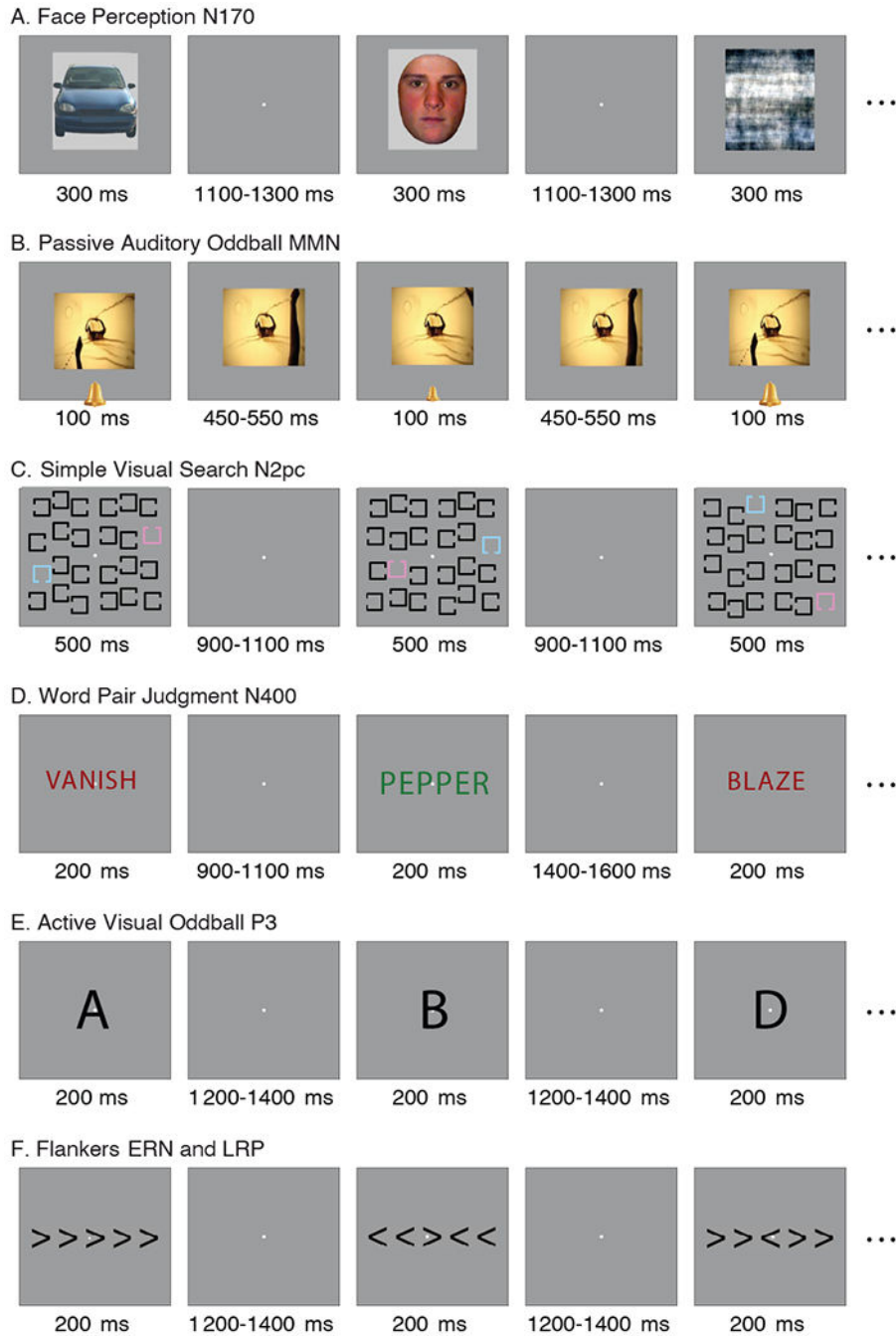
Figure 1.

Examples of a subset of the trials in each of the six tasks. The stimuli are not drawn to scale; see Supplementary Materials and Methods for actual sizes. (A) Face perception task used to elicit the N170. On each trial, an image of a face, car, scrambled face, or scrambled car was presented in the center of the screen, and participants indicated whether a given stimulus was an "object" (face or car) or a "texture" (scrambled face or scrambled car). (B) Passive auditory oddball task used to elicit the mismatch negativity (MMN). Standard tones (80 dB, p = .8) and deviant tones (70 dB, p = .2) were presented over speakers while participants

watched a silent video and ignored the tones. (C) Simple visual search task used to elicit the N2pc. Either pink or blue was designated the target color at the beginning of a trial block, and participants indicated whether the gap in the target color square was on the top or bottom. (D) Word pair judgment task used to elicit the N400. Each trial consisted of a red prime word followed by a green target word, and participants indicated whether the target word was semantically related or unrelated to the prime word. (E) Active visual oddball task used to elicit the P3. The letters A, B, C, D, and E were presented in random order (p = .2 for each letter). One of the letters was designated the target for a given block of trials, and participants indicated whether each stimulus was the target or a non-target for that block. Thus, the probability of the target category was .2, but the same physical stimulus served as a target in some blocks and a nontarget in others. (F) Flankers task used to elicit the lateralized readiness potential (LRP) and the error-related negativity (ERN). The central arrowhead was the target, and it was flanked on both sides by arrowheads that pointed in the same direction (congruent trials) or the opposite direction (incongruent trials). Participants indicated the direction of the target arrowhead on each trial with a left- or right-hand buttonpress.
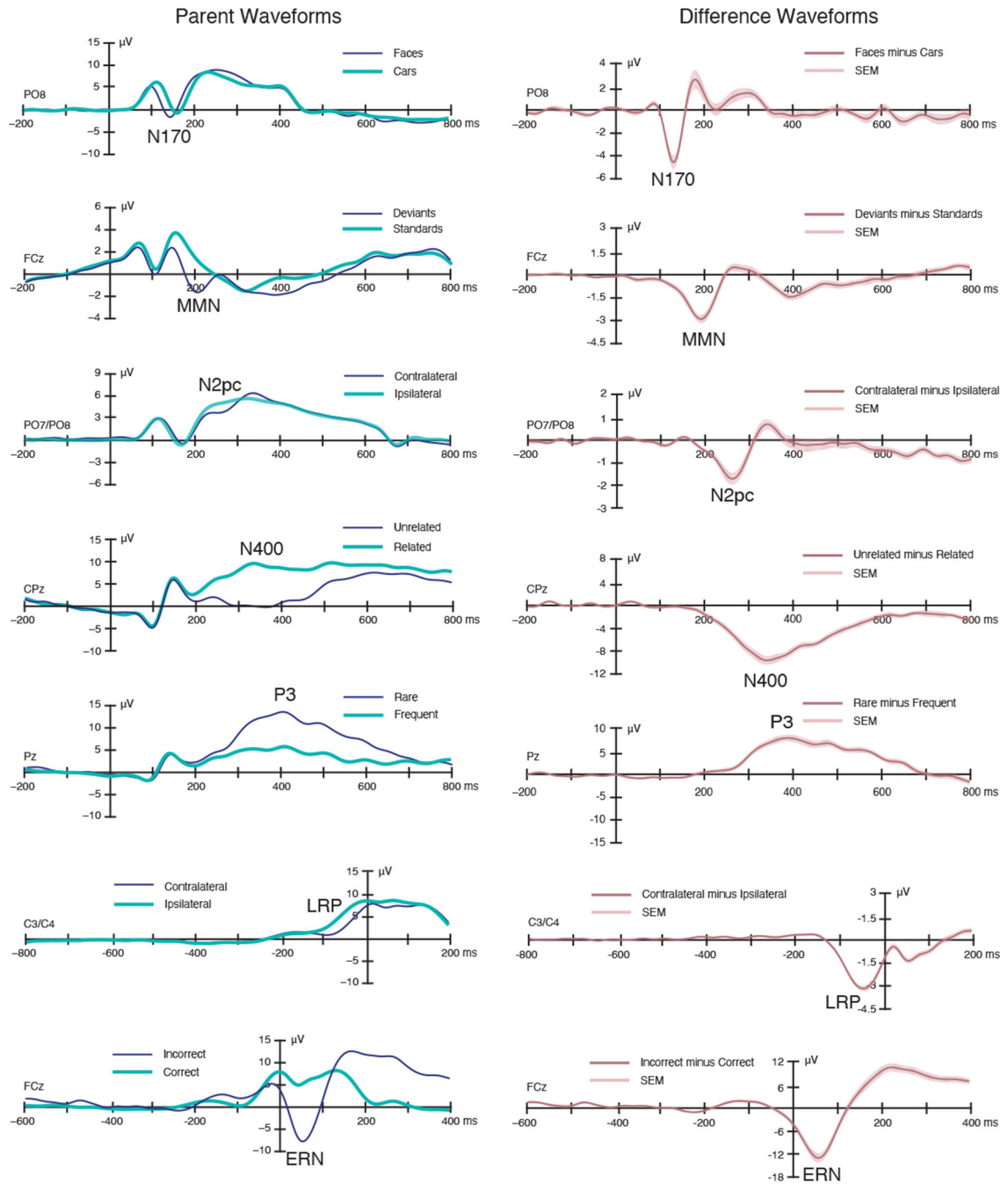
Figure 2.
Grand average parent ERP waveforms (left) and difference waveforms (right). The shading surrounding the difference waveforms indicates the region that fell within ±1 SEM at a given time point (which reflects both measurement error and true differences among participants). A digital low-pass filter was applied offline before plotting the ERP waveforms (Butterworth impulse response function, half-amplitude cutoff at 20 Hz, 48 dB/oct roll-off).
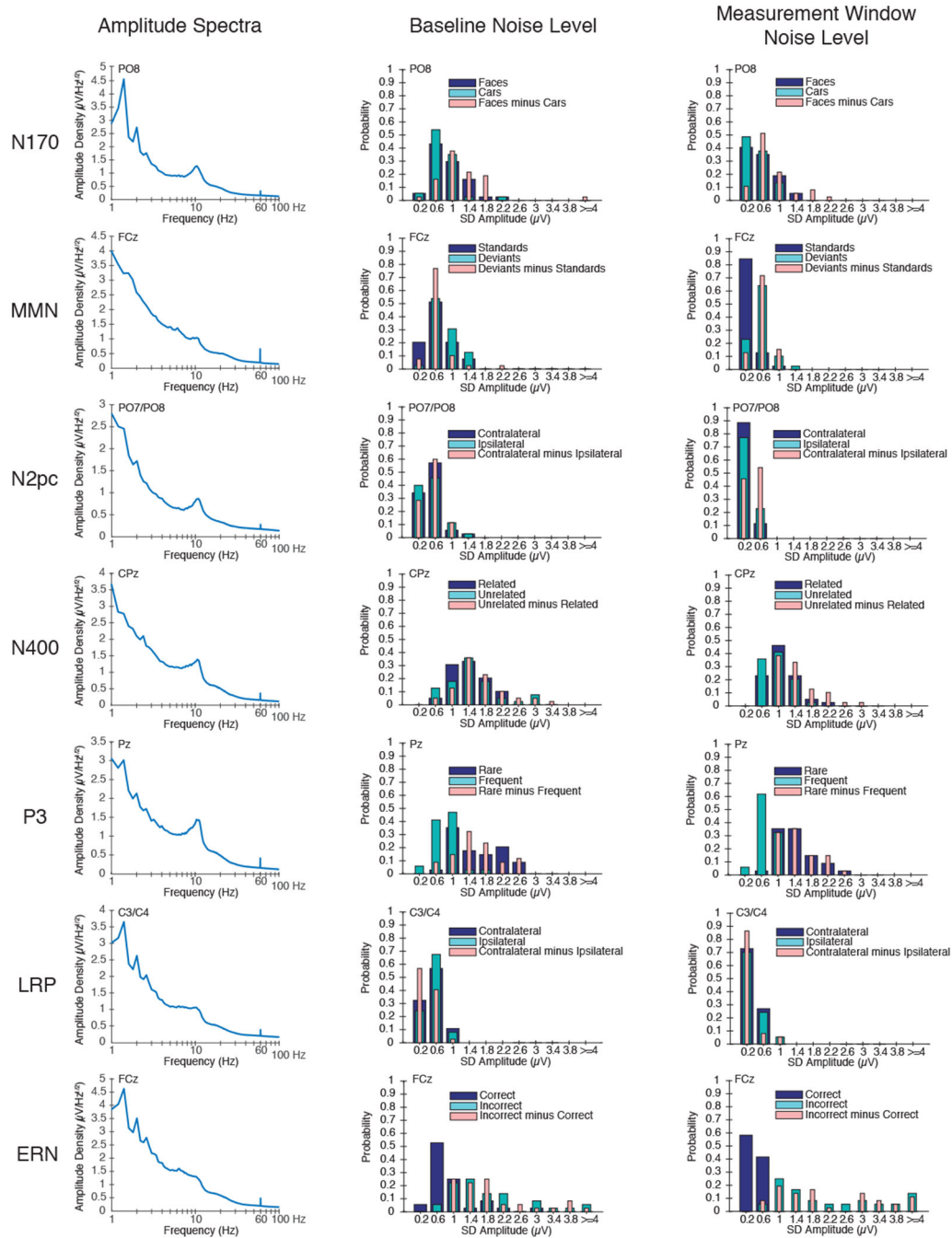
Figure 3.
Quantification of the EEG signal and the ERP noise. (Left) Amplitude density as a function of frequency (on a log scale) for each ERP component at the electrode site where that component was maximal, calculated from individual participants and then averaged. Note that, although LRP and ERN were isolated in the same task, the spectra were obtained at different electrode sites and therefore differ slightly. (Middle) Probability histograms of the noise levels during the baseline period for the averaged ERP parent waveforms and difference waveforms. Bins are 0.4 µV in width, and the x-axis indicates the midpoint value

for each bin. (Right) Probability histograms of the noise levels during the measurement time window of the plus-minus average parent waveforms and difference waveforms. Bins are 0.4 μV in width, and the x-axis indicates the midpoint value for each bin.

Figure 4.
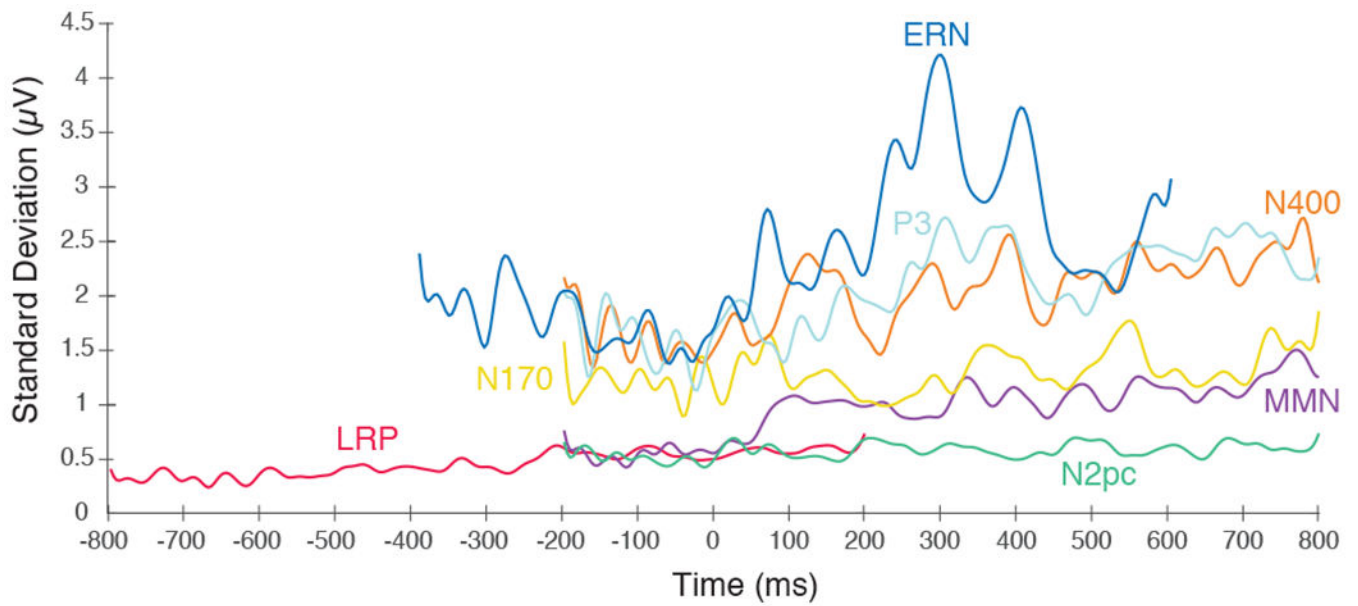Noise level at each time point for each component at the electrode site where that component was maximal, measured as the standard deviation across participants of the plus-minus ERP difference waveforms at a given time point. Time zero represents the time-locking point for each ERP component (i.e., the onset of the stimulus for the N170, MMN, N2pc, N400, and P3, and the buttonpress for the LRP and ERN).

Table 1

Sample size (after excluding subjects with too many artifacts), electrode site, time-locking event, epoch window, and baseline period for each component

ERP Component	Sample Size (N)	Electrode Site	Time-Locking Event	Epoch Window (ms)	Baseline Period (ms)
N170	37	PO8	Stimulus-locked	−200 to 800	−200 to 0
MMN	39	FCz	Stimulus-locked	−200 to 800	−200 to 0
N2pc	35	PO7/PO8	Stimulus-locked	−200 to 800	−200 to 0
N400	39	CPz	Stimulus-locked	−200 to 800	−200 to 0
P3	34	Pz	Stimulus-locked	−200 to 800	−200 to 0
LRP	37	C3/C4	Response-locked	−800 to 200	−800 to −600
ERN	36	FCz	Response-locked	−600 to 400	−400 to −200

Table 2

Time windows of statistically significant differences (based on the mass univariate approach) and recommended measurement windows, relative to stimulus onset (for N170, MMN, N2pc, N400, and P3) or response onset (for LRP and ERN).

ERP Component	Statistically Significant Time Windows (ms)	Recommended Measurement Windows (ms)
N170	105.47 to 148.44	110 to 150
MMN	113.28 to 230.47	125 to 225
N2pc	191.41 to 292.97	200 to 275
N400	183.59 to 750.00	300 to 500
P3	253.91 to 664.06	300 to 600
LRP	−125.00 to 15.63	−100 to 0
ERN	−27.34 to 109.38	0 to 100

Table 3

ERP difference waveform measures, averaged across participants (standard deviations in parentheses), along with effect size (Cohen's d$_z$) of the difference in amplitude from 0 μV.

ERP Component	Mean Amplitude (μV)	Peak Amplitude (μV)	Peak Latency (ms)	50% Area Latency (ms)	Onset Latency (ms)
N170	−3.37 (2.71) $d_z = 1.24$	−5.52 (3.32) $d_z = 1.66$	131.44 (12.56)	131.84 (8.58)	95.76 (29.05)
MMN	−1.86 (1.22) $d_z = 1.52$	−3.46 (1.71) $d_z = 2.02$	187.60 (19.13)	185.20 (14.44)	146.94 (30.33)
N2pc	−1.14 (1.15) $d_z = 1.00$	−1.86 (1.60) $d_z = 1.16$	253.24 (18.51)	246.43 (8.81)	213.63 (30.85)
N400	−7.61 (3.27) $d_z = 2.33$	−11.04 (4.65) $d_z = 2.38$	370.09 (49.43)	387.72 (17.85)	284.86 (44.92)
P3	6.29 (3.39) $d_z = 1.86$	10.15 (4.53) $d_z = 2.24$	408.89 (70.48)	436.47 (32.77)	327.44 (61.98)
LRP	−2.40 (0.94) $d_z = 2.56$	−3.41 (1.15) $d_z = 2.97$	−49.94 (15.66)	−49.09 (9.72)	−96.50 (19.85)
ERN	−9.26 (5.90) $d_z = 1.57$	−13.86 (7.01) $d_z = 1.98$	54.47 (12.63)	54.36 (9.99)	2.50 (27.53)

Table 4

Average standard deviation (SD) of the voltage in the baseline time window of the ERP waveform and in the measurement time window of the plus-minus averages

		Baseline Time Window		Measurement Time Window	
ERP Component	**Trial Type**	**Unfiltered SD Amplitude (μV)**	**Filtered SD Amplitude (μV)**	**Unfiltered SD Amplitude (μV)**	**Filtered SD Amplitude (μV)**
	Faces	0.937	0.824	0.595	0.411
N170	Cars	0.799	0.677	0.489	0.306
	Faces-Cars	1.214	1.044	0.833	0.590
	Deviants	0.811	0.743	0.555	0.452
MMN	Standards	0.664	0.639	0.315	0.254
	Deviants-Standards	0.684	0.577	0.616	0.496
	Contralateral	0.500	0.417	0.316	0.200
N2pc	Ipsilateral	0.542	0.449	0.304	0.191
	Contralateral-Ipsilateral	0.559	0.446	0.410	0.281
	Unrelated	1.505	1.394	0.945	0.811
N400	Related	1.423	1.311	1.051	0.912
	Unrelated-Related	1.662	1.468	1.414	1.221
	Rare	1.570	1.415	1.383	1.238
P3	Frequent	0.823	0.750	0.689	0.622
	Rare-Frequent	1.566	1.378	1.516	1.362
	Contralateral	0.507	0.420	0.378	0.272
LRP	Ipsilateral	0.524	0.439	0.387	0.279
	Contralateral-Ipsilateral	0.388	0.291	0.324	0.212
	Incorrect	1.848	1.625	2.250	2.055
ERN	Correct	0.918	0.884	0.375	0.290
	Incorrect-Correct	1.909	1.662	2.264	2.050