High-pass filtering artifacts in multivariate classification of neural time series data

Joram van Driel, Christian N.L. Olivers[*] and Johannes J. Fahrenfort[*]

Institute for Brain and Behaviour Amsterdam

Department of Experimental and Applied Psychology - Cognitive Psychology

Faculty of Behavioural and Movement Sciences

Vrije Universiteit Amsterdam

Keywords: EEG, MEG, high-pass filtering, preprocessing, multivariate pattern classification, decoding

Address for correspondence:
Dr. Johannes J Fahrenfort
fahrenfort.work@gmail.com
Van der Boechorststraat 7
1081 BT Amsterdam

Author note:
*CNLO and JJF share senior authorship. JvD and CNLO designed the experiment. JvD and JJF designed and conducted the simulations and analyses. CNLO, JvD and JJF wrote the paper. This work was supported by ERC-CoG-2013-615423 grant from the European Research Council, and NWO Vici grant 453-16-002 awarded to CNLO.

0. Abstract

The application of time-resolved multivariate pattern classification analyses (MVPA) to EEG and MEG data has become increasingly popular. Traditionally, such time series data are high-pass filtered before analyses, in order to remove slow drifts. Here we show that high-pass filtering should be applied with extreme caution in MVPA, as it may easily create artifacts that result in displacement of decoding accuracy, leading to statistically significant above-chance classification during time periods in which the source is clearly not in brain activity. This is particularly problematic in paradigms that have long trial durations, such as working memory experiments with long retention intervals, where the signal of interest may reside in low parts of the frequency spectrum and thus is more likely to be affected by high-pass filters. In both real and simulated EEG data, we show that spurious decoding may emerge with filter cut-off settings from as modest as 0.1 Hz. We provide an alternative method of removing slow drift noise, referred to as robust detrending (de Cheveigne & Arzounian, 2018), which, when applied in concert with masking of cortical events does not result in the temporal displacement of information. We show that temporal generalization may benefit from robust detrending, without any of the unwanted side effects introduced by filtering. However, we conclude that for sufficiently clean data sets, no filtering or detrending at all may work sufficiently well. Implications for other types of data are discussed, followed by a number of recommendations.

1. Introduction

Recent years have seen an upsurge in the application of time-resolved multivariate pattern classification analyses (MVPA) – also referred to as *decoding* – to electro- and magnetoencephalographic data (EEG/MEG; see Table 1 for an extensive list of references). MVPA allows researchers to uncover the active sensory and mnemonic representations underlying cognitive processes as wide-ranging as perception, attention, categorization, language, working memory, and long-term memory. Many researchers therefore now prefer the information-rich multivariate approach over traditional univariate event-related potential (ERP) or event-related field (ERF) analyses based on signals averaged over epochs. Moreover, toolboxes have recently emerged to facilitate these types of analyses (e.g. Bode, Feuerriegel, Bennett, & Alday, 2018; Fahrenfort, van Driel, van Gaal, & Olivers, 2018; Hanke et al., 2009; Meyers, 2013; Oosterhof, Connolly, & Haxby, 2016).

However, as the field is making the transition from univariate to multivariate approaches, some of the standard data processing procedures remain, raising the question whether these procedures are actually optimal, or perhaps even harmful, for decoding. One of the most common processing steps is *high-pass filtering*. Given the slow drifts in especially EEG data (less so in MEG data), high-pass filtering has become a crucial component in extracting ERPs and improving signal-to-noise. However, it is well known that high-pass filtering can lead to artifacts. Specifically, too high cut-off values (typically 0.1 Hz or more) may cause the signal enhancement to result in spurious local ringing effects[1] around the event-related responses – artifacts which may be misinterpreted as real components in the event-related signal (Acunzo, MacKenzie, & van Rossum, 2012; Kappenman & Luck, 2010; Luck, 2005; Tanner, Morgan-Short, & Luck, 2015; Tanner, Norton, Morgan-Short, & Luck, 2016; Widmann, Schroger, & Maess, 2015). Nevertheless, high-pass filtering is generally still considered a crucial step for extracting meaningful ERPs (for which drift correction is necessary), and therefore continues to be part of the recommendations with regards to EEG data preprocessing (with appropriate cut-off values, e.g. Maess, Schroger, & Widmann, 2016; Tanner et al., 2016; Widmann & Schroger, 2012; Widmann et al., 2015).

[1] Ringing effects are rippling artifacts near sharp edges as a result of filtering out high-frequency information

Perhaps less well known is that, depending on the specific cut-off value and frequency of the ERP, high-pass filtering may also lead to quite diffuse, but still spurious, activity differences both well before and well after the event-related response (Tanner et al., 2016). Even with modest cut-off settings, these slower components may emerge as subtle overall baseline shifts. A not uncommon step for ERP researchers is to correct for these shifts (whether apparent or real), thus potentially obscuring any artifacts. Thus, although sufficiently powered ERP studies could still show such artifacts, subtle baseline differences are often thought to be remedied by ensuing baseline corrections in ERP analyses (though see Tanner et al., 2016). However, multivariate analyses may be more sensitive to spuriously transposed information present in the topographical landscape. So far, little is known about the effects of high-pass filtering on multivariate pattern classification, and to what extent it leads to artifacts in decoding.

The potential for spurious temporal displacement of information is particularly worrisome when testing hypotheses on neural activity in the absence of stimulation, for example in the field of working memory. Indeed, after extensively analyzing one of our own EEG-based working memory experiments, we had to conclude that the above-chance decoding of the memoranda during the blank delay period was at least partly caused by the (modest) high-pass filter applied during preprocessing. As Table 1 shows, we have not been the only ones applying high-pass filtering prior to MVPA, as filtering has remained part of the pre-processing pipeline in a wide range of studies. Moreover, the same table also shows a wide range of cut-off values used when high-pass filtering is applied, from as low as 0.03 Hz to as high as 2 Hz, with 0.1 Hz being the most typical[2]. We thus decided to conduct a systematic exploration of high-pass filtering-related artifacts in MVPA, the results of which are presented here. First, we show how high-pass filtering led to clear signs of spurious decoding in one of our own EEG experiments, which involved a working memory task illustrated in Figure 1. The task contained an initial presentation of a cue, a blank delay period during which the cue had to be retained, and a test stimulus in which observers

[2] Note that Table 1 is only intended to illustrate the wide-ranging use of high-pass filters in EEG/MEG, and not to suggest that anything is necessarily wrong with these studies. For example, different studies may use different filter types: online (causal) or offline (either causal or acausal), Finite Impulse Response (FIR) or Infinite Impulse Response (IIR), different filter lengths and so forth, and each of these filter types may have different effects on the data that do not necessarily have to be problematic in the scientific context in which they are applied. Here we investigate only one particular common type of high-pass filter to assess its influence on MVPA of EEG. We return to this issue in the discussion.

searched for the cued object. To uncover the cause of the artifacts, and because empirical data does not come with a ground truth, we subsequently chose to create a simulated data set that allowed us to assess how decoding of filtered signals compares against decoding a known raw signal.

In addition to testing the effects of high-pass filtering, we tested an alternative method of detrending the data that was recently advocated by de Cheveigné and Arzounian (2018), and is referred to as *robust detrending*. Robust detrending involves fitting an nth order polynomial to the data and subtracting the fit, thereby removing slow-fluctuating drifts. Because the fit can be sensitive to either meaningful (ERPs, oscillations) or meaningless deviations (glitches) from the slow trend, ringing artifacts may also occur here. Furthermore, overfitting with a higher-order polynomial may result in the removal of the meaningful effects, thus obscuring real effects in an attempt to remove artifacts. In robust detrending, one therefore first presets a mask on parts of the data that are deemed to contain relevant events and thus should not be captured by the polynomial fit. In addition, through an iterative weighting procedure, outliers to the polynomial trend, such as glitches, are masked as well. We show that robust detrending leads to modest improvements in decoding, but more importantly, it does so without the artifacts.

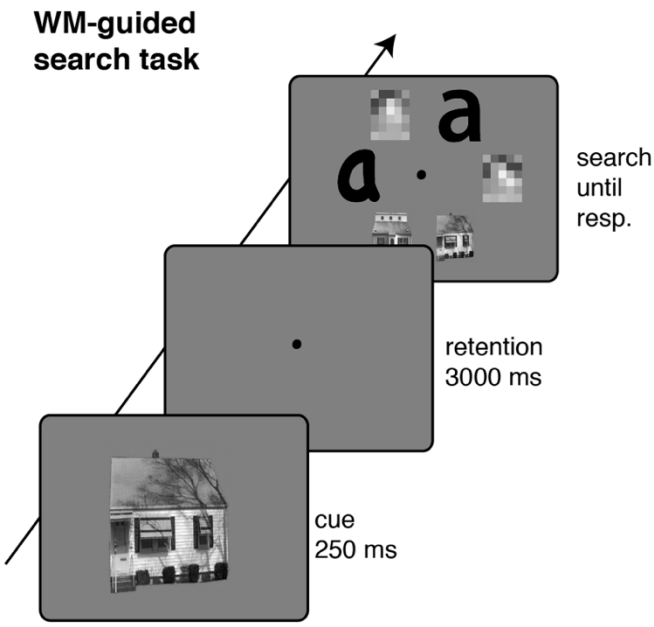

Figure 1. Example trial for the real experiment. Observers remembered a cued house, face or letter target for a subsequent visual search task presented after a 3000 ms blank retention interval. Observers then indicated with one of two button presses whether the memorized target was present or absent in a visual search display which also contained a number of nontarget objects. The faces in the search display were replaced with downsampled versions in this illustration to make them unidentifiable for reasons of privacy.

2. Methods

For both the real and the simulated data set, stimuli, data, code and analyses scripts are available from the Open Science Framework, at https://osf.io/t9rkz/

2.1. Real data

We report data from an experiment that is illustrated in Figure 1. On every trial, observers were presented with a face, house, or letter (the cue), which they had to remember for a visual search task presented 3 seconds later. The task was to determine the presence or absence of the cued target. The experiment included other conditions, but to simplify matters here we report on the condition that best serves the current purpose.

2.1.1. Participants. Twenty-five students from the Vrije Universiteit Amsterdam participated for course credits or monetary payment (€9 per hour). All subjects reported

normal or corrected to normal vision. The protocol complied with ethical guidelines as approved by the Scientific and Ethical Review Committee of the Faculty of Behavioural and Movement Sciences, and with the Declaration of Helsinki. Data of two subjects were removed from further analyses, one due to excessive high frequency noise reflecting muscle artifacts, and another due to a very strong but poorly understood artefact in the ERPs that is most likely due to equipment failure.

2.1.2. Stimuli and task. Subjects were asked to memorize at each trial a briefly presented picture (250 ms), which could be of the category face, house or letter (width: ~4° visual angle; height: ~5°). After a retention interval of 3 seconds (with only a white dot at the center of the screen as fixation point), a search display appeared, consisting of six pictures (two exemplars of each category; ~2.5° in size) randomly arranged along a hexagon array (radius of 4.5°; three pictures per hemifield; white fixation dot remained at the center of the screen). Subjects were asked to indicate whether the target picture they memorized at the start of the trial was present (left index finger) or absent (right index finger) by pressing a button on a button box connected to the EEG acquisition computer via a parallel port. Probability of target present/absent was 50%. The search array disappeared upon the subject's response (which changed the color of the fixation dot to black for 500 ms), or when 5 seconds had passed (after which the warning "respond faster!" appeared at the center of the screen for 500 ms). The inter-trial interval was set to 1 second ± 500 ms jitter. Low-level image properties of face and house pictures were controlled with the SHINE toolbox (Willenbockel et al., 2010). Subjects performed a short practice block of 12 trials with feedback on accuracy (words "correct!" and "wrong…" presented centrally for 500 ms), after which EEG recording started for 252 trials (84 per picture category), without feedback (except for slow responses). Prior to participating, subjects signed an informed consent form. Each unique picture within a category was only presented once as target, while it could be used more than once as distractors within the search arrays. Furthermore, when the target was a face, the two face stimuli in the search display were of same gender, encouraging subjects to memorize facial features rather than category. We randomly selected face stimuli out of 100 face pictures (from Endl et al., 1998, 50 male, 50 female). Similarly, when the target was a letter, the two letter stimuli in the search display where of same identity and capitalization, encouraging subjects to memorize the specific font. House

stimuli were randomly sampled from 100 exemplars of pictures used in Egner, Monti, and Summerfield (2010).

2.1.3. Data acquisition. EEG data from 64 Biosemi ActiveTwo (biosemi.com) electrodes placed according to the 10-20 positions were acquired at 512 Hz sampling rate. The ActiveTwo system is DC-coupled, and thus has no online (hardware) high-pass filter. On such DC-coupled systems, drifts are common due to non-brain artefacts such as sweating. Further, the data was down-sampled offline to 256 Hz and re-referenced to the average of signals recorded from both earlobes. Error trials, trials without a response, or with responses slower than 3 seconds were not included in the analyses. Continuous, raw data was first inspected for malfunctioning electrodes, which were interpolated after the below preprocessing steps. We did not perform any oculomotor artifact correction.

2.2. Simulated data

We describe the creation of an artificial dataset for a task with a similar but simplified structure, involving the presentation of a to-be-remembered stimulus, a retention phase, and a test stimulus.

2.2.1. Creating class-specific topographical patterns. Figure 2A illustrates the creation of the underlying spatial pattern of evoked responses. The features fed into a linear discriminant classifier are typically activity values at a given time point (or averaged over a time window) for each of N electrodes that cover the scalp or part thereof. Here we simulated activity of 64 electrodes with positions placed according to the international 10-20 system. From these we selected a fixed set of electrodes to represent one stimulus class, and another, partially overlapping set of electrodes to represent another stimulus class. Stimulus-related class-specific activity was thus associated with different multivariate spatial patterns, such that multivariate classification trained and tested on the channel features over time would be able to reproduce the stimulus-related activity.

To simulate stimulus-related activity assigned to these sets of electrodes, we first created an event-related potential (ERP, shown in Figure 2B) that mimicked three phases of a working memory task: encoding the stimulus into the visual system, retaining a representation of the stimulus in working memory, and recognizing the stimulus upon the presentation of a probe. A “trial” consisted of an array of data containing the entire ERP, and lasted from -2 seconds to 4.5 seconds surrounding the “event” (what would be the

onset of the to-be-encoded stimulus). Starting at *t* = .05 seconds, we used a Weibull function to create a typical encoding response with a steep rising slope and a shallower falling slope, returning to baseline around *t* = 0.5 seconds. To mimic memory maintenance during the retention interval, shortly after encoding, activity increased again with a steep logistic curve (starting at 0.5 seconds, then plateaued for about 0.7 seconds, after which it dropped with a shallower logistic curve to baseline at around *t* = 2 seconds). The response to the final memory probe display was simulated with a similar Weibull function as for encoding. After about 2.5 seconds, the event-related activity remained at baseline for the remainder of the trial. Note that each of the different phases (encoding, retention, recall) returned to baseline before entering the next phase.

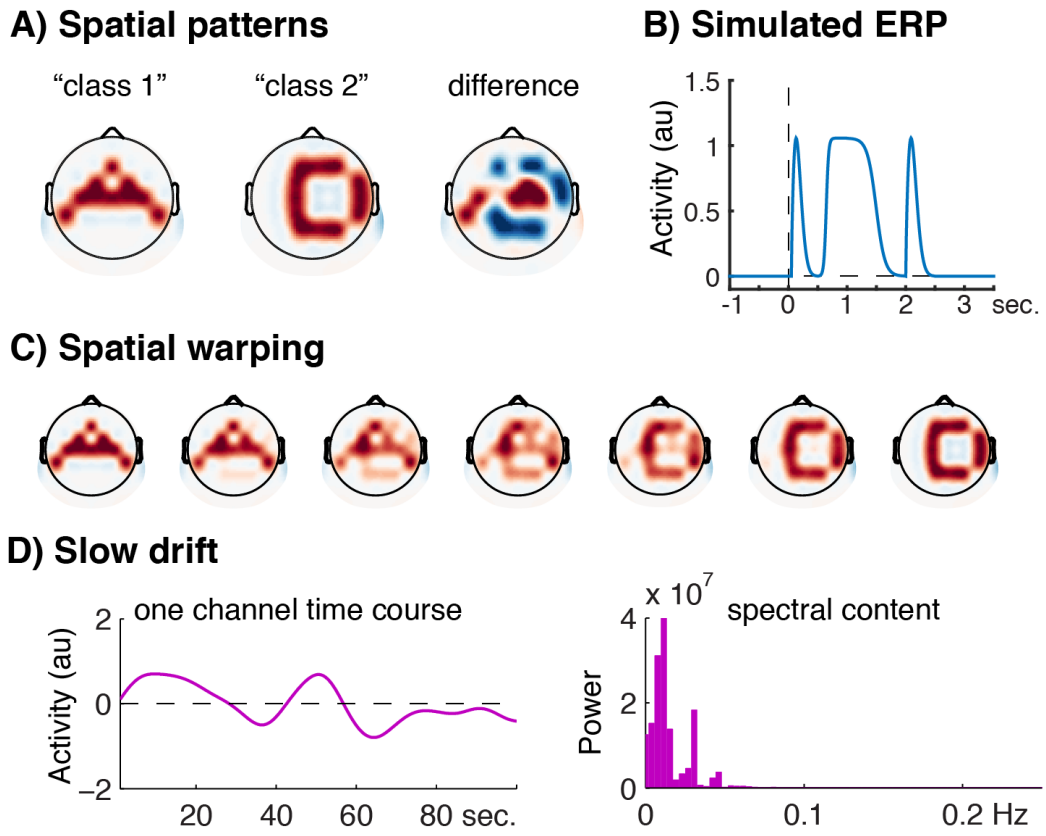

Figure 2. Creation of simulated data. A) Two different electrode topographies representing the two stimulus classes, plus their difference. Red as positive, blue as negative. B) The underlying simulated ERP time series as injected into each electrode of the topographical patterns. C) A fuzzy decision boundary was then created by different degrees of warping between the two patterns. D) Example time course of 1/f pink noise slow drift as was added to the data (left panel), and its spectral content (right panel).

2.2.2. Decision boundary. We created one class of patterns by injecting the ERP into each of the electrodes in one of the two sets described above for 100 trials, and another class by injecting the ERP to each of the electrodes in the other set for 100 trials, thus creating two different underlying spatiotemporal landscapes of activity. Trial order was randomized before injecting noise, and again before training and testing the classifier. With such two highly different patterns, a classifier would produce nearly perfect classification accuracy. To avoid such a ceiling effect, we created a "decision boundary space" by warping one spatial pattern into the other pattern in 80 linearly spaced transitions, where transitions 40 and 41 are closest to the exact middle between the two patterns. Figure 2C illustrates the warping in seven steps. This warping resulted in non-overlapping channels now showing a *relatively* stronger ERP for one stimulus class than the other (where this was binary prior to warping). Another way of describing the effect of this warping procedure is that the multivariate patterns of the two stimulus classes become more similar, thereby moving them closer to the decision boundary of a multivariate classifier. For the three working memory stages of the trial, we selected different degrees of warping: for the encoding phase we used relatively divergent patterns (transition 31 vs transition 50); for the retention phase we used a near-complete mixture of the two patterns (transition 39 vs transition 42); for the recall phase, we used a warping-degree in between encoding and retention (transition 35 vs transition 46). Note that these simulated ERPs and spatial distributions were purely meant to illustrate decoding under different degrees of separability, and not intended as an exact model of brain mechanisms of working memory. At the same time, the classification result yielded a pattern that can be observed in real data (e.g. Myers et al., 2015; Wolff, Ding, Myers, & Stokes, 2015): A transient sweep of high decoding accuracy during encoding, low (yet significant) sustained accuracy during retention, and high (yet somewhat reduced) accuracy during recall/search.

2.2.3. Adding low-frequency pink drifts. As high-pass filtering and robust detrending are used to remove low-frequency non-stationary drifts, the next step was to add such drifts to the simulated activity, as illustrated in Figure 2D. To this end, we created different time series of pink noise (1/f, power spectral density is inversely related to frequency) for each electrode separately. Specifically, we first created Fourier coefficients of random phase angles multiplied by random amplitudes that showed an exponential decay over frequency. The real part of the inverse fast Fourier transform of this simulated power spectrum

produced “continuous data” of low frequency noise, with length equal to 500 concatenated trials of 4.5 seconds, plus 10 seconds before and after as buffer zones for filtering edge artifacts. In our main analyses, power was relatively strong for the lowest (<0.01 Hz) ultra-slow frequencies, and reduced to roughly zero at around 0.1 Hz. Figure 2D shows an example 100 second snippet of the time series of one channel, together with its spectral content. The channel-specific ERPs created in the previous step were then added to the simulated drift in varying ratios. In supplementary analyses, we also investigated a shallower decay setting, in wich there was still observable power around 0.25 Hz (Supplementary Figure 4C and 4D). The rationale for the slow and fast decay setting was that the speed of drifts/noise will typically differ between datasets, and cannot be known a-priori. A typical cut-off of 0.1 Hz might not remove all fast drifts, yet a more rigorous cut-off of 0.5 Hz might also affect frequencies in which there was no drift at all (i.e. between ~0.25 and 0.5 Hz).

In addition, we also explored a broad range of signal-to-noise (SNR) ratios between ERP and drift by multiplying the normalized drift between 1 (weak noise; high SNR) and 20 (strong noise; low SNR) times with 20 integer steps (SNRs from 1:1 to 1:20). This larger SNR parameter space yielded no qualitatively interesting results, so these too are presented as Supplementary material. We randomly generated 24 subjects using the above procedure, so as to be able to run a standard group analysis using the ADAM toolbox (Fahrenfort et al., 2018).

2.3. Data preprocessing and analyses

Before applying MVPA analyses, slow drifts were either not removed at all ('raw data'), or removed through either high-pass filtering, or robust detrending, both of which are described in more detail below.

2.3.1. Removing low-frequency drift noise with high-pass filtering. To investigate the effect of drift removal using high-pass filtering, we high-pass filtered the continuous data (both the simulated and the re-referenced real time series) according to typical M/EEG preprocessing pipeline settings. We used a two-way sinc FIR filter with a Kaiser window type, with a maximum passband deviation of 0.1% (recommended by Widmann et al., 2015), by using EEGLAB’s pop_firws() function (Delorme & Makeig, 2004), with all parameters set to default except filter order, which was set to correspond to 3 cycles of the cut-off frequency. In case of the real data, we show the decoding of "raw" unfiltered data,

as well as the effect of high-pass filters on decoding for three cut-off values: 0.05 Hz, 0.1 Hz, 0.25 Hz and 0.5 Hz. In case of the simulated data, we show the decoding of the "raw" unfiltered simulated data as well as the effect of cut-off frequency on decoding in 9 semi-logarithmically spaced steps of 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.25 and 0.5 (so 10 steps in total when including the raw unfiltered data).

2.3.2. Removing low-frequency drift noise with robust detrending. In a second procedure, we applied an alternative to high-pass filtering, which involves fitting an nth order polynomial to the data and subtracting the fit, thereby detrending the data to remove slow-fluctuating drifts (de Cheveigne & Arzounian, 2018). Because the fit can be sensitive to sudden deviations from the slow trend ("glitches"; muscle, motion or electrode-specific artifacts; but also sharp, transient ERPs or synchronized oscillations such as posterior alpha band), two unwanted side effects of detrending can occur. First, the glitch can impose ringing artifacts, similar to what happens in filtering. Second, if a signal of interest is largely or fully captured by a high-order polynomial, one risks removing real effects in an attempt to remove artifacts. A solution is *robust* detrending, where first a pre-set mask can be set on parts of the data that are deemed to reflect experimentally relevant (e.g. cognitive) events and thus should not be captured by the polynomial fit (for an illustration of the procedure see Figure 3, for an illustration on real data of a representative subject, see Supplementary Figure S1); second, through an iterative weighting procedure other parts of the data, which are recognized as outliers of the polynomial trend, are masked as well (see de Cheveigne & Arzounian, 2018, for details). The final fit is then done on the non-masked data, and subtracted from all data (masked and non-masked). For the simulated dataset we used a pre-set mask to remove the ERPs occurring in current trial from the detrending operation, so including the encoding, retention and recall phases (i.e. we set a mask that runs from t = 0 to t = 2.5 seconds); all other surrounding data were left unmasked. Similarly, for the real dataset we used a pre-set mask to remove the current trial from the fit, going from stimulus onset (t = 0 seconds) until 0.75 seconds after the response, as to not include any meaningful perceptual, cognitive and/or motor-related dynamics into the polynomial fit.

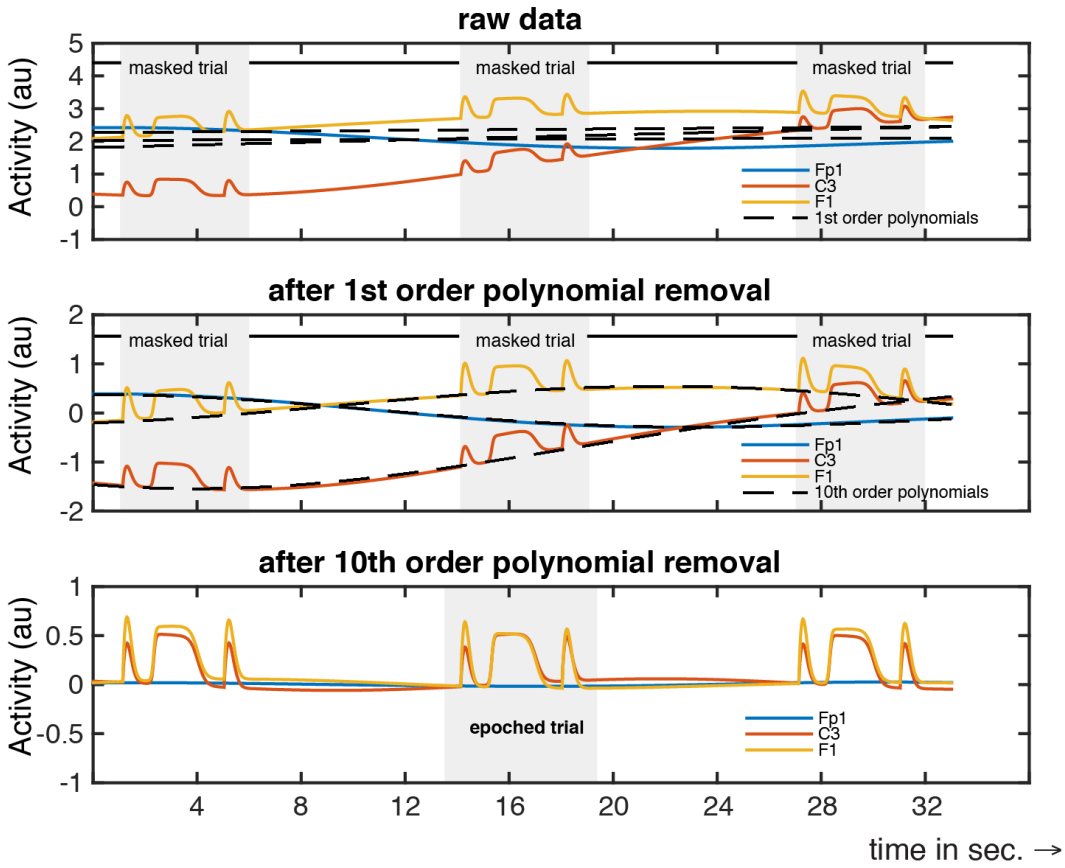

Figure 3. Illustration of the procedure for removing low-frequency drift noise with robust detrending based on simulated data. Top panel: raw data for three electrodes: Fp1 (no ERP), C3 and F1 (both of which contain an ERP). Also shown as dotted lines are the polynomial fits on the raw data from which the trial events were masked out (grey panels in the background). Middle panel: data after removing 1st order polynomials fits from the top panel. Also shown are the 10th order polynomial fits on these data, from which the trial events were masked out (grey panels in the background). Bottom: data after removing the 10th order polynomial fits. Finally, the middle trial is segmented out for further analysis. Note that this figure serves as illustration only. The width of the time window of data on which the polynomials were fitted was much wider than what is shown in this illustration. Further, depending on the length of the intertrial interval one may choose to mask out only the trial that will be epoched trial (as was done in the analyses presented in this paper), or also mask out neighboring trials (which may also work well, as was done in the above illustration). Finally, the robust detrending algorithm will iteratively mask out additional sharp transients from the data that otherwise disturb smooth fits, see main text as well as (de Cheveigne & Arzounian, 2018) for details. A similar figure is produced for real data when using the detrending function included in the ADAM toolbox, an illustration of which can be found in Supplementary Figure S1.

High-pass filters are usually applied to continuous data with sufficient buffer zones before and after the experimental recording, because a low-frequency cut-off results in long lasting edge artifacts that may enter the task-related data. However, robust detrending of a whole recording session of typically more than an hour can be suboptimal: the non-stationary slow trend may be too complex, requiring a high polynomial order that is difficult to select a priori. Although de Cheveigné and Arzounian provide no clear recommendation as to how long data epochs should be for optimal detrending, the examples given in their paper show segments in the range of a few hundreds of seconds. Because we did not know a priori what the length of the drifts were in our experimental data, we segmented into liberal (wide) padded epochs of 207.5 seconds (i.e. trial-related epochs of 7.5 seconds with 100 seconds of trial data pre-/post-padded). To be able to include all trials, the continuous data were symmetrically mirror-padded with 100 seconds prior to segmentation. For the simulated data, we segmented into less extensive 56.5 second epochs, each consisting of the duration of one trial padded with 25 seconds on both sides, approximately coinciding with the length where the injected drift power was maximal. Note however, that the duration of a padded trial during detrending does not directly impinge on the frequency of the drift that can be removed (as is the case for filter lengths), as a polynomial can easily fit onto a small portion of an oscillation. Preliminary testing showed that detrending on padded windows as small as 50 seconds seems to work quite well.

Similar to varying the cut-off frequency for filtering, we varied the polynomial order for detrending in 9 steps, using the orders: 1, 2, 3, 4, 5, 10, 15, 20 and 30 in case of the simulated data (so 10 steps in total when including the raw data). For the real data, we ran four detrending procedures with 1st order only, 10th order, 20th order, and 30th order. For all polynomial orders higher than 1, the data were first detrended with a 1st order polynomial (i.e. in fact removing a linear trend over the entire epoch) to improve the fit of the higher order polynomial (as recommended by de Cheveigne & Arzounian, 2018), also see Figure 3 for an illustration of the procedure from top to bottom for a 10th order polynomial. Because of the robust, iterative fitting procedure, the first detrending step also updates the mask with additional time-channel-specific outliers; this updated mask is then used as a pre-mask for the next detrending step. As can be observed in electrode C3 in Figure 3, the fit is not necessarily perfect (middle panel) and the drift is not perfectly removed (bottom panel). Detrending is not guaranteed to produce perfect fits, as noise can occur in many frequency

spectra that are not necessarily always captured by a polynomial of a given order. For this reason, it can be advantageous to try out different filter orders during drift removal. However, by the analytic logic of the mask procedure, the fit cannot be affected by the ERP (the signal) that occurs in the current trial. Therefore, any remaining effect on MVPA can be regarded as imperfect noise removal, which by the same logic is evenly distributed across trials and conditions.

Robust detrending was done with the Noise Tools toolbox (http://audition.ens.fr/adc/NoiseTools), using the nt_detrend() function. Note that we have also added a detrending function to version 1.06 of the ADAM toolbox (Fahrenfort et al., 2018) that allows one to easily perform a detrending and epoching operation on EEG data in EEGLAB format while masking out 'cognitive' events, by internally making use of the nt_detrend function. The ADAM function is called adam_detrend_and_epoch(), and takes as inputs EEG data, a specification of the epoch window, the window in which event take place that should be masked out, and some other parameters. Its output can then be used directly for MVPA first level analyses. The function also produces a plot of the detrending procedure on a trial in the middle of the dataset, using some illustrative channels with strong drifts, as well as a butterfly plot of the ERP data before and after detrending. An example of such a plot can be found in Supplementary Figure S1. See the help of adam_detrend_and_epoch for further details on how to execute the function.

2.3.3. MVPA Analyses. We performed multivariate pattern analyses (MVPA) on both the real and the simulated data, with the use of version 1.06 of the ADAM toolbox (Fahrenfort et al., 2018) – a freely available script-based Matlab analysis package for both backward decoding and forward encoding modeling of M/EEG data. The latest release of the toolbox is available from Github, through http://www.fahrenfort.com/ADAM.htm. A linear discriminant classifier was trained and tested on each time point either using 10-fold cross-validation (real data) or 2-fold cross-validation (simulated data). As classification accuracy metric we used the Area Under the Curve (AUC), in which the curve refers to the Receiver Operating Characteristic (ROC, Hand & Till, 2001).

For the real dataset, the three image categories of faces, houses and letters were used as classes to train the classifier. As features we pre-selected 9 occipital channels (PO7, PO3, O1, Iz, Oz, POz, PO8, PO4, and O2) to increase the signal-to-noise ratio (Fahrenfort, van Leeuwen, Olivers, & Hogendoorn, 2017). Classes comprised the three balanced picture

categories (Fahrenfort et al., 2018). Prior to MVPA, EEG data were down-sampled to 32 Hz sampling rate using MATLAB's resample function, and baseline-corrected using a window of -0.2 to 0 s. Here we tested the classifier not only on the same time point at which it was trained (as we did for the simulation analysis), but also across all other time points. This generates temporal generalization matrices, which are informative as to whether a pattern of neural activity underlying classification performance is stable, or whether it dynamically evolves over time (King & Dehaene, 2014). In the context of the current analyses they are informative with respect to the degree to which patterns are artificially distorted over time. At the group level, subject-specific AUC as accuracy measure of multivariate classification was statistically compared against chance for the raw data, as well as for the different cut-offs and polynomial order values using t-tests. We corrected for multiple comparisons using cluster-based permutation ($p<0.05$).

For the simulated dataset, all 64 channels served as features, and the condition labels assigned to the trials as described in section 2.2.1 served as classes. Prior to MVPA, data were down-sampled to 25 Hz using MATLAB's function resample. We tried out various pre-stimulus baseline windows for baseline correction to determine whether this could explain spurious decoding effects throughout the trial window. The reported results were obtained using a –0.5 to –0.25 s pre-stimulus interval as baseline, but other baseline windows produced near identical results. Other than that, analyses were identical for the simulated and the real data.

3. Results

3.1. Empirical EEG data

Figure 4 shows classifier performance for the working memory task. We were able to reliably dissociate multivariate patterns of broadband EEG activity across the nine included occipital channels, during encoding, retention, and the search period for the face, house and letter stimuli. Classification increased transiently during the presentation of the initial target cue, after which it decreased yet remained at above chance levels for up to two seconds during the delay period, before it dropped to near-chance levels. Classifier accuracy then increased again during presentation of the search display, presumably upon selecting the target category. Moreover, Figure 5 shows that the multivariate pattern was relatively

stable during the encoding/retention phase (outlined by the dotted square), but for the raw data did not significantly cross-generalized across time to the search phase.

Important for the present purpose, Figure 4A reveals how decoding accuracy was affected by different high-pass filter cut-off values. Most notably, filtering contributed little to overall decoding accuracy during the encoding/retention phase, except towards the cut-off of 0.5 Hz, where decoding appeared to improve considerably during the WM delay period. However, we have no way of ascertaining whether this improvement is real or a displacement of information from other periods in the trial. In fact, it is likely to be spurious, since similar effects emerged prior to cue onset and after search offset – suggestive of artificial increases in decoding accuracy in periods in which one expects baseline activity. To further underpin how small spurious effects of filtering can be missed, we also computed the uncorrected plots, which can be found in Supplementary Figure S2. Here one can observe significant decoding in the baseline period already at 0.25 Hz.

Figure 5B (left panel at a cutoff of 0.25 Hz and 0.5 Hz) shows how these filter-related artifacts may also lead to suggestions of very stable representations that generalize across time. Here too, cross-generalization to the pre- and post-trial baseline periods indicates that these patterns are spurious. It can be easy to miss the ostensibly spurious nature of such classification performance if the plotted baseline period is too narrow, missing the fact that above chance performance also occurs prior to stimulus onset. For example, note that classification performance during the baseline period just prior to t=0 neatly drops to chance (plausibly due to baseline correction), while this correction does not resolve or correct for the spurious performance that was introduced to other (generalization) periods in the trial. To highlight once more how small spurious effects of filtering can be missed, we also computed uncorrected temporal generalization plots which can be found in Supplementary Figure S3, showing above chance classification performance in the baseline window for values as low as 0.1 Hz.

Figure 4B shows results for the same classification, but now after robust detrending, at three levels of complexity (i.e. 1st, 10th, 20th and 30st order). Here too, as the raw data was relatively clean to begin with, the detrending actually contributed relatively little to overall decoding accuracy, compared to decoding of the raw signal. However (and more important in the current context), there were no signs of any artifacts, as the detrending results follow the decoding of the raw signal. The same is true when we assessed temporal generalization,

as shown in Figure 5C. The overall pattern of generalization was the same across different orders of detrending, and remained comparable to raw, although it did appear to convey some benefits with respect to temporal generalization both within the encoding/retention phase as well as in the generalization from the encoding phase to the search phase. In contrast to high-pass filtering, these improvements occurred without similar increases during baseline periods.

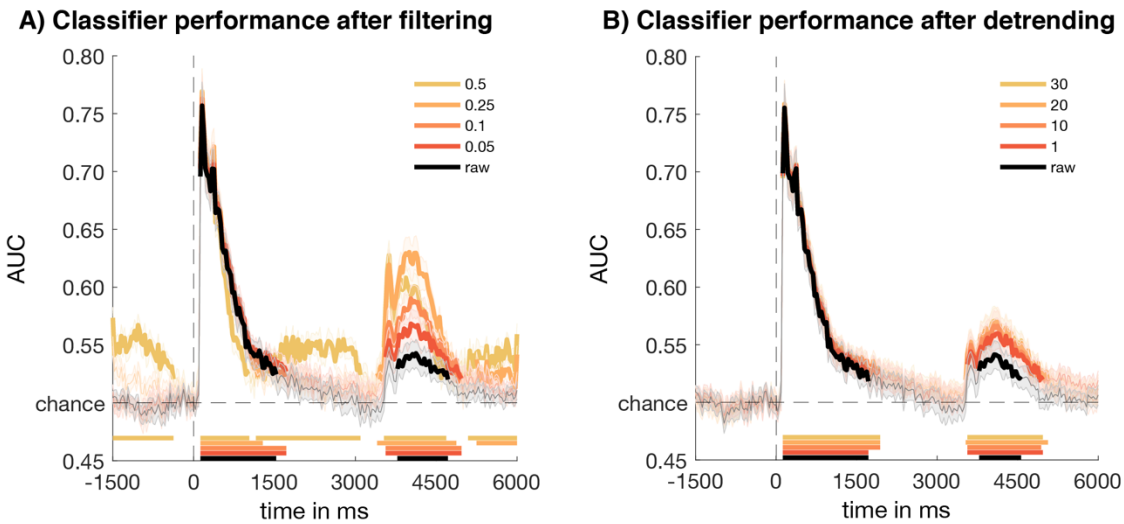

Figure 4. Results for empirical data from a working memory guided search task. A) Decoding performance (AUC) at each time point for different high-pass filter cut-off frequencies (shades of orange), and for raw data (black). All thick colored lines denote p<0.05 under cluster-based permutation testing. Note the marked above-chance decoding prior to stimulus onset, after filtering at 0.5 Hz. This in turn raises doubts about the above-chance decoding during the delay period. A similar artifact already emerges with cut-off values at 0.25 Hz, showing above chance decoding after the search display (between 5000 and 6000 ms). The spurious effects of filtering at 0.25 Hz can be seen even more strongly when inspecting the temporal generalization plots (see Figure 5B) or when inspecting the uncorrected plots in Supplementary Figure S2 in which effects also occur in the baseline window. B) The same, but now after robust detrending at various polynomial orders. Here no artifacts occur. Colored bars indicate reliable difference from chance.

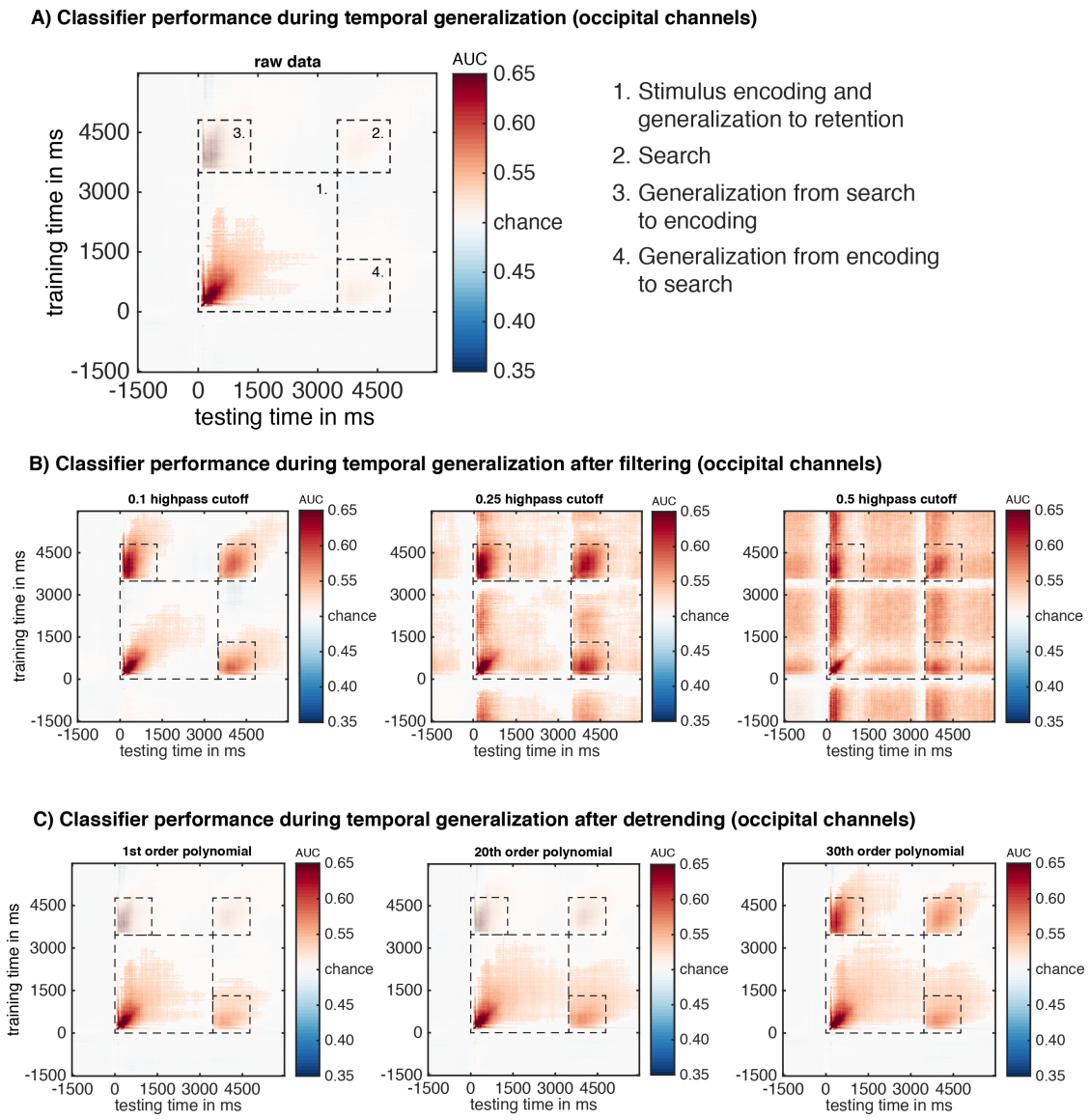

Figure 5. Temporal generalization results for the empirical data. A) Temporal generalization plot for the raw data. Significance was evaluated using cluster-based permutation ($p<0.05$, saturated colors, non-significant as non-saturated colors). The regions outlined by the dotted square indicates various phases (referred to by numbers using the in-figure legend). These suggest a relatively stable representation during roughly the first two seconds after stimulus onset but which does not significantly generalize to the end of the retention phase (1), or to the search phase (3 and 4). B) Temporal generalization, but now after high-pass filtering at three different cut-off levels. Note that the encoding phase generalizes better to the search phase (for all frequencies). Worryingly though, strong and ostensibly spurious generalization to these and other time windows occurs after a slightly more rigorous filter of 0.25 Hz and even more strongly so at a filter of 0.5 Hz, so that we cannot be sure that anything observed at lower cutoffs is real or (partly) spurious. When inspecting the

uncorrected temporal generalization plots in Supplementary Figure S3, one can see spurious decoding in the baseline window at the even lower cutoff of 0.1 Hz. C) The same temporal generalization analyses, but now after robust detrending at different orders. Here, there is no sign of spurious generalization, while still obtaining better generalization of the encoding to the retention phase, as well as generalization from the encoding to the search phase when compared to raw.

Thus, we found that high-pass filtering can result in clear artifacts in decoding, while contributing little to overall decoding accuracy. In contrast, robust detrending showed no clear artifacts, while it did modestly enhance temporal generalization across time. For subtle cognitive and/or perceptual phenomena, such a small yet significant increase may be very valuable. However, without a ground truth, one may always wonder whether observed improvements are real or spurious. We therefore turned to simulated data, as described next.

3.2. Simulated data

Figure 6A illustrates the effect of filtering (left panel) and robust detrending (right panel) on a simulated single trial ERP, for three different cut-offs (0.05, 0.1 and 0.5) and orders (1, 10, 30) respectively. This clearly shows attenuated drift for higher level removals. However, filtering with higher cut-off values also comes with ringing artifacts surrounding the ERPs, some of which extend up to seconds prior to and after the events (cf. Tanner et al., 2015; Tanner et al., 2016). As with high-pass filtering, the removal of drift by detrending is clearly better for higher orders. Importantly, it did not contain the typical filtering artifacts. As discussed next, this difference between filtering and detrending has consequences for decoding too.

Figure 6B shows how decoding performance follows the simulated ERP, for the various high-pass filtering steps. This was true even when no preprocessing at all was applied to remove the drift. In fact, the drift appears to have relatively little negative effects on decoding of the raw signal (black line), unless the classes are relatively close to the decision boundary (as during the retention phase). The red to yellow colored lines show how filtering affects decoding. Figure 6C shows the same, but now after robust detrending of the data with the various polynomial orders. Neither method has much to add to the encoding phase of the trial, where raw decoding was already at 100%. But both methods lead to clear and

substantial improvements for the weaker signals underlying the later phases of retention and recall, where accuracy improved moderately to strongly for the highest levels of drift removal.

However, and importantly, while decoding after detrending correctly returned to baseline in between the different task-related phases of the trial, high-pass filtering led to artificially elevated decoding (~70%) throughout the trial, including the episodes of null activity prior to, between, and after the different phases. These vivid artifactual increases in classifier performance emerged for cut-offs starting at 0.1 Hz. When plotting the topographical maps of the forward-transformed classifier weights for the most extreme cases of filtering (0.5 Hz and 30th order polynomial, Figure 6D), we observed that after filtering the differential patterns of the two classes was indeed temporally displaced onto pre- and post-trial time points (left panel), while this did not occur after robust detrending (right panel).

However, these results may partially be caused by relatively low SNR levels and/or by the fact that so far we have only inspected simulated data that contain very slow drifts (<.1 Hz), rather than fast drifts. Therefore, we also explored the extent to which drifts in different frequency spectra and for different SNR levels are affected by different filter and polynomial settings. The results of these analyses are shown in Supplementary Figure S4. This figure contains two analyses: one containing the low frequency drifts that are reported in Figure 6 (Figure S4A and S4B), and one containing faster drifts (Figure S4C and S4D). The graphs in S4A and S4C show drifts for illustration purposes (left panel) and the spectral profile of these drifts (right panel). The graphs in S4B and S4D explore the relationship between the degree of noise on the x-axis (SNR: low to high) and filter cutoff / polynomial order on the y-axis, and color shows the difference between decoding performance in raw EEG and decoding performance. This is shown for all relevant periods in a trial (pre-stim, encoding phase, retention and so forth). These analyses provide two clear results: (1) somewhat unsurprisingly, the spurious effects of high-pass filtering in the pre-stim window occur for lower filter settings when the drifts are slower (compare figure S4B to S4D) and (2) spurious effects seem to occur across a wide range of SNR values. Especially the second finding is relevant, as it shows that even under relatively high SNR, high-pass filtering can potentially produce spurious effects in for example pre-stimulus or retention time windows.

In sum, the simulations show that the topographical information on which the classifier performance relies can inadvertently be transposed onto baseline time windows when applying high-pass filters prior to decoding, thus resulting in artificially inflated and extended “above chance decoding” epochs, and that this occurs for a wide range of noise frequency spectra and SNR values. We found that artifacts potentially already start emerging at the often-used cut-off of 0.1 Hz. Robust detrending does not suffer from such displacements when the ERPs are masked out. Instead, it comes with few artifacts and improves decoding for components where there is a real underlying signal.

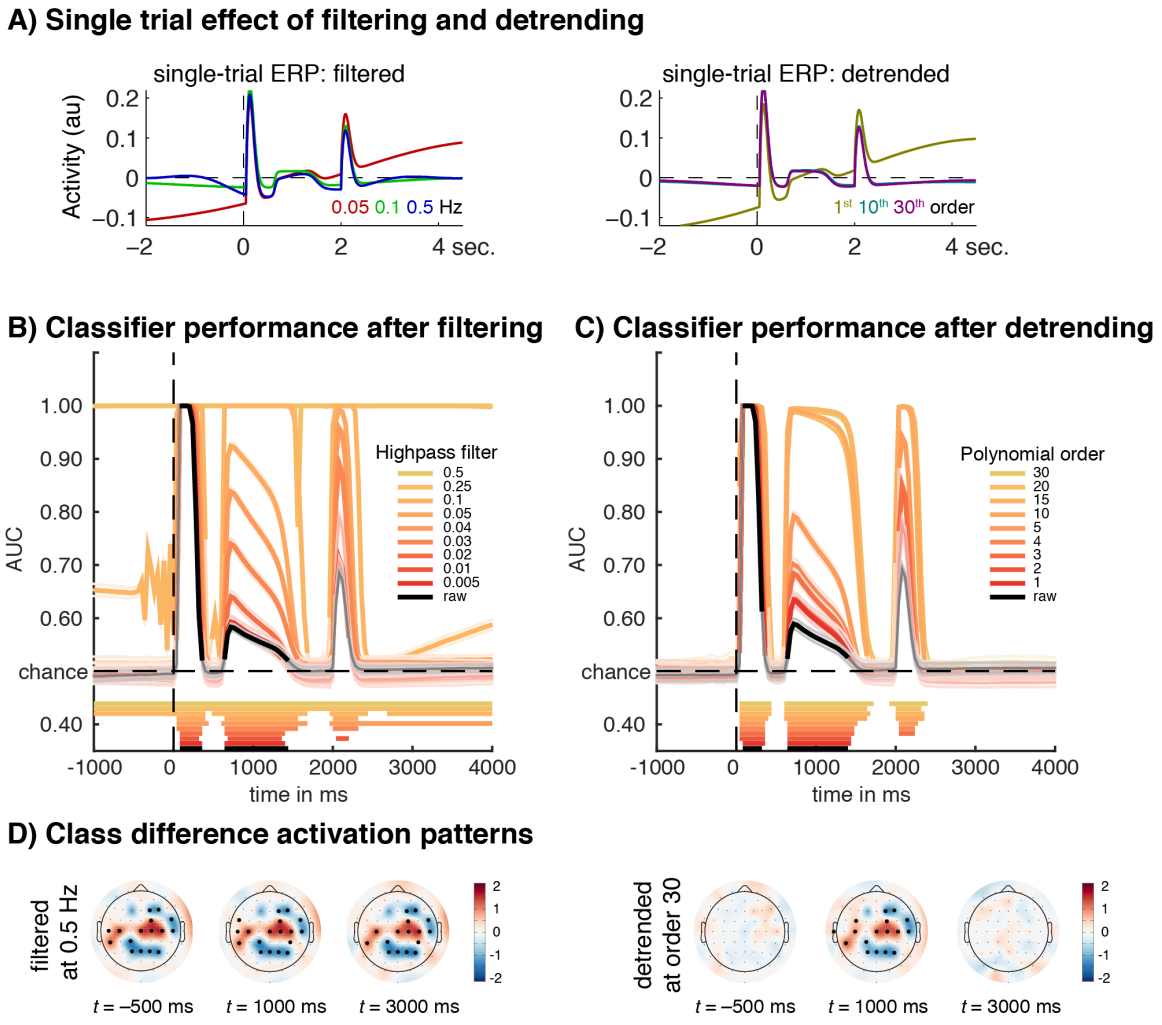

Figure 6. Results for simulated data. A) Average single-trial ERPs as a function of three different high-pass filter cut-offs (left panel), or different polynomial orders of detrending (right panel). B) Classification over time for raw versus high-pass filtered data, for 9 different cut-off values. Thick colored lines denote p<.05 under cluster-based permutation testing. Note the spurious above-chance decoding at activity-silent intervals pre-stimulation for cut-off values as low as 0.1 Hz, and

post-stimulation as low as 0.04 Hz. C) Classification over time for raw versus robustly detrended data for 9 different orders of polynomial fit (see methods for details). Note that no spurious classification occurs. D) Average class-separability maps for different time points in the trial (pre-stimulus, delay, post-trial) for data that was high-pass filtered at 0.5 Hz (left panel) and detrended using 30th order polynomials (right panel). Individual subject patterns were spatially z-scored prior to averaging, color denotes z-value. Thick black dots indicate significant clusters under cluster-based permutation testing at $p<.05$. Clear class-related topographical patterns emerge after high-pass filtering during time windows when there was actually no information. Note that these are shown for illustration purposes only, not to claim that effects of similar strength can also be obtained in empirical data.

4. Discussion

For a long time, high-pass filtering has been a standard step in processing EEG and MEG data, as it has clear benefits when analyzing event-related potentials. However, we show here that one should be careful, if not distrustful, when applying any high-pass filtering in service of multivariate pattern classification, because it can easily lead to spurious above-chance decoding effects. We demonstrate these artifacts to clearly emerge for both real EEG and simulated data, and from cut-off values as low as 0.04 Hz (in the simulated data) and 0.1 Hz (in the uncorrected empirical data). As Table 1 indicates, 0.1 Hz remains a popular cut-off value in cognitive neurophysiology studies, with only six out of 38 studies using a lower threshold.

It is important to point out that the enhancement of decoding accuracy was particularly strong for time windows where the real underlying signals were actually weak, in particular the “delay period activity” of our (simulated) working memory task. In working memory experiments, this is an interval during which the memorandum is no longer present, and the classifier is supposed to pick up on purely mnemonic representations. Any enhancement of such EEG- or MEG-based “mind-reading” capabilities would be very attractive to researchers, and thus extra caution is necessary. We show that above-chance decoding can easily extend to time points where there was actually no signal, not only during the delay period between phases where the simulated signals contained no information, but also during pre-stimulus and post-trial intervals. Combine this with the fact that the reverse may also happen (i.e. high-pass filtering may actually destroy a real sustained signal, de

Cheveigne & Arzounian, 2018), and the risk of drawing false conclusions on the presence or absence of sustained mental representations becomes more than real.

The cause of the spurious decoding appears to lie in small yet reliable artifacts caused by the interaction between the filter and the ERPs (Acunzo et al., 2012; Kappenman & Luck, 2010; Tanner et al., 2015; Widmann et al., 2015). Depending on the nature of the ERPs and filter settings, these artifacts may have quite diffuse effects, extending for seconds prior to and after the relevant events. Although these issues have been pointed out before in the context of ERP analyses, the problem becomes even more salient when performing decoding analyses, as classifiers may be sensitive enough to pick up on subtle, distributed patterns of displaced information that might only be noticed in averaged univariate ERPs when a study has sufficiently high power. The difference between the filtering effects on ERPs and on decoding becomes especially apparent when considering what may be seen as "baseline shifts", which are then often corrected for in standard ERP analyses. While quite subtle in our simulated ERPs, these shifts actually led to very strong above-chance decoding prior to stimulus onset. Baseline correction does not help here, since the baseline applies to the average of an idiosyncratically chosen pre-stimulus period. This may abolish the (displaced) *average* multivariate pattern within a certain temporal region, but this is not guaranteed to capture and remove all displaced information, as can clearly be seen in Figure 4A, which shows remaining pre-baseline decoding artefacts despite baseline correction and correction for multiple comparisons. Similarly, the simulations show that even within the baseline period itself, the average may not capture and remove all local (sample-by-sample) multivariate patterns, as can be seen in figure 6B. Here, filter cutoffs as low as 0.1 Hz result in spurious classification accuracy throughout the entire pre-stimulus time window (including the baseline), despite subtracting the average baseline activity from every sample and electrode in the trial. This is a clear warning that relatively subtle artifactual effects in ERP studies can have large undesirable effects on decoding.

The reason for these artefacts lies in the nature of the filtering operation. When filtering the data, one assumes that the noise (drifts) that one attempts to remove through the filtering operation occur in a different part of the frequency spectrum than the signals of interest. However, in practice this is often an unwarranted assumption, especially when trials have a long duration (as for example in working memory experiments that have a retention interval). In such cases, the frequency spectrum in which the signal resides may

overlap with the frequency range in which the filter operates. As a result, the filter may end up distorting the signal of interest and displacing information to periods where nothing occurred in reality.

A better way to remove slow trends from the data appears to be robust detrending (de Cheveigne & Arzounian, 2018), in concert with carefully defined masks that remove potentially relevant (cortical) sources of information from the fits. This method is advantageous when low frequency noise contributions that occur in the same frequency spectrum as the signal of interest can still be separated temporally. In such cases the signal of interest cannot affect the denoising operation because it is masked out. We found that robust detrending led to reliable improvements in decoding, while avoiding the artifacts that come with high-pass filtering. Nevertheless, here too there are choices to make, and pitfalls to avoid. One important drawback is the search space for optimally detrending the results, where polynomial order and data segment length seem to interact in ways that in turn depend on the spectral content of the noise one wants to remove. In real data, this will often be unpredictable and highly study- and subject-specific, further complicating choices as to which detrending options to employ. Moreover, too high an order polynomial might result in increased risks of fitting effects of interest. Masking important epochs will prevent this, but relies on additional assumptions as to which time windows are important. For our particular empirical data set, the improvement achieved with robust detrending, relative to the raw data, was relatively modest, and may not outweigh the extra decisions and assumptions. Of course, this depends on the quality of the data and the conclusions one is after.

Our findings may have wider implications beyond those for EEG decoding analyses. First and foremost, although here we focused on both simulated and real EEG data, our demonstrations may naturally apply to MEG data too, given its similar time series structure. Although slow-drift is usually much less of a problem in MEG, similar high-pass filtering procedures are typically being applied (see Table 1). Second, the spurious displacements of information patterns will not only affect MVPA-based decoding of EEG or MEG data, but also analyses using inverted or forward encoding models that rely on the same type of information (e.g. Herbst, Fiedler, & Obleser, 2018). Finally, there may be important implications for fMRI analyses too. Here is where MVPA took off, with numerous studies demonstrating sustained mental representations beyond the initial stimulus presentation.

High-pass filtering is a standard step also in preprocessing fMRI data, and although event-related BOLD responses evolve at a much slower scale than typical EEG or MEG responses, high-pass filter cut offs are scaled accordingly. Notably, where in EEG or MEG typically combine trials with event structures in the order of about 2 seconds with high-pass cut-off values in the order of 0.1 Hz, in fMRI event structures are typically in the order of 20 seconds, while cut-offs used are in the order of 0.01 Hz. Interestingly, after pointing out disadvantages of high-pass filtering in fMRI time series (unrelated to decoding), Kay et al. (2008) similarly proposed detrending through polynomial regressors as a solution.

We also note that the decision on whether and how to apply high-pass filtering adds to a list of design and data processing factors that may all affect decoding results, including transformation into source space, dimensionality reduction, subsampling, aggregating signals across time, artifact rejection, trial averaging, specific classifier selection, and the specific cross-validation design used (Grootswagers, Wardle, & Carlson, 2017). Most notably within the current context, Grootswagers et al. argued for caution when applying *low*-pass filtering (see also Vanrullen, 2011). With too low cut-off values, low-pass filtering too can cause significant decoding to emerge when in fact no signal exists in the original data.

Further, it may be the case that some filters are more problematic than others, depending on the scientific context (i.e. paradigm, research question, and outcome measure). Here we have explored the impact of a two-way sinc FIR filter with a Kaiser window, but other options, such as the common 4th order Butterworth filter (Tanner et al., 2015), may produce different results. In doing so, we have also not considered fundamental differences between filter types. Causal filters for example (such as online filters) only take samples from the past and the present into consideration. Naturally, these can never lead to displacement of information backward in time as observed here, although they can still lead to displacement forward in time. Acausal filters on the other hand (such as the offline filter we used here), take into account information from the future and the past. These types of filters are particularly popular when filtering EEG, because they are able in principle to filter the data without changing the underlying phase of the signal (such filters that combine forward and backward filtering are also called zero-phase filters). However, as we have seen here, the promise not to affect the phase of the signal can come at a significant cost, which is that the causal chain of events that the EEG signal attempts to capture can be

compromised. How problematic various filter types are in the context of MVPA remains a question for future research, and depends on the research question that one is trying to answer.

In conclusion, filtering of neural time series data may be problematic in more than one respect, but here we show that it becomes particularly troublesome in the advent of modern decoding methods, as it can create widespread displacement of information onto time points where no information was present. We also show that while robust detrending provides a solution, no detrending at all may often be good enough. Based on our current findings we therefore recommend extreme caution with regards to high-pass filtering EEG and MEG time series data for MVPA purposes, in particular when using slow paradigms such as found in working memory tasks, and in particular when looking at temporal generalization where spurious results were very pronounced in our empirical dataset. More specifically, we recommend the following steps:

1. Assess the general data quality (unspecific to condition differences). If the quality is good, consider not doing any form of detrending at all – whether through high-pass filtering or other methods (Luck, 2005). As our own results show, when the data is good baseline correction is often sufficient, so that decoding is likely to work just fine without removing slow trends.
2. This might not be sufficient when the relevant signal extends over longer periods of time. In working memory tasks for example, the retention interval is relatively long, and therefore easily affected by slow drifts. In such cases, one might consider the method of robust detrending (de Cheveigne & Arzounian, 2018) while masking out the ERP and other potentially cognitive events such as an ensuing retention interval. When signal and noise are likely to reside in the same (low) frequency bands, explicitly masking out the signal during detrending precludes the risk that the detrending operation is affected by it, as would certainly be the case when high-pass filtering on the continuous signal. A related advantage is that this decreases the risk of throwing out real effects. Using this method, we found a modest improvement in decoding accuracy compared to decoding the raw data, in particular when looking at temporal generalization.

Still, this method requires one to be aware of several parameters that may affect the results.

3. If there are good reasons to dismiss steps 1 and 2 and to still prefer standard high-pass filtering, then systematically explore the cut-off parameter space to assess when spurious enhancement of decoding starts to emerge, and pick a cut-off value well below that (see also Tanner et al., 2015; Tanner et al., 2016). Given that we found artifacts emerging with cut-off values as low as 0.1 Hz, our choice would be in the range of 0.05 and lower, but this may be different for different event structures and spacing, as there may also be interactions with the inter-trial interval (a topic that we chose not to explore in the current study). But even under conservative filter settings, one should be aware not to overinterpret the precise timing of decoding onsets and offsets when using any kind of filter.

Table 1. A non-exhaustive list of EEG and MEG studies that have used MVPA decoding techniques after applying band-pass filtering. High-pass and low-pass cut-off values are provided. Note this table is only intended illustrate the wide-ranging use of high-pass filters in EEG/MEG, and not to suggest that anything is necessarily wrong with these studies. For example, different studies may use different filter types: online (causal) or offline (either causal or acausal), Finite Impulse Response (FIR) or Infinite Impulse Response (IIR), different filter lengths and so forth, and each of these filter types may have different effects on the data that do not necessarily have to be problematic in the scientific context in which they are applied.

Publication	*EEG*	*MEG*	*High pass cut-off (Hz)*	*Low pass cut-off (Hz)*
(Alizadeh, Jamalabadi, Schonauer, Leibold, & Gais, 2017)	•		0.1	40
(Bae & Luck, 2018)	•		0.1	80
(Bae & Luck, 2019)	•		0.1	80
(Barragan-Jason, Cauchoix, & Barbeau, 2015)	•		0.1	40
(Borst, Ghuman, & Anderson, 2016)		•	0.5	50
(Borst, Schneider, Walsh, & Anderson, 2013)	•		0.5	30
(Brandmeyer, Farquhar, McQueen, & Desain, 2013)	•		1	25
(Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011)		•	-	-
(Carlson, Tovar, Alink, & Kriegeskorte, 2013)		•	0.1	200
(Cauchoix, Barragan-Jason, Serre, & Barbeau, 2014)			0.1	40
(Chan, Halgren, Marinkovic, & Cash, 2011)	•	•	0.1	200
(Cichy & Pantazis, 2017)	•	•	0.03	300
(Cichy, Pantazis, & Oliva, 2014)		•	0.03	330
(Cichy, Ramirez, & Pantazis, 2015)		•	0.03	330
(Clarke, Devereux, Randall, & Tyler, 2015)		•	0.03	40
(Correia, Jansma, Hausfeld, Kikkert, & Bonte, 2015)	•		0.1	100
(Fahrenfort, Grubert, Olivers, & Eimer, 2017)	•		0.1	-
(Fahrenfort, van Leeuwen, et al., 2017)	•		0.1	-
(Herrmann, Maess, Kalberlah, Haynes, & Friederici, 2012)		•	2	10
(Hogendoorn & Burkitt, 2018)	•		-	-
(Hogendoorn, Verstraten, & Cavanagh, 2015)	•		-	-
(Isik, Meyers, Leibo, & Poggio, 2014)		•	2 (0.01)	100
(Kaiser, Azzalini, & Peelen, 2016)		•	1	330
(Kaiser, Oosterhof, & Peelen, 2016)		•	1	300
(King, Pescetelli, & Dehaene, 2016)		•	0.1	30
(LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013)	•		1	55
(Marti & Dehaene, 2017)		•	0.1	30
(Marti, King, & Dehaene, 2015)		•	0.1	330
(Mohsenzadeh, Qin, Cichy, & Pantazis, 2018)		•	0.03	330

(Mostert, Kok, & de Lange, 2015)		●	-	30
(Myers et al., 2015)	●	●	0.03	300
(Nemrodov, Niemeier, Mok, & Nestor, 2016)	●		0.1	40
(Nemrodov, Niemeier, Patel, & Nestor, 2018)	●		0.1	40
(Peters, Bledowski, Rieder, & Kaiser, 2016)		●	0.1	150
(Rose et al., 2016)	●		1	60
(Simanova, van Gerven, Oostenveld, & Hagoort, 2010)	●		1	30
(Sudre et al., 2012)		●	0.1	50
(Trubutschek et al., 2017)		●	0.1	330
(Turner, Johnston, de Boer, Morawetz, & Bode, 2017)	●		0.1	70
(Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016)		●	0.1	200
(Wolff et al., 2015)	●		0.1	40
(Wolff, Jochim, Akyurek, & Stokes, 2017)	●		0.1	40

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 (vol 209, pg 212, 2012). *Journal of Neuroscience Methods, 211*(2), 309-309. doi:10.1016/j.jneumeth.2012.08.023

Alizadeh, S., Jamalabadi, H., Schonauer, M., Leibold, C., & Gais, S. (2017). Decoding cognitive concepts from neuroimaging data using multivariate pattern analysis. *Neuroimage, 159*, 449-458. doi:10.1016/j.neuroimage.2017.07.058

Bae, G. Y., & Luck, S. J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *Journal of Neuroscience, 38*(2), 409-422. doi:10.1523/JNEUROSCI.2860-17.2017

Bae, G. Y., & Luck, S. J. (2019). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *Neuroimage, 184*, 242-255. doi:10.1016/j.neuroimage.2018.09.029

Barragan-Jason, G., Cauchoix, M., & Barbeau, E. J. (2015). The neural speed of familiar face recognition. *Neuropsychologia, 75*, 390-401. doi:10.1016/j.neuropsychologia.2015.06.017

Bode, S., Feuerriegel, D., Bennett, D., & Alday, P. M. (2018). The Decision Decoding ToolBOX (DDTBOX) - A Multivariate Pattern Analysis Toolbox for Event-Related Potentials. *Neuroinformatics*. doi:10.1007/s12021-018-9375-z

Borst, J. P., Ghuman, A. S., & Anderson, J. R. (2016). Tracking cognitive processing stages with MEG: A spatio-temporal model of associative recognition in the brain. *Neuroimage, 141*, 416-430. doi:10.1016/j.neuroimage.2016.08.002

Borst, J. P., Schneider, D. W., Walsh, M. M., & Anderson, J. R. (2013). Stages of processing in associative recognition: evidence from behavior, EEG, and classification. *J Cogn Neurosci, 25*(12), 2151-2166. doi:10.1162/jocn_a_00457

Brandmeyer, A., Farquhar, J. D., McQueen, J. M., & Desain, P. W. (2013). Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. *PLoS One, 8*(7), e68261. doi:10.1371/journal.pone.0068261

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *J Vis, 11*(10). doi:10.1167/11.10.9

Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *J Vis, 13*(10). doi:10.1167/13.10.1

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *Journal of Neuroscience, 34*(3), 846-854. doi:10.1523/JNEUROSCI.3030-13.2014

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage, 54*(4), 3028-3039. doi:10.1016/j.neuroimage.2010.10.073

Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *Neuroimage, 158*, 441-454. doi:10.1016/j.neuroimage.2017.07.023

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nat Neurosci, 17*(3), 455-462. doi:10.1038/nn.3635

Cichy, R. M., Ramirez, F. M., & Pantazis, D. (2015). Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *Neuroimage, 121*, 193-204. doi:10.1016/j.neuroimage.2015.07.011

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. *Cereb Cortex, 25*(10), 3602-3612. doi:10.1093/cercor/bhu203

Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., & Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Frontiers in Psychology, 6*, 71. doi:10.3389/fpsyg.2015.00071

de Cheveigne, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *Neuroimage, 172*, 903-912. doi:10.1016/j.neuroimage.2018.01.035

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. doi:10.1016/j.jneumeth.2003.10.009

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience, 30*(49), 16601-16608. doi:10.1523/JNEUROSCI.2770-10.2010

Endl, W., Walla, P., Lindinger, G., Lalouschek, W., Barth, F. G., Deecke, L., & Lang, W. (1998). Early cortical activation indicates preparation for retrieval of memory for faces: an event-related potential study. *Neurosciene Letters, 240*(1), 58-60.

Fahrenfort, J. J., Grubert, A., Olivers, C. N. L., & Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports, 7*(1), 1886. doi:10.1038/s41598-017-01911-0

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). *Frontiers in Neuroscience, 12*. doi:ARTN 368
10.3389/fnins.2018.00368

Fahrenfort, J. J., van Leeuwen, J., Olivers, C. N., & Hogendoorn, H. (2017). Perceptual integration without conscious access. *Proc Natl Acad Sci U S A, 114*(14), 3744-3749. doi:10.1073/pnas.1617268114

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J Cogn Neurosci, 29*(4), 677-697. doi:10.1162/jocn_a_01068

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning, 45*(2), 171-186.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Frund, I., Rieger, J. W., . . . Pollmann, S. (2009). PyMVPA: A Unifying Approach to the Analysis of Neuroscientific Data. *Front Neuroinform, 3*, 3. doi:10.3389/neuro.11.003.2009

Herbst, S. K., Fiedler, L., & Obleser, J. (2018). Tracking Temporal Hazard in the Human Electroencephalogram Using a Forward Encoding Model. *eNeuro, 5*(2). doi:10.1523/ENEURO.0017-18.2018

Herrmann, B., Maess, B., Kalberlah, C., Haynes, J. D., & Friederici, A. D. (2012). Auditory perception and syntactic cognition: brain activity-based decoding within and across subjects. *Eur J Neurosci, 35*(9), 1488-1496. doi:10.1111/j.1460-9568.2012.08053.x

Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. *Neuroimage, 171*, 55-61. doi:10.1016/j.neuroimage.2017.12.063

Hogendoorn, H., Verstraten, F. A., & Cavanagh, P. (2015). Strikingly rapid neural basis of motion-induced position shifts revealed by high temporal-resolution EEG pattern classification. *Vision Res, 113*(Pt A), 1-10. doi:10.1016/j.visres.2015.05.005

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *J Neurophysiol, 111*(1), 91-102. doi:10.1152/jn.00394.2013

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *J Neurophysiol, 115*(4), 2246-2250. doi:10.1152/jn.01074.2015

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *Journal of Neuroscience, 36*(41), 10522-10528. doi:10.1523/Jneurosci.1385-16.2016

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology, 47*(5), 888-904. doi:10.1111/j.1469-8986.2010.01009.x

Kay, K. N., David, S. V., Prenger, R. J., Hansen, K. A., & Gallant, J. L. (2008). Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp, 29*(2), 142-156. doi:10.1002/hbm.20379

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences, 18*(4), 203-210. doi:10.1016/j.tics.2014.01.002

King, J. R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron, 92*(5), 1122-1134. doi:10.1016/j.neuron.2016.10.051

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: an EEG study. *J Cogn Neurosci, 25*(1), 127-142. doi:10.1162/jocn_a_00305

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

Maess, B., Schroger, E., & Widmann, A. (2016). High-pass filters and baseline correction in M/EEG analysis-continued discussion. *Journal of Neuroscience Methods, 266*, 171-172. doi:10.1016/j.jneumeth.2016.01.016

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications, 8*. doi:ARTN 1955
10.1038/s41467-017-02079-x

Marti, S., King, J. R., & Dehaene, S. (2015). Time-Resolved Decoding of Two Processing Chains during Dual-Task Interference. *Neuron, 88*(6), 1297-1307. doi:10.1016/j.neuron.2015.10.040

Meyers, E. M. (2013). The neural decoding toolbox. *Front Neuroinform, 7*, 8. doi:10.3389/fninf.2013.00008

Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *Elife, 7*. doi:10.7554/eLife.36329

Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports, 5*, 18253. doi:10.1038/srep18253

Myers, N. E., Rohenkohl, G., Wyart, V., Woolrich, M. W., Nobre, A. C., & Stokes, M. G. (2015). Testing sensory evidence against mnemonic templates. *Elife, 4*, e09000. doi:10.7554/eLife.09000

Nemrodov, D., Niemeier, M., Mok, J. N. Y., & Nestor, A. (2016). The time course of individual face recognition: A pattern analysis of ERP signals. *Neuroimage, 132*, 469-476. doi:10.1016/j.neuroimage.2016.03.006

Nemrodov, D., Niemeier, M., Patel, A., & Nestor, A. (2018). The Neural Dynamics of Facial Identity Processing: Insights from EEG-Based Pattern Analysis and Image Reconstruction. *eNeuro, 5*(1). doi:10.1523/ENEURO.0358-17.2018

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front Neuroinform, 10*, 27. doi:10.3389/fninf.2016.00027

Peters, B., Bledowski, C., Rieder, M., & Kaiser, J. (2016). Recurrence of task set-related MEG signal patterns during auditory working memory. *Brain Res, 1640*(Pt B), 232-242. doi:10.1016/j.brainres.2015.12.006

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136-1139. doi:10.1126/science.aah7011

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS One, 5*(12), e14465. doi:10.1371/journal.pone.0014465

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., & Mitchell, T. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *Neuroimage, 62*(1), 451-463. doi:10.1016/j.neuroimage.2012.04.048

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. doi: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. doi:10.1016/j.jneumeth.2016.01.002

Trubutschek, D., Marti, S., Ojeda, A., King, J. R., Mi, Y., Tsodyks, M., & Dehaene, S. (2017). A theory of working memory without consciousness or sustained activity. *Elife, 6*. doi:10.7554/eLife.23871

Turner, W. F., Johnston, P., de Boer, K., Morawetz, C., & Bode, S. (2017). Multivariate pattern analysis of event-related potentials predicts the subjective relevance of everyday objects. *Consciousness and Cognition, 55*, 46-58. doi:10.1016/j.concog.2017.07.006

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

Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S. M., & Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *Neuroimage, 132*, 59-70. doi:10.1016/j.neuroimage.2016.02.019

Widmann, A., & Schroger, E. (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. *Frontiers in Psychology, 3*. doi:ARTN 233 10.3389/fpsyg.2012.00233

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

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavioral Research Methods, 42*(3), 671-684. doi:10.3758/BRM.42.3.671

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers Systems Neuroscience, 9*, 123. doi:10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyurek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience, 20*(6), 864-871. doi:10.1038/nn.4546

SUPPLEMENTARY FIGURES

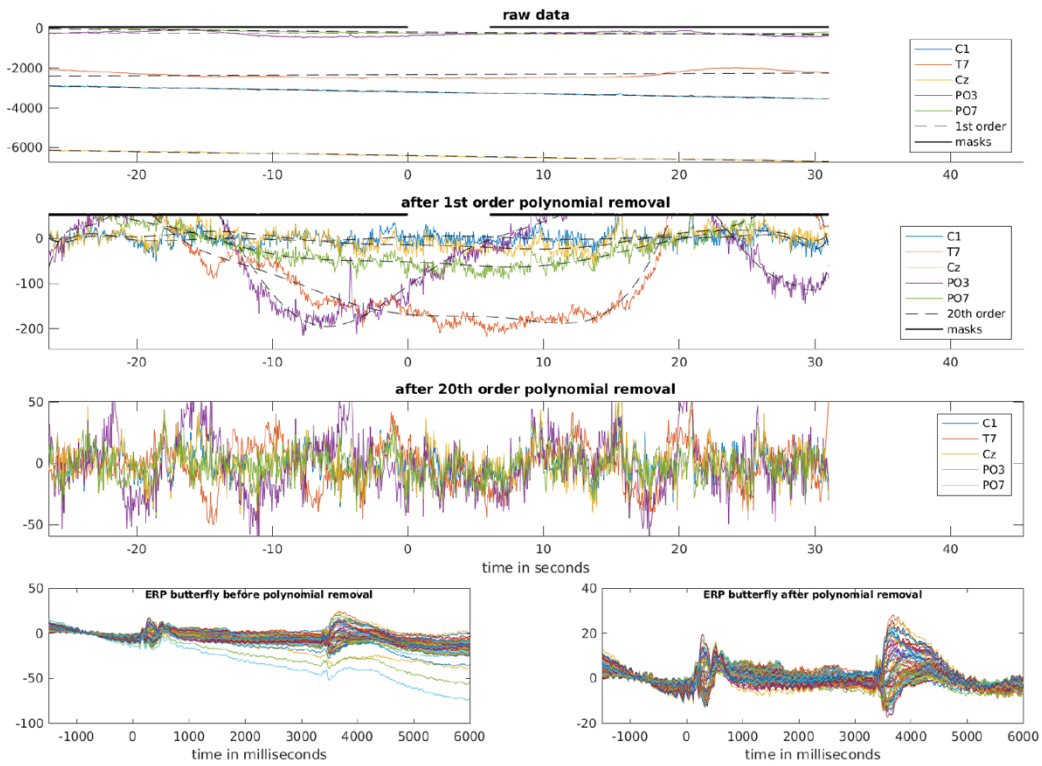

Figure S1. Example figure of a representative subject of the empirical data, generated by the function adam_detrend_and_epoch. The MATLAB function adam_detrend_and_epoch that is included with the ADAM toolbox (Fahrenfort et al., 2018) is able to detrend and epoch continuous data in EEGLAB format. For every subject, it produces a plot like the above. The top three panels show the results of the detrending operation for a trial in the middle of the dataset, for five electrodes that show the largest deviation from center. These three panels are analogous to the panels that are shown in Figure 3 of the main manuscript, but now for empirical data. The bottom two panels show a butterfly plot of the ERPs before robust detrending (left panel) and after robust detrending (right panel).

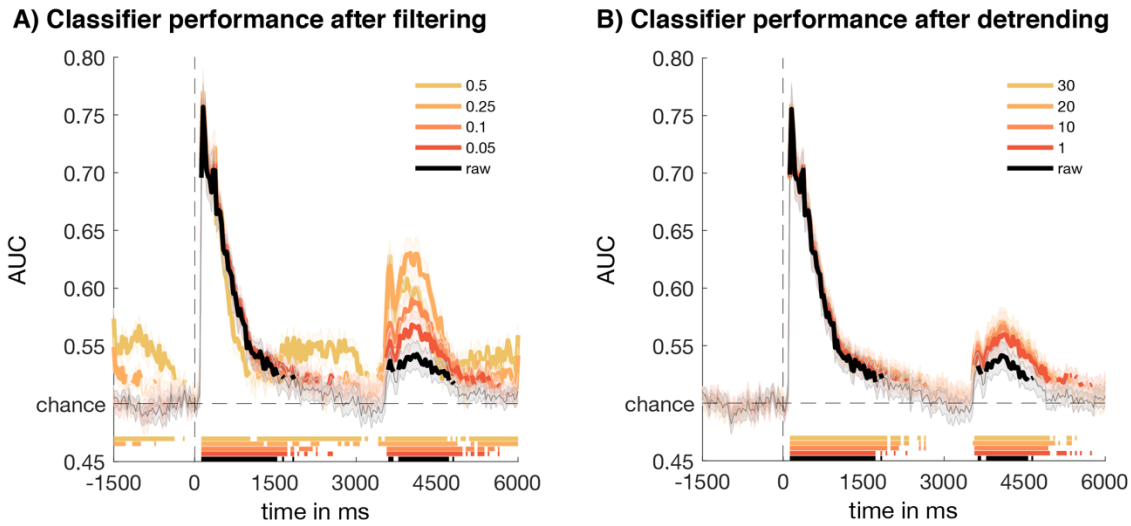

Figure S2. Results for empirical data from a working memory guided search task. This figure shows the same results as in Figure 4 of the main manuscript, but now showing the uncorrected significance values prior to cluster-based permutation.

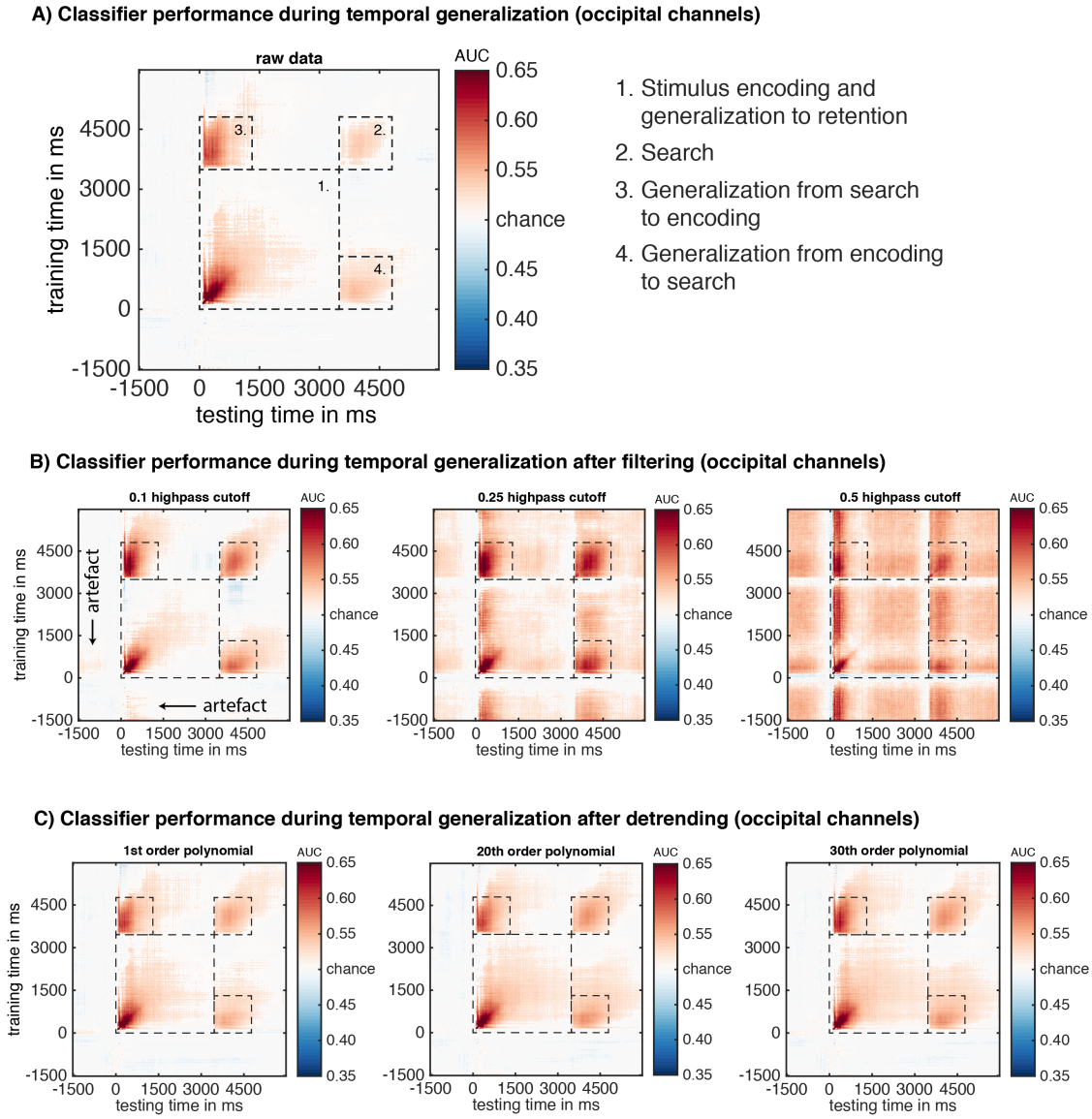

Figure S3. Temporal generalization results for the empirical data. This figure shows the same results as in Figure 5 of the main manuscript, but now showing the uncorrected significance values prior to cluster-based permutation. Note that artefacts emerge with high-pass filtering already for the 0.1 Hz cutoff (B, leftmost panel). Detrending (C) does not results in these artefacts and the pattern resembles more closely the results for the raw data (A). Moreover, a clear performance benefit can be observed in the retention interval when comparing decoding after 30th order polynomial removal to decoding on the raw data (better generalization).

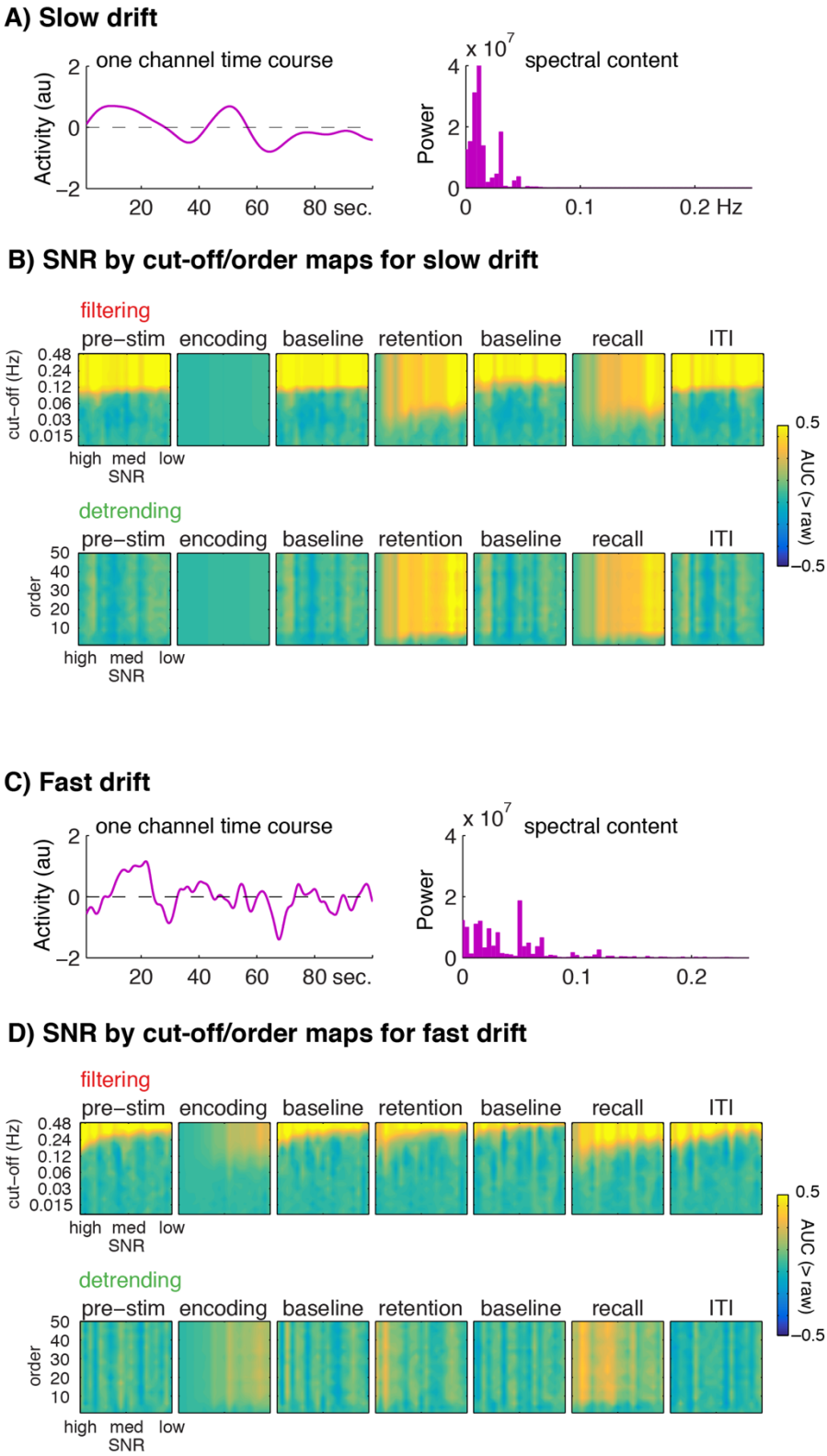

Figure S4. Slow and fast drifts in simulated data. A) Example time course of 1/f pink noise slow drift as was added to the data (left panel), and its spectral content (right panel). B) The relationship

between slow drift signal to noise ratio (on the x-axis), filter cutoff / detrending polynomial (on the y-axis) and decoding (in color). C) Example time course of 1/f pink noise fast drift as was added to the data (left panel), and its spectral content (right panel). D) same as in B, but now for fast drifts. Note that the graphs in B and D show the difference in decoding between raw data and filtered/detrended data.