Research Article

Artifact Subspace Reconstruction (ASR) for electroencephalography artifact removal must be optimized for each unique dataset

Patrick Bloniasz[1,2]

1. Boston University, United States; 2. Program in Neuroscience, Bowdoin College, United States

Artifact subspace reconstruction (ASR) is an automatic artifact reject method that can effectively remove transient or large-amplitude artifacts found in electroencephalographic (EEG) data. There is little systematic evidence on the effective parameter choice of ASR in real EEG data. No existing study has evaluated ASR's performance in functional connectivity analysis, such as renormalized Partial Directed Coherence (rPDC). This paper systematically evaluates ASR on 31 EEG recordings taken during a source episodic memory retrieval task. Independent component analysis (ICA) and an independent component classifier, ICLabel, are applied to separate artifacts from brain signals to quantitatively assess the effectiveness of ASR. The effectiveness of ASR was quantified on the following metrics: the number of dipolar independent components, model order for multivariate autoregressive modeling, and the number of preserved trials. Results showed that ASR is either as effective or more effective than manual rejection of artifacts. Contrary to previous literature, the present study shows that the optimal ASR parameter could be substantially higher than 20 to 30 and could be as high as 120, depending on experimenter decisions for what to preserve. As such, ASR parameter choice should be justified in each study using quantitative preliminary analysis. This is the first study to systematically analyze ASR's effectiveness in rPDC-based functional connectivity research. NOTE: This is the first draft; several methodological changes might occur at a later time upon further analysis.

Electroencephalography (EEG) is a relatively low-cost, non-invasive method to measure cortical brain activity in human participants. It has found wide appeal in research settings, due to it being well equipped to answer questions about the temporal dynamics of high-level cognitive functions. Relative to other techniques, such as positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), EEG allows for the investigation of higher-order cognitive processes with high temporal resolution on the scale of milliseconds.

Analyzing these dynamics allows for the discernment of 1) a temporal order of processes in a given construct, such as episodic memory (e.g., Zion-Golumbic et al., 2010), or 2) the presence of neural dysfunction, such as mood and anxiety disorders (e.g., Al Zoubi et al., 2019). When coupled with fMRI, which has high spatial resolution, EEG allows for the characterization of functional networks in the brain with high spatiotemporal resolution (e.g., Shu et al., 2021).

EEG also has promising applications in the clinical sphere including, but not limited to, personalized medicine (Keizer, 2021), anesthesia administration (Sun et al., 2020), and diagnostic procedures (Adamou et al., 2020). However, EEG is still rightfully resisted in a wide-reaching clinical setting because many analysis techniques are not standardized (Robbins et al., 2020) and both the algorithmic and manual preprocessing steps used to clean the data are often insufficiently reported or contextualized (Clayson et al., 2019). As a result, research can differ even between studies that use the same dataset, simply from minor changes in the order of preprocessing steps or the parameters used.

Calls for more standardized preprocessing pipelines are not new (Gross et al., 2013; Keil et al., 2014; Pernet et al., 2018). In fact, several attempts at standardizing pipelines have already been released, with one popular example being the Harvard Automated Processing Pipeline for Electroencephalography (HAPPE) (Gabard-Durnam et al., 2018). However, even with these pipelines, there is an enormous amount of space for the researcher to alter parameters or algorithms used that make results non-replicable. Further, there is a serious question of whether standardization necessarily produces robust analyses.

It is reasonable to doubt whether it is possible for a single, or even a few, universal pipeline(s) to exist that work across EEG experiential designs (Robbins et al., 2020). Analyses need to become more transparent in order to explicitly state and justify the parameter choices made for a given study, rather than relying on choices from previous research that might not apply in a different setting. While it is not a desirable goal to 'take out' the autonomy of the expert researcher in designing a methodological pipeline–nor it is not possible–it is essential for the integrity of the field to build more robust analyses. One way to approach more robust research, as measured through measures of validity and replicability, is through using automated algorithms to remove the researcher-to-researcher variability introduced at each analysis step.

 doi.org/10.32388/SMEI50

There are many open-source or freely available algorithmic toolboxes available, including those packaged in, or compatible with, the popular GUI MATLAB program EEGLAB. One algorithm that comes preinstalled with EEGLAB 2021 is artifact subspace reconstruction (ASR). ASR was first proposed by Mullen et al. (2013). It is an automatic, component-based artifact removal method. It effectively removes transient, large-amplitude or uniquely variable artifacts in multi-channel EEG recordings. It does so by first automatically selecting 'clean' portions of data to serve as a reference and then applying a principle component analysis (PCA) to remove artifacts relative to the selected reference. The intuition for this approach is that meaningful brain signals across channels, especially those clustered in a particular scalp region, should have low variation relative to artifacts. As such, artifacts are detectable relative to regions with minimized variance. For more technical discussion on the function of ASR, see Chang et al. (2019).

ASR has been adopted widely across contexts (e.g., Bulea et al., 2014; Mullen et al., 2015; Perera et al., 2016; Artoni et al., 2017; Blum et al., 2019). However, because of this widespread adoption, methodological research has not been able to effectively keep up with understanding the best-practice use and parameter choices of ASR in a specific area. The best attempt thus far is by Chang et al. (2019) where they convincingly demonstrate that ASR is a powerful tool for removing artifacts both online and offline and that it has an optimal parameter choice of 20-30 standard deviation (SD). This parameter range, they argue, was the best choice for their real data to balance the removal of non-brain signals and retaining brain signals. While Chang et al. (2019) conducted crucial work, it is worth testing whether this recommendation remains ideal in other types of data sets (e.g., muscle movement tasks, cognitive tasks) where artifacts look very similar to meaningful brain signals. It is also worth testing the performance of ASR in areas that rely heavily on the low variability of data, such as in functional connectivity analysis.

Recently, ASR has been used in coordination with information-theoretic techniques such as renormalized Partial Directed Coherence (rPDC) to infer the directional flow of information between brain regions–a useful step in functional connectivity analysis (Loo et al., 2019; Koshiyama et al., 2020a; Koshiyama et al., 2020b; Koshiyama et al., 2020c; Miyakoshi et al., 2021; Jurgiel et al., 2021). rPDC and other similar techniques such as Granger-Geweke causality analysis are based around using multivariate autoregressive models (MVAR) to describe empirical data, and then those models are used to determine the directional flow of information between brain regions (for full discussion on each, see Schelter et al., 2009 and Seth et al., 2015, respectively).

As a brief aside, the following questions can be used to understand the logic of the simplest form of information-theoretic measures: Granger-Geweke causality analysis. Suppose an EEG channel worth of data *X* and another simultaneously recorded channel worth of data *Y*. First, how well can we predict X_{t+1} based on the preceding data points in *X*? In other words, if we only knew the characteristics of one of the channels, how well does the time series predict itself based on how it behaved in the past? This is heavily dependent on how variable the data are. Second, how well can X_{t+1} be predicted based on the preceding data points in *X and* the past terms of the other channel *Y* ? In other words, how well can the model predict the behavior of channel *X* if we have information both about *X* and *Y*? Finally, how do these predictions compare? If the model that takes both channels into account better predicts the overall behavior, *Y* is sending information to *X*.

rPDC essentially uses the same logic as Granger-Geweke causality analysis, except with some mathematical differences that do not necessarily drastically change the interpretation of the results (i.e., some brain region is sending information to another brain region). For full discussion on rPDC, see Schelter et al. (2009) and for a comparison on the intricate differences between most information-theoretic techniques including the non-renormalized version of rPDC (i.e., PDC) and Granger-Geweke causality analysis, see Gourévitch et al. (2006).

These techniques are powerful for functional connectivity analysis, but have drawbacks, mainly related to how data are altered in preprocessing steps, such as artifact rejection or filtering. For example, Granger-Geweke causality is theoretically invariant to filtering, but practically can be disrupted by high- and low-pass filtering since it will alter the success of MVAR model fitting (Florin et al., 2010; Barnett and Seth, 2011). Thus, it is important to understand how ASR impacts key measures of success in MVAR fitting for a given information theoretic technique.

The present study set out to investigate how different parameters used in ASR affect several widely accepted measures of success when manually rejecting data and when fitting MVAR models for rPDC. Specifically, model order,[1] number of preserved trails, and number of brain-related independent components were analyzed for 7 different parameter conditions:[2] ASR parameters of 20 SD, 40 SD, 60 SD, 80 SD, 100 SD, and 120 SD, along with a control condition of expert manual rejection. These ASR parameters were selected because Chang et al. (2018) demonstrated that the optimal empirical range is between 10 SD and 100 SD and Anders et al. (2020) similarly demonstrated that there is no value in picking ASR values below 10. The present preliminary results from using ASR vs hand rejection suggested that values as high as 120 SD could produce favorable effects. Thus, the analysis expanded beyond Chang et al. (2018)'s recommendation to a maximum of 120 SD.

The present study was an exploratory, benchmark analysis on the best-practice parameters to use on a specific data set. As such, no specific hypotheses were generated. However, a few general trends were expected. First, higher SD values result in rejecting fewer data; in this case, it would be expected that fewer trials would be rejected. Second, for the same reason, as SD increased, there would be more data to decompose in independent

component analysis (ICA), which would then potentially result in a better chance of brain ICs being identified. This expectation might be complicated, however, because it has been shown that the higher the variance of the data to which ICA is applied, the worse ICA decomposition becomes (Delorme et al., 2005); as such, it is possible the number of brain related ICs detected might decrease if the data are still highly variable after ASR is applied. Finally, it is expected that the lower the lower the SD for ASR, the lower the model order of the MVAR fitting. In the most basic sense, if there are fewer data left after more aggressive (lower) parameter selection for ASR, then necessarily the data that is left for MVAR fitting has a relatively low variance. If that is the case, it should be easier to construct a model that requires fewer parameters (i.e., model order).

Method

Data Source and Computational Design

Data were collected as part of a larger experimental design that was studying source episodic memory retrieval in the context of the "old/new" effect; data correspond to experiment one in Nyhus (2010). The details of that experiment have been redescribed below for clarity. These data have been used for several subsequent studies, such as Bloniasz (2022) and Patel (2020). For full discussion of the theoretical background, see Nyhus (2010), Patel (2020), and Bloniasz (2022).

Participants completed two sessions of the experimental design. As such, there were seven total conditions for testing the ASR algorithm's effects on the data (described in "Artifact subspace reconstruction (ASR)" section) across two sessions (session 1, session 2) in two different epoch conditions (Correct rejection and Hit, described below). Thus, in-line with the expected use of ASR, the algorithm's performance was analyzed in real experimental conditions on real data.

Participants

Participants were made up of 17 males and 15 females, ranging between 19-29 years old; one participant was rejected because of corrupted data. All were native English speakers and were right-handed. All participants were pre-screened for neurological and psychiatric disorders via a questionnaire and all participants had normal or corrected to normal vision. All participants gave informed consent for the study. No personal identifiable information was available to the present author and all data was collected under the Nyhus (2010) protocol; as such, no IRB approval was required to work with this data.

Electroencephalography Methods

Brain activity was collected over a 128-channel scalp EEG system and was sampled at 250Hz/ch using a high-input impedance NetAmps amplifier (Electrical Geodesics Inc., Eugene, OR). The impedances of the individual channels were adjusted until they were less than 50 kΩ. EEG data were measured with respect to a vertex reference (Cz). However, to minimize the effects of reference-site activity, a cross-channel average-reference transform was used after channel rejection (Dien, 1998). EEGLAB 2021.0 (http://sccn.ucsd.edu/eeglab/) was used to epoch data, apply the ASR algorithm, perform independent component analysis (ICA), fit dipoles via DIPFIT, and characterize dipoles as brain related; each of these will be fleshed out further below.

Data were downsampled to 100 Hz/ch in EEGLAB. A linear, high-pass, FIR filter at 1 Hz was applied to the EEG data in order to improve model order and ICA performance (Klug and Gramann, 2020). The data were epoched from 500 ms pre-stimulus onset to 1500 ms post-stimulus onset and saved by condition of interest (Correct Rejection, abbreviated as 'CR,' and Hit). The theory behind the interest in CR and Hit conditions is not relevant to the present computational study. However, epoching across certain behavioral conditions is a core feature of many studies interested in information flow, so the analysis was done for each subject on each condition for practicality reasons.

Channels were removed by hand if roughly 40% of the data were unusable. After downsampling, filtering epoching, and channel rejection, EOG (eye blink) artifacts were still present in the data along with discontinuous data. For the full preprocessing pipeline, see the appendix.

Artifact subspace reconstruction (ASR)

A full technical description of ASR can be found in Chang et al. (2019) and is omitted here for brevity. Similar to Chang et al. (2018), ASR varied the SD for artifact rejection, referred to in the documentation as *k*, stepwise. In the present study, *k* was increased by 20 SD with each iteration, starting with it being inactive. When ASR was active, the following parameters were used: 20 SD, 40 SD, 60 SD, 80 SD, 100 SD, 120 SD. ASR can be run on any data set loaded into EEGLAB using the following example code for 80 SD:

```
EEG = pop_clean_rawdata(EEG, 'FlatlineCriterion' , 'off' , 'ChannelCriterion' , 'off' , 'LineNoiseCriterion' , 'off' ,
'Highpass' , 'off' , 'BurstCriterion' , 80 , 'WindowCriterion' , 'off' , 'BurstRejection' ,'on' , 'Distance' ,'Euclidian' ,
'channels', []);
```

When ASR was inactive, artifacts were rejected via manual inspection. Specifically, waveforms more positive than 100 mV or more negative than -100 mV were rejected, except for EOG artifacts. EOG artifacts left in by hand were later removed using ICA decomposition, which is standard practice. Leaving EOG artifacts in the data for later ICA decomposition is considered to preserve the largest number of trials. In theory the brain signals at a particular oscillatory frequency (e.g., theta oscillations at 4-8 Hz) are present under the EOG activity, and using ICA to remove non-brain related frequencies can salvage those trials.

Independent Component Analysis and ICLabel

The extended version infomax ICA algorithm, called in EEGLAB via (CODE), was used to decompose data. The extended parameter takes a subgaussian approach to decomposing data, which can often remove line noise at 50 Hz/60 Hz. This is of particular interest in preparing data for rPDC or Granger-Geweke causality analysis, because low-pass or notch filters, which typically remove line noise, severely impact these information-theoretic techniques (Florin et al., 2010; Barnett and Seth, 2011). The extended infomax ICA (Lee et al., 1999), which is one of the most widely used algorithms for ICA and is available in EEGLAB, was used because other studies investigating ASR have used it (Chang et al., 2018; Chang et al., 2019). There is some logic to this selection, since the extended infomax appears to be one of the best algorithms for preserving mean mutual information reduction (i.e., the information in a signal) and the dipolarity of a signal (Delorme et al., 2012).

There are two drawbacks to the extended infomax algorithm in the context of using information-theoretic techniques. First, because of the way the extended infomax ICA decomposes data, it can return different solutions on the same data (Delorme & Makeig, 2004). If getting the same solution is the end goal every time, a better algorithm would be SOBI because it does not involve random partitioning of data (Belouchrani et al., 1997). Rather, it uses cross-correlation for joint diagonalization to separate source information. Other algorithms for replicable decomposition are JADE (Rutledge and Jouan-Rimbaud Bouveresse, 2013) and PICARD (Ablin et al., 2018). Both have tradeoffs for their ability to preserve signal information and dipole fitting, though the effects can be considered marginal.

Another drawback is that the infomax ICA can drastically alter the EEG signal when used to remove EOG artifacts (Pontifex et al., 2017). This would alter the ultimate goal of MVAR model fitting and could potentially create false relationships in the data or erase real relationships in the data. As such, eyeblinks should be rejected prior to ICA for information theoretic techniques. In the present paper, eye blinks were not rejected in the manual condition prior to ICA because that is standard practice in the field and is used as a demonstration.

The extended infomax algorithm was run on epoched data in EEGLAB using the following code:

```
Ncomp = sum(eig(cov(double(EEG.data([1:EEG.nbchan],:)'))) > 1E-7);

EEG = pop_runica(EEG, 'icatype', 'runica','concatcond', 'on','extended',1,'interupt','on','pca', Ncomp, 'stop', 1e-7);
```

The ‘Ncomp’ variable is used to fix an unaddressed problem in the literature where a bug in MATLAB’s code around the ‘rank’ function can produce a phenomena called ‘ghost ICs’, which is when data rank is under-determined.[3]

During ICA decomposition, components were determined as ‘brain related’ if they were detected as being greater than 70% brain activity using the EEGLAB toolbox ICLabel, which is a widely used automatic IC classifier (Pion-Tonachini et al., 2019); ICLabel has been used in other ASR focused papers (Chang et al., 2019) and papers using rPDC in EEG data (e.g., Lou et al., 2019; Koshiyama et al. 2020). The remaining components were rejected.

Dipole localization was performed using the Template Boundary Element Model (Montreal Neurological Institute standard brain) in EEGLAB. After dipole localization, dipoles were rejected if their residual variance was greater than 15% and if their centralized dipole density was outside of the canonical head, as done in with Lou et al. (2019) and Koshiyama et al. (2020). Remaining ICs after these rejection criteria are considered ‘dipolar’ in the present study. Dipolarity is “the number of component scalp maps matching the projection of a single equivalent dipole with less than a given residual variance” (Delorme et al., 2012). In short, a dipolar IC is considered to be a biological meaningful signal. Chang et al. (2019) used the criteria of 5% residual variance; however, more recent papers using this definition of dipolarity use 15% as a cut-off (e.g., Loo et al., 2019). The resulting ICs were recorded for each subject.

Multivariate autoregressive (MVAR) model fitting

The most important step prior to conducting the hypothesis test for rPDC is MVAR model fitting. The Source Information Flow Toolbox (SIFT) (https://sccn.ucsd.edu/githubwiki/files/eeglab2011_tm_sift.pdf) is a GUI-based plugin in EEGLAB that can easily generate MVAR models. The broader function of SIFT is for modeling and visualizing dynamical interactions between electrophysiological signals (e.g., rPDC). SIFT has recently been extended into groupSIFT (https://github.com/sccn/groupSIFT) which allows for the group analysis of information flow between dipoles across many participants. As such, all EEG data underwent MVAR fitting using the SIFT plugin within groupSIFT, since groupSIFT is quickly becoming a popular option for functional

connectivity analysis (Loo et al., 2019; Koshiyama et al., 2020a; Koshiyama et al., 2020b; Koshiyama et al., 2020c; Miyakoshi et al., 2021; Jurgiel et al., 2021).

SIFT used a hamming window technique that minimizes the optimum model order. The sliding window length was at 0.5 seconds and moved stepwise from initial time t_0 to final time t_n in intervals of 0.048 seconds. The entire frequency analysis range of interest was from 2 to 30 Hz. The frequencies were broken down into 28 individual frequency bins. The full width at half maximum of the Gaussian distribution was set to 20 mm (FWHM/2.355*3)—the default parameter in groupSIFT. The order of the resulting MVAR model was recorded for each subject.

Results

Experiment 1. Trails modified across ASR parameter choices

A concern for artifact rejection, whether using algorithm-based or manual techniques, is balancing the amount of data removed with having a high enough trial count to maximize statistical power. As such, it is relevant to see how the number of trials are affected with each parameter choice for ASR. It was expected that the exploratory analysis would show a relationship between the number of trials and the parameter choice of ASR. It was predicted that as the parameter choice of ASR became more aggressive (i.e., a lower SD threshold), the fewer trials would be preserved. This general hypothesis was corroborated across all conditions. For the remainder of the results an ASR *k* (SD) value will be combined with the name (e.g., ASR SD = 40 will be "ASR 40").

Experiment 1: Session 1, Correct Rejection condition

In session 1 of the correct rejection condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 1; $f(6,210)=10.240$, MSE = 1256.819, $p=6.148 \times 10^{-10}$). This effect has medium power ($1-\beta=0.7837$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used (Table 1). There were seven statistically significant differences based on a Tukey's critical mean of 26.81.

Hand rejection preserved 55.10 more trials on average than ASR 20 ($p = 9.53 \times 10^{-8}$). Hand rejection preserved 37.97 more trials than the ASR 40 condition ($p = 7.22 \times 10^{-4}$). ASR 60 preserved 29.971 more trials on average than ASR 20 ($p = 0.0182$). ASR 80 preserved 37.065 more trials on average than ASR 20 ($p = 0.001073$). ASR 100 preserved 42.8065 more trials on average than ASR 20 ($p = 7.513 \times 10^{-5}$). ASR 120 preserved 56.387 more trials on average than ASR 20 ($p = 4.43 \times 10^{-8}$). ASR 120 preserved 39.258 more trials on average than ASR 40 ($p = 4.039 \times 10^{-4}$).

When checking the variances across each condition, a Levene's test to assess the equality of variance determined that the variances were, in fact, not all equal ($F = 3.731$, $p = 0.00151$). ASR 20 had significantly higher variance than ASR 120, with ASR 20 having a variance 19.581 trials higher on average than ASR 120 ($p = 0.00513$). ASR 40 had significantly higher variance than ASR 120, with ASR 40 having a variance 20.161 trials higher on average than ASR 120 ($p = 0.00347$). See appendix A for full analysis. As such, an additional analysis was done to account for this reality.

The Kruskal-Wallis H test indicated that there is a significant difference in the number of preserved trials between the different ASR parameters, $\chi^2(6) = 45.78$, $p < .001$, with a mean rank score of 142.6 trials for Hand Rejection, 60.56 trials for ASR 20, 83.68 trials for ASR 40, 98.65 trials for ASR 60, 109.71 trails for ASR 80, 120.56 trials for ASR 100, and 147.24 trials for ASR 120. The Post-Hoc Dunn's test using a Bonferroni corrected alpha of 0.0024 indicated that seven mean ranks were significantly different (Table 2).

Hand rejection preserved 82.032 more trials on average than ASR 20 ($p = 2.685 \times 10^{-7}$). Hand rejection preserved 58.919 more trials on average than ASR 40 ($p = 2.2 \times 10^{-4}$). ASR 80 preserved 49.145 more trials on average than ASR 20 ($p = 0.00206$). ASR 100 preserved 60 more trials on average than ASR 20 ($p = 1.681 \times 10^{-4}$). ASR 120 preserved 86.677 more trials on average than ASR 20 ($p = 5.46 \times 10^{-8}$). ASR 120 preserved 63.565 more trials on average than ASR 40 ($p = 6.714 \times 10^{-5}$). ASR 120 preserved 48.597 more trials on average than ASR 60 ($p = 0.00231$).

All but two results are consistent between the two different tests. First, with the post-hoc test for the one-way ANOVA, ASR 60 preserved statistically more trials compared to ASR 20; this did not occur in the post-hoc test for the Kruskal-Wallis ANOVA. Second, with the post-hoc test for the one-way ANOVA, ASR 120 was not significantly different from ASR 60, though it was close ($p = 0.0547$); there was a significant different in the post-hoc test for the Kruskal-Wallis ANOVA ($p = 0.00231$).

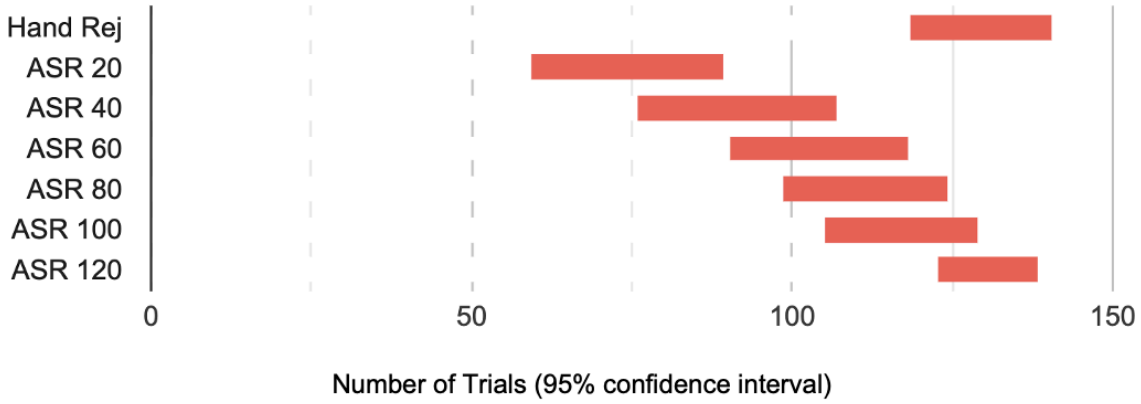

Figure 1. Shown are the mean number of trials preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 Correct Rejection data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,210)=10.240$, $MSE = 1256.819$, $p=6.148 \times 10^{-10}$). There were seven significant differences across conditions. ASR 120 preserves similar numbers of trails as the hand rejection condition. ASR 20 and ASR 40 are significantly more variable compared to ASR 120.

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	55.0968	8.6531	28.2884	81.9052	9.53E-08***
Hand Rej - ASR 40	37.9677	5.9629	11.1593	64.7761	0.000722***
Hand Rej - ASR 60	25.2258	3.9618	-1.5826	52.0342	0.08008
Hand Rej - ASR 80	18.0323	2.832	-8.7761	44.8407	0.4162
Hand Rej - ASR 100	12.2903	1.9302	-14.5181	39.0987	0.8198
Hand Rej - ASR 120	1.2903	0.2026	-25.5181	28.0987	1
ASR 40- ASR 20	17.129	2.6902	-9.6794	43.9374	0.4811
ASR 60 - ASR 20	29.871	4.6913	3.0626	56.6794	0.0182*
ASR 80 - ASR 20	37.0645	5.8211	10.2561	63.8729	0.001073**
ASR 100 - ASR 20	42.8065	6.7229	15.998	69.6149	0.00007513***
ASR 120 - ASR 20	56.3871	8.8557	29.5787	83.1955	4.43E-08***
ASR 60 - ASR 40	12.7419	2.0012	-14.0665	39.5503	0.7933
ASR 80 - ASR 40	19.9355	3.1309	-6.8729	46.7439	0.2925
ASR 100 - ASR 40	25.6774	4.0327	-1.131	52.4858	0.07027
ASR 120 - ASR 40	39.2581	6.1656	12.4497	66.0665	0.0004039***
ASR 80 - ASR 60	7.1935	1.1298	-19.6149	34.002	0.9849
ASR 100 - ASR 60	12.9355	2.0315	-13.8729	39.7439	0.7814
ASR 120 - ASR 60	26.5161	4.1644	-0.2923	53.3245	0.05469
ASR 100 - ASR 80	5.7419	0.9018	-21.0665	32.5503	0.9955
ASR 120 - ASR 80	19.3226	3.0347	-7.4858	46.131	0.33
ASR 120 - ASR 100	13.5806	2.1329	-13.2278	40.3891	0.7397

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 26.8084. The SE was 6.3673. N = 31 values.

Table 1. *Post-hoc Tukey Honest Significant Difference test comparing the impact of each ASR parameter condition on the average number of persevered trials with a 95% confidence interval (Session 1, Correct Rejections)*

Pair	Difference	Z-value	p-value
Hand Rej - ASR 20	82.0323	5.1443	2.69E-07***
Hand Rej - ASR 40	58.9194	3.6949	0.00022***
Hand Rej - ASR 60	43.9516	2.7562	0.005847
Hand Rej - ASR 80	32.8871	2.0624	0.03917
Hand Rej - ASR 100	22.0323	1.3817	0.1671
Hand Rej - ASR 120	4.6452	0.2913	0.7708
ASR 40- ASR 20	23.1129	1.4494	0.1472
ASR 60 - ASR 20	38.0806	2.3881	0.01694
ASR 80 - ASR 20	49.1452	3.0819	0.002057**
ASR 100 - ASR 20	60	3.7627	0.0001681***
ASR 120 - ASR 20	86.6774	5.4356	5.46E-08***
ASR 60 - ASR 40	14.9677	0.9386	0.3479
ASR 80 - ASR 40	26.0323	1.6325	0.1026
ASR 100 - ASR 40	36.8871	2.3132	0.02071
ASR 120 - ASR 40	63.5645	3.9862	0.00006714***
ASR 80 - ASR 60	11.0645	0.6939	0.4878
ASR 100 - ASR 60	21.9194	1.3746	0.1693
ASR 120 - ASR 60	48.5968	3.0476	0.002307**
ASR 100 - ASR 80	10.8548	0.6807	0.4961
ASR 120 - ASR 80	37.5323	2.3537	0.01859
ASR 120 - ASR 100	26.6774	1.673	0.09433
Note. ** Denotes $p < 0.0024$ *** Denotes $p < 0.001$. The critical mean for the comparisons was 48.4456. The SE was 15.9462. N = 31 values.			

Table 2. *Post-hoc Dunn's test comparing the impact of each ASR parameter condition on the average number of persevered trials (Session 1, Correct Rejections)*

Experiment 1: Session 2, Correct Rejection condition

In session 2 of the correct rejection condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 2; $f(6,210)=7.225$, MSE = 11427.7742, $p=5.036 \times 10^{-7}$). This effect has medium power ($1-\beta=0.7837$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used (Table 3). There were six statistically significant differences based on a Tukey's critical mean of 28.5736. Unlike in session 1, the variances across conditions were considered to be equal ($F = 1.79383$, $p = 0.101713$).

Hand rejection preserved 56.871 more trials on average than ASR 20 ($p = 2.631 \times 10^{-7}$). Hand rejection preserved 35.226 more trials than the ASR 40 condition ($p = 0.005614$). ASR 60 preserved 32.581 more trials on average than ASR 20 ($p = 0.0142$). ASR 80 preserved 39.677 more trials on average than ASR 20 ($p = 0.001$). ASR 100 preserved 45.129 more trials on average than ASR 20 ($p = 9.432 \times 10^{-5}$). ASR 120 preserved 37.1935 more trials on average than ASR 20 ($p = 0.002683$).

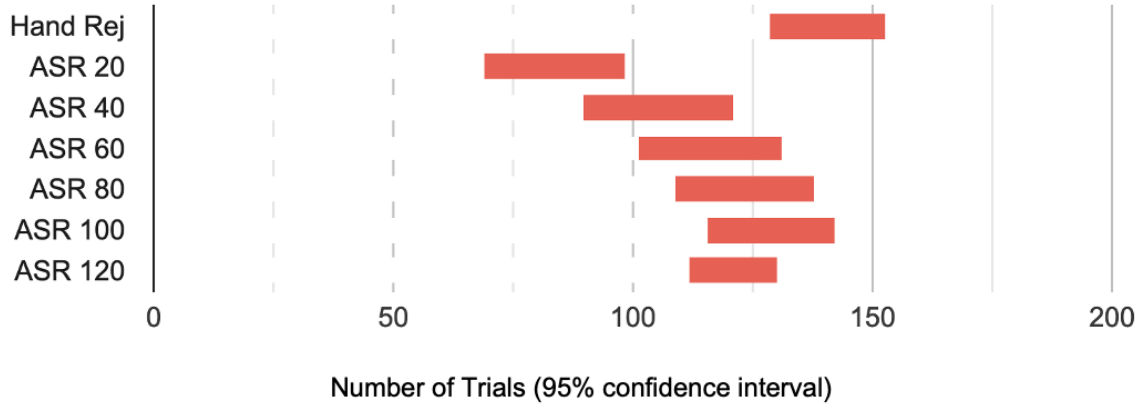

Figure 2. Shown are the mean number of trials preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 2 Correct Rejection data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,210)=7.225$, $MSE = 11427.7742$, $p=5.036 \times 10-7$). There were six significant differences across conditions. ASR 120 preserves similar numbers of trails as the hand rejection condition.

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	56.871	8.38E+00	2.83E+01	85.4445	0.0000002631 ***
Hand Rej - ASR 40	35.2258	5.1905	6.6522	63.7994	0.005614**
Hand Rej - ASR 60	24.2903	3.5792	-4.2832	52.8639	0.1537
Hand Rej - ASR 80	17.1935	2.5335	-11.38	45.7671	0.5553
Hand Rej - ASR 100	11.7419	1.7302	-16.8316	40.3155	0.8844
Hand Rej - ASR 120	19.6774	2.8995	-8.8961	48.251	0.3865
ASR 40- ASR 20	21.6452	3.1894	-6.9284	50.2187	0.271
ASR 60 - ASR 20	32.5806	4.8008	4.0071	61.1542	0.01422*
ASR 80 - ASR 20	39.6774	5.8465	11.1039	68.251	0.001**
ASR 100 - ASR 20	45.129	6.6498	16.5555	73.7026	0.00009432***
ASR 120 - ASR 20	37.1935	5.48E+00	8.62E+00	65.7671	0.002683**
ASR 60 - ASR 40	10.9355	1.6113	-17.6381	39.509	0.9151
ASR 80 - ASR 40	18.0323	2.6571	-10.5413	46.6058	0.4967
ASR 100 - ASR 40	23.4839	3.4604	-5.0897	52.0574	0.1847
ASR 120 - ASR 40	15.5484	2.2911	-13.0252	44.1219	0.6696
ASR 80 - ASR 60	7.0968	1.0457	-21.4768	35.6703	0.9899
ASR 100 - ASR 60	12.5484	1.849	-16.0252	41.1219	0.8479
ASR 120 - ASR 60	4.6129	0.6797	-23.9607	33.1865	0.9991
ASR 100 - ASR 80	5.4516	0.8033	-23.1219	34.0252	0.9976
ASR 120 - ASR 80	2.4839	0.366	-26.0897	31.0574	1
ASR 120 - ASR 100	7.9355	1.1693	-20.6381	36.509	0.9819

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 28.5736. The SE was 6.7865. N = 31 values.

Table 3. *Post-hoc Tukey Honest Significant Difference test comparing the impact of each ASR parameter condition on the average number of persevered trials with a 95% confidence interval (Session 1, Correct Rejections)*

Experiment 1: Session 1, Hit condition

In session 1 of the Hit condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 1; f(6,210)=20.6679, MSE = 703.5462, p = 0). This effect has medium power (1-β=0.7837), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used (Table 4). There were 11 statistically significant differences based on a Tukey's critical mean of 20.0577.

Hand rejection preserved 62.29 more trials on average than ASR 20 ($p = 1.073 \times 10^{-10}$). Hand rejection preserved 41.871 more trials than the ASR 40 condition ($p = 5.705 \times 10^{-8}$). ASR 120 preserved 26.968 more trials on average than ASR 60 ($p = 0.001662$). ASR 60 preserved 35.323 more trials on average than ASR 20 ($p = 7.932 \times 10^{-6}$). ASR 80 preserved 43 more trials on average than ASR 20 ($p = 2.307 \times 10^{-8}$). ASR 100 preserved 50.161 more trials on average than ASR 20 ($p = 1.589 \times 10^{-10}$). ASR 120 preserved 55.064 more trials on average than ASR 20 ($p = 1.079 \times 10^{-10}$). ASR 80 preserved 22.581 more trials on average than ASR 40 ($p = 0.01632$). ASR 100 preserved 29.742 more trials on average than ASR 40 ($p = 0.0003219$). ASR 120 preserved 34.645 more trials on average than ASR 40 ($p = 1.277 \times 10^{-5}$).

When checking the variances across each condition, a Levene's test to assess the equality of variance determined that the variances were, in fact, not all equal (F = 3.251009, p = 0.00446). ASR 20 had significantly higher variance than ASR 120, with ASR 20 having a variance 14.161 trials higher on average than ASR 120 (p = 0.013). See appendix B for full analysis. As such, an additional analysis was done to account for this reality.

The Kruskal-Wallis H test indicated that there is a significant difference in the dependent variable between the different groups, $\chi^2(6) = 35.79$, $p < .001$, with a mean rank score of 147.53 trials for Hand Rejection, 60.81 trials for ASR 20, 90.63 trials for ASR 40, 105.81 trials for ASR 60, 117.94 trials for ASR 80, 125.92 trials for ASR 100, and 114.37 trials for ASR 120. The Post-Hoc Dunn's test using a Bonferroni corrected alpha of 0.0024 indicated that five mean ranks were significantly different (Table 5).

Hand rejection preserved 86.726 more trials on average than ASR 20 ($p = 5.371 \times 10^{-8}$). Hand rejection preserved 56.903 more trials on average than ASR 40 ($p = 3.592 \times 10^{-4}$). ASR 80 preserved 57.129 more trials on average than ASR 20 ($p = 3.402 \times 10^{-4}$). ASR 100 preserved 65.113 more trials on average than ASR 20 ($p = 4.441 \times 10^{-5}$). ASR 120 preserved 53.565 more trials on average than ASR 20 ($p = 7.822 \times 10^{-4}$).

Five of the results were preserved from the one-way ANOVA test to the Kruskal-Wallis ANOVA, whereas six significance were erased when accounting for variance. The five that were consistent across ANOVAs were the following: Hand rejection - ASR 20, Hand rejection - ASR 40, ASR 80 - ASR 20, ASR 100 - ASR 20, ASR 120 - ASR 20.

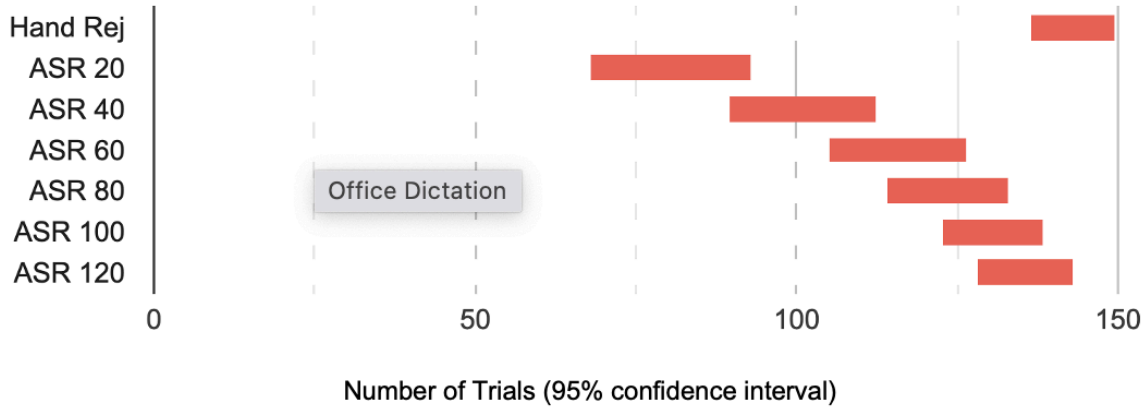

Figure 3. Shown are the mean number of trials preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 Hit data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions (f(6,210)=20.6679, MSE = 703.5462, p = 0). There were 11 significant differences across conditions. ASR 120 preserves similar numbers of trails as the hand rejection condition. ASR 20 is significantly more variable compared to ASR 120.

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	62.2903	13.0754	4.22E+01	82.348	1.073E-10***
Hand Rej - ASR 40	41.871	8.7892	21.8133	61.9287	5.705E-8***
Hand Rej - ASR 60	26.9677	5.6608	6.9101	47.0254	0.001662**
Hand Rej - ASR 80	19.2903	4.0492	-0.7674	39.348	0.06813
Hand Rej - ASR 100	12.129	2.546	-7.9287	32.1867	0.5494
Hand Rej - ASR 120	7.2258	1.5168	-12.8319	27.2835	0.9355
ASR 40- ASR 20	20.4194	4.2862	0.3617	40.477	0.04299*
ASR 60 - ASR 20	35.3226	7.4146	15.2649	55.3803	7.932E-6***
ASR 80 - ASR 20	43	9.0262	22.9423	63.0577	2.307E-8***
ASR 100 - ASR 20	50.1613	10.5294	30.1036	70.219	1.589E-10***
ASR 120 - ASR 20	55.0645	11.5586	35.0068	75.1222	1.079E-10***
ASR 60 - ASR 40	14.9032	3.1283	-5.1545	34.9609	0.2935
ASR 80 - ASR 40	22.5806	4.7399	2.523	42.6383	0.01632*
ASR 100 - ASR 40	29.7419	6.2431	9.6842	49.7996	0.0003219***
ASR 120 - ASR 40	34.6452	7.2724	14.5875	54.7028	1.277E-5***
ASR 80 - ASR 60	7.6774	1.6116	-12.3803	27.7351	0.9151
ASR 100 - ASR 60	14.8387	3.1148	-5.219	34.8964	0.2986
ASR 120 - ASR 60	19.7419	4.144	-0.3158	39.7996	0.05689
ASR 100 - ASR 80	7.1613	1.5032	-12.8964	27.219	0.9381
ASR 120 - ASR 80	12.0645	2.5325	-7.9932	32.1222	0.5558
ASR 120 - ASR 100	4.9032	1.0292	-15.1545	24.9609	0.9907

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 20.0577. The SE was 4.7639. N = 31 values.

Table 4. *Post-hoc Tukey Honest Significant Difference test comparing the impact of each ASR parameter condition on the average number of persevered trials with a 95% confidence interval (Session 1, Hit)*

Pair	Difference	Z-value	p-value
Hand Rej - ASR 20	86.7258	5.4386	5.37E-08***
Hand Rej - ASR 40	56.9032	3.5684	3.59E-04***
Hand Rej - ASR 60	41.7258	2.6166	0.00888
Hand Rej - ASR 80	29.5968	1.856	0.06345
Hand Rej - ASR 100	21.6129	1.3553	0.1753
Hand Rej - ASR 120	33.1613	2.0795	0.03757
ASR 40- ASR 20	29.8226	1.8702	0.06146
ASR 60 - ASR 20	45	2.8219	0.004773**
ASR 80 - ASR 20	57.129	3.5826	3.40E-04***
ASR 100 - ASR 20	65.1129	4.0832	4.44E-05
ASR 120 - ASR 20	53.5645	3.359	7.82E-04
ASR 60 - ASR 40	15.1774	0.9518	0.3412
ASR 80 - ASR 40	27.3065	1.7124	0.08683
ASR 100 - ASR 40	35.2903	2.2131	0.02689
ASR 120 - ASR 40	23.7419	1.4889	0.1365
ASR 80 - ASR 60	12.129	0.7606	0.4469
ASR 100 - ASR 60	20.1129	1.2613	0.2072
ASR 120 - ASR 60	8.5645	0.5371	0.5912
ASR 100 - ASR 80	7.9839	0.5007	0.6166
ASR 120 - ASR 80	3.5645	0.2235	0.8231
ASR 120 - ASR 100	11.5484	0.7242	0.4689
Note. ** Denotes $p < 0.0024$ *** Denotes $p < 0.001$. The critical mean for the comparisons was 48.4463. The SE was 15.9464. N = 31 values.			

Table 5. *Post-hoc Dunn's test comparing the impact of each ASR parameter condition on the average number of persevered trials (Session 1, Hit)*

Experiment 1: Session 2, Hit condition

In session 2 of the correct rejection condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 4; $f(6,210)=11.0705$, MSE = 887.5143, $p=1.02 \times 10^{-10}$). This effect has medium power ($1-\beta=0.7837$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used (Table 6). There were seven statistically significant differences based on a Tukey's critical mean of 22.528. Unlike in session 1, the variances across conditions were considered to be equal ($F = 1.068148$, $p = 0.382747$).

Hand rejection preserved 52.226 more trials on average than ASR 20 ($p = 1.36 \times 10^{-9}$). Hand rejection preserved 32.097 more trials than the ASR 40 condition ($p = 0.0006525$). ASR 60 preserved 29.839 more trials on average than ASR 20 ($p = 0.002082$). ASR 80 preserved 36.129 more trials on average than ASR 20 ($p = 6.853 \times 10^{-5}$). ASR 100 preserved 42 more trials on average than ASR 20 ($p = 1.774 \times 10^{-6}$). ASR 120 preserved 46.387 more trials on average than ASR 20 ($p = 0.002683$). ASR 120 preserved 26.258 more trials on average than ASR 40 ($p = 0.01111$).

doi.org/10.32388/SMEI50

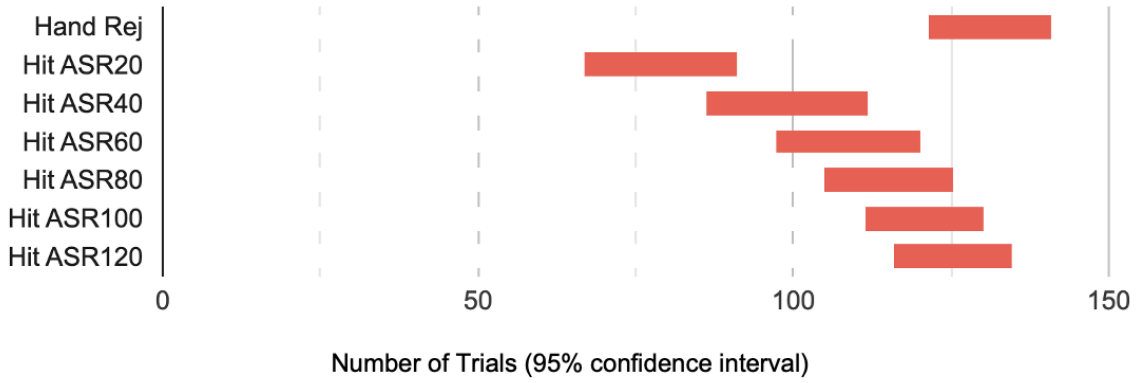

Figure 4. Shown are the mean number of trials preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 2 Hit data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,210)=11.0705$, MSE = 887.5143, $p=1.02 \times 10^{-10}$). There were seven significant differences across conditions. ASR 120 preserves similar numbers of trails as the hand rejection condition.

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	52.2258	9.76E+00	2.97E+01	74.7538	1.36E-09***
Hand Rej - ASR 40	32.0968	6.00E+00	9.5688	54.6248	0.0006525***
Hand Rej - ASR 60	22.3871	4.184	-0.1409	44.9151	0.05265
Hand Rej - ASR 80	16.0968	3.0084	-6.4312	38.6248	0.3406
Hand Rej - ASR 100	10.2258	1.9111	-12.3022	32.7538	0.8266
Hand Rej - ASR 120	5.8387	1.0912	-16.6893	28.3667	0.9874
ASR 40- ASR 20	20.129	3.762	-2.3989	42.657	0.1138
ASR 60 - ASR 20	29.8387	5.5766	7.3107	52.3667	0.002082**
ASR 80 - ASR 20	36.129	6.75E+00	13.6011	58.657	6.853E-05***
ASR 100 - ASR 20	42	7.85E+00	19.472	64.528	1.774E-06***
ASR 120 - ASR 20	46.3871	8.67E+00	23.8591	68.9151	8.97E-08***
ASR 60 - ASR 40	9.7097	1.8147	-12.8183	32.2377	0.859
ASR 80 - ASR 40	16	2.9903	-6.528	38.528	0.348
ASR 100 - ASR 40	21.871	4.0875	-0.657	44.3989	0.06339
ASR 120 - ASR 40	26.2581	4.9075	3.7301	48.786	0.01111*
ASR 80 - ASR 60	6.2903	1.1756	-16.2377	28.8183	0.9814
ASR 100 - ASR 60	12.1613	2.2729	-10.3667	34.6893	0.6779
ASR 120 - ASR 60	16.5484	3.0928	-5.9796	39.0764	0.3071
ASR 100 - ASR 80	5.871	1.0972	-16.657	28.3989	0.987
ASR 120 - ASR 80	10.2581	1.9172	-12.2699	32.786	0.8245
ASR 120 - ASR 100	4.3871	0.8199	-18.1409	26.9151	0.9973

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 22.528. The SE was 5.3507. N = 31 values.

Table 6. *Post-hoc Tukey Honest Significant Difference test comparing the impact of each ASR parameter condition on the average number of persevered trials with a 95% confidence interval (Session 1, Hit)*

Experiment 1 Discussion

In line with expectations for the exploratory analysis, more 'aggressive' parameters for ASR remove more trials than less aggressive parameters. Throughout, ASR 120 performed similarly to the manual rejection condition. Surprisingly, in both session 2 conditions there was unequal variance across certain conditions that suggest some data sets ASR is applied to can result in significantly more variable data rejection compared to manual rejection. Since the number of preserved trials that have high variance occurring at the more 'aggressive' end of the parameter spectrum relative to higher values with ASR 120 (which is comparable to manual rejection in the present case), researchers should be aware that more aggressive choices for ASR might remove certain signals/artifacts unpredictably, relative to if a more conservative approach is taken. As such, researchers should investigate what type of artifacts exist in their data set *a priori* and keep track of whether those artifacts are being rejected or if more brain signals are being rejected.

The number of trials preserved on its own is a neutral metric on its own. In other words, having many trials preserved or rejected does not suggest better or worse performance of ASR. Rather, it depends on the application. For example, it might make sense in event-related potential (ERP) studies to retain more trials, since averaging elevates meaningful signals since random noise across trials theoretically cancels out. In the present case where data is being prepared for MVAR models, without taking any other metrics into account, cleaner data might be a better direction (i.e., more aggressive

parameters). However, the way to test whether having fewer trials and cleaner data involves looking at the number of ICs and the model order *in context* of the number of trials preserved, which will be done in the following sections. For full discussion on this point, see 'Full Discussion.'

Experiment 2. Model order of MVAR fitting modified by ASR parameter choices

When trying to improve the performance of rPDC, it is important to have the lowest model order of data possible without minimizing it artificially (Barnett and Seth, 2014). In other words, over- and underspecified models are not informative to infer directional flow of information (Cohen, 2014; Seth, 2010). Downsampling, which was applied in the present study, is one popular way to minimize model order. However, low variance brain signals can also produce low model orders because fewer parameters can capture the curvature of data. Thus, finding an optimal approach for artifact rejection is potentially an effective way to reduce model order, since it is not alternating data, but merely removing data that appears to be too variable to show meaningful brain signals. It was expected that more aggressive parameters for ASR would decrease model order. This expectation was, in general, corroborated, with an important caveat. It appears that more aggressive parameters decrease the number of participants with more than 1 ICs.

Experiment 2: Session 1, Correct Rejection condition

In session 1 of the correct rejection condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 5; $f(6,195)=2.5769$, MSE = 0.4959, $p=0.02002$). This effect has medium power ($1-\beta=0.747$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used. There was one statistically significant difference based on a Tukey's critical mean of 0.564. The hand rejection condition average model order was significantly greater than ASR 20, specifically by 0.61 units on average ($p = 0.02305$). The variances across conditions were considered to be equal in every condition.

Upon regression analysis, there was a strong association between the model order in each condition and the number of participants left in the analysis after accounting for the participants with less than 2 ICs, $r = 0.9203$, $p = 0.003298$. In fact, 84.7% of the residual variance in model order is explained by the number of participants left in the analysis. For every additional participant included in the analysis, the average model order increases by 0.0994 units.

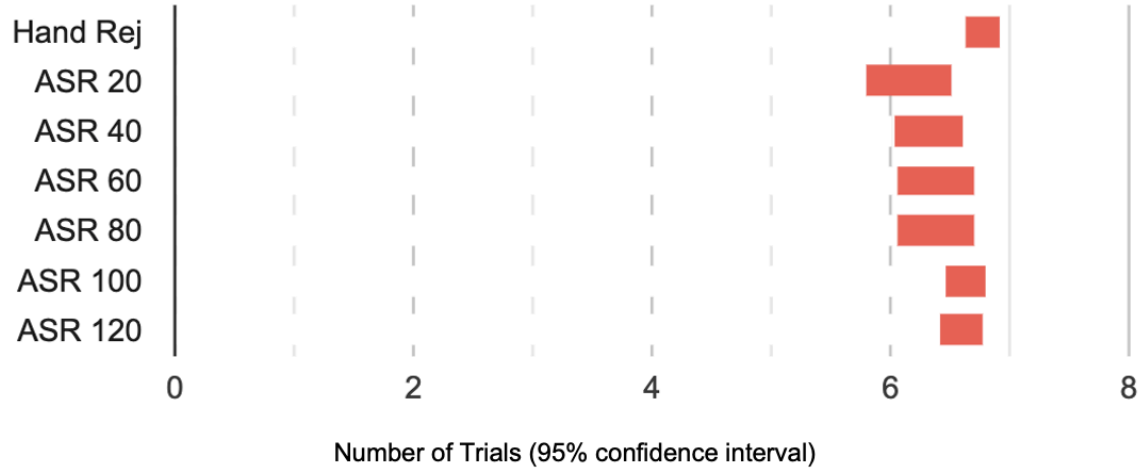

Figure 5. Shown is the mean model order preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 CR data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,195)=2.5769$, MSE = 0.4959, $p=0.02002$). ASR 20 had a significantly lower model order on average compared to Hand rejection. All other conditions are considered identical.

Experiment 2: Session 2, Correct Rejection condition

In session 2 of the correct rejection condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 6; $f(6,182)=2.9993$, MSE = 0.4139, $p=0.008$). This effect has medium power ($1-\beta=0.712$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used. There were three statistically significant differences. The variances across conditions were considered to be equal in every condition.

The hand rejection condition had a significantly higher model order than ASR 20, with a 0.64 unit difference on average (p = 0.005816). ASR 80 had a significantly higher model order than ASR 20, with a 0.54 unit difference on average (p = 0.0404). ASR 120 had a significantly higher model than ASR 20, with a 0.62 unit difference on average (p = 0.0122). There was no significant association between the model order and the number of participants included in the analysis (p = 0.221).

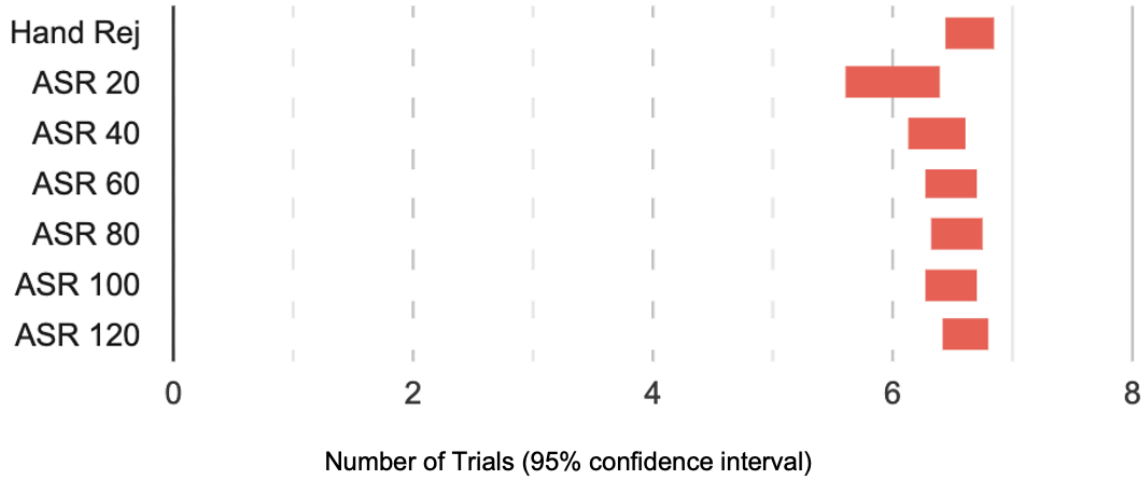

Figure 6. Shown is the mean model order preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 2 CR data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions (f(6,182)=2.9993, MSE = 0.4139, p=0.008). There were three significant differences.

Experiment 2: Session 1, Hit condition

In session 1 of the Hit condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 7; f(6,201)=4.5583, MSE = 0.3586, p=0.000233). This effect has medium power (1-β=0.762), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used. There were four statistically significant differences. The variances across conditions were considered to be equal in every condition.

The hand rejection condition had a significantly higher model order than ASR 20, with a 0.63 unit difference on average (p = 0.0023). The hand rejection condition had a significantly higher model order than ASR 40, with a 0.54 unit difference on average (p = 0.00994). ASR 80 had a significantly higher model order than ASR 20, with a 0.49 unit difference on average (p = 0.04178). ASR 120 had a significantly higher model than ASR 20, with a 0.53 unit difference on average (p = 0.02272). There was no significant association between the model order and the number of participants included in the analysis (p = 0.106).

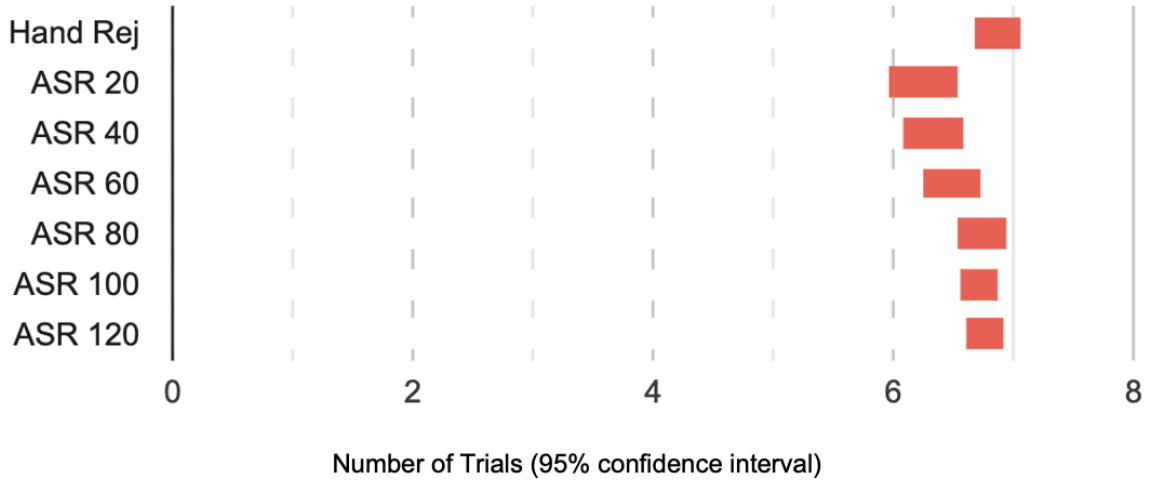

Figure 7. Shown is the mean model order preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 hit data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,201)=4.5583$, MSE = 0.3586, $p=0.000233$). There were four significant differences.

Experiment 2: Session 2, Hit condition

In session 2 of the Hit condition, a one-way ANOVA determined that there was a significant difference between conditions (Figure 8; $f(6,194)=3.6203$, MSE = 0.4552, $p=0.001993$). This effect has medium power ($1-\beta=0.745$), which further increases confidence in this result. To characterize which conditions were significantly different, a post-hoc Tukey's Honest Significant Difference test was used. There were three statistically significant differences. The variances across conditions were considered to be equal in every condition.

The hand rejection condition had a significantly higher model order than ASR 20, with a 0.64 unit difference on average ($p = 0.01076$). ASR 100 had a significantly higher model order than ASR 20, with a 0.59 unit difference on average ($p = 0.01999$). ASR 120 had a significantly higher model than ASR 20, with a 0.63 unit difference on average ($p = 0.01107$).

Upon regression analysis, there was a strong association between the model order in each condition and the number of participants left in the analysis after accounting for the participants with less than 2 ICs, $r = 0.825$, $p = 0.0223$. In fact, 68.1% of the residual variance in model order is explained by the number of participants left in the analysis. For every additional participant included in the analysis, the average model order increases by 0.116 units.

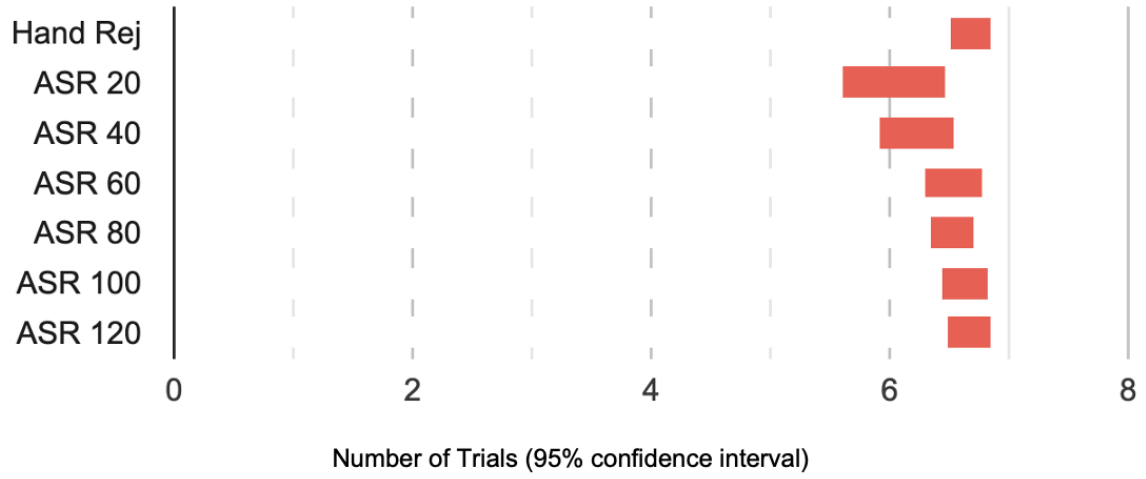

Figure 8. Shown is the mean model order preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 hit data. A one-way ANOVA demonstrated that there was a significant difference somewhere across the conditions ($f(6,194)=3.6203$, MSE = 0.4552, $p=0.001993$). There were three significant differences.

Experiment 2: Discussion

In line with expectations for the exploratory analysis, more 'aggressive' parameters for ASR produces a lower model order than less aggressive parameters. This is because fewer data allows for few parameters to be used during the MVAR fitting process; in other words, the data that is left is the least variable and is the easiest to model with lower dimensional models. However, the results are complicated by an additional regression analysis. In CR session 1 and Hit session 2, much of the effect was explained by the number of participants included in the analysis. More aggressive parameters of ASR produce lower model order because it preserves fewer data; since fewer data is preserved, there is a lower ability for ICs to be fit (i.e., more participants are excluded due to having only one IC, which means MVAR cannot be fitted to that data). This did not occur to a statistically significant degree in CR session 2 and Hit session 1. However, it is reasonable to expect that, in general, the lower the model order produced from higher parameters of ASR, the lower the overall power of the final analysis of interest (e.g., rPDC). Thus, there is incentive to not make model order as low as possible from artifact rejection alone.

In terms of the performance of specific ASR parameters, the only consistently significant result was that hand rejection produced a statistically higher model order than ASR 20. However, because of the aforementioned issue of decreasing overall study power from forcing more participants to be excluded, ASR 20 might be too aggressive for some types of data sets. Thus, since there were so few statistical differences across parameters choices of ASR, it is reasonable to conclude that, more or less, ASR parameters choices from 20-120 perform similarly well for fitting model order. The only consistent exception in the present study is ASR 20, which in two out of four cases was too aggressive and is at risk of lowering overall study power. ASR 20 can be used in studies with a similar result if the research has good reasons for trading power for low variable data.

Experiment 3:Independent Component Generation via Extended Infomax ICA

Typically, the goal of decomposition is to maximize the number of dipolar ICs produced by ICA; in fact, the number of ICs is considered to be a direct measure of the quality of ICA decomposition (Medaglia et al., 2011). Since dipolarity is a relatively strict criteria based on residual variance, there is no reason to worry if an increase in ICs is being produced by non-brain related components.

There was no clear predict on the relationship between ASR parameter choice and the

number of ICs produced. In one sense, more aggressive ASR parameters could produce fewer ICs. This is because there would be fewer data, which would provide fewer data to effectively sift through and identify ICs. However, more aggressive ASR parameters could also produce more ICs, because less variable data increases the effectiveness of ICA decomposition. As such, this experiment was considered to be fully exploratory.

Experiment 3: Combined results

There was no significant difference across ASR parameters in any of the study conditions. For completeness, results for each condition are shown: correct rejection session 1 (Figure 9), correct rejection session 2 (Figure 10), hit session 1 (Figure 11), hit session 2 (Figure 12).

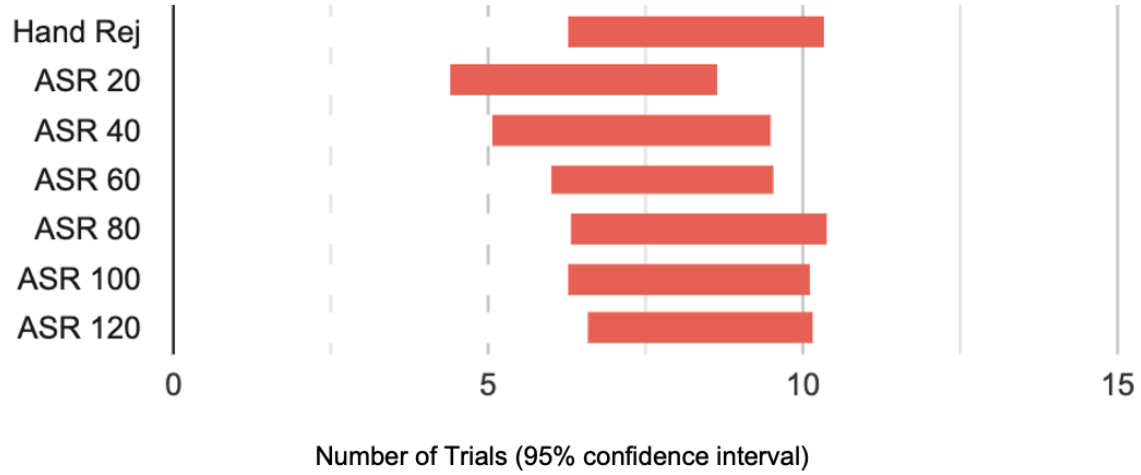

Figure 9. Shown are the mean number of ICs preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 correct rejection data. There was no significant difference across conditions.

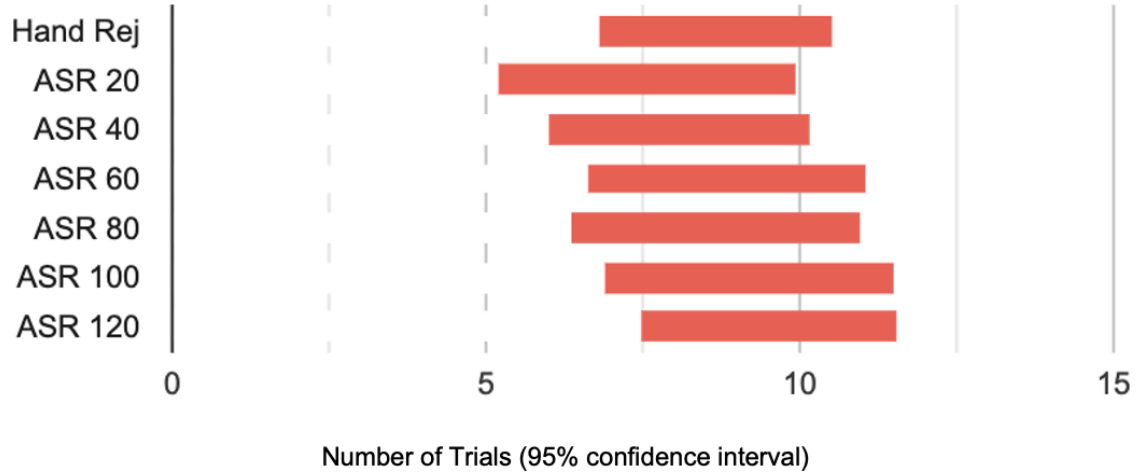

Figure 10. Shown are the mean number of ICs preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 2 correct rejection data. There was no significant difference across conditions.

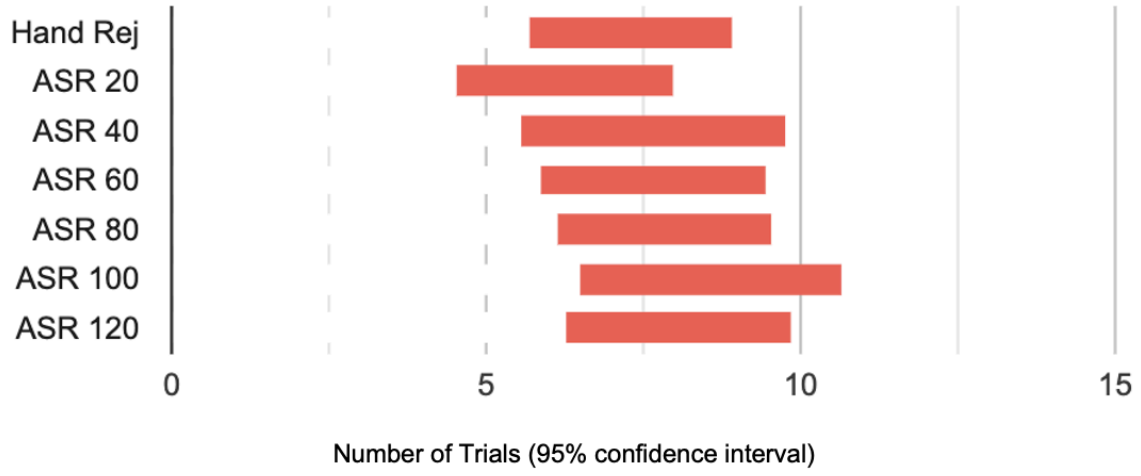

Figure 11. Shown are the mean number of ICs preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 1 hit data. There was no significant difference across conditions.

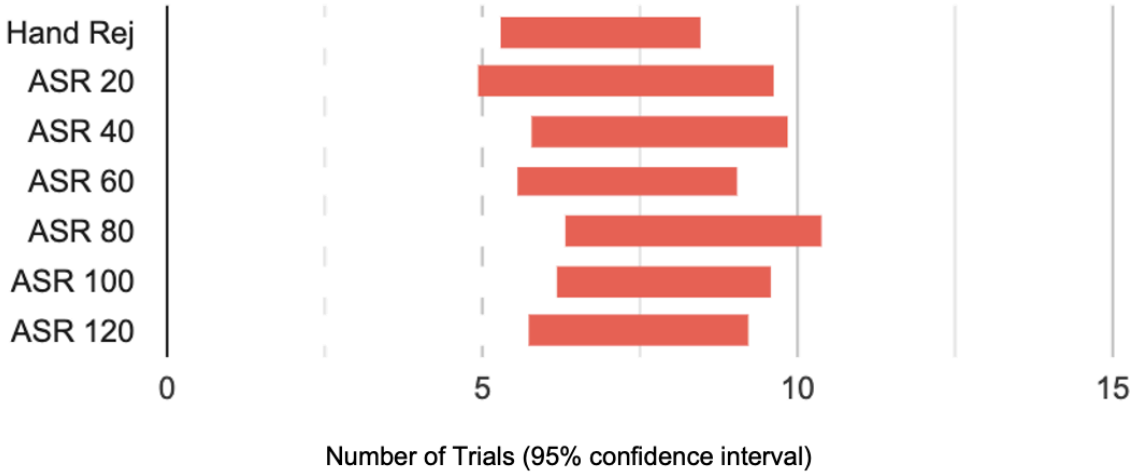

Figure 12. Shown are the mean number of ICs preserved after different artifact rejection strategies, with the edge of the bars showing the upper and lower bound of the 95% confidence interval for session 2 hit data. There was no significant difference across conditions.

Experiment 3: Discussion

No clear prediction was made, since at least two opposing, equally valid predictions were positive. Ultimately, there was no significant difference between the number of ICs produced across all conditions. As such, it appears that there is no difference in the overall number of ICs produced by different ASR parameter choices between 20 and 120. The variance across all conditions was the same, as well.

Overall Recommendations

After experiments 1-3, the results show that, in general, ASR is as effective or better than hand rejection of artifacts in EEG data. Specifically, there is no difference in the choice between ASR with a parameter of 20 to 120 and hand rejection for the average number of ICs produced. As such, secondary characteristics become more important when picking a parameter.

It was found, as expected, that higher or less aggressive values of ASR (e.g., 100, 120) preserved more trials on average than more aggressive parameters of ASR (e.g., 20, 40). However, it appears that more aggressive parameters of ASR (e.g., 20, 40) produce lower model orders in MVAR model

fitting, which is an important characteristic of rPDC and other information-theoretic techniques. As discussed, more aggressive parameters produce fewer data, which in turn requires fewer parameters to fit an MVAR model to the data. However, as it was also noted, in one half of the conditions the lower model order was explained primarily by there being fewer participants included in the analysis. A participant was removed when there was only one IC identified in decomposition, because two ICs are required for MVAR model fitting, by definition. Thus, a parameter choice of 20 or 40 for ASR, if it removes a large number of participants, is too aggressive given the number of ICs preserved is the same across conditions.

Taking all quantitative features into account, higher parameters of ASR (e.g., 120) performed as well as hand rejection with only a mild impact in model order. However, as mentioned, the trade off is that all participants remained included in the analysis which maximizes the ultimate statistical power of rPDC to find an effect. This is the first evidence that parameters outside of 10-100 for ASR might be effective and also is the first clear evidence to show that the optimal parameter choice of ASR is not 20 - 30 as Chang et al. (2019) suggest. It is recommended that authors do a similar quantitative analysis as the one in this study to justify ASR parameter choice. Even though it is mildly time consuming, ASR increases the replicability of studies and allows researchers to fully inform scientists of the choices and trade offs made by selecting a particular parameter.

Future research should increase the types of data sets that ASR is tested on. This is the first study to look at ASR's use in cognitive tasks. Because the present results are different compared to previously published research, it appears that ASR's effectiveness depends on the type of data set it is applied to (e.g., muscle movement, cognitive tasks). This will better inform "best practice" research.

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	12.806452	3.418423	-2.966698	28.579602	0.196607
Hand Rej - ASR 40	13.387097	3.573414	-2.386053	29.160247	0.155109
Hand Rej - ASR 60	7.096774	1.89434	-8.676376	22.869924	0.832512
Hand Rej - ASR 80	3.322581	0.886896	-12.450569	19.095731	0.995867
Hand Rej - ASR 100	1.225806	0.327204	-14.547344	16.998956	0.999987
Hand Rej - ASR 120	6.774194	1.808234	-8.998956	22.547344	0.861067
ASR 40- ASR 20	0.580645	0.154991	-15.192505	16.353795	1
ASR 60 - ASR 20	5.709678	1.524083	-10.063472	21.482828	0.934037
ASR 80 - ASR 20	9.483871	2.531527	-6.289279	25.257021	0.556263
ASR 100 - ASR 20	11.580646	3.091219	-4.192504	27.353796	0.307663
ASR 120 - ASR 20	19.580646	5.226657	3.807496	35.353796	0.00513158 **
ASR 60 - ASR 40	6.290323	1.679074	-9.482827	22.063473	0.898341
ASR 80 - ASR 40	10.064516	2.686519	-5.708634	25.837666	0.482848
ASR 100 - ASR 40	12.161291	3.24621	-3.611859	27.934441	0.251128
ASR 120 - ASR 40	20.161291	5.381648	4.388141	35.934441	0.00346557 **
ASR 80 - ASR 60	3.774193	1.007444	-11.998957	19.547343	0.991745
ASR 100 - ASR 60	5.870968	1.567136	-9.902182	21.644118	0.925102
ASR 120 - ASR 60	13.870968	3.702574	-1.902182	29.644118	0.125803
ASR 100 - ASR 80	2.096775	0.559692	-13.676375	17.869925	0.999696
ASR 120 - ASR 80	10.096775	2.69513	-5.676375	25.869925	0.478822
ASR 120 - ASR 100	8	2.135438	-7.77315	23.77315	0.738623

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 15.77315. The SE was 3.746304. N = 31 values.

Appendix A. *Levene's test for the equality of variances comparing the variances of the data produced by each ASR parameter condition for the average number of persevered trials with a 95% confidence interval (Session 1, Correct Rejections)*

			Confidence interval		
Pair	**Difference**	**Critical-Q**	**Lower**	**Upper**	**p-value**
Hand Rej - ASR 20	14.16129	4.84E+00	1.840949	2.65E+01	0.013009*
Hand Rej - ASR 40	11.709677	4.001636	-0.610664	24.030018	0.0744356
Hand Rej - ASR 60	8.741935	2.987447	-3.578406	21.062276	0.349196
Hand Rej - ASR 80	6.903225	2.359091	-5.417116	19.223566	0.638026
Hand Rej - ASR 100	2.451612	0.837808	-9.868729	14.771953	0.996982
Hand Rej - ASR 120	2.32258	0.793713	-9.997761	14.642921	0.997767
ASR 40- ASR 20	2.451613	0.837808	-9.868728	14.771954	0.996982
ASR 60 - ASR 20	5.419355	1.851997	-6.900986	17.739696	0.846913
ASR 80 - ASR 20	7.258065	2.480353	-5.062276	19.578406	0.580662
ASR 100 - ASR 20	11.709678	4.001636	-0.610663	24.030019	0.0744355
ASR 120 - ASR 20	11.83871	4.05E+00	-0.481631	24.159051	0.0685786
ASR 60 - ASR 40	2.967742	1.014189	-9.352599	15.288083	0.991445
ASR 80 - ASR 40	4.806452	1.642545	-7.513889	17.126793	0.90764
ASR 100 - ASR 40	9.258065	3.163828	-3.062276	21.578406	0.280324
ASR 120 - ASR 40	9.387097	3.207923	-2.933244	21.707438	0.264451
ASR 80 - ASR 60	1.83871	0.628356	-10.481631	14.159051	0.999407
ASR 100 - ASR 60	6.290323	2.149639	-6.030018	18.610664	0.732547
ASR 120 - ASR 60	6.419355	2.193735	-5.900986	18.739696	0.713365
ASR 100 - ASR 80	4.451613	1.521283	-7.868728	16.771954	0.934592
ASR 120 - ASR 80	4.580645	1.565378	-7.739696	16.900986	0.925481
ASR 120 - ASR 100	0.129032	0.0440951	-12.191309	12.449373	1

Note. * Denotes $p < 0.05$, ** Denotes $p < 0.01$, *** Denotes $p < 0.001$. The critical mean for the comparisons was 12.320341. The SE was 2.926222. N = 31 values.

Appendix B. *Levene's test for the equality of variances comparing the variances of the data produced by each ASR parameter condition for the average number of persevered trials with a 95% confidence interval (Session 1, Hit)*

Footnotes

[1] Model order acts essentially as a parameter that allows the constructed MVAR model to not over or underfit the data by specifying how many data points to include in the model.

[2] Power spectrum of four different channels was measured, but was not analyzed in this draft due to time constraints. This will be addressed in the summer.

[3] Please see link for description on Ghost ICs https://sccn.ucsd.edu/wiki/Makoto%27s_preprocessing_pipeline#Tips_for_copying_ICA_results_to_other_datasets_.2806.2F26.2F2018_updated.

References

- Ablin, P., Cardoso, J.-F., & Gramfort, A. (2018). Faster Independent Component Analysis by Preconditioning With Hessian Approximations. *IEEE Transactions on Signal Processing: A Publication of the IEEE Signal Processing Society*, *66*(15), 4040–4049.

- Adamou, M., Fullen, T., & Jones, S. L. (2020). EEG for Diagnosis of Adult ADHD: A Systematic Review With Narrative Analysis. *Frontiers in Psychiatry / Frontiers Research Foundation*, *11*, 871.
- Al Zoubi, O., Mayeli, A., Tsuchiyagaito, A., Misaki, M., Zotev, V., Refai, H., Paulus, M., Bodurka, J., & Tulsa 1000 Investigators. (2019). EEG Microstates Temporal Dynamics Differentiate Individuals with Mood and Anxiety Disorders From Healthy Subjects. *Frontiers in Human Neuroscience*, *13*, 56.
- Artoni, F., Fanciullacci, C., Bertolucci, F., Panarese, A., Makeig, S., Micera, S., & Chisari, C. (2017). Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking. *NeuroImage*, *159*, 403–416.
- Barnett, L., & Seth, A. K. (2011). Behaviour of Granger causality under filtering: theoretical invariance and practical application. *Journal of Neuroscience Methods*, *201*(2), 404–419.
- Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68.
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing: A Publication of the IEEE Signal Processing Society*, *45*(2), 434–444.
- Bloniasz, P. F. (2022). *Directed Information Flow During Episodic Memory Retrieval at Theta Frequency* [Bowdoin College]. IN-PREPARATION.
- Blum, S., Jacobsen, N. S. J., Bleichner, M. G., & Debener, S. (2019). A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling. *Frontiers in Human Neuroscience*, *13*, 141.
- Bulea, T. C., Prasad, S., Kilicarslan, A., & Contreras-Vidal, J. L. (2014). Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution. *Frontiers in Neuroscience*, *8*, 376.
- Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2019). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. *IEEE Transactions on Bio-Medical Engineering*, *67*(4), 1114–1121.
- Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2018). Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1242–1245.
- Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, *56*(11), e13437.
- Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.
- Delorme, A., Jung, T.-P., Sejnowski, T., & Makeig, S. (2005). Improved rejection of artifacts from EEG data using high-order statistics and independent component analysis. *NeuroImage*. https://www.academia.edu/download/51274215/IEEE2005.pdf
- 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.
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS One*, *7*(2), e30135.
- Florin, E., Gross, J., Pfeifer, J., Fink, G. R., & Timmermann, L. (2010). The effect of filtering on Granger causality based multivariate causality measures. *NeuroImage*, *50*(2), 577–588.
- Gabard-Durnam, L. J., O'Muircheartaigh, J., Dirks, H., Dean, D. C., 3rd, Tottenham, N., & Deoni, S. (2018). Human amygdala functional network development: A cross-sectional study from 3 months to 5 years of age. *Developmental Cognitive Neuroscience*, *34*, 63–74.
- Gourévitch, B., Bouquin-Jeannès, R. L., & Faucon, G. (2006). Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biological Cybernetics*, *95*(4), 349–369.
- Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J. R., van Wassenhove, V., Wibral, M., & Schoffelen, J.-M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363.
- Jurgiel, J., Miyakoshi, M., Dillon, A., Piacentini, J., Makeig, S., & Loo, S. K. (2021). Inhibitory control in children with tic disorder: aberrant fronto-parietal network activity and connectivity. *Brain Communications*, *3*(2), fcab067.
- Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*(1), 1–21.
- Keizer, A. W. (2021). Standardization and Personalized Medicine Using Quantitative EEG in Clinical Settings. *Clinical EEG and Neuroscience: Official Journal of the EEG and Clinical Neuroscience Society* , *52*(2), 82–89.
- Koshiyama, D., Miyakoshi, M., Joshi, Y. B., Molina, J. L., Tanaka-Koshiyama, K., Sprock, J., Braff, D. L., Swerdlow, N. R., & Light, G. A. (2020). Abnormal Effective Connectivity Underlying Auditory Mismatch Negativity Impairments in Schizophrenia. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *5*(11), 1028–1039.

- Koshiyama, D., Miyakoshi, M., Tanaka-Koshiyama, K., Joshi, Y. B., Molina, J. L., Sprock, J., Braff, D. L., & Light, G. A. (2020). Neurophysiologic Characterization of Resting State Connectivity Abnormalities in Schizophrenia Patients. *Frontiers in Psychiatry / Frontiers Research Foundation*, *11*, 608154.
- Koshiyama, D., Miyakoshi, M., Thomas, M. L., Joshi, Y. B., Molina, J. L., Tanaka-Koshiyama, K., Nungaray, J. A., Sprock, J., Braff, D. L., Swerdlow, N. R., & Light, G. A. (2020). Auditory-Based Cognitive Training Drives Short- and Long-Term Plasticity in Cortical Networks in Schizophrenia. *Schizophrenia Bulletin Open*, *1*(1).
- Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, *11*(2), 417–441.
- Loo, S. K., Miyakoshi, M., Tung, K., Lloyd, E., Salgari, G., Dillon, A., Chang, S., Piacentini, J., & Makeig, S. (2019). Neural activation and connectivity during cued eye blinks in Chronic Tic Disorders. *NeuroImage. Clinical*, *24*, 101956.
- Miyakoshi, M., Gehrke, L., Gramann, K., Makeig, S., & Iversen, J. (2021). The AudioMaze: An EEG and motion capture study of human spatial navigation in sparse augmented reality. *The European Journal of Neuroscience*, *54*(12), 8283–8307.
- Mullen, T., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Cauwenberghs, G., & Jung, T.-P. (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *Conference Proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2013*, 2184–2187.
- Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., & Cauwenberghs, G. (2015). Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. *IEEE Transactions on Bio-Medical Engineering*, *62*(11), 2553–2567.
- Nyhus, E. M. (2010). *Theta Oscillations in Top-down Control of Episodic Memory Retrieval* (T. Curran (ed.)) [PhD, University of Colorado Boulder]. http://scholar.colorado.edu/downloads/707957952
- Patel, R. J. (2020). *Directed interactions during episodic memory* [Bowdoin College]. https://digitalcommons.bowdoin.edu/honorsprojects/179/
- Perera, H., Shiratuddin, M. F., & Wong, K. W. (2016). A Review of Electroencephalogram-Based Analysis and Classification Frameworks for Dyslexia. *Neural Information Processing*, 626–635.
- Pernet, C. R., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2018). *Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG*. https://doi.org/10.31219/osf.io/a8dhx
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197.
- Pontifex, M. B., Gwizdala, K. L., Parks, A. C., Billinger, M., & Brunner, C. (2017). Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts. *Psychophysiology*, *54*(3), 386–398.
- Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., & Bigdely-Shamlo, N. (2020). How Sensitive Are EEG Results to Preprocessing Methods: A Benchmarking Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *28*(5), 1081–1090.
- Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2013). Independent Components Analysis with the JADE algorithm. *Trends in Analytical Chemistry: TRAC*, *50*, 22–32.
- Schelter, B., Timmer, J., & Eichler, M. (2009). Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*, *179*(1), 121–130.
- Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, *186*(2), 262–273.
- Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(8), 3293–3297.
- Shu, H., Gu, L., Yang, P., Lucas, M. V., Gao, L., Zhang, H., Zhang, H., Xu, Z., Wu, W., Li, L., & Zhang, Z. (2021). Disturbed temporal dynamics of episodic retrieval activity with preserved spatial activity pattern in amnestic mild cognitive impairment: A simultaneous EEG-fMRI study. *NeuroImage. Clinical*, *30*, 102572.
- Sun, Y., Wei, C., Cui, V., Xiu, M., & Wu, A. (2020). Electroencephalography: Clinical Applications During the Perioperative Period. *Frontiers of Medicine*, *7*, 251.
- Zion-Golumbic, E., Kutas, M., & Bentin, S. (2010). Neural dynamics associated with semantic and episodic memory for faces: evidence from multiple frequency bands. *Journal of Cognitive Neuroscience*, *22*(2), 263–277.

Declarations

Funding: The author(s) received no specific funding for this work.

Potential competing interests: The author(s) declared that no potential competing interests exist.