On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis

Article *in* Journal of Neuroscience Methods · January 2016

DOI: 10.1016/j.jneumeth.2016.01.002

4 authors:

Darren Tanner
University of Illinois, Urbana-Champaign

James Norton
University of Illinois, Urbana-Champaign

Kara Morgan-Short
University of Illinois at Chicago

Steven J Luck
University of California, Davis

Journal of Neuroscience Methods xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis

Darren Tanner[a,b,*], James J.S. Norton[b], Kara Morgan-Short[c], Steven J. Luck[d]

[a] Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[b] Neuroscience Program and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[c] Departments of Hispanic and Italian Studies and Psychology, University of Illinois at Chicago, Chicago, IL, United States
[d] Center for Mind & Brain and Department of Psychology, University of California, Davis, CA, United States

In Tanner, Morgan-Short and Luck (2015; henceforth TMSL) we demonstrated how commonly-used high-pass filter settings can distort ERP (and analogously ERMF) data, and that these distortions can lead to spurious conclusions about the nature of the cognitive processes engaged during the experimental task. We appreciate Maess, Schröger, and Widmann's interest in our work, and we thank them for their thoughtful commentary. Indeed, we feel that open discussion of these issues – and importantly empirical demonstration of the benefits and pitfalls of high-pass filtering, baseline correction, and other issues – will benefit the field by helping establish a set of best practices for signal processing in ERP research. Establishing a consistent best-practices approach to filtering and ERP analysis more generally will help ensure cross-study comparability within sub-fields of ERP research and lead more reliable, consistent, and replicable results.

Maess et al. raise two major points in response to our article. First, they argue that our original test data were not optimally suited to show the benefits of high-pass filtering because they simply did not contain enough low-frequency noise. Second, they argue that high-pass filtering should replace the common practice of baseline correction in ERP research, contra our recommendations. We will respond to both of these arguments here, as well as a point they raise about criteria for detecting filter-induced artifacts.

1. Were our data *too good*?

First, Maess et al. suggest that our data had too little low-frequency noise to show benefits of high-pass filtering, and instead show only the pitfalls of filtering-namely induced artifactual effects. They base this claim on visual inspection of our DC, 0.01 Hz and 0.1 Hz high-pass filtered ERP waveforms, and note that there was very little difference between them. They therefore argue that there must have been nearly no low-frequency noise in the data, such that these modest filters had nearly no effect on the quality of the data. However, inspection of grand mean ERPs is not a valid means of assessing the presence of low-frequency noise in the data. Low-frequency components that are truly noise should have a random phase with respect to the onset of any given stimulus, so that averaging will attenuate the low-frequency noise. This will be especially true with a grand average that is based on many hundreds of total trials. Thus, it is not possible to assess the amount of low-frequency noise by examining the effects of high-pass filtering on averaged data.

There are three clear pieces of evidence that substantial low-frequency noise was indeed present in the EEG. The first piece of evidence was presented in TMSL: Our Monte Carlo simulations showed that high-pass filtering at 0.01 or 0.1 Hz increased the statistical power of the N400 and P600 analyses. This would not have been possible in the absence of low-frequency noise in the EEG.

A second piece of evidence is that the raw EEG contained easily visible low-frequency drifts when viewed with a long time scale. Fig. 1A illustrates this by showing a 1000-s interval of raw EEG data for five randomly-selected individual participants from TMSL at electrode Pz. It is quite clear that the voltage is drifting slowly over time in all five cases. To formalize this and provide a third piece of evidence, we used the Fourier transform to compute the amplitude spectral density (ASD) of the unfiltered EEG data and the data after high-pass filtering with half amplitude cutoffs of 0.1 and 1 Hz for all 24 participants in TMSL. As shown in Fig. 1B, substantial low-frequency activity was present in the unfiltered data. Indeed, the low-frequency noise was several times greater than the 60-Hz noise. It is also clear that even a modest high-pass frequency cutoff produced a marked attenuation of the amplitudes in these very low frequencies. Thus, the EEG in TMSL included substantial low-frequency activity; high-pass filtering attenuated this noise, which in turn led to improved statistical power. These findings provide overwhelming evidence against the proposal of Maess et al. that our data did not contain substantial low-frequency activity.

Maess et al. note that some EEG/ERP studies – such as those involving children – may yield higher levels of low-frequency noise than observed in TMSL. This is certainly true. For example, Kappenman and Luck (2010) showed that recording with

* Corresponding author at: Department of Linguistics University of Illinois at Urbana-Champaign FLB 4080, MC-168 707 S. Mathews Ave. Urbana, IL 61801 USA.
E-mail address: dstanner@gmail.com (D. Tanner).

http://dx.doi.org/10.1016/j.jneumeth.2016.01.002

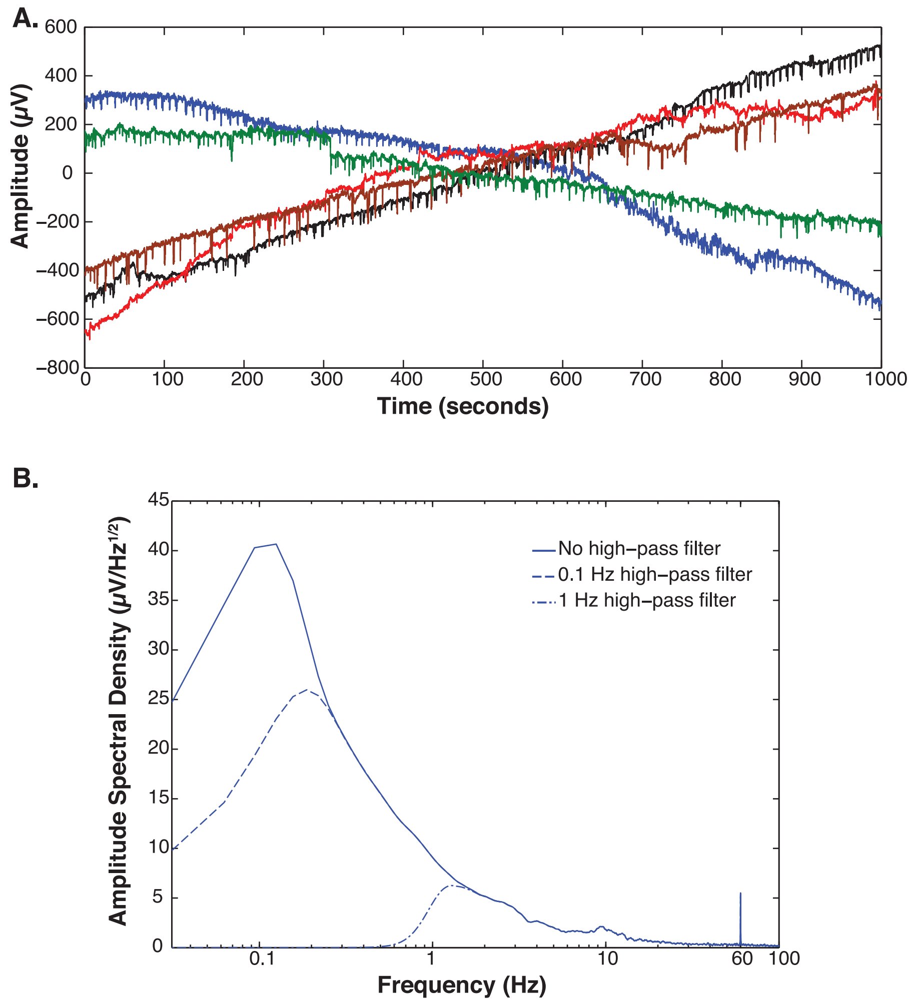

Fig. 1. (A) Raw EEG from five randomly selected participants from TMSL. 1000 s of time is depicted. To highlight low frequencies in the data, a 5 Hz low-pass filter (−6 dB cut-off, 24 dB/octave roll-off) was applied. The mean value of each individual's 1000 s segment was removed prior to plotting. (B) Amplitude spectral density (ASD) of the raw EEG for three filter settings (DC, 0.1 Hz, and 1 Hz) at electrode Pz, averaged across all 24 participants in TMSL. To compute the ASD, first the mean DC value was removed from each individual's data. A fourth order high-pass IIR Butterworth filter (−6 dB cut-off) was then applied to the continuous EEG data for the 0.1 Hz and 1 Hz filter conditions. Individuals' data were then split into 10-s segments with no overlap and the mean DC value was removed from each segment for the no-filter condition. Following the recommendations in Maess et al.'s commentary, as well as Widmann et al. (2015), we did not re-DC-correct the filtered data after segmenting, as filters should suppress DC offsets. Before converting to the frequency domain, a Hanning window was applied to each segment to reduce spectral leakage. The segments were transformed into the frequency domain using the Fast Fourier Transform, multiplied by their complex conjugate to obtain a measure of power, and normalized. These power estimates were then averaged across segments within an individual. After averaging, the positive frequencies were doubled to obtain the single-sided power spectral density and the square root was taken to obtain an ASD for each participant. The individual ASD estimates were then averaged across all 24 participants from TMSL. Our method is equivalent to the MATLAB *pwelch* function, with the exception that the DC component was removed from the individual segments in the no filter condition. The *x*-axis is plotted on a log scale to enhance visibility of the low frequencies.

high electrode impedances in a warm, humid recording environment can dramatically increase the amount of low-frequency noise. Kappenman and Luck (2010) further showed that high-pass filters can improve statistical power for large, easily measurable effects like the P300 in these less-than-ideal conditions. However, they also showed that severe high-pass filtering (with cutoff frequencies ≥0.5 Hz) can dramatically distort the resulting ERP waveforms, reducing P300 amplitude and producing an artifactual negative deflection before the P300. Thus, even when high levels of low-frequency noise are present, high-pass filters with cutoffs greater than approximately 0.1 Hz may create more problems than they solve.

2. Is there a good way to identify filter distortions?

Maess et al. argue that filter cutoffs should be selected by considering both the nature of the noise to be filtered and the nature of the signal; the general goal is to choose a filter that is the best compromise between maximal noise reduction and minimal distortion of the signal. In principle, we agree. However, this assumes that we know the properties of the signal *a priori*. In most experiments the signal is not known and the goal of the experiment is to determine the properties of the signal. Moreover, the observed data will be a mixture of signal and noise, so it will be difficult to use the observed data to determine the nature of the signal. However, it may be possible to filter artificial waveforms to determine whether an observed pattern of results in filtered data could potentially be explained by filter artifacts (see, e.g., Fig. 2 in TMSL). This approach was used very effectively by Yeung et al. (2007) to demonstrate that previous conclusions regarding the role of theta oscillations in the error-related negativity could potentially be explained by filter artifacts.

A second issue with the post hoc waveform comparison approach advocated for by Maess et al. is that it can lead to problems

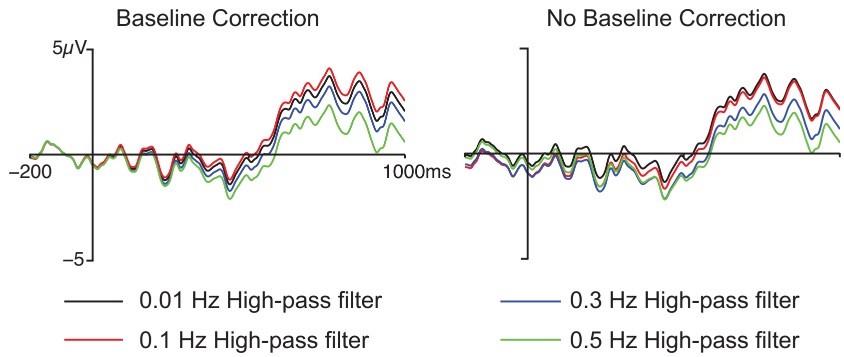

Fig. 2. Ungrammatical minus grammatical difference waves showing the effects of filtering and baseline correction on data from the syntactic condition reported in TMSL. Difference waves on the left-hand side were baseline corrected using a 200 ms prestimulus interval. Difference waves on the right-hand side were not baseline corrected. Note that in the data without baseline correction, there were no systematic differences between conditions in the pre-stimulus interval in the 0.01 Hz filtered data. The 0.1 and 0.3 Hz high-pass filters, however, caused systematic differences between conditions in the pre-stimulus interval.

of researcher bias. As discussed above, the true ERP signal is usually unknown to the researcher, and in TMSL we showed that high-pass filter artifacts can lead to theoretically viable (but bogus) ERP effects. Therefore, inspection of the data after filtering in multiple different ways could in some cases lead the researcher to choose the filter setting that either best fits her/his *a priori* hypotheses or that provides the most novel outcome, even if this outcome is due simply to filtering artifacts. Unless extensive simulations are conducted for a given experiment, we argue that it is safer to use less severe filtering, with the balance tilted toward minimal waveform distortion and minimal experimenter bias, even if this leads to a cost in statistical power.

3. Does high-pass filtering overcome the problems of baseline correction?

Maess et al. correctly point out that baseline correction can lead to spurious effects (see Handy, 2005; Luck, 2014, for examples). As an alternative, Maess et al. (see also Widmann et al., 2015) suggest using a sufficiently low high-pass cutoff frequency (i.e., low enough to avoid spurious filter effects) in lieu of baseline correction. In theory, the filter will remove DC offsets and slow drifts, bringing the signal to the true zero level during the prestimulus period. However, as we will demonstrate here, there are notable problems with this approach as well – even when relatively modest high-pass filters are used.

One problem is that the bidirectional, noncausal filters usually used in ERP research will cause effects that occur after stimulus onset to be pushed backward in time, potentially into the prestimulus interval. As we will demonstrate here, this can happen even when relatively modest high-pass filters are used. Thus, the baseline and early poststimulus interval may become contaminated by poststimulus effects, even in well-controlled experimental designs that contain no prestimulus differences in the raw data (see also Acunzo et al., 2012).

This is illustrated in Fig. 2, which shows difference waves (ungrammatical minus grammatical) from the syntactic condition in TMSL with different high-pass filter cutoffs, both with and without baseline correction. With baseline correction, the mean voltage during the prestimulus period did not vary as a function of the filter cutoff; this is a good approximation of the raw, unfiltered data, which showed little differential prestimulus activity between the grammatical and ungrammatical conditions (see Fig. 3 and Fig. 4 in TMSL). Without baseline correction, however, the P600 effect led to a negative offset prior to stimulus onset that continued into the immediate poststimulus interval when the high-pass filter cutoff was ≥0.1 Hz, and this offset was not present with the milder 0.01 Hz filter. This could lead to the (obviously false) conclusion that the participants had a precognition about whether the upcoming word would be grammatical or ungrammatical. Thus, the application of a high-pass filter without baseline correction can lead to artifactual differences between conditions in the prestimulus baseline period, even when none was present in the raw data.

As mentioned previously, it can be dangerous to use real data to assess filter artifacts because the truth is not known for the real data. We therefore performed some simple simulations to demonstrate how high-pass filters can lead to artifactual effects during the prestimulus period when used without baseline correction. Specifically, we simulated P600 effects of varying magnitudes and applied several high-pass filters to the simulated waveforms with and without baseline correction. As shown in Fig. 3, even with the relatively mild 0.1 Hz filter recommended by Widmann et al. (2015), the filters led to contamination of the prestimulus interval when applied without baseline correction (even though there was no contamination in the raw data), and the amplitude of the prestimulus contamination varied as a function of the effect magnitude. This fact makes it difficult to use high-pass filters without baseline correction to test for possible problems in the experimental design, as recommended by Maess et al., since it is impossible to know whether any differences in the prestimulus interval were caused by design problems or were caused by filter artifacts. Moreover, the filter artifacts were magnified in amplitude when no baseline correction was applied. The amplification of artifacts with increasing effect amplitudes would have consequences for researchers studying other components like the P300, which can have effect magnitudes in the tens of microvolts. The distortion of the prestimulus baseline was not as severe with higher cutoff frequencies, but these higher cutoffs led to more extreme distortion of the poststimulus waveform. Together, these simulations suggest that mild filtering (e.g., 0.1 Hz) combined with baseline correction is the best solution for well-designed experiments in which there are no systematic differences across conditions in the prestimulus voltage, and that even mild filtering without baseline correction can introduce spurious effects in the prestimulus interval.

The baseline contamination shown in Figs. 2 and 3 would not have occurred if we had used unidirectional, causal filters. However, it is widely known that causal filters produce undesirably latency shifts. Moreover, as shown by Acunzo and colleagues (Acunzo et al., 2012; see also Widmann et al., 2015), causal high-pass filters introduce even more warping and distortions of the poststimulus waveforms than do noncausal filters. Thus, causal

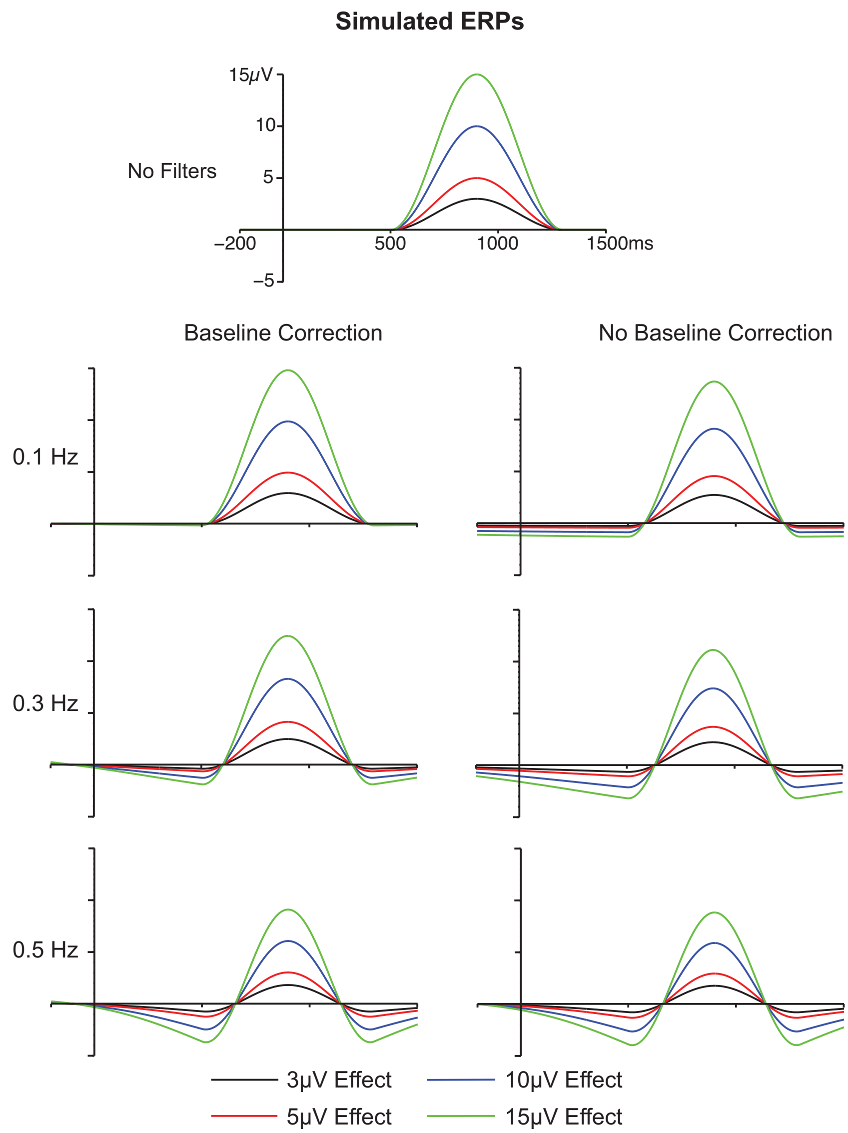

Fig. 3. Simulated ERPs showing the impacts of filtering, baseline correction, and ERP effect amplitude. Raw simulated waveforms for four different effect amplitudes (3 μV, 5 μV, 10 μV, and 15 μV) are depicted at the top. Waveforms in the left column were baseline corrected using a 200 ms prestimulus interval. Waveforms in the right column were not baseline corrected. Filters were fourth order (-6 dB cut-off; 24 dB/octave roll- off) Butterworth IIR filters.

filters are not usually a viable alternative to the noncausal filters used here.

4. Summary and conclusion

The additional analyses and simulations presented here provide clear evidence that (1) the conclusions of TMSL were not a result of a lack of low-frequency noise in the data, (2) it is nontrivial to determine *a priori* whether a given filter setting will be able to minimize noise while avoiding distortion of the averaged ERPs, and (3) high-pass filtering is not an adequate substitute for baseline correction, at least for the kinds of well-controlled experiments assessed in our data and simulations. We therefore conclude that the best approach in most experiments looking at late ERP components such as P300, N400, and P600 is a combination of: (1) well-controlled experimental designs, (2) consideration of the participant population and recording conditions when designing the experiment so that statistical power can be maximized without the use of damaging high-pass filters, (3) application of modest high-pass filters (0.1 Hz or lower) to the continuous EEG to maximize statistical power while not introducing spurious components into the ERP waveforms, and (4) baseline correction of the ERPs to eliminate spreading effects of DC offset suppression that can contaminate the prestimulus and immediate poststimulus interval with spurious effects. We also encourage ERP researchers to become acquainted with the relatively simple mathematics underlying filtering (see Luck, 2014) and to test the effects of filters on simulated data to assess the artifacts that might be produced by the filters.

References

Acunzo DJ, MacKenzie G, van Rossum MCW. Systematic biases in early ERP and ERF components as a result of high-pass filtering. J Neurosci Methods 2012;209:212–8, http://dx.doi.org/10.1016/j.jneumeth.2012.06.011.

Handy. Basic principles of ERP quantification. In: Handy, editor. Event-related potentials: a methods handbook. Cambridge, MA: MIT Press; 2005. p. 33–56.

Kappenman ES, Luck SJ. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology 2010;47:888–904, http://dx.doi.org/10.1111/j.1469-8986.2010.01009.x.

Luck SJ. An introduction to the event-related potential technique. 2nd ed. Cambridge, MA: MIT Press; 2014.

Tanner D, Morgan-Short K, Luck SJ. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology 2015;52:997–1009, http://dx.doi.org/10.1111/psyp.12437.

Widmann A, Schröger E, Maess B. Digital filter design for electrophysiological data—a practical approach. J Neurosci Methods 2015;250:34–46, http://dx.doi.org/10.1016/j.jneumeth.2014.08.002.

Yeung N, Bogacz R, Holroyd CB, Nieuwenhuis S, Cohen JD. Theta phase resetting and the error-related negativity. Psychophysiology 2007;44:39–49, http://dx.doi.org/10.1111/j.1469-8986.2006.00482.x.