ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Optimal filters for ERP research II: Recommended settings for seven common ERP components

Guanghui Zhang | **David R. Garrett** | **Steven J. Luck**

Center for Mind & Brain, University of California-Davis, Davis, California, USA

Correspondence
Guanghui Zhang, Center for Mind & Brain, University of California-Davis, 267 Cousteau Place, Davis, CA 95618, USA.
Email: zhang.guanghui@foxmail.com

Abstract

In research with event-related potentials (ERPs), aggressive filters can substantially improve the signal-to-noise ratio and maximize statistical power, but they can also produce significant waveform distortion. Although this tradeoff has been well documented, the field lacks recommendations for filter cutoffs that quantitatively address both of these competing considerations. To fill this gap, we quantified the effects of a broad range of low-pass filter and high-pass filter cutoffs for seven common ERP components (P3b, N400, N170, N2pc, mismatch negativity, error-related negativity, and lateralized readiness potential) recorded from a set of neurotypical young adults. We also examined four common scoring methods (mean amplitude, peak amplitude, peak latency, and 50% area latency). For each combination of component and scoring methods, we quantified the effects of filtering on data quality (noise level and signal-to-noise ratio) and waveform distortion. This led to recommendations for optimal low-pass and high-pass filter cutoffs. We repeated the analyses after adding artificial noise to provide recommendations for data sets with moderately greater noise levels. For researchers who are analyzing data with similar ERP components, noise levels, and participant populations, using the recommended filter settings should lead to improved data quality and statistical power without creating problematic waveform distortion.

KEYWORDS
ERP, noise level, optimal filtering, scoring methods, SNR, wave distortion

1 | INTRODUCTION

Almost all ERP studies involve filtering, and optimal filter settings are important to maximize statistical power and avoid waveform distortions that might lead to erroneous conclusions (Acunzo et al., 2012; Rousselet, 2012; Tanner et al., 2015, 2016; Van Driel et al., 2021; VanRullen, 2011; Widmann et al., 2015; Yeung et al., 2007). However, there is no widespread consensus about what filter settings are optimal, and the optimal settings are likely to vary across scoring methods, experimental paradigms, subject populations, and recording setups.

Filter settings vary widely across published studies, even studies within the same research domain. For example, in a review of emotional face processing studies, Schindler and Bublatzky (2020) found that studies used high-pass cutoff frequencies ranging from 0.01 to 1.0 Hz and low-pass cutoff frequencies ranging from 16 to 200 Hz. Formal recommendations for filter settings also vary widely. For example, a set of recommendations for

clinical ERP studies (Duncan et al., 2009) recommended a bandpass of 1 to 30 Hz for the mismatch negativity and 0.1 to 100 Hz for the P3 and N400 components, Widmann et al. (2015) recommended a bandpass of ≤0.1 to ≥40 Hz for most studies, and Luck (2014) recommended a bandpass of 0.1 to 30 Hz for most cognitive and affective studies. Given this variability in recommendations, what cutoffs should a researcher use? Different filter cutoffs lead to different levels of statistical power and waveform distortion, so the field would be well served by a set of careful, mathematically rigorous, and empirically justified recommendations for filter settings that take into account the details of a given study (e.g., the component being studied and the scoring method).

In a companion paper (Zhang et al., 2024), we presented a new approach for determining the optimal filter settings for a given situation. This approach involves filtering a data set with many different filters and then assessing the resulting noise level and the signal-to-noise ratio (SNR) for each filter. Each filter is also applied to noise-free simulated ERP waveforms to quantify the amount of waveshape distortion produced by the filter. The optimal filter is chosen as the one that produces the best data quality without exceeding a preset criterion for waveshape distortion. The present article applied this approach to an open EEG/ERP data set—the ERP CORE (Kappenman et al., 2021)—to determine optimal filter settings for a broad set of ERP components obtained from a relatively large set of neurotypical young adults. These settings should also be nearly optimal for similar paradigms, participant populations, and recording conditions. In addition, we repeated the analyses after adding more noise to the data to determine optimal filters for modestly noisier data sets. The recommended filter cutoffs are shown in Table 1.

1.1 | Overview of the approach

The companion paper (Zhang et al., 2024) provides a detailed description and justification of our approach. This section provides a brief overview.

Our approach is based on the observation that the effects of a particular type of noise depend on the details of the study, including the shape of the component being scored (e.g., a narrow N170 or a broad P3), the scoring method (e.g., mean amplitude vs. peak latency), and the types of noise present in the data (e.g., muscle noise vs. skin potentials). When the SNR is computed with our approach, the specific amplitude or latency score for a specific component defines the signal, and the expected error in this score defines the noise. The signal can be quantified by obtaining the score of interest from a grand average waveform. The noise that is relevant for that specific score can be assessed using a new metric of ERP data quality called the *standardized measurement error* (SME; Luck et al., 2021; Zhang & Luck, 2023). The SME is obtained from individual participants and then aggregated across participants by computing the root mean square (RMS) of the single-participant SME values. The SNR is then calculated as the score divided by the RMS(SME) of that score. We use the term SNR_{SME} to indicate this specific definition of the SNR.

Filters typically reduce the amplitude of an ERP component (the signal) as well as the noise, so it is essential to use the SNR_{SME} when determining the optimal filter settings for amplitude scores. However, most filters have only a modest effect on latency values (see Zhang et al., 2024), so we do not assess the ratio of signal to noise for each filter but instead assess the noise level directly using the RMS(SME).

TABLE 1 Recommended filter settings (in Hz, with a slope of 12 dB/octave) for each combination of scoring method and ERP component.

	Mean amplitude		Peak amplitude		Peak latency		50% Area latency	
	High pass	Low pass	High pass	Low pass	High pass	Low pass	High pass	Low pass
N170	0.9	≥30 or none	0.9	≥30 or none (or 30[a])	≤0.9	10-20	≤0.9	10–20
MMN	0.5	≥20 or none	0.5	≥20 or none (or 30[a])	≤0.5	10 (or 5[b])	≤0.5	≥20 or none
N2pc	0.5	≥20 or none	0.5	20–40 (or 20[a])	0.01 or none	10	0.01 or none	5 (or 10[c])
P3	0.2	≥10 or none	0.2	≥10 (or 10[a])	0.2	10 (or 5[a])	0.2	10 (or 5[a])
N400	0.2	≥10 or none	0.2	≥10 (or 10[a])	0.2	10 (or 5[a])	0.2	10 (or 5[a])
LRP	0.3	≥30 or none	0.3	30 (or 20[a])	0.3	5 (or 10[c])	0.3	5 (or 10[c])
ERN	0.4	≥20 or none	0.4	≥20 or none	0.4	10	0.4	10

[a]For data with moderate amounts of high-frequency noise.

[b]For data with moderate amounts of broadband noise (if distortion of onset and offset times can be tolerated).

[c]To avoid distortion of onset and offset times.

Low-pass filters typically spread the components in time, whereas high-pass filters often produce artifactual opposite-polarity peaks before and after a true peak (De Cheveigné & Nelken, 2019; Widmann et al., 2015). Our approach to filter selection involves applying filters to noise-free simulated ERP waves and quantifying the size of the artifactual peaks. Simulated data must be used for this purpose because ground truth is not known for real data. The size of the artifactual peaks produced by a given filter is quantified as the amplitude of the artifactual peak relative to the amplitude of the true peak (the *artifactual peak percentage*).

The amount of waveform distortion that can be tolerated depends on the scientific goals of a given study. In most cases, artifactual opposite-polarity peaks are the most significant kind of distortion because they can cause a modulation of one component to be misinterpreted as a modulation of some other component. We propose that the selected filter should produce an artifactual peak that is less than 5% of the size of the true peak, at least for studies conducted in highly cooperative participant populations using high-quality recording systems (Zhang et al., 2024). Thus, the optimal filter in such studies is defined as the one that yields the best SME or SNR_{SME} while producing an artifactual peak percentage of <5%.

The fact that different filter parameters may be optimal for different ERP components or different scoring methods creates complications when multiple components or scores are analyzed in the same data set. Recommendations for these scenarios are provided in the companion paper (Zhang et al., 2024).

1.2 | Goals of the present study

Although our approach is straightforward, and we have provided tools to make it easy to implement in ERPLAB Toolbox (Lopez-Calderon & Luck, 2014), applying this approach to a broad set of data requires an investment of time. Moreover, to avoid double dipping, the optimal filter settings should be determined using a previous data set rather than the data set currently being analyzed. The present study therefore applied this approach to an existing public data set that contains a broad set of ERP components, providing researchers with recommended filter settings that they can use in future research with these components.

Specifically, we used the data from the ERP CORE (Compendium of Open Resources and Experiments; Kappenman et al., 2021), which includes data from 40 young adults who performed six standardized paradigms that yielded seven commonly studied ERP components: P3b, N400, N170, N2pc, mismatch negativity (MMN),

error-related negativity (ERN), and lateralized readiness potential (LRP). We provide recommended filter settings for each of these components, separately for four different scoring methods: mean amplitude, peak amplitude, peak latency, and 50% area latency (Clayson et al., 2013; Luck, 2014). The recommended filter settings should be useful for a broad range of researchers who record ERPs from similar subject populations using similar recording setups. Note that peak amplitude is often a problematic scoring method (Clayson et al., 2013; Zhang & Luck, 2023), but we include it here because it is still widely used in some domains.

The ERP CORE data contain relatively low noise levels because the participants were highly compliant young adults and the EEG was recorded using a high-quality recording system with active electrodes, a gel conductor, and a climate-controlled shielded chamber. To demonstrate that our recommendations generalize to data that have modestly greater noise levels, we provide additional analyses in which different types of noise are added to the data. However, we do not expect the results to generalize to highly different subject populations (e.g., infants), very dissimilar recording setups (e.g., dry electrodes), or components with very different waveshapes (e.g., the auditory brainstem response or contralateral delay activity).

2 | METHOD

All data processing was conducted in Matlab using EEGLAB Toolbox (Delorme & Makeig, 2004) and ERPLAB Toolbox (Lopez-Calderon & Luck, 2014). The data and scripts used in our analyses are available at https://osf.io/z3hfp/. If desired, the processing steps can be carried out using the graphical user interface rather than using scripts (using version 9.20 or higher of ERPLAB Toolbox).

2.1 | ERP CORE data and preprocessing

The ERP CORE (Kappenman et al., 2021) data were downloaded from the online repository (https://doi.org/10.18115/D5JW4R). Details of the participants, paradigms, recording methods, and analysis procedures can be found in the original paper. Here, we provide a brief overview of the participants, recording methods, and preprocessing procedures. A brief description of each individual ERP paradigm is provided in the corresponding portion of the Results section.

The ERP CORE contains data from 40 neurotypical college students (25 females) who were recruited from the University of California, Davis community. We used all 40 participants in the present analysis, irrespective of the

number of artifacts or behavioral errors. The one exception is that one participant had zero usable trials for one condition of the N2pc experiment and was excluded from the analyses of that experiment.

The EEG was recorded using a DC-coupled Biosemi ActiveTwo recording system (Biosemi B.V., Amsterdam, the Netherlands) with active electrodes, an antialiasing filter (fifth-order sinc filter, half-power cutoff at 204.8 Hz), and a 1024 Hz sampling rate. Although data are available for 30 scalp sites, along with horizontal and vertical electrooculogram electrodes, the present analyses focused on the single site where a given component was largest (see Table 2).

The data used here had already passed through several steps of the ERP CORE preprocessing pipeline. This pipeline began by shifting the event codes to reflect the intrinsic delay of the video monitor and resampling the data at 256 Hz with an antialiasing filter at 115 Hz. The data were then referenced to the average of the P9 and P10 electrodes, except that the average of all scalp sites was used as the reference for the N170 paradigm. Independent component analysis (ICA) was used to correct the data for eyeblinks and eye movements. A noncausal Butterworth high-pass filter (half-amplitude cutoff of 0.1 Hz, 12 dB/oct roll-off) was applied to optimize the ICA decomposition, and the component weights were then transferred back to the original data so that the artifacts could be corrected in the unfiltered data (See Chapter 9 in Luck, 2022).

2.2 | Filtering and additional preprocessing

To assess the effects of filtering on the SME and SNR_{SME}, we separately filtered the continuous EEG data resulting from the preceding steps using a variety of different combinations of low-pass and high-pass cutoffs. All cutoff frequencies listed in this article indicate the half-amplitude point in the frequency response function. We used noncausal Butterworth filters (Hamming, 1998) because they are efficient, flexible, well behaved, and widely used for EEG and ERP signals. The companion paper (Zhang et al., 2024) indicated that a relatively gentle slope of 12 dB/octave was optimal for minimizing

waveform distortions, so we used this roll-off for all analyses. However, it would be straightforward to repeat the analyses with a different slope using the data and scripts we have provided at https://osf.io/z3hfp/ . It would also be straightforward to repeat the analyses with other families of filters (e.g., finite impulse response filters).

After the data were filtered, we implemented the remaining preprocessing steps used in the original ERP CORE paper. First, we epoched and baseline corrected the EEG using the time windows shown in Table 2. Then, we averaged the EEG epochs, excluding trials with artifacts or incorrect behavioral responses.

It is possible that filtering could influence the SNR of a given data set indirectly by impacting the artifact rejection and correction processes. However, it is possible and even advisable to use different filtering parameters for determining which trials to reject, for performing the ICA decomposition, and for averaging and scoring the ERP amplitudes and latencies. Thus, the question of what filter is optimal for the primary ERP analyses is independent of the question of what filter is optimal for artifact rejection or correction. To keep these issues separate, and to focus on the effects of filtering on the primary ERP scores, we performed artifact rejection and correction in a manner that was not influenced by the different filter settings being examined in the main analyses. Specifically, we applied the rejection flags and ICA decompositions from the original ERP CORE data to the present data, independent of the filter settings being tested. The ERP CORE repository (https://doi.org/10. 18115/D5JW4R) provides the artifact rejection and correction parameters, along with the number of included and excluded trials.

2.3 | Computing the signal, noise, and signal-to-noise ratio

We computed the signal, noise, and signal-to-noise ratio separately for each combination of the seven ERP components and four scoring methods: mean amplitude, peak amplitude, peak latency, and 50% area latency. In an attempt to isolate a single ERP component (Zhang et al., 2024), the signal, noise, and signal-to-noise ratio

TABLE 2 Epoch window, baseline period, electrode site, and measurement window used for each ERP component.

	N170	MMN	N2pc	P3	N400	LRP	ERN
Epoch (ms)	−200 to 800	−200 to 800	−200 to 800	−200 to 800	−200 to 800	−800 to 200	−600 to 400
Baseline period (ms)	−200 to 0	−200 to 0	−200 to 0	−200 to 0	−200 to 0	−800 to −600	−400 to −200
Electrode site	PO8	FCz	PO7/PO8	Pz	CPz	C3/C4	FCz
Measurement window (ms)	110 to 150	125 to 225	200 to 275	300 to 600	300 to 500	−100 to 0	0 to 100

were computed from the difference waves for a given component (e.g., the faces-minus-cars difference wave for the N170 component). However, only the noise values were relevant for the latency measures (see Section 1.1).

Table 2 shows the time windows and electrode sites that were used for each component. Mean amplitude was defined as the mean voltage during the time window. Peak amplitude was defined as the maximal positive voltage (for the P3 component) or the maximal negative voltage (for the other components) in that time window, and peak latency was defined as the latency at which that peak occurred. The 50% area amplitude was computed by taking the integral of the voltages during that time window and finding the point that bisected that integral into equal halves. To increase temporal precision, spline interpolation was used to upsample the waveforms by a factor of 10 before the latencies were scored (see Luck, 2014, for the rationale). The signal was estimated by obtaining a given score from the grand average difference waveform for a given component.

We used the RMS(SME) for a given score to estimate the noise. The SME was obtained separately from each participant, and values were aggregated across participants by computing the root mean square (RMS) of the single-participant values (see the companion paper for the rationale). The aggregate SNR_{SME} was then computed as the score from the grand average divided by the RMS(SME) of the single-participant scores.

For the N170, MMN, P3, N400, and ERN mean amplitude scores, we used built-in ERPLAB functions to compute the SME for each condition (e.g., separately for faces and cars). The mean amplitude SME values for the difference scores (e.g., faces minus cars) were then calculated as $SME_{A-B} = \sqrt{SME_A{}^2 + SME_B{}^2}$. In this equation, SME_{A-B} is the SME of the difference between conditions A and B, and SME_A and SME_B are the SMEs of the two individual conditions (see the companion paper for details).

This approach is not valid for differences between electrode sites, such as the contralateral-minus-ipsilateral difference waves used to isolate the N2pc component and the lateralized readiness potential. It is also inappropriate for nonlinear scoring methods such as peak amplitude, peak latency, and 50% area latency. For these cases, we used bootstrapping with 1000 iterations to estimate the SME directly from the difference wave (see the companion paper Zhang et al., 2024). The scripts used to compute SME values and the associated SNR_{SME} values for the present study are provided at https://osf.io/z3hfp/.

Note that ERPLAB uses a simple bootstrapping scheme that does not take into account systematic changes over the course of a recording session (i.e., autocorrelation). Most of the autocorrelation is a result of low-frequency drifts in the EEG signal, which was minimized by the high-pass filter and the baseline correction. In addition, each experiment was brief (~10 min), so psychological sources of systematic changes would be minimal. In the future, researchers may wish to use a blocked bootstrapping procedure (Kunsch, 1989) to deal with any remaining autocorrelations.

2.4 | Effects of increased noise

The ERP CORE data are quite clean, especially with respect to line noise. To determine whether the results of the present study generalize to noisier data, we repeated all the analyses after adding a random-phase 60 Hz sinusoidal oscillation (to simulate line noise), white noise (to simulate muscle activity and other kinds of spiky noise), and pink noise (to simulate movement artifacts and other kinds of nonstationary noise). The noise was added to the continuous EEG prior to filtering. The 60 Hz oscillating noise had a peak-to-peak amplitude of 20 μV. The white noise and pink noise had a mean of 0 and a standard deviation of 7.07 μV (which was the same as the standard deviation of the line noise). Supplementary Figure S1 provides an example of the EEG after the addition of these types of noise, showing that the noise was moderate relative to the original EEG. The scripts at https://osf.io/z3hfp/ can be easily modified to see the effects of more extreme noise. The results of adding noise are provided in the Supplementary Materials, and the main manuscript indicates when different filter settings are appropriate for noisier data.

2.5 | Quantifying waveform distortion

To quantify the distortion produced by a given filter, we applied the filter to an artificial waveform that was designed to simulate the grand average difference wave for a given component (e.g., the faces-minus-cars difference wave for the N170). The artificial waveforms were Gaussian or ex-Gaussian functions created using ERPLAB's *Create Artificial ERP Waveform* tool and designed to match the grand average difference wave for a given component (see parameters in Table 3). We applied each candidate filter to the simulated waveform and calculated the *artifactual peak percentage*, defined as 100 times the amplitude at the time of the largest artifactual peak divided by the amplitude at the time of the true peak (with the amplitudes obtained from the filtered waveform).

TABLE 3 Parameters used to create the artificial waveforms.

	N170	MMN	N2pc	P3	N400	LRP	ERN
Function	Gaussian	Gaussian	Gaussian	Ex-Gaussian	Ex-Gaussian	Gaussian	Ex-Gaussian
Peak amplitude (μV)	−4.6	−2.82	−1.62	8.6	−9.65	−3.2	−12.5
Gaussian mean (ms)	129	190	257	310	280	−47	76
Gaussian SD (ms)	14	33	28	58	65	36	26
Exponential tau (ms)	0	0	0	2000	1400	0	−300

3 | RESULTS

Each of the following sections provides the details of the filter optimization process for a given component. These sections are written so that readers can focus on the section for a single component rather than reading the sections for all seven components. However, we recommend that all readers start with the section for the N170 component, which provides a more detailed description of how the optimal filter settings were obtained. The sections are ordered according to ascending latency for the stimulus-locked components (N170, MMN, N2pc, P3, and N400), followed by the two response-locked components (LRP and ERN). The optimal filter settings resulting from this process are listed for all seven components in Table 1.

3.1 | The N170 component

As illustrated in Figure 1a, the stimuli in the N170 paradigm consisted of a randomized sequence of faces, cars, scrambled faces, and scrambled cars (see Kappenman et al., 2021, for details). Participants were instructed to press one button for intact stimuli (faces or cars) and a different button for scrambled stimuli (scrambled faces or scrambled cars). Here, we focus solely on the ERPs elicited by faces and cars. Figure 1b shows the grand average ERPs elicited by the faces and the cars, and Figure 1c shows the faces-minus-cars difference wave overlaid with the artificial N170 waveform.

3.1.1 | N170 waveform distortion

Figure 1d shows the effects of several low-pass filters on the simulated N170 waveform. As the low-pass cutoff frequency was reduced, the amplitude of the N170 was reduced, and the N170 became broader. With a 5 Hz cutoff frequency, the amplitude was reduced by more than 50%. Note that all cutoff frequencies listed in this article indicate the half-amplitude point in the frequency response function, and the results would be different for half-power

cutoffs at these frequencies. In addition, all filters examined here have a roll-off of 12 dB/octave, and the filter distortion would be greater for steeper roll-offs.

Figure 1e shows the effects of several high-pass filters on the simulated N170 waveform. Even with a 2 Hz cutoff, high-pass filtering produced very little amplitude reduction. However, higher cutoff frequencies led to clear artifactual peaks before and after the N170 peak. If these peaks were large enough to be statistically significant, they could lead to invalid conclusions (e.g., that P1 amplitude is greater for faces than for cars). However, this is unlikely to occur if the artifactual peak is less than 5% of the size of the true peak (Zhang et al., 2024). This 5% threshold for artifactual peak percentage was exceeded for the 1 Hz and 2 Hz cutoffs. Supplementary Figure S2 provides artifactual peak percentage values for a denser sampling of cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies that exceeded 0.9 Hz. Thus, 0.9 Hz (with a 12 dB/octave roll-off) is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE N170 experiment.

3.1.2 | Optimal filters for N170 mean amplitude and peak amplitude

Figure 2 shows the signal, noise, and signal-to-noise ratio for several combinations of low-pass and high-pass cutoffs (see additional cutoffs in Supplementary Figure S2).

These figures show that the N170 mean and peak amplitude scores were dramatically reduced when the low-pass cutoff was less than 20 Hz but with little impact of the high-pass cutoff. These effects are consistent with the effects of filtering on the simulated N170 shown in Figure 1. The noise—quantified as RMS(SME)—was also progressively reduced as the low-pass cutoff frequency decreased and as the high-pass cutoff increased. Decreasing the low-pass cutoff frequency reduced the signal more than it reduced the noise, so decreasing the low-pass cutoff reduced the signal-to-noise ratio (SNR_{SME}), especially for cutoffs of less than 30 Hz. This indicates that low-pass filtering is not helpful and can
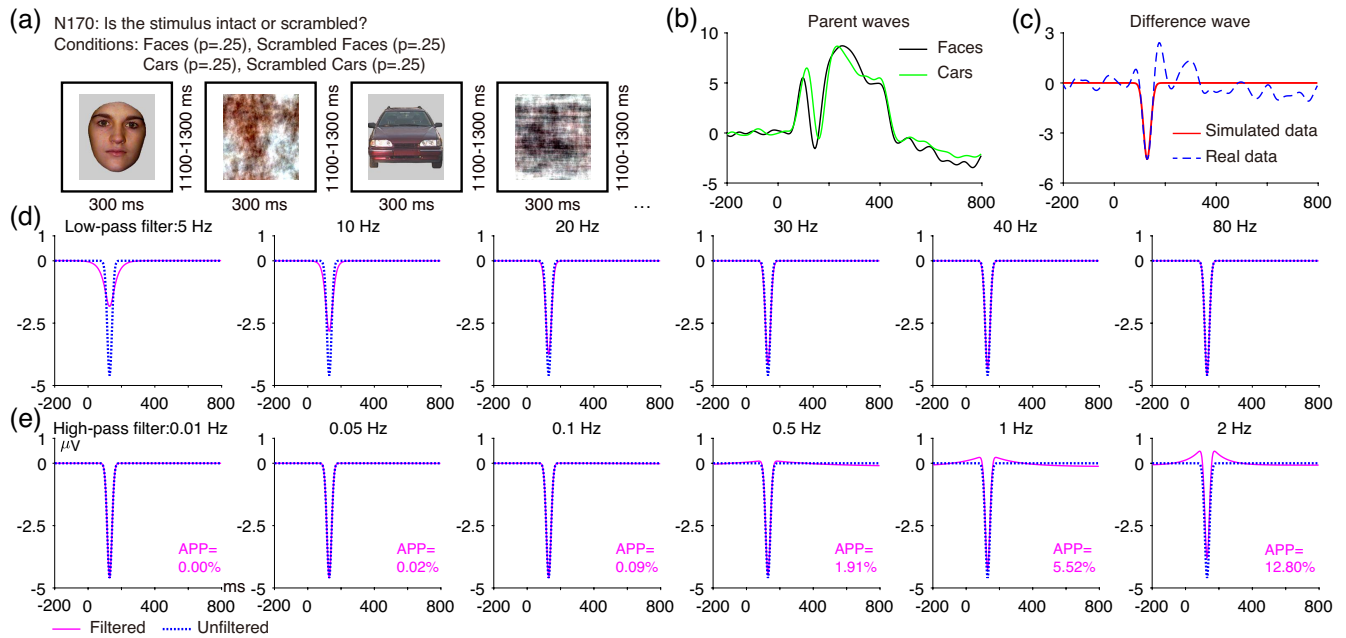
FIGURE 1 (a) N170 face perception paradigm. Only the data from the face and car trials were used in the present analyses. (b) Grand average ERP waveforms from the PO8 electrode site for the face and car trials. (c) Grand average face-minus-car difference wave at PO8, along with the simulated N170 difference wave (Gaussian function, mean = 129 ms, SD = 14 ms, and peak amplitude = −4.6 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the simulated waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

actually be harmful to N170 amplitude scores. For data sets like the ERP CORE N170 experiment, we therefore recommend no low-pass filtering, or a low-pass filter with a cutoff of 30 Hz or higher if desired to minimize the "fuzz" in the waveforms so that differences between groups or conditions are easier to visualize (see Table 1). For noisier data, a low-pass cutoff at 30 Hz is helpful for N170 peak amplitude scores (see Supplementary Figures S3–S5).

By contrast, increasing the high-pass cutoff frequency reduced the noise more than it reduced the signal, resulting in a progressive increase in the SNR_{SME} value as the cutoff frequency increased. However, high-pass cutoffs of greater than 0.9 Hz produced artifactual peaks that exceeded our 5% threshold (see Figure S2). Thus, a high-pass cutoff of 0.9 Hz provides the best SNR_{SME} without producing excessive waveform distortion, so this is our recommended cutoff for N170 amplitude scores.

3.1.3 | Optimal filters for N170 peak latency and 50% area latency

The bottom row of Figure 2 shows how the noise level varied across filter settings for the N170 peak latency and

50% area latency scores. As noted in Section 1.1, filters can substantially reduce amplitude values but typically produce minimal bias for latency scores, so we used the RMS(SME) rather than the SNR_{SME} to quantify data quality for latency scores.

For peak latency, the RMS(SME) was best with a low-pass cutoff of 10–20 Hz. With higher cutoff frequencies (e.g., 30 Hz), high-frequency noise in the data was not attenuated and added variability to the latency of the peak. With lower cutoff frequencies (e.g., 5 Hz), the amplitude of the N170 was substantially reduced by the low-pass filter, making it more difficult to reliably determine the peak latency. The same pattern was observed for the 50% area latency score, but with a smaller impact of the filtering. Consequently, we recommend a low-pass cutoff frequency of 10–20 Hz for N170 peak latency and 50% area latency scores.

With a low-pass cutoff at 20 Hz, there was very little impact of the high-pass filter on the noise for either peak latency or 50% area latency, except that the noise started to rise when the cutoff exceeded 0.9 Hz. Because the noise was largely the same for high-pass cutoffs between 0 and 0.9 Hz (see Supplementary Figure S2), any high-pass cutoff below 1 Hz would be justified for N170 latency scores.
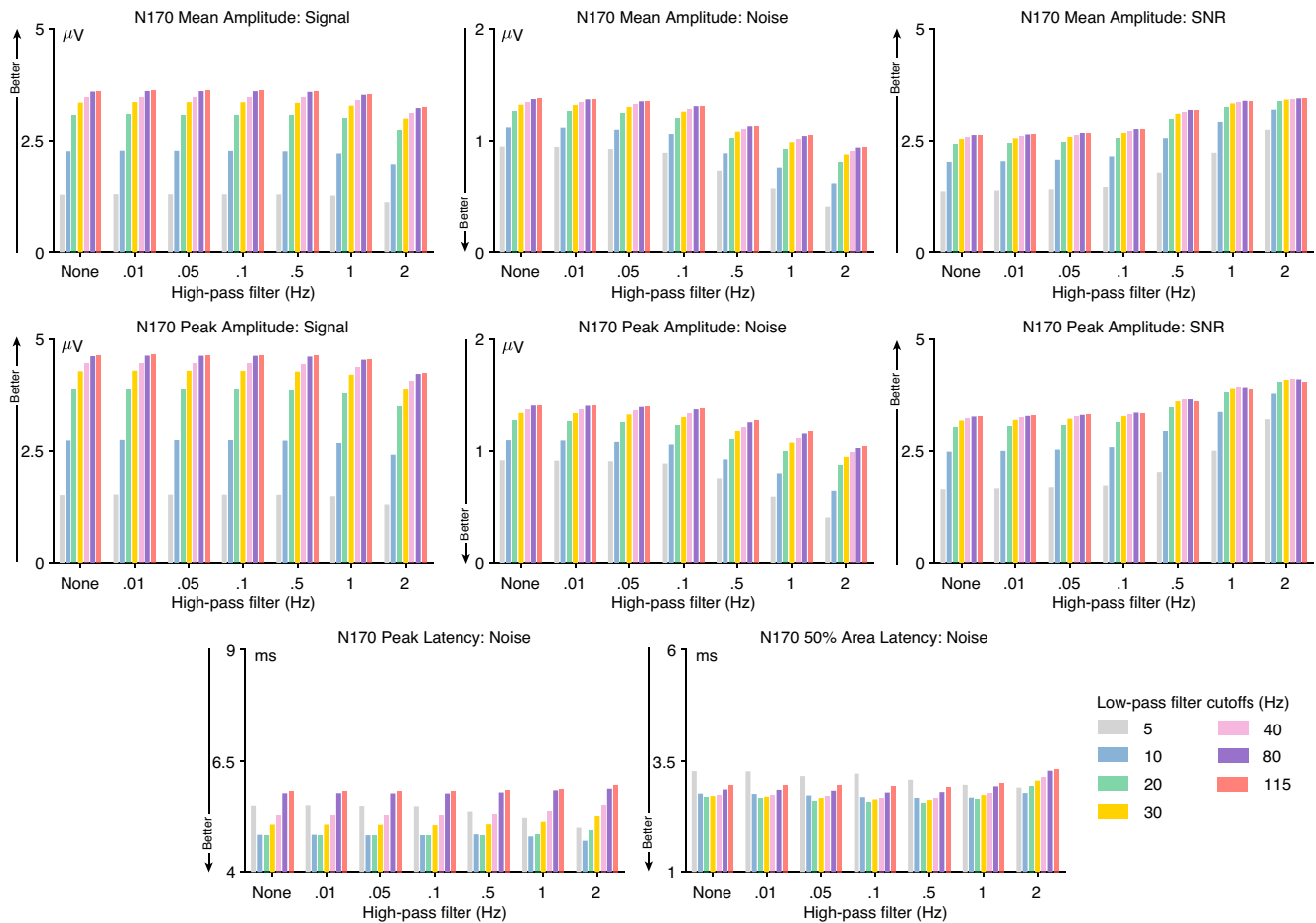
FIGURE 2 N170 data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (faces minus cars). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

3.2 | The mismatch negativity

As shown in Figure 3a, the mismatch negativity (MMN) was elicited using a passive auditory oddball task. Participants watched a silent video while they were presented with a task-irrelevant sequence of standard tones (1000 Hz, 80 dB, $p = .8$) and deviant tones of lower intensity (1000 Hz, 70 dB, $p = .2$). Each participant was presented with 800 standards and 200 deviants. Figure 3b shows the grand average waveforms for standard and deviant stimuli at the FCz electrode site, and Figure 3c shows the deviant-minus-standard difference wave overlaid with the simulated MMN.

3.2.1 | MMN waveform distortion

Figure 3 shows the effects of several low-pass and high-pass filters on the simulated MMN waveform. For low-pass filters, the peak amplitude of the MMN was progressively

reduced as the low-pass cutoff decreased, and the MMN became progressively broader (see Figure 3d). For high-pass filters, cutoffs above 0.5 Hz strongly distorted the MMN waveform, producing large opposite-polarity peaks before and after the MMN peak (see Figure 3e). Figure S6 provides results for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.5 Hz, so this is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE MMN experiment.

3.2.2 | Optimal filters for MMN mean amplitude and peak amplitude

As shown in Figure 4, the SNR_{SME} for low-pass filters was largely constant when the cutoff frequency was 20 Hz and higher. For data sets like the ERP CORE MMN
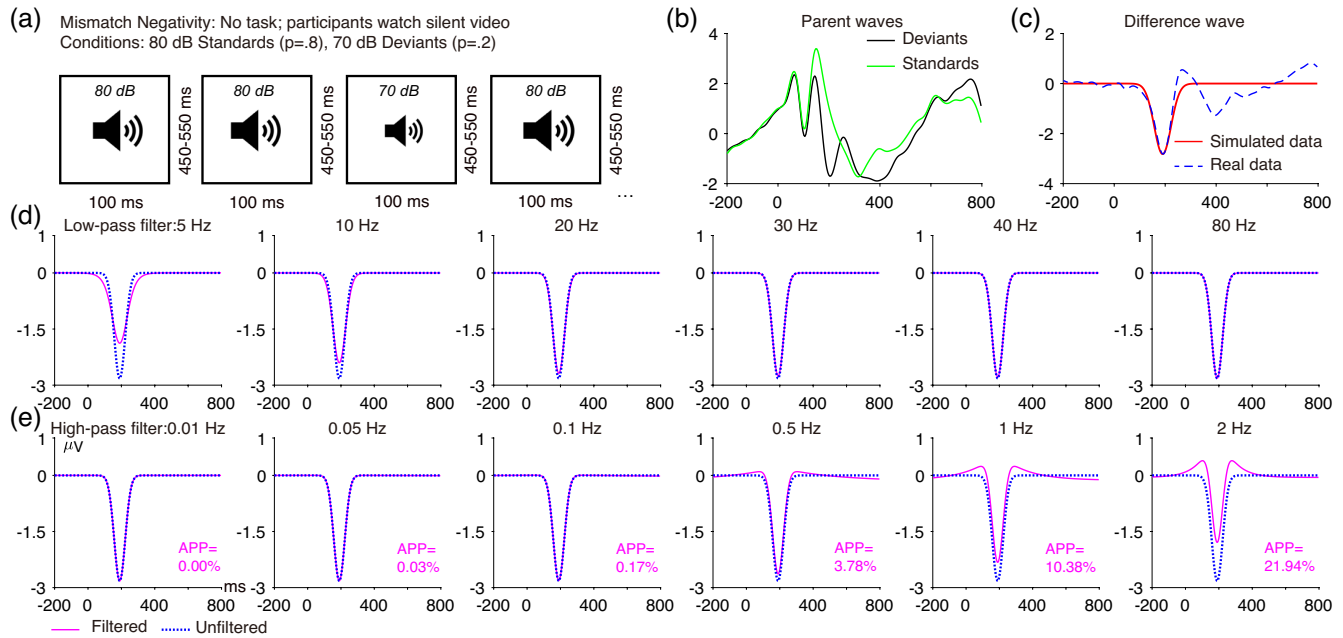
FIGURE 3 (a) MMN passive auditory oddball task. (b) Grand average ERP waveforms at FCz electrode site for deviant and standard trials. (c) Grand average deviant-minus-standard wave at FCz along with its simulated MMN difference wave (Gaussian function, mean = 190 ms, SD = 33 ms, and peak amplitude = −2.82 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the simulated waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

experiment, we therefore recommend no low-pass filtering, or a low-pass filter with a cutoff of 20 Hz or higher if desired to make it easier to visualize the ERP waveforms (see Table 1). For noisier data, a low-pass cutoff of 30 Hz yields the highest SNR_{SME} for peak amplitude scores (see Supplementary Figures S7–S9).

For high-pass filtering, the SNR_{SME} increased as the high-pass cutoff increased, particularly at cutoffs of 1 Hz and higher. However, waveform distortion also increased as the high-pass cutoff increased. A high-pass cutoff of 0.5 Hz yielded the highest SNR_{SME} while remaining below our 5% threshold for artifactual peak distortion, so this is our recommended high-pass cutoff frequency for MMN amplitude scores.

### 3.2.3 | Optimal filters for MMN peak latency and 50% area latency

The bottom row of Figure 4 shows how the RMS(SME) values varied across filter settings for the MMN peak latency and 50% area latency scores. For peak latency, the data quality was largely unaffected by high-pass filtering, but low-pass filtering had a substantial impact on the RMS(SME) values. The combination of a 0.5 Hz high-pass filter and a 10 Hz low-pass filter produced the lowest noise without exceeding our 5% threshold for

artifactual peak percentage (see also Supplementary Figure S6), so this is our recommendation for MMN peak latency. A 5 Hz cutoff improves the RMS(SME) for data containing moderate amounts of broadband noise (Supplementary Figures S8 and S9). However, this cutoff will make the MMN appear to onset quite a bit earlier than the true onset time (see Figure 3), so a 10 Hz cutoff is preferred when onset times are important for addressing the scientific goals of a study.

The results for 50% area latency were quite different. High-pass filtering again had little effect, but low-pass filtering actually increased the noise levels slightly (see Figure 4 and Supplementary Figure S6). We therefore recommend a high-pass cutoff of 0.5 Hz or lower and no low-pass filtering (or a low-pass filter with a cutoff of 20 Hz or higher to make it easier to visualize the ERP waveforms).

3.3 | The N2pc component

As illustrated in Figure 5a, the N2pc component was obtained using a simple visual search task. For a given block, either pink or blue was designated the target color, with the other color being the nontarget. Each stimulus array contained 1 pink square, 1 blue square, and 22 black squares. The side containing the target color was unpredictable,

ZHANG ET AL.
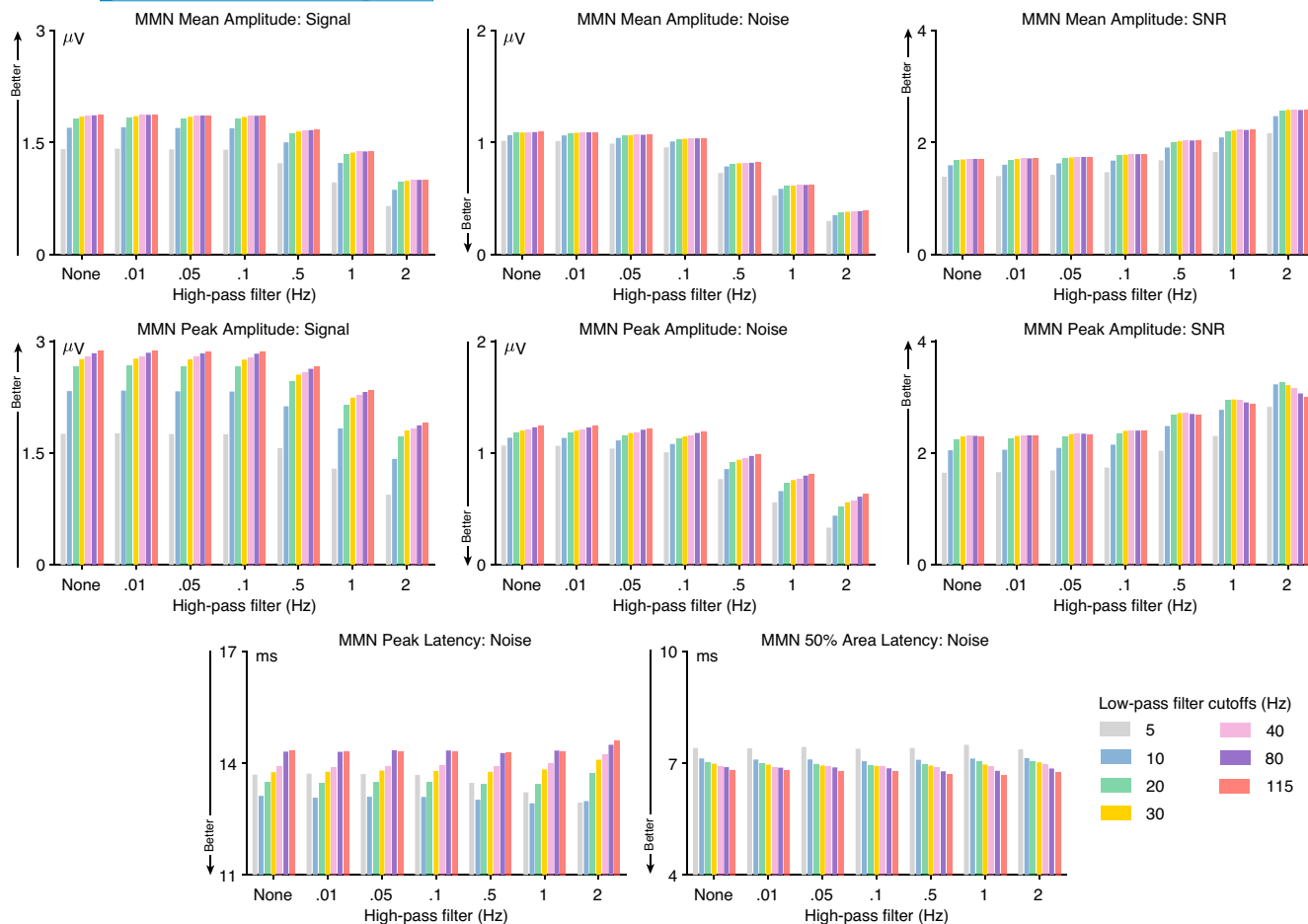
FIGURE 4 MMN data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (deviants minus standards). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

but the target and nontarget color were always on opposite sites. For each array, participants pressed one of two buttons to indicate the location (top or bottom) of a gap in the attended-color square (while maintaining central fixation). Each participant completed a total of 320 trials (160 trials with the target on each side). Figure 5b shows the grand average ERPs at the PO7/PO8 electrode site of the hemisphere contralateral or ipsilateral to the target. Figure 5c shows the grand average contralateral-minus-ipsilateral difference wave overlaid with the simulated N2pc.

3.3.1 | N2pc waveform distortion

Figure 5d,e shows the effects of several low-pass and high-pass filters on the simulated N2pc waveform. For low-pass filters, the N2pc became substantially smaller and broader when the cutoff was below 20 Hz. For high-pass

filters, cutoffs above 0.5 Hz strongly distorted the N2pc waveform, producing large opposite-polarity peaks before and after the N2pc peak. Figure S10 provides results for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.5 Hz, so this is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE N2pc experiment.

3.3.2 | Optimal filters for N2pc mean amplitude and peak amplitude

As shown in Figure 6, low-pass filters had little impact on the N2pc except when the cutoffs were less than 20 Hz. Peak amplitude was more impacted by filtering, with reductions in SNR_{SME} for frequencies less
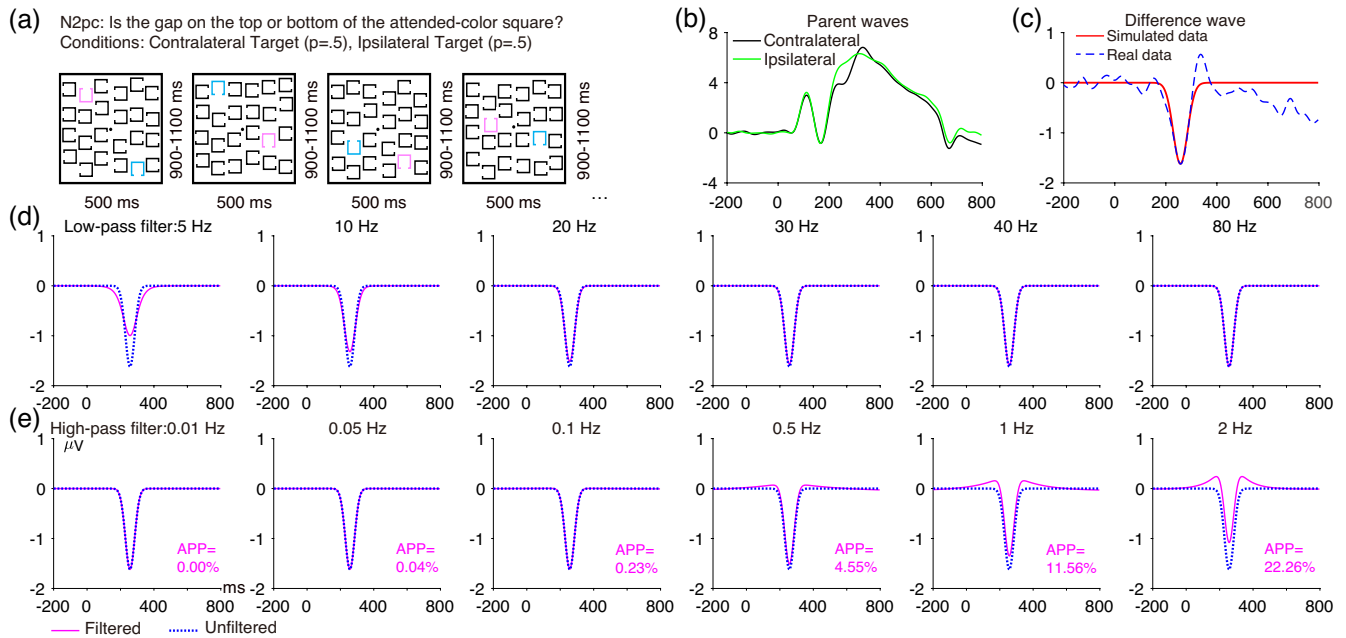
FIGURE 5 (a) N2pc simple visual search task. (b) Grand average waves from the PO7/PO8 electrode sites for contralateral and ipsilateral conditions. (c) Grand average contralateral-minus-ipsilateral difference wave at PO7/PO8, along with the simulated N2pc difference wave (Gaussian function, mean = 257 ms, $SD = 28$ ms, and peak amplitude = −1.62 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the simulated waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

than 20 Hz and also for frequencies greater than 40 Hz. For data sets like the ERP CORE N2pc experiment, we, therefore, recommend a low-pass cutoff of ≥20 Hz (or no low-pass filtering) for mean amplitude, and 20–40 Hz for peak amplitude. For noisier data, a low-pass cutoff of 20 Hz is optimal for peak amplitude scores (see Supplementary Figures S11–S13; this assumes a high-pass cutoff of 0.5 Hz).

For high-pass filtering, the SNR_{SME} increased as the high-pass cutoff increased, particularly at cutoffs of 0.5 Hz and higher (see Figure 6 and Supplementary Figure S10). However, waveform distortion also increased as the high-pass cutoff increased. A high-pass cutoff of 0.5 Hz yielded the highest SNR_{SME} while remaining below our 5% threshold for artifactual peak distortion, so this is our recommended cutoff.

3.3.3 | Optimal filters for N2pc peak latency and 50% area latency

The bottom row of Figure 6 shows how the RMS(SME) values varied across filter settings for N2pc peak latency and 50% area latency scores. For peak latency, low-pass filtering had a substantial impact on the RMS(SME) values but high-pass filtering had very little impact. The

combination of a 10 Hz low-pass filter with either no high-pass filter or a .01 Hz high-pass cutoff produced the lowest noise without exceeding our 5% threshold for artifactual peak percentage (see also Supplementary Figure S10), so this is our recommendation for N2pc peak latency.

For 50% area latency, a 5 Hz low-pass filter produced the lowest RMS(SME) values, combined with either no high-pass filter or a .01 Hz high-pass cutoff, so this is our recommendation. Note, however, that a 5 Hz low-pass filter will make the N2pc appear to onset quite a bit earlier than the true onset time (see Figure 5), so a 10 Hz low-pass cutoff is preferred when onset times are important for addressing the scientific goals of a study.

3.4 | The P3 component

As illustrated in Figure 7a, an active visual oddball task was used to elicit the P3 component. Participants were shown a randomized sequence of five letters (A, B, C, D, and E), with a probability of $p = .2$ for each letter. One of the five letters was designed as the target for a given block, and participants were instructed to press one button when the target was presented and a different button for any nontarget. For example, if D was the target, participants pressed the target button for D and the non-target button for A, B, C, or
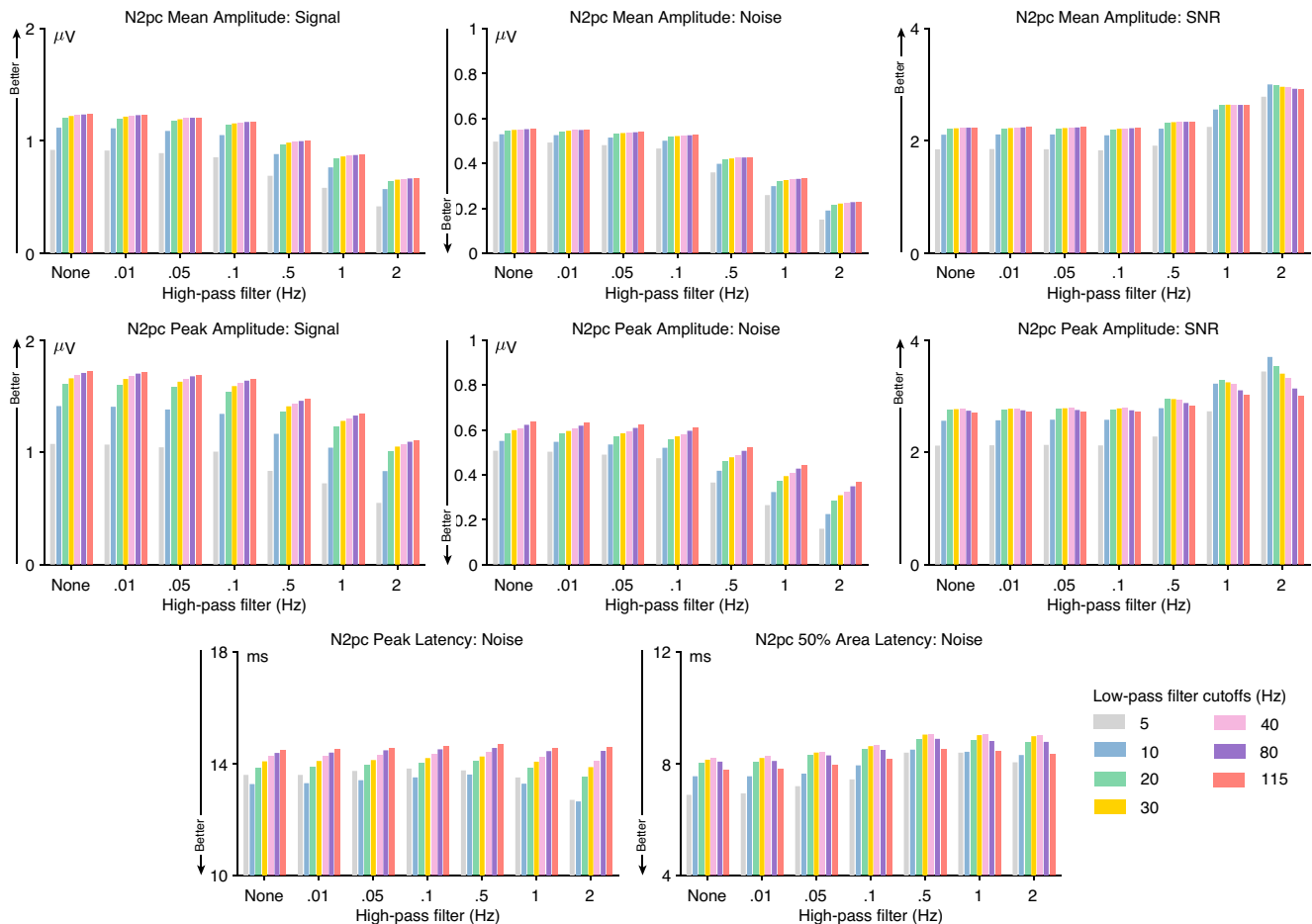
FIGURE 6 N2pc data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (contralateral minus ipsilateral). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

E. Each participant received 40 target trials (referred to as *rare* trials) and 160 nontarget trials (referred to as *frequent* trials). Figure 7b shows the grand average ERP waveforms at the Pz electrode site for the rare and frequent conditions, and Figure 7c shows the rare-minus-frequent difference wave overlaid with the simulated P3.

3.4.1 | P3 waveform distortion

Figure 7d,e shows the effects of several low-pass and high-pass filters on the simulated P3 waveform. Low-pass filters produced minimal distortion of this waveform for cutoffs of 10 Hz or higher. However, high-pass filters with cutoffs of 0.5 Hz or higher created substantial distortion, with a large negative artifactual peak preceding the true P3 wave that could easily be mistaken for an N2 effect. Figure S14 provides results for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the

artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.2 Hz, so this is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE P3 experiment.

3.4.2 | Optimal filters for P3 mean amplitude and peak amplitude

As shown in Figure 8, low-pass filters had minimal impact on the SNR_{SME} for mean and peak P3 amplitude scores (unless an inadvisably strong high-pass filter was also used). For data sets like the ERP CORE P3 experiment, we therefore recommend no low-pass filtering, or a low-pass filter with a cutoff of 10 Hz or higher if desired to make it easier to visualize the ERP waveforms. For noisier data, a 10 Hz low-pass filter for peak amplitude scores is slightly better (see Supplementary Figures S15–S17; this assumes a high-pass cutoff of <0.5 Hz).
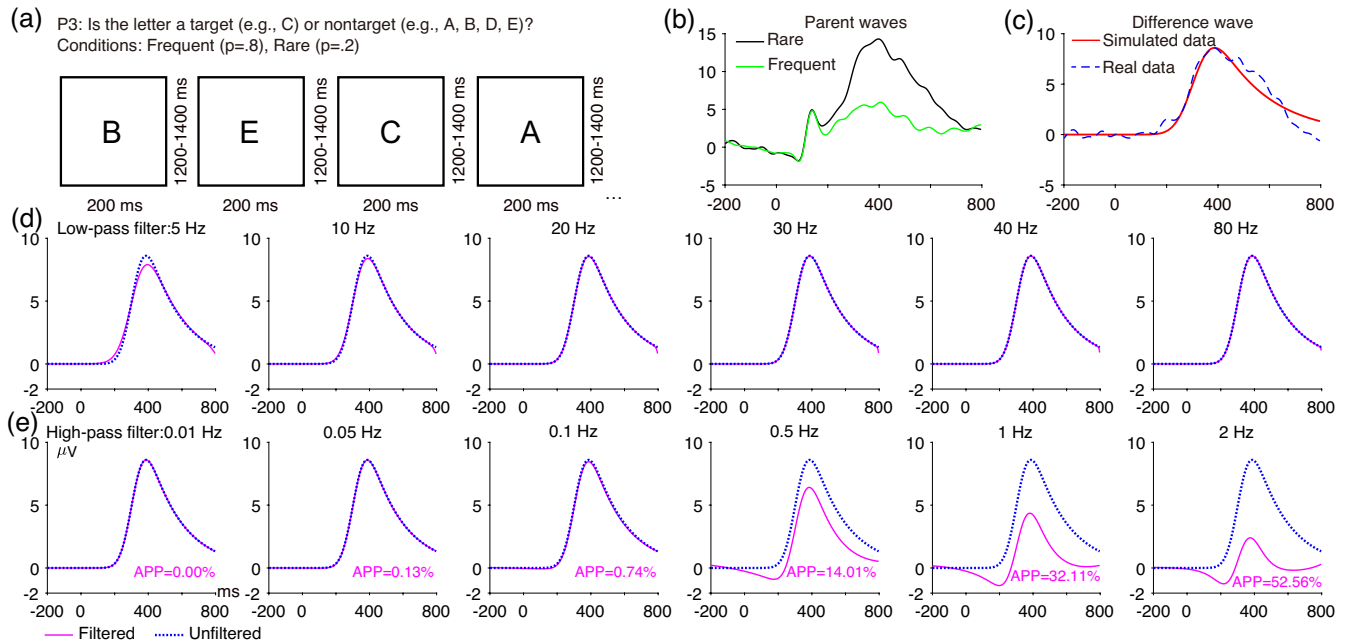
FIGURE 7 (a) P3 active visual oddball paradigm. (b) Grand average ERP waveforms from the Pz electrode site for the frequent and rare trials. (c) Grand average rare-minus-frequent difference wave at Pz, along with the simulated P3 difference wave (Ex-Gaussian function, mean = 310 ms, SD = 58 ms, λ = 2000 ms, and peak amplitude = 8.6 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the artificial waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

For high-pass filtering, the SNR$_{SME}$ increased as the high-pass cutoff increased up to 0.5 Hz (because the reduction in noise outweighed the small reduction in signal) but then decreased for higher cutoffs (because the large reduction in signal outweighed the reduction in noise; see Figure 8 and Supplementary Figure S14). A high-pass cutoff of 0.2 Hz yielded the highest SNR$_{SME}$ while remaining below our 5% threshold for artifactual peak distortion, so this is our recommended cutoff.

3.4.3 | Optimal filters for P3 peak latency and 50% area latency

The bottom row of Figure 8 shows how the RMS(SME) values varied across filter settings for P3 peak latency and 50% area latency scores. For both peak latency and 50% area latency, a low-pass cutoff of 5 Hz produced the lowest noise, so that is our recommendation. However, this cutoff also produces a modest decrease in onset latency (Figure 7), so a low-pass cutoff of 10 Hz may be preferred if onset times are important. A high-pass cutoff of 0.2 Hz yielded the lowest noise while remaining below our 5% threshold for artifactual peak distortion (see Supplementary Figure S14), so this is our recommended high-pass cutoff.

3.5 | The N400 component

As depicted in Figure 9a, the N400 component was elicited using a word pair judgment paradigm. Each trial consisted of a red prime word followed by a green target word. Participants were required to report whether the green word was semantically related (p = .5) or unrelated (p = .5) to the preceding red word by pressing one of two buttons. A total of 120 trials were presented to each participant (60 trials for related word pairs and 60 trials for unrelated word pairs). Figure 9b displays the grand average waveforms for the unrelated and related words at the CPz electrode site, and Figure 9c shows the unrelated-minus-related difference wave overlaid with the simulated N400.

3.5.1 | N400 waveform distortion

Figure 9d,e shows the effects of several low-pass and high-pass filters on the simulated N400 waveform. As was observed for the P3, low-pass filters produced minimal distortion of the N400 waveform as long as the cutoff was 10 Hz or higher. However, high-pass filters with cutoffs of 0.5 Hz or higher created substantial distortion, with a large positive artifactual peak preceding the true N400 wave that could easily be mistaken for a P2 effect. Figure S18 provides results
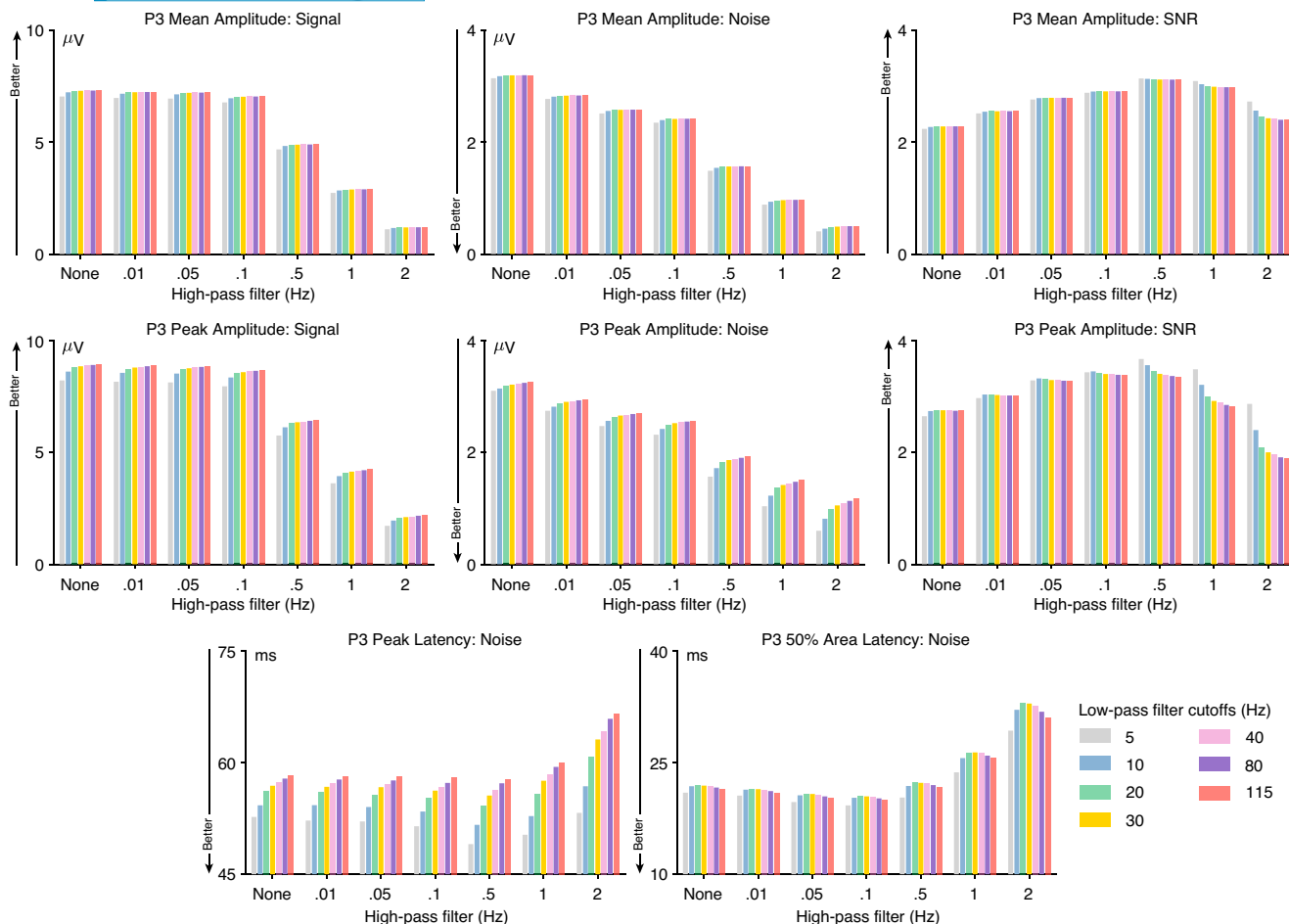
FIGURE 8 P3 data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (frequent minus rare). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.3 Hz, so this is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE N400 experiment.

### 3.5.2 | Optimal filters for N400 mean amplitude and peak amplitude

As shown in Figure 10, low-pass filters had minimal impact on the SNR_{SME} for mean and peak N400 amplitude scores (unless an inadvisably strong high-pass filter was also used). For data sets like the ERP CORE N400 experiment, we therefore recommend no low-pass filtering, or a low-pass filter with a cutoff of 10 Hz or higher if desired to make it easier to visualize the ERP waveforms. For noisier

data, there is a small benefit to applying a 5 or 10 Hz low-pass filter (see Supplementary Figures S19–S21).

For high-pass filtering, the SNR_{SME} increased as the high-pass cutoff increased up to 0.2 Hz (presumably reflecting noise reduction) but then decreased for higher cutoffs (presumably because the reduction in signal outweighed the reduction in noise; see Figure 10 and Supplementary Figure S18). A high-pass cutoff of 0.2 Hz was also below our 5% threshold for artifactual peak distortion, so this is our recommended high-pass cutoff.

3.5.3 | Optimal filters for N400 peak latency and 50% area latency

The bottom row of Figure 10 shows how the RMS(SME) values varied across filter settings for the N400 peak
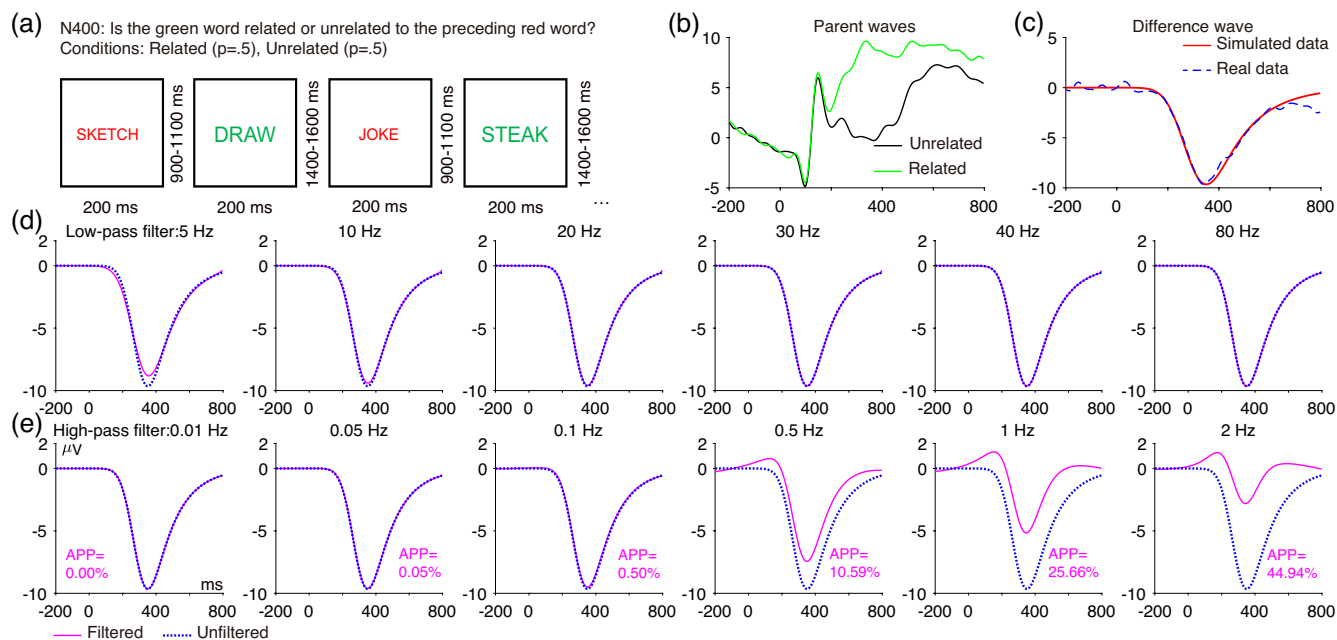
FIGURE 9 (a) N400 word pair judgment paradigm. (b) Grand average ERP waveforms at CPz electrode site for unrelated and related trials. (c) Grand average unrelated-minus-related difference wave along with its simulated N400 difference wave (Ex-Gaussian function, mean = 280 ms, SD = 65 ms, λ = 1400 ms, and peak amplitude = −9.65 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the artificial waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

latency and 50% area latency scores. For both peak latency and 50% area latency, a low-pass cutoff of 5 Hz produced the lowest noise, so that is our recommendation. However, this cutoff also produces a modest decrease in onset latency (Figure 9), so researchers may prefer a low-pass cutoff of 10 Hz when onset latencies are important. A high-pass cutoff of 0.2 Hz yielded the lowest noise while remaining below our 5% threshold for artifactual peak distortion (see supplementary Figure S18), so this is our recommended high-pass cutoff.

## 3.6 | The lateralized readiness potential

As depicted in Figure 11a, the lateralized readiness potential (LRP) was elicited using a flankers paradigm. In this paradigm, each stimulus consisted of a central arrow surrounded by flanking arrows, which could either point in the same direction as the central arrow or in the opposite direction. On each trial, participants indicated whether the central arrow was pointing leftward (p = .5, 200 trials) or rightward (p = .5, 200 trials) by pressing a button with the corresponding hand. Figure 11b shows the grand average ERPs at the C3/C4 electrode site of the hemisphere contralateral or ipsilateral to the responding hand (on trials with correct

responses). Figure 11c shows that the grand average contralateral-minus-ipsilateral difference wave overlaid with the simulated LRP.

The LRP can be analyzed either time locked to the stimulus or time locked to the response. Here, we used response-locked averages. The LRP is typically broader in stimulus-locked averages than in response-locked averages, so different filter parameters may be appropriate for stimulus-locked averages. It would be straightforward for researchers to modify the LRP scripts at https://osf.io/z3hfp/ to analyze stimulus-locked averages.

3.6.1 | LRP waveform distortion

Figure 11d,e shows the effects of several low-pass and high-pass filters on the simulated LRP waveform. For low-pass filters, the LRP became noticeably smaller and broader for cutoffs below 20 Hz. For high-pass filters, cutoffs above 0.1 Hz clearly distorted the LRP waveform, producing large opposite-polarity peaks before and after the LRP peak. Figure S22 provides results for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.3 Hz, so this is the highest high-pass cutoff frequency
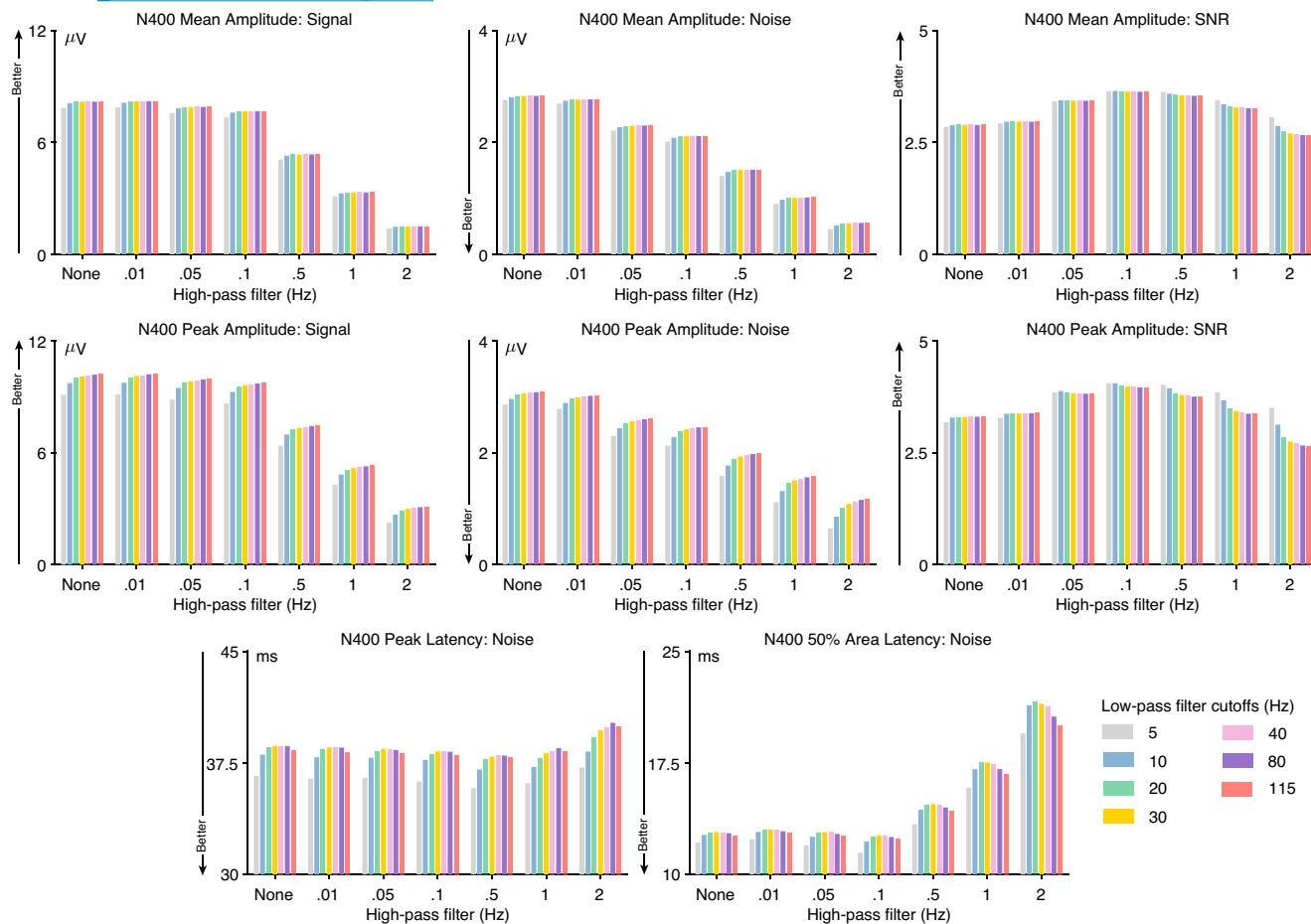
FIGURE 10 N400 data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (unrelated minus related). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

that we would recommend for data sets like the ERP CORE LRP experiment.

3.6.2 | Optimal filters for LRP mean amplitude and peak amplitude

As shown in Figure 12, low-pass filters had little impact on the SNR_{SME} for LRP mean amplitude scores except that the SNR_{SME} was reduced for cutoffs less than 30 Hz. For peak amplitude scores, the SNR_{SME} was also reduced for low-pass cutoff frequencies greater than 30 Hz (except when combined with an inadvisably high high-pass cutoff).

For high-pass filtering, the SNR_{SME} increased as the high-pass cutoff increased for both mean and peak amplitude (see also Supplementary Figure S22). However, waveform distortion also increased as the high-pass cutoff increased. A high-pass cutoff of 0.3 Hz yielded the highest

SNR_{SME} while remaining below our 5% threshold for artifactual peak distortion. At this high-pass cutoff frequency, a low-pass cutoff of 30 Hz yielded the highest SNR_{SME} for peak amplitude. Thus, for data sets like the ERP CORE LRP experiment, we recommend a high-pass cutoff of 0.3 Hz combined with a low-pass cutoff of 30 Hz for peak amplitude or ≥30 Hz for mean amplitude. For noisier data, decreasing the low-pass cutoff to 20 Hz would slightly improve the SNR_{SME} for peak amplitude scores (see Supplementary Figures S23–S25).

3.6.3 | Optimal filters for LRP peak latency and 50% area latency

The bottom row of Figure 12 shows how the RMS(SME) values varied across filter settings for LRP peak latency and 50% area latency scores. For peak latency, low-pass filtering had a substantial impact on the RMS(SME)
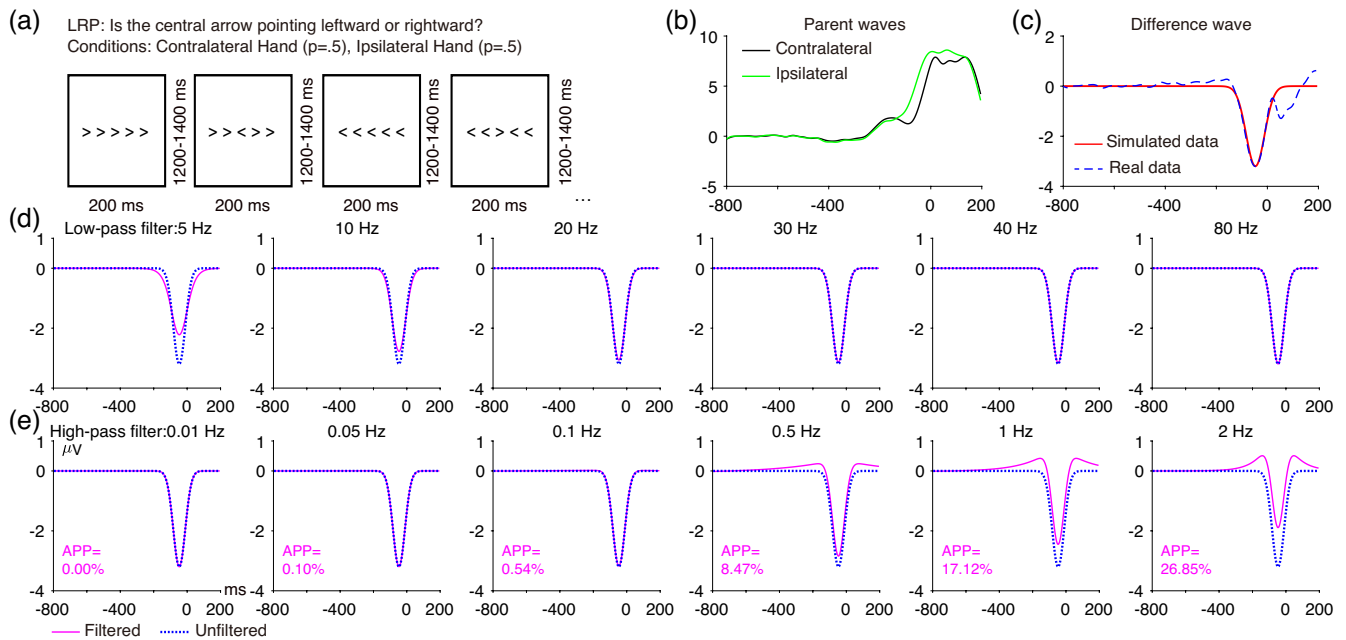
FIGURE 11 (a) LRP flankers task. (b) Grand average ERP waveforms from the C3/C4 electrode sites for the contralateral and ipsilateral trials. (c) Grand average contralateral-minus-ipsilateral difference wave at C3/C4, along with the simulated LRP difference wave (Gaussian function, mean = −47 ms, $SD = 36$ ms, and peak amplitude = −3.2 μV). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the artificial waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

values but high-pass filtering had very little impact. The combination of a 5 Hz low-pass cutoff and a high-pass cutoff of 0.3 Hz produced the lowest noise without exceeding our 5% threshold for artifactual peak percentage (see also Supplementary Figure S22), so this is our recommendation for LRP peak latency.

For 50% area latency, the filter cutoffs had remarkably little effect on the RMS(SME) metric of noise (except when the high-pass cutoff exceeded the 5% artifactual peak threshold). The lowest noise value was obtained for a 0.3 Hz high-pass cutoff combined with a 5 Hz low-pass cutoff, so that is our recommendation. For both peak latency and area latency, a 5 Hz low-pass filter leads to a noticeably earlier LRP onset and later LRP offset, so a 10 Hz high-pass cutoff may be more appropriate when onset times are important for addressing the scientific goals of a study.

3.7 | The error-related negativity

As shown in Figure 13a, the error-related negativity (ERN) was obtained from the same flankers paradigm used for the LRP (as described in Section 3.6). Whereas the LRP was defined as the difference between the

contralateral and ipsilateral hemispheres (relative to the response hand), the ERN was defined as the difference between correct trials and error trials at a single midline electrode site (FCz). Each participant made correct responses on approximately 320–360 trials and incorrect responses on approximately 40–80 trials. Response-locked averages were used for the ERN. Figure 13b presents the grand average ERP waveforms for trials with correct versus incorrect responses, and Figure 13c shows the incorrect-minus-correct difference wave overlaid with the simulated ERN.

3.7.1 | ERN waveform distortion

Figure 13d,e shows the effects of several low-pass and high-pass filters on the simulated ERN waveform. For low-pass filters, the ERN became noticeably smaller and broader for cutoffs below 10 Hz. For high-pass filters, cutoffs of 0.5 Hz and higher clearly distorted the LRP waveform, producing distinct positive peaks before and after the ERN peak. Figure S26 provides results for a denser sampling of high-pass cutoff frequencies. The 5% threshold for the artifactual peak percentage was exceeded for high-pass cutoff frequencies greater than 0.4 Hz (when
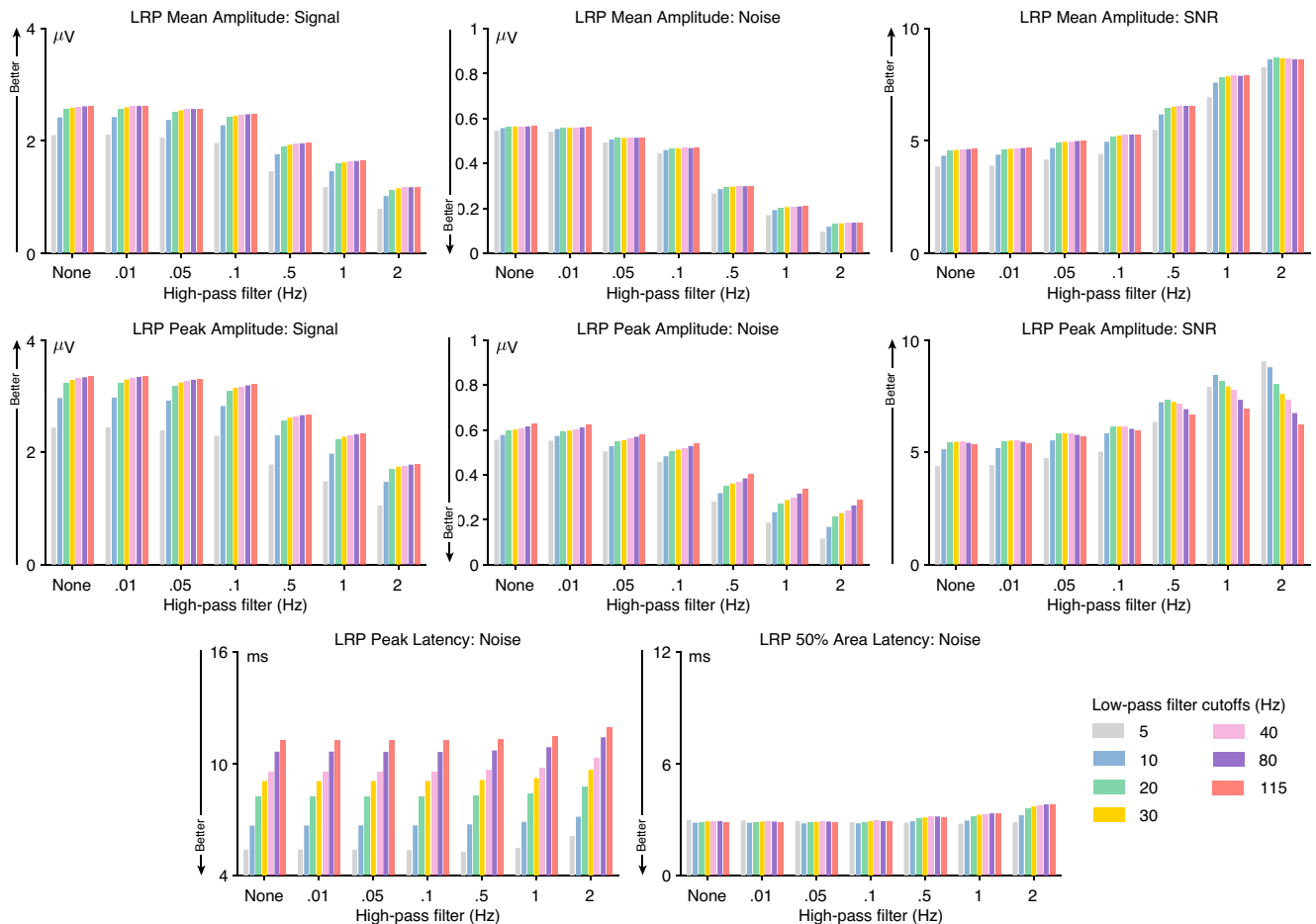
FIGURE 12 LRP data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (contralateral minus ipsilateral). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

combined with most low-pass cutoffs), so this is the highest high-pass cutoff frequency that we would recommend for data sets like the ERP CORE ERN experiment.

### 3.7.2 | Optimal filters for ERN mean amplitude and peak amplitude

As shown in Figure 14, low-pass filters had relatively little impact on the SNR_{SME} for ERN mean amplitude scores, except that the SNR_{SME} was reduced for cutoffs less than 20 Hz. For peak amplitude scores, the SNR_{SME} was maximal at 20 Hz (except when combined with an inadvisably high high-pass cutoff). For noisier data, a low-pass cutoff of 20 Hz yields the highest SNR_{SME} for peak amplitude scores (see Supplementary Figures S27–S29).

For high-pass filtering, the SNR_{SME} increased as the high-pass cutoff increased for both mean and peak

amplitude (see also Supplementary Figure S26). However, waveform distortion also increased as the high-pass cutoff increased. A high-pass cutoff of 0.4 Hz yielded the highest SNR_{SME} for both mean and peak amplitude while remaining below our 5% threshold for artifactual peak distortion. Thus, for data sets like the ERP CORE LRP experiment, we recommend a high-pass cutoff of 0.4 Hz combined with a low-pass cutoff of 20 Hz for peak amplitude or ≥20 Hz for mean amplitude.

3.7.3 | Optimal filters for ERN peak latency and 50% area latency

The bottom row of Figure 14 shows how the RMS(SME) values varied across filter settings for the ERN peak latency and 50% area latency scores. For both peak latency and 50% area latency, the lowest noise levels were
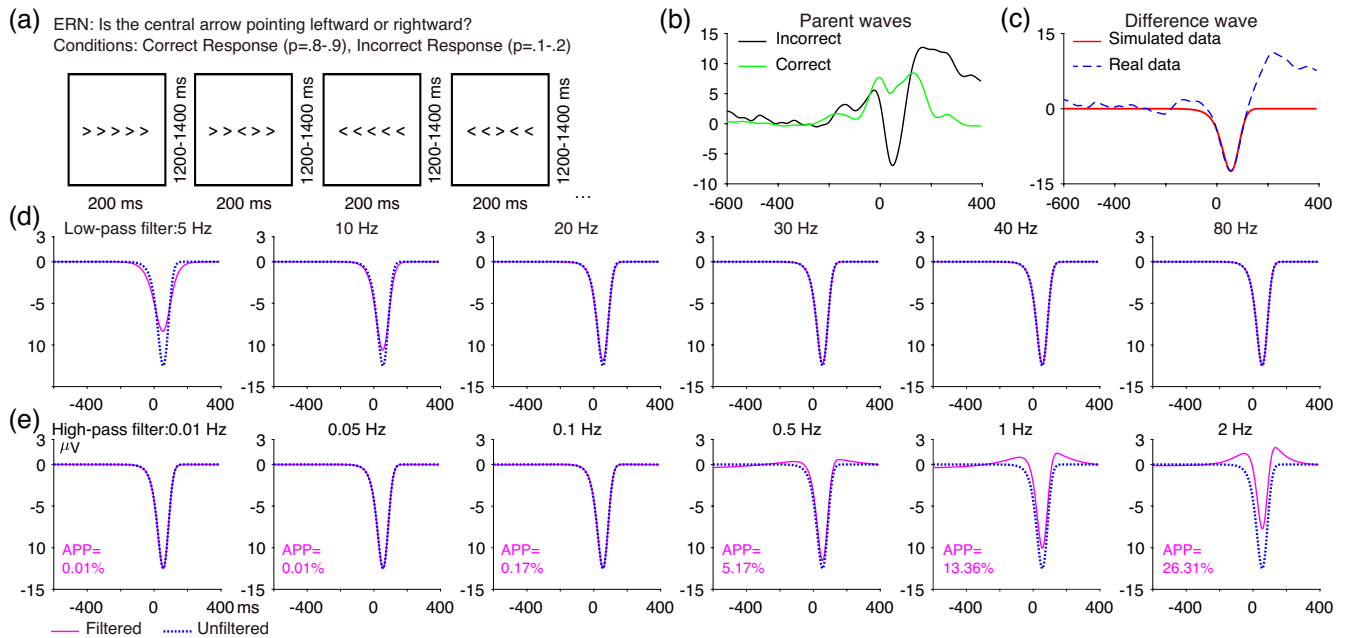
FIGURE 13 (a) ERN flankers task. (b) Grand average ERP waveforms from the FCz electrode site for correct and incorrect responses. (c) Grand average incorrect-minus-correct difference wave at FCz, along with its simulated ERP difference wave (Ex-Gaussian function, mean $= 76$ ms, $SD = 26$ ms, $\lambda = -300$ ms, and peak amplitude $= -12.5\,\mu V$). (d) Artificial waveform overlaid with the low-pass filtered version of that waveform for several different filter cutoffs. (e) Artificial waveform overlaid with the high-pass filtered version of that waveform for several different filter cutoffs. The number next to each high-pass filtered waveform is the artifactual peak percentage (APP). Note that the artificial waveforms were preceded and followed by 1000 ms of zero values to avoid edge artifacts. All the filters used here were noncausal Butterworth filters with a slope of 12 dB/octave, and cutoff frequencies indicate the half-amplitude point.

produced by a low-pass cutoff of 5 Hz combined with no high-pass filtering (see also Supplementary Figure S26), so this is our recommendation. However, a 5 Hz low-pass filter produces a noticeably earlier ERN onset and later ERN offset, so a 10 Hz high-pass cutoff may be more appropriate when onset times are important for addressing the scientific goals of a study.

4 | DISCUSSION

The present study examined the impact of a broad range of low-pass and high-pass filter cutoffs on seven common ERP components recorded from a set of neurotypical young adults. We quantitatively assessed the effects of a broad range of filter cutoffs on waveform distortion and data quality (signal, noise, and signal-to-noise ratio). The observed effects led to recommendations for optimal filter cutoffs that maximize data quality while producing minimal waveform distortion (Table 1). By repeating the analyses after adding artificial noise to the data, we were also able to provide recommendations for data with moderately greater noise levels. These recommendations were based on objective, data-driven criteria, and similar studies that follow these recommendations should have improved data quality

and therefore greater statistical power without inducing problematic waveform distortion. The data and the analysis scripts are available online at https://osf.io/z3hfp, making it straightforward for researchers to repeat these analyses with different filters, different noise levels, or different thresholds for waveform distortion. In addition, the companion paper (Zhang et al., 2024) describes how the present approach can be applied to other data sets using tools available in version 9.2 and higher of ERPLAB Toolbox.

We would like to emphasize that the present results demonstrate that previous recommendations (Acunzo et al., 2012; Duncan et al., 2009; Luck, 2014; Widmann et al., 2015) were less than optimal, especially recommendations of high-pass cutoffs of ≤0.1 Hz for relatively narrow components such as N170 and MMN. This shows the value of using a quantitative, data-driven approach for selecting filter settings.

We would also like to emphasize that the optimal filter settings for a given experiment depend on the waveshape of the ERP components. Thus, researchers should use the recommended filter settings shown in Table 1 only for studies that yield similar waveshapes to those in ERP CORE data. It is also important to note that the present approach to identifying optimal filter parameters requires that the component of interest be isolated from the ERP
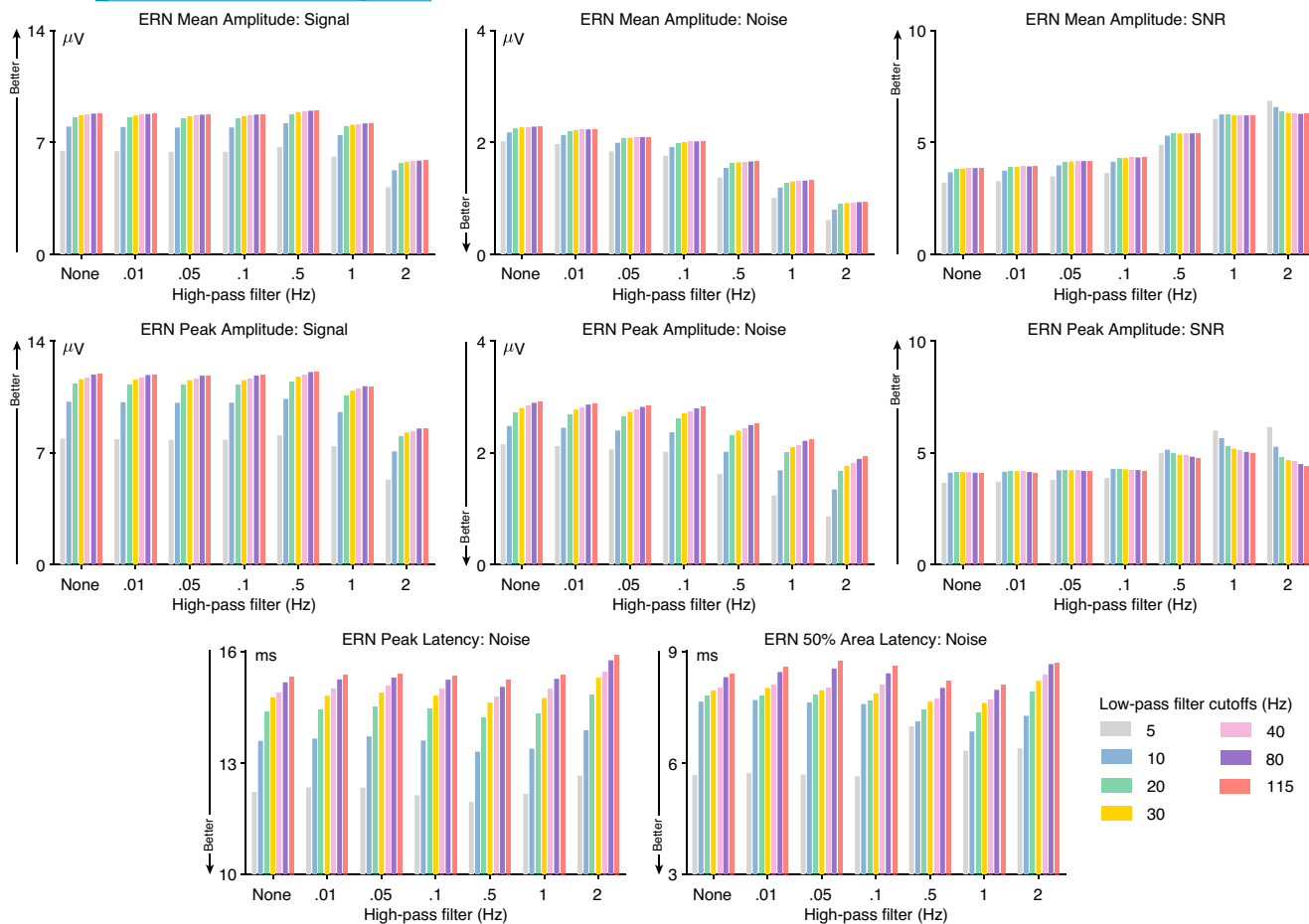
FIGURE 14 ERN data quality metrics for four different scoring methods and several combinations of high-pass filter cutoffs (0, 0.01, 0.05, 0.1, 0.5, 1, and 2 Hz) and low-pass filter cutoffs (5, 10, 20, 30, 40, 80, and 115 Hz). The signal was defined as the score (e.g., peak amplitude) obtained from the grand average ERP difference wave (correct minus incorrect). The noise was defined as the root mean square (RMS) of the single-participant standardized measurement error (SME) for that score. The signal-to-noise ratio (SNR) was computed as the signal divided by the noise. SNR is unitless. For latency scores, the signal is not consistently reduced by filtering, so only the RMS(SME) value is provided for the peak latency and 50% area latency scores.

waveform. The present study used difference waves for this purpose. If multiple components are present, additional steps must be taken to ensure that the filtering of one component does not create unacceptable levels of distortion in other components (for details, see the companion paper [Zhang et al., 2024]).

4.1 | Effects of filter cutoffs on data quality and waveform distortion

In some cases, we found fairly large differences in noise levels and SNR for different filter settings. For example, the SNR for N170 peak amplitude was reduced by 50% when a 5 Hz low-pass filter was applied relative to when the cutoff was 30 Hz. Similarly, the SNR for P3 mean amplitude was 20% greater for the high-pass cutoff of 0.2 Hz recommended here than for the cutoff of 0.01 Hz recommended by Duncan et al. (2009) and Luck (2005).

Moreover, increasing the high-pass cutoff from 0.2 Hz to 1 Hz for the P3 increased the size of the artifactual peak by a factor of almost 10 (see Supplementary Figure S14). These results provide quantitative evidence that some filter cutoffs are much better than others.

On the other hand, small changes in cutoff frequencies typically produced only small changes in data quality and waveform distortion, so slight deviations from the recommendations shown in Table 1 would be expected to have only a minor impact. For example, our recommended high-pass filter cutoff for P3 mean amplitude was 0.2 Hz because the SNR was higher at 0.2 Hz than at 0.1 Hz and because cutoffs above 0.2 Hz led to an artifactual peak percentage that exceeded our 5% cutoff (see supplemental Figure S14). However, the SNR was only slightly better at 0.2 Hz than at 0.1 Hz (3.0 vs. 2.9), which is unlikely to be a meaningful difference in most studies. Moreover, it is possible that the SNR would be slightly better at 0.1 Hz than at 0.2 Hz in a different sample of participants. Similarly,

a cutoff of 0.2 Hz might exceed our 5% threshold for the artifactual peak percentage in a study with a broader P3 waveform. However, the 0.2 Hz cutoff for P3 mean amplitude and the other values shown in Table 1 represent the current best guesses for optimal filter cutoffs for studies like the ERP CORE, and it would be sensible to use these cutoffs unless there is a good reason to use something else or until quantitative recommendations become available from a larger data set.

4.2 | Deviating from the present recommendations

One such reason for deviating from these recommendations would be the presence of very different noise levels. By adding noise to the ERP CORE data, we were able to demonstrate that more aggressive filtering of high frequencies (i.e., lower low-pass cutoffs) becomes optimal when there is more high-frequency or broadband noise in the data. However, the addition of moderately large noise did not impact our recommendations for high-pass filtering. This reflects the fact that our high-pass cutoff recommendations were mainly driven by the artifactual peak percentage, not by the SNR. For example, the best SNR for P3 mean amplitude was produced by a 0.5 Hz high-pass cutoff, but we recommended a 0.2 Hz cutoff to avoid exceeding the 5% threshold for artifactual peak percentage. For studies with very high levels of low-frequency noise and/or small numbers of trials, the increased SNR resulting from a 0.5 Hz high-pass cutoff might make it worthwhile to increase the threshold for artifactual peak percentage. In these cases, an artifactual peak of even 10% would be unlikely to be statistically significant and lead to incorrect conclusions.

Researchers might also choose to use the same filter cutoffs for different scoring methods, opting for simplicity over optimal data quality. For example, a researcher who is measuring both the mean amplitude and the peak latency of the ERN might use a bandpass of 0.4–10 Hz for both scores, even though a bandpass of 0.4–20 Hz is recommended for ERN mean amplitude. This would produce only a 2% drop in SNR (see Supplementary Figure S26), which is unlikely to change the results of a single study. By contrast, if the optimal LRP peak latency bandpass of 0.3–5 Hz is also applied to LRP mean amplitude data, this would produce a 17% reduction in SNR relative to the optimal mean amplitude bandpass of 0.3–30 Hz (see Supplementary Figure S22). This is more likely to be problematic.

In the end, decisions about filter cutoffs should reflect the scientific goals of a given study along with the nature of the data being filtered. The present approach to filter selection allows researchers to make these decisions in an informed manner.

4.3 | General patterns

Several general patterns can be seen in the optimal filter values shown in Table 1. First, the optimal filter settings depend on the specific scoring method. Mean amplitude scores are largely insensitive to high-frequency noise, whereas high-frequency noise can distort peak amplitude and peak latency (Clayson et al., 2013). Consequently, a low-pass filter is ordinarily unnecessary for mean amplitude, but it can improve data quality for peak amplitude and peak latency. For example, whereas low-pass filtering does not improve the SNR for N2pc mean amplitude, applying a 10 Hz low-pass filter reduces the noise level of N2pc peak latency scores by approximately 20% when the data are contaminated by substantial broadband noise (see Supplementary Figure S12). However, low-pass filtering had much less effect on N2pc 50% area latency scores.

Our results also revealed that the optimal filter settings depend on the shape of the ERP component that is being analyzed. As a component becomes broader, it suffers less amplitude reduction from low-pass filtering but more amplitude reduction from high-pass filtering. In addition, high-pass filters produce larger artifactual peaks in broader components than in narrower components. Consequently, relatively high low-pass and high-pass cutoffs are optimal for narrower components (e.g., N170), and relatively low low-pass and high-pass cutoffs are optimal for broader components (e.g., P3). Researchers can use this finding to make informed guesses about optimal cutoffs for components that were not examined in the present study.

4.4 | Limitations and future directions

The present study has some important limitations. First, our recommendations may not be applicable for recording setups that yield very different types or levels of noise. For example, data collected with dry electrodes often contain substantially more noise (Kam et al., 2019; Li et al., 2020; Shad et al., 2020). Similarly, our recommendations may not generalize to mobile experiments, for which a previous study has shown that higher high-pass filter cutoffs are beneficial (Klug & Gramann, 2021). However, it would be important for future research on optimal filter values for these systems to use an approach like we have used, which quantifies the data quality for the specific score of interest and also quantifies the amount of waveform distortion.

PSYCHOPHYSIOLOGY SPR

Another limitation is that the data used in the present study were obtained from highly cooperative college students, and the optimal filter settings may differ considerably in other populations, such as infants or individuals with clinical disorders. For example, in infant EEG recordings, higher high-pass filter cutoffs (e.g., 1 Hz) are currently viewed as appropriate (Debnath et al., 2020; Gabard-Durnam et al., 2018). However, such high cutoffs may induce problematic waveform distortion and may attenuate the signal more than the noise. Consequently, a careful quantitative analysis of a broad set of paradigms and scoring methods would be needed to determine the optimal filter settings for infants, specific clinical populations, and other groups that are likely to be quite different from the college students tested in the ERP CORE.

Finally, the present study assessed only four scoring methods, and it did not consider onset latencies or mass univariate approaches. Moreover, it focused only on the Butterworth family of infinite impulse response filters, and it is possible that the results would be different for other filter families (e.g., finite impulse response filters). It would be valuable for future research to determine the optimal filter settings for other families of filters and other ERP scoring algorithms.

AUTHOR CONTRIBUTIONS

Guanghui Zhang: Conceptualization; methodology; software; validation; visualization; writing – original draft; writing – review and editing. **David R. Garrett:** Methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Steven J. Luck:** Conceptualization; funding acquisition; methodology; project administration; software; supervision; validation; visualization; writing – original draft; writing – review and editing.

FUNDING INFORMATION

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

https://osf.io/z3hfp.

ORCID

Guanghui Zhang https://orcid.org/0000-0003-0134-3614
Steven J. Luck https://orcid.org/0000-0002-3725-1474

REFERENCES

Acunzo, D. J., MacKenzie, G., & Van Rossum, M. C. W. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. https://doi.org/10.1016/j.jneumeth.2012.06.011

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*(2), 174–186. https://doi.org/10.1111/psyp.12001

De Cheveigné, A., & Nelken, I. (2019). Filters: When, why, and how (not) to use them. *Neuron*, *102*(2), 280–293. https://doi.org/10.1016/j.neuron.2019.02.039

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*(6), e13580. https://doi.org/10.1111/psyp.13580

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L., & Levin, A. R. (2018). The Harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, *12*, 97. https://doi.org/10.3389/fnins.2018.00097

Hamming, R. W. (1998). *Digital filters*. Dover Publications.

Kam, J. W. Y., Griffin, S., Shen, A., Patel, S., Hinrichs, H., Heinze, H.-J., Deouell, L. Y., & Knight, R. T. (2019). Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes. *NeuroImage*, *184*, 119–129. https://doi.org/10.1016/j.neuroimage.2018.09.012

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, *225*, 117465. https://doi.org/10.1016/j.neuroimage.2020.117465

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *European Journal of Neuroscience*, *54*(12), 8406–8420. https://doi.org/10.1111/ejn.14992

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*(3), 1217–1241. https://doi.org/10.1214/aos/1176347265

Li, G.-L., Wu, J.-T., Xia, Y.-H., He, Q.-G., & Jin, H.-G. (2020). Review of semi-dry electrodes for EEG recording. *Journal of Neural Engineering*, *17*(5), 051004. https://doi.org/10.1088/1741-2552/abbd50

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luck, S. J. (2022). *Applied event-related potential data analysis*. LibreTexts. https://doi.org/10.18115/D5QG92

Luck, S. J., Stewart, A. X., Simmons, A. M., & Rhemtulla, M. (2021). Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology*, *58*(6), e13793. https://doi.org/10.1111/psyp.13793

Rousselet, G. A. (2012). Does filtering preclude us from studying ERP time-courses? *Frontiers in Psychology*, *3*, 131. https://doi.org/10.3389/fpsyg.2012.00131

Shad, E. H. T., Molinas, M., & Ytterdal, T. (2020). Impedance and noise of passive and active dry EEG electrodes: A review. *IEEE Sensors Journal*, *20*(24), 14565–14577. https://doi.org/10.1109/JSEN.2020.3012394

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Tanner, D., Norton, J. J. S., Morgan-Short, K., & Luck, S. J. (2016). On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *Journal of Neuroscience Methods*, *266*, 166–170. https://doi.org/10.1016/j.jneumeth.2016.01.002

Van Driel, J., Olivers, C. N. L., & Fahrenfort, J. J. (2021). High-pass filtering artifacts in multivariate classification of neural time series data. *Journal of Neuroscience Methods*, *352*, 109080. https://doi.org/10.1016/j.jneumeth.2021.109080

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 365. https://doi.org/10.3389/fpsyg.2011.00365

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—A practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S., & Cohen, J. D. (2007). Theta phase resetting and the error-related negativity. *Psychophysiology*, *44*(1), 39–49. https://doi.org/10.1111/j.1469-8986.2006.00482.x

Zhang, G., Garrett, D. R., & Luck, S. J. (2024). Optimal filters for ERP research I: A general approach for selecting filter settings. *Psychophysiology*, *00*, e14531. https://doi.org/10.1111/psyp.14531

Zhang, G., & Luck, S. J. (2023). Variations in ERP data quality across paradigms, participants, and scoring procedures. *Psychophysiology*, *60*(7), e14264. https://doi.org/10.1111/psyp.14264

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Example of waves and spectra for data with or without additional noise.

Figure S2. N170 artifactual peak percentage and data quality metrics.

Figure S3. Data quality metrics for the N170 component with 60 Hz line noise.

Figure S4. Data quality metrics for the N170 component with white noise.

Figure S5. Data quality metrics for the N170 component with pink noise.

Figure S6. MMN artifactual peak percentage and data quality metrics.

Figure S7. Data quality metrics for the MMN component with 60 Hz line noise.

Figure S8. Data quality metrics for the MMN component with white noise.

Figure S9. Data quality metrics for the MMN component with pink noise.

Figure S10. N2pc artifactual peak percentage and data quality metrics.

Figure S11. Data quality metrics for the N2pc component with 60 Hz line noise.

Figure S12. Data quality metrics for the N2pc component with white noise.

Figure S13. Data quality metrics for the N2pc component with pink noise.

Figure S14. P3 artifactual peak percentage and data quality metrics.

Figure S15. Data quality metrics for the P3 component with 60 Hz line noise.

Figure S16. Data quality metrics for the P3 component with white noise.

Figure S17. Data quality metrics for the P3 component with pink noise.

Figure S18. N400 artifactual peak percentage and data quality metrics.

Figure S19. Data quality metrics for the N400 component with 60 Hz line noise.

Figure S20. Data quality metrics for the N400 component with white noise.

Figure S21. Data quality metrics for the N400 component with pink noise.

Figure S22. LRP artifactual peak percentage and data quality metrics.

Figure S23. Data quality metrics for the LRP component with 60 Hz line noise.

Figure S24. Data quality metrics for the LRP component with white noise.

Figure S25. Data quality metrics for the LRP component with pink noise.

Figure S26. ERN artifactual peak percentage and data quality metrics.

Figure S27. Data quality metrics for the ERN component with 60 Hz line noise.

Figure S28. Data quality metrics for the ERN component with white noise.

Figure S29. Data quality metrics for the ERN component with pink noise.